

Fuzzy Classification Methods Based Diagnosis of Parkinson's disease from Speech Test Cases



Niousha Karimi Dastjerd¹, Onur Can Sert¹, Tansel Ozyer¹ and Reda Alhajj^{2,3,*}

¹TOBB University of Economics and Technology, Sogutozu, Ankara, 06560 Turkey; ²Department of Computer Science, University of Calgary, Calgary, Alberta, Canada; ³Department of Computer Engineering, Istanbul Medipol University, Istanbul, Turkey

Abstract: Background: Together with the Alzheimer's disease, Parkinson's disease is considered as one of the two serious known neurodegenerative diseases. Physicians find it hard to predict whether a given patient has already developed or is expected to develop the Parkinson's disease in the future. To overcome this difficulty, it is possible to develop a computing model, which analyzes the data related to a given patient and predicts with acceptable accuracy when he/she is anticipated to develop the Parkinson's disease.

Objectives: This paper contributes an attractive prediction framework based on some machine learning approaches for distinguishing people with Parkinsonism from healthy individuals.

Methods: Several fuzzy classifiers such as Inductive Fuzzy Classifier, Fuzzy Rough Classifier and two types of neuro-fuzzy classifiers have been employed.

Results: The fuzzy classifiers utilized in this study have been tested using the "Parkinson Speech Dataset with Multiple Types of Sound Recordings Data Set" of 40 subjects available on the UCI repository.

Conclusion: The results achieved show that FURIA, MLP- Bagging - SGD, genfis2 and scg1 performed the best among the fuzzy rough, WEKA, adaptive neuro-fuzzy and neuro-fuzzy classifiers, respectively. The worst performance belongs to nearest neighborhood, IBK, genfis3 and scg3 among the formerly mentioned classifiers. The results reported in this paper are better in comparison to the results reported in Sakar *et al.*, where the same dataset was used, with utilization of different classifiers. This demonstrates the applicability and effectiveness of the fuzzy classifiers used in this study as compared to the non-fuzzy classifiers used by Sakar *et al.*

ARTICLE HISTORY

Received: February 07, 2019
Revised: May 14, 2019
Accepted: May 22, 2019

DOI:
10.2174/1874609812666190625140311



CrossMark

Keywords: Parkinson's disease, data mining, machine learning, fuzzy classification, neuro fuzzy classification, adaptive neuro fuzzy classification.

1. INTRODUCTION

Parkinson's Disease (PD) is one of the central nervous system's neurodegenerative-disease. It is considered to be the second prevalent neurological disease after the Alzheimer's disease. It is anticipated that there are about 10 million individuals worldwide who suffer from the symptoms of PD [1, 2]. PD may lead to a full or a partial loss in some vital functions, behavior, motor reflexes, mental processing, and speech [1]. PD is usually caused by a decrease which occurs in the dopamine levels in the brain [3]. James Parkinson, after whom the disease was named, suggests that the first symptom of PD is a slight sense of weakness with a tendency to trembling in a specific body part, commonly one of the hands or arms [4]. The progress of PD is extremely slow, and it is hard for a patient to recollect the precise period of the disease commencement. Afterwards, a PD victim confronts deterioration of speech, *i.e.*, difficulty in articulating sound, reduced volume and pitch range. During the disease progres-

sion, the symptoms start to develop and engage sleeping ability along with increasing the risk of insanity.

Despite the existence of a significant number of medical treatments for decreasing the difficulties caused by PD, there is no known cure yet [5, 6]. Besides, PD diagnosis is generally performed using invasive methods which convolute the procedure of diagnosis and treatment [7]. Thus, there is an increasing necessity to develop a noninvasive diagnosis system. The side effects considered by studies carried on PD diagnosis include shaking or firmness occurred in some parts of the body, poor balance, gradualness in movements and specifically problems related to victim's voice [8-11]. The most important rationale in the popularity gained by speech test utilization in PD diagnosis is self-use easiness of tele-diagnosis and telemonitoring systems and the associated cost reduction [6, 12]. Additionally, the mentioned systems cause a decrease in the inconvenience and the costs occurring when the PD patients visit the medical centers. They empower the early analysis of the malady, and lessen the workload of medical staff [6, 12-14]. PWP (People With Parkinsonism) experience effects of discourse ruinations like hypophonia (lessens volume), dysphonia (inadequate utilization of the

*Address correspondence to this author at the Department of Computer Science, University of Calgary, Calgary, Alberta, Canada; Tel: 403-284-4707; E-mail: rsalhajj@gmail.com

voice), dysarthria (trouble with verbalization of sounds or syllables), and monotone (diminished pitch gauge).

The target of the research described in this paper is to use speech test data for diagnosing PWP (people with Parkinsonism) and differentiating them from healthy people. To achieve this, we employed a variety of machine learning techniques which utilize fuzziness as part of their engines. The dataset used in this study was contributed by Sakar *et al.* [1]. It is publicly available and has been downloaded from the UCI Machine Learning Repository. In total, the dataset is composed of 40 test subjects, which are distributed into two groups. The first group consists of 6 females and 14 males who are diagnosed with PD, and the second group consists of 10 females and 10 males who are healthy people. The test subjects were checked by the Neurology Department in Cerrahpasa School of Medicine at Istanbul University in Turkey. The machine learning-based techniques used in this paper are fuzzy rough classification, inductive fuzzy classification, neuro fuzzy classification, adaptive neuro fuzzy classification and some classical classification algorithms. They are available as part of WEKA software. The results have been validated by employing MSE (mean square error), sensitivity, accuracy, Matthews Correlation Coefficient, and specificity metrics.

The rest of this paper is organized as follows. Section 2 covers the related work. Section 3 describes the problem tackled in this paper. The fuzzy based classifiers used in this study are described in Section 4. The test results are reported and analyzed in Section 5. The comparative analysis is presented in Section 6. Section 7 is conclusion.

2. THE LITERATURE REVIEW

The target of the research conducted in this paper is to distinguish PWP from healthy individuals based on the features extracted from speech tests taken from PWP and healthy individuals. For this purpose, we employed three fuzzy classification methods for PD diagnosis. Thus, to better understand the methodology described in this paper, we review in this section PD together with its diagnosis and the utilized fuzzy classification methods.

2.1. Parkinson's Disease and its Diagnosis

According to Chakraborty *et al.* [15] "Parkinsonism is one of the prominent neurological disorders of old age and its prevalence is rising as the geriatric population is on the rise". As it is the case with every disease, early and exquisite detection of PD might be extremely useful in proficient treatment and inability restriction of casualties. To satisfy this goal, several researchers have focused on diagnosis methods and tests for PD.

Sakar *et al.* [1] produced the dataset utilized in this study. Their dataset is related to real patients who were diagnosed and checked in Istanbul, Turkey. They shared their dataset by adding it to the publicly available UCI Dataset Repository. They analyzed their data using some of the known classical classification algorithms to distinguish PWP from healthy test subjects. They used Leave One Subject Out (LOSO) and Summarized Leave One Out (SLOO) in the validation process. The classification results showed that

sustained vowels carry more PD- discriminative information. Fortunately, their results are not better than our results reported in this paper which we determined by employing some fuzzy classification techniques.

Shahsavari *et al.* [2] studied PD by employing machine learning techniques, and in particular hybrid particle swarm optimization method. Some researchers investigated the possibility of monitoring PD patients remotely, *e.g.*, [6, 8-10, 16, 17]. For instance, researchers used a set of wearable sensors to assess and classify the tremor activity in PD [9, 12]. Sama *et al.* [17] used a waist-worn sensor to analyze gait and to estimate bradykinesia severity in PD. Parisi *et al.* [18] developed a technique for handling and evaluating gait in PD patients. Other researchers also applied machine learning techniques, including a variety of classification algorithms, *e.g.*, [10, 11, 15, 16, 19-23].

The study conducted by Murdoch *et al.* [19] presented various impacts of PD in discourse and correspondence. Effects of various levels of PD on muscles of voice and signal generation were analyzed by Fox *et al.* [24]. Many researchers have focused on utilizing voice and speech features for PD diagnosis. Indeed, voice disorders form the main emphasis of neurologists and researchers in case of PD diagnosis. As mentioned before, PD gradually affects the voice and speech signals producing muscles and weakens them during different levels of PD severity. Periodic vibrations in the voice are used to measure voice disorders while utilizing acoustic devices. Clinical voice disorder diagnosis systems can be emphasized by utilizing some sound properties, including turbulent, complex nonlinear aperiodicity, non-Gaussian irregularity of sound and aeroacoustics [12]. Little *et al.* [6] analyzed the phase of sickness by measuring dysphonia caused by PD. The dataset they considered included 31 subjects of whom 23 were diagnosed with PD. In the speech test, they recorded sustained vowel "a" phonation. Then, the disease grade was identified by telemonitoring using dysphonia measures adopted from phonations. Tsanas *et al.* [16] predicted the tendency of PD progression by means of speech data utilization. They evaluated 6000 samples of 42 PWP using signal processing algorithms for feature extraction. They identified helpful features. Regression and classification techniques were utilized for voice features projection to a unified PD rating scale. In another study, Tsanas *et al.* [13] selected dysphonia measure subsets by means of a Support Vector Machine (SVM) classifier.

2.2. Fuzzy Data Mining Methods

Several studies have used fuzzy classification or clustering for data analysis because most real-world data encapsulate indisputable amounts of inexactness and ambiguity. Chakraborty *et al.* [15] utilized fuzzy c-mean clustering and subtractive clustering methods for PD detection. According to the reported results, their detection accuracy is about 96-97% with reasonable efficiency. Additionally, it was found that the accuracy of FIS based on Fuzzy C-Means (FCM) is higher comparable to the other methods. An efficient and effective PD diagnosis method is presented in [25]. The method described in [25] is a fuzzy k-nearest neighbor approach. PD progression is assessed using a fuzzy inference system and artificial neural networks which are based on the adaptive networks described in [20]. Parisi *et al.* [18] used

artificial neural networks and adaptive neuro fuzzy classifier for PD recognition from sustained phonation tests. They developed a fuzzy expert system for PD diagnosis from speech tests. Gürüler [21] diagnosed PD using an artificial neural network that has a complex value with k-means clustering feature weighting.

To the best of our knowledge, the dataset used in our paper was only utilized by Sakar *et al.* [1] for PD diagnosis based merely on classical classification techniques. For the study described in this paper, we have applied various other classification methods which combine fuzziness in the process and hence produced more interesting results related to PD diagnosis.

3. MATERIALS AND METHODS

3.1. The Dataset

Our target in this study is to distinguish PWP from healthy people using fuzzy classifiers based methods. The dataset used here has been adopted from [1]. It consists of speech test of 40 persons, 20 of whom are PWP (6 females, 14 males), and the rest 20 are healthy people (10 females, 10 males). Tests were taken in the Neurology Department in Cerrahpasa School of Medicine at Istanbul University. Patients are mainly people who suffered from PD for a period from 0 to 6 years. The age of PWP patients who were included in the tests varied between 43 and 77 years with an average of 64.86 and a standard deviation of 8.97, while the age of healthy people ranges between 45 and 83 years, with an average of 62.55 and a standard deviation of 10.79. Twenty-six voice samples were recorded from all the subjects. These voices include words, short sentences, sustained vowels and numbers. As a result of this test, a set of exercises of speeches were prepared. The considered samples opted for a more powerful sound of people with parkinsonism [22]. Samples were selected by a group of neurologists.

As mentioned in [1], the frequency range of recording devices is 50Hz- 13kHz, and their model is Trust MC- 1500. For taking voice samples, the device frequency was set to 96 kHz, 30 dB and it was placed 10 cm away from the subjects. Recordings were carried while subjects were reading or repeating the prepared texts. The subjects were asked to count from 1 to 10, read 4 rhymed sentences, read 9 words and say 'a', 'o', 'u' letters. According to [1], 26 features were extracted from these voice tests. These features have been distinguished into different categories of parameters: (1) frequency parameters, including various types of jitter, (2) pulse parameters, including features related to the number of pulses and periods existing in voice signals, (3) amplitude parameters, including different shimmer versions, (4) voicing parameters, (5) pitch parameters, and (6) harmonicity parameters.

3.2. FUZZY LOGIC

The fuzzy set theory introduced by Lotfi Zadeh in 1963 is an effective mechanism for managing imprecision in problems which deal with decision making, including instability and ambiguity of real-world applications [23]. Actually, fuzzy inference can be defined as the procedure of projecting a specific input to an output dataset by means of the fuzzy set theory where knowledge is codified by unequivocal lin-

guistic rules utilization which is understandable by people who do not have technical proficiency.

The main characteristic which differentiates fuzzy logic from Boolean logic is that fuzzy logic assumes a degree for a fact to be true or false, while according to Boolean logic each fact is either completely true or completely false. In other words, there is a smooth transition from completely true to completely false. In real life scenarios, most of the cases cannot be classified to belong to one of the two extremes. They have a degree of membership in each of the given sets. This degree ranges from 0 to 1, inclusive.

As depicted in Fig. (1), the primary segments of fuzzy logic are: (1) *Fuzzification* which is the process of translating crisp inputs into fuzzy values, (2) *Rule base reasoning* which is the process of applying a fuzzy reasoning mechanism to get a fuzzy yield by fuzzy rule utilization, and (3) *Defuzzification* which is the process of translating the latter output into a crisp value.

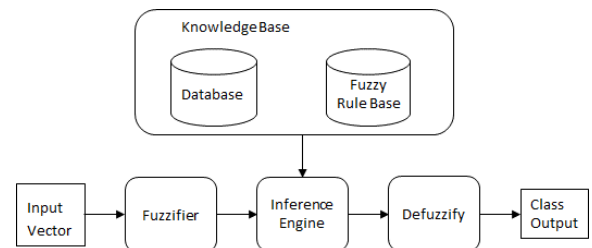


Fig. (1). Modified diagram of a Mamdani-FRBS (Fuzzy Rule Based System) [26].

The objective of fuzzification is the projection of system input values to the interval $[0, 1]$. This projection is realized by defining membership functions which produce membership degrees for the inputs. Rule-based reasoning is the process of projecting fuzzy membership degrees of fuzzy inputs in order to classify the output by utilizing if-then rules, which are presented as logical expressions, *i.e.*, if p then q . Here p is the antecedent and q is the consequent of the rule [27]. Defuzzification produces a single system output by applying a defuzzification formula on a fuzzy output.

3.3. Fuzzification

The main purpose of fuzzification is selecting a membership function for transforming numerical valued inputs into corresponding membership values. Every input value may have one membership value corresponding to each linguistic set. The input is always characterized as a crisp value which is restricted by cosmos of the input variable and the output is a fuzzy membership degree in the appropriate linguistic set(s). Membership function μ_B of a fuzzy set B is defined as:

$\mu_B: \chi \rightarrow [0, 1]$, where χ is the universal set. Here, μ_B may be triangular, trapezoidal, Gaussian function, *etc.*

3.4. Rule-Based Reasoning

According to a study [28] “fuzzy sets are an aid in providing symbolic knowledge information in a more human understandable or natural form. They can hold uncertainty at various levels”. The steps of the fuzzy rule-based system

manipulation are: (1) proper if-then fuzzy rules derivation, (2) dividing the universe into parts, and (3) deciding on membership functions for the mapping. Linguistic rules are a form of fuzzy rules generated by experts [17]. The connection between antecedents and consequents in the rules is accomplished by If-Then rules which are generated using linguistic variables. An antecedent can be defined as a fuzzy clause which has a certain degree of membership in the interval $[0,1]$. It is possible for a fuzzy rule to have more than one component in the antecedent connected using "and"/"or" operators, where all parts are considered at the same time and transformed into a single number. The same is true for the consequent. Several parts of the rule consequent are gathered into a single output of a fuzzy set [29].

An adaptive Neuro Fuzzy Inference System (ANFIS) can be expressed as a combination of neural networks and fuzzy logic principles. Advantages offered by ANFIS are: (1) smoothness, which is a property of the fuzzy principle, and (2) adaptability which comes from the neural networks training structure [30].

3.5. Defuzzification

Defuzzification is a process in which a quantifiable result in crisp logic is produced for a given fuzzy set and the corresponding membership degrees. Max-membership, max-max, centroid method, center of largest area, and center of sums are some defuzzification approaches.

4. THE DEVELOPED APPROACH

As mentioned previously, the objective of this study is to distinguish PWP from healthy individuals using a variety of fuzzy classification methods. For further evaluation, the classification methods available in WEKA have been utilized. The techniques employed in this study are described next in this section.

4.1. Neuro-Fuzzy Classification

This method relies on the neuro-fuzzy classifier introduced by Jang [31], *i.e.*, the scaled conjugate gradient method (SCG), where adaptive networks are employed for solving a classification problem. In this method, system parameters, including membership functions, are specified for each feature. Parametrized t-norms which are used for combining conjunctive conditions are calibrated using the backpropagation method. Parameter optimization and rule weights used in this study are different from the ones used by Sakar *et al.* [31]. The adaptation of rule weights is done using a number of rule samples. The main objective of utilizing the suggested algorithm is determining optimal values for nonlinear parameters. The reason for SGD utilization is that it is faster in comparison to some second order derivative based and steepest descent methods. It is appropriate for large scale problems [32].

The SCG method is categorized under supervised learning methods for feedforward neural networks. This method belongs to the category of conjugate gradient approaches. Conjugate gradient methods (CGMs) are known as a type of second order approaches which help in minimizing the goal function of diverse variables based on a valid theoretical foundation. Second order methods are named so because

they utilize the second derivatives of the goal function, while the first order methods use first derivatives. Utilizing second order derivative methods may have some pros and cons. They are advantageous in finding a better way to a minimum in comparison to that of first order derivative techniques. Furthermore, they usually run at a higher computational cost. The logic here is similar to that of standard backpropagation, *i.e.*, CGMs try to approach the minimum at each iteration. The difference between backpropagation and CGMs comes from the direction of movement. In backpropagation, the direction is down the gradient of the error function, while in CGMs the direction is towards a vector which is conjugate to the movement direction of the former steps. As a result of this movement strategy, minimization at any step is not undone by the operations of a subsequent step.

Let $p = \exp\left(\frac{-\Delta E}{T}\right)$ be a vector from space $T = T.\text{deg}$; and assume that E is the error function to be minimized. The differences between SCG and CGMs are listed below:

- In each iteration k , w_i is computed using CGM. Here, R^N is defined as a new direction which is the conjugate to the direction of the previous step, N represents the sum of the number of weights and the number of biases constituting the network. The step size in this direction is determined by $w_{k+1} = w_i + \alpha_k \cdot p_k$. Here, parameter p_k is a function of α_k . Parameter α_k is the Hessian matrix of the error function. SCG uses a simple approximation method to calculate the term $E''(w_k)$, which is an essential factor in the computation of $\alpha_k: s_k$.
- As Hessian is not always positive definite, which prevents the algorithm from achieving good performance, SCG uses a scalar α_k which is supposed to regulate the indefiniteness of Hessian. It is a kind of Levenberg-Marquardt method and is done by setting s_k as follows:

$$s_k = E''(w_k) \cdot p_k \cong \frac{E'(w_k + \sigma_k \cdot p_k) - E'(w_k)}{\sigma_k}, \quad 0 < \sigma_k \ll 1$$

and regulating λ_k at each iteration. This last point can be considered as the main contribution of SCG to the neural learning and optimization theory. SCG has been proved to be faster than the standard backpropagation and other CGMs.

Here, fuzzy rules are initialized using the k-means algorithm and fuzzy set descriptions are derived by the Gaussian membership function.

In addition to the first method presented above, another version of SCG based classifier is described. Here, the Least squares estimation approaches are utilized. They are used for estimating the gradient without utilization of all the training samples. In the third variation, linguistic hedges are adapted by the SCG algorithm. They are applied to the sets of rules which are fuzzified. In this manner, using power values emphasizes some particular features and damps. Linguistic hedge utilization increases the recognition rates.

4.2. Adaptive Neuro-Fuzzy Classification Method

The proposed method has been coded using the fuzzy logic toolbox available in MATLAB. The first step involves

creating the fuzzy inference system using three different methods: (1) subtractive clustering, (2) fuzzy c-means clustering, and (3) grid partitioning. Then, the generated system is trained by the Adaptive Neuro Fuzzy Inference System (ANFIS).

Each data point is considered as a potential center for a cluster in a subtractive clustering approach. The likelihood of each data point being the cluster center is calculated according to the density of the neighboring data points. The steps of the algorithm are described next:

- Data points with the highest possibility to be considered as the center of the first cluster are selected.
- Subsequently, the data cluster and the location of its center is determined by removing the neighbors of the first cluster center. These neighbors are determined by the radii value.
- The process is repeated until no data points are out of the cluster center radii.

GENFIS2 function is part of the fuzzy inference system toolbox in MATLAB. The mentioned function generates a Sugeno type fuzzy inference system. The calculations are done based on the subtractive clustering method. The first step is determining rules and antecedent membership function numbers using the subclust function. Then, consequent equations of each rule are estimated by least squares method. The mentioned function yields a fuzzy inference system structure which contains a set of fuzzy rules for covering the feature space.

Mostly, a grid partitioning method is used in designing a fuzzy controller. In general, the only variables involved in this process are state variables which are fed to the controller as input values. As a strategy, for each input, a small number of membership functions is required by the partitioning operation. This method faces some problems when the number of inputs is relatively large. For example, in a fuzzy model which has 10 inputs and two membership functions, each input would result in 2^{10} fuzzy if-then rules, which is prohibitively large. This problem is referred to as the *curse of dimensionality*.

In MATLAB, genfis1 is used to create a single-output Sugeno-type fuzzy inference system based on grid partitioning.

Fuzzy C- Means (FCM) is a data clustering method in which the dataset under study is grouped into n clusters. Data points in the dataset belong to every cluster with a certain degree ranging from 0 to 1, inclusive. For instance, a specific data point which is close to the center of a cluster will have a high degree of membership in the cluster and another data point which is at a distance far from the cluster center will have a lower degree of membership in the cluster.

In MATLAB, genfis3 creates a fuzzy inference system using fuzzy c-means clustering by extracting a set of rules

which can model the behavior of the data.

4.3. Inductive Fuzzy Classification

Inductive Fuzzy Classification with Normalized Likelihood Ratios (IFCNLR) was first suggested by Kaufmann in 2009 [33]. As mentioned in [34], the method has been implemented to WEKA; it can be found under the supervised attribute filters.

According to [33] “IFC is introduced as inducing membership functions to fuzzy classes and assigning individuals to those classes”. Any function capable of mapping crisp valued input into output data in the range 0 and 1, inclusive, can be considered as a membership function. Inductive fuzzy classification can be described as the procedure of assigning individuals to fuzzy sets, and inductive inference is the basis of generated membership functions for these fuzzy sets. The process for multivariate inductive fuzzy class induction is shown in Fig. (1). Modified Diagram of a Mamdani-FRBS (Fuzzy Rule Based System) [26].

Fig. (2). The proposed procedure starts with data preparation. Next, univariate membership functions should be induced for the attributes. Univariate membership functions can be induced in two ways: namely supervised and unsupervised induction. In this study, the first method is utilized. In this method, induction is done based on a target variable. It is required to normalize the differences and ratios when trying to obtain the membership functions for inductive fuzzy classes for a variable according to the distribution of a second variable. For instance, a membership degree to an inductive fuzzy class can be represented as a normalized likelihood ratio.

IFC-NLR is based on the transformation of the inductive support of the target class membership into a membership function with a specific property, *i.e.*, the membership degree of I in class y' increases as the likelihood of $i \in y$ goes higher compared to $i \notin y$. For instance, the calculation of the membership degree of a given attribute X in predictive class y' is done using the normalized likelihood ratio function. The calculated membership degree (x) should be in the domain of X , *i.e.*, $x \in \text{dom}(X)$. The derivation of the normalized likelihood ratio function is done based on the probability of the target class membership. In fact, the relation between all values in $\text{dom}(X)$ and their likelihood ratios in a normalized form constitutes the resulting membership function.

According to the principle of likelihood for a pair of incompatible hypotheses H_1 and H_2 , evidence E supports H_1 over H_2 , if and only if $p(E|H_1) > p(E|H_2)$. The likelihood ratio (LR) measures the strength of evidence for H_1 over H_2 [3]:

$$LR(H_1 > H_2 | E) := \frac{L(H_2|E)}{L(H_1|E)}$$

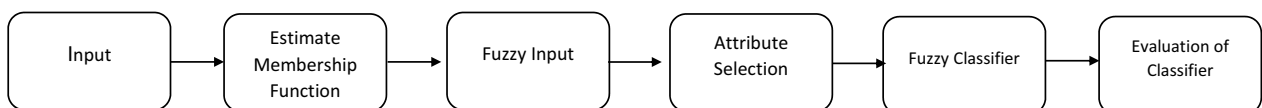


Fig. (2). Procedure of inducing multivariate inductive fuzzy class [33].

The epistemological problem of induction can be solved by utilizing the likelihood ratio. In this case, the likelihood ratios are used as the measures of support for inductive inferences. If the data contains fuzziness, the likelihood can be calculated as:

$$L(\tilde{H}|\tilde{E}) = p(\tilde{E}|\tilde{H}) = \frac{\sum_{i=1}^n \mu_{\tilde{H} \cap \tilde{E}}(i)}{\sum_{i=1}^n \mu_{\tilde{H}}(i)}$$

In the likelihood principle stated above, the index of the degree of support for the conclusion of $ii \in yy$ is the ratio between the two probabilities, while the evidence of $(ii) = xx$ is given. The transformation of the likelihood ratios into a fuzzy set membership function necessitates a normalization in the range [0, 1]. Fortunately, for each ratio defined as $R = A/B$, a normalization of the form $N = A/(A+B)$ exists and it has the following properties:

- An R value close to 0 results in an N value near 0.
- An R value equal to 1 results in an N value equal to 0.5.
- A large R value results in an N value close to 1.

The NLR function is derived by applying the previously mentioned normalization to the calculated likelihood ratios. Accordingly, the membership degree of an attribute like x in the target class prediction y' can be expressed by utilizing the corresponding NLR given below:

$$\mu_{y'} = \text{NLR}(y|x) = \frac{L(y|x)}{L(y|x) + L(-y|x)}$$

Once univariate membership functions are induced, attribute values should be transformed into univariate target membership degrees. Classification is done by aggregating fuzzified attributes into a multivariate fuzzy classification and the last step is the evaluation of the predictive performance of the resulting model.

The idea in this procedure is to build up a fuzzy classification which gradually positions an inductive membership individual i in target class y. The multivariate model $\mu_{y'}$ is utilized by the mentioned fuzzy classification to accomplish the procedure of allocating individuals with an inductive membership degree in the predictive inductive fuzzy class y'. The inductive support for class membership in target class y depends on the inductive degree of $\mu_{y'}(i)$ of an individual in y'. The increase in $\mu_{y'}(i)$ results in an increase in the inductive support for class membership in the mentioned target class.

4.4. Fuzzy Rough Classification

4.4.1. Rough Set Theory

Rough Set Theory (RST) [35] makes it possible to accurately adopt knowledge from a domain. Furthermore, RST provides a tool that can keep up data content while decreasing the amount of information included. Indiscernibility may be expressed as the most critical point in rough set theory. Let (U, A) represent a data system in which U is a non-discharge set of limited objects (cosmos), and let A be a non-exhaust limited arrangement of attributes with the end goal that $a: U \rightarrow V_a$ for each $a \in A$. V_a is the set of qualities that attribute a may take. With any $B \subseteq A$, there is a related proportionality connection R_B :

$$R_B = \{(x, y) \in U^2 \mid \forall a \in B, a(x) = a(y)\}$$

According to the definition given above, for any $(x, y) \in R_B$, x and y are said to be indiscernible by attributes from B. The representation of B- discernibility relation can be illustrated as $[x]_B$. Assuming that $A \subseteq U$, the development of B-upper and B-lower approximations for A makes it possible to use the information in B for approximations about A:

$$R_B \downarrow A = \{x \in U \mid [x]_B \subseteq A\}$$

$$R_B \uparrow A = \{x \in U \mid [x]_B \cap A \neq \emptyset\}$$

A rough set can be described as a tuple $\langle R_B \downarrow A, R_B \uparrow A \rangle$.

A decision system $(X, \mathcal{A} \cup \{d\})$ can be defined as a special kind of information system used in the context of classification, where d ($d \notin \mathcal{A}$) is an assigned attribute called the decision attribute. Decision classes of the decision attribute are defined as its equivalence classes $[x]_{R_d}$. For $B \subseteq \mathcal{A}$, the region which includes objects in X, where B values allow the prediction of the decision classes unambiguously, is included in the B- positive region denoted by POS_B :

$$POS_B = \bigcup_{x \in X} R_B \downarrow [x]_{R_d}$$

In fact, whenever $x \in POS_B$ holds, it can be claimed that at any point a given object (which has the same values as x for the attributes included in B) belongs to the same decision class as x. The following formula is used to measure the predictive ability of the attributes included in B:

$$\gamma_B = \frac{|POS_B|}{|X|}$$

In case that $\gamma_A = 1$, $(X, \mathcal{A} \cup \{d\})$ is called consistent. A decision reduct for a subset B of \mathcal{A} occurs when it satisfies the equation $POS_B = POS_A$. This equation means that the decision-making power of \mathcal{A} is preserved by B and no further reduction is possible. The further reduction is the existence of a proper subset B' of B such that $POS_{B'} = POS_A$. B is called a decision superreduct if the latter constraint (which states that it is essential for B) to be minimal is lifted.

4.4.2. Fuzzy Sets Theory

The fuzzy sets theory [36] makes it possible for objects to belong to a single set or multiple sets with a given degree. Remember that a fuzzy set X can be defined with values transformed into membership degrees in the interval [0,1] and the fuzzy relation in X is defined as a fuzzy set in $X \times X$. The R- foreset of y is the fuzzy set denoted by R_y . It is defined as below for all y in X:

$$R_y(x) = R(x, y) \quad \forall x \in X$$

In case R is a reflexive and symmetric fuzzy relation, we have:

$$R(x, x) = 1$$

$$R(x, y) = R(y, x)$$

Which hold for all x and y in X. For this condition, a fuzzy tolerance relation is defined by R. Given a fuzzy tolerance relation R, the fuzzy tolerance class of y is denoted by R_y ; for fuzzy sets A and B in X, $A \subseteq B \Leftrightarrow (\forall x \in X)(A(x) \leq B(x))$. For a finite X, $|A|$ is defined below:

$$|A| = \sum_{x \in X} A(x)$$

The role of fuzzy logic connectives in the development of the fuzzy rough set theory is inevitable. Thus, we review here some vital definitions. A triangular norm (t-norm for short) T is defined as an increasing, commutative and associative $[0,1]^2 \rightarrow [0,1]$ mapping which satisfies the relation $T(1, x) = x$, for all x in $[0,1]$.

In this paper, T_M and T_L are used. These are defined as follows: $T_M(x, y) = \min(x, y)$ and $T_L(x, y) = \max(0, x + y - 1)$ (Lukasiewicz t-norm), for $x, y \in [0, 1]$. On the other hand, an implicator is described as any transformation in the form $[0,1]^2 \rightarrow [0,1]$ which maps T and satisfies $T(0, 0) = 1$, $T(1, x) = x$, for all $x \in [0,1]$. Additionally, T must be decreasing in its first, and increasing in its second component.

4.4.3. Theory of the Fuzzy- Rough Sets

The procedure illustrated above can be effective when the considered datasets contain discrete values. In case the dataset includes real valued attributes, a discretization operation is required hereafter. Modelling the approximate equality among objects with continuous attribute values by utilizing fuzzy relation R in U is a more natural and adaptable approach. Here, U is defined as a 0-1 projection that assigns object's degree of similarity to them.

5. NUMERICAL RESULTS AND ANALYSIS

To test the effectiveness and applicability of the proposed framework for distinguishing PWP from healthy people, we utilized a dataset of speech tests taken from 40 subjects [1]. The dataset consists of a total of 1040 instances and 26 features.

5.1. Data Preprocessing

Before applying the fuzzy classification algorithms described in Section 5, the data was cleaned from outliers and extreme values using WEKA. Interquartile Range unsupervised filter was applied to the dataset. The raw dataset contained 189 outliers and 62 extreme values in addition to some outliers which were removed using the remove with values filter. This process decreased the number of instances to 817.

5.2. WEKA Classification Results and Analysis

The classification process was conducted using WEKA. Three types of model validations have been considered: (1) the k-fold cross validation technique with $k=10$, (2) leave-one-out (LOO) cross validation, and (3) validation by dividing the dataset into a 70 to 30 percent training and test datasets. The metrics used for evaluating the performance of the classifiers are accuracy, sensitivity, specificity, and Matthew's Correlation Coefficient (MCC).

The model validation methods used in this paper are described next.

- In 10-fold cross validation, the dataset is randomly divided into 10 parts and each time one of the parts is left as the test set while the other nine parts are used together as the training set.

- Leave-one-out cross validation requires that the number of folds is equal to the number of instances (817 here). In this method, data is divided into n random parts. Each time, the model is trained by a combination of $n-1$ parts forming the training set and the model is tested using one part as the test set.
- In the third method, the model is trained with 70% of the dataset and tested using the left 30%.

We used four evaluators to assess the performance of the classification models employed in this study. These evaluation methods are described next based on prediction versus the actual observation regarding whether instances belong to a given class C or not. An instance which is predicted and observed in class C is classified as a True Positive (TP) case. An instance which is not predicted but observed in class C is classified as a False Positive (FP) case. An instance which is predicted but not observed in class C is classified as a False Negative (FN) case. An instance which is neither predicted nor observed in class C is classified as a True Negative (TN) case.

- Accuracy is defined as the rate of the instances correctly classified, $\frac{TP+TN}{(TP+FP+FN+TN)}$
- Sensitivity (alternatively called true positive rate, recall, or probability of detection) measures the ratio $\frac{TP}{(TP+FN)}$
- Specificity (alternatively called true negative rate) is computed as $\frac{TN}{(TN+FP)}$
- The correlation coefficient between the predicted and the observed binary classifications is known as Matthew's Correlation Coefficient (MCC). Values for MCC are in the interval $[-1,+1]$. As for any correlation coefficient, +1 means perfect prediction, 0 represents random prediction and -1 indicates total disagreement between the observed and the predicted cases.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Six classifiers have been utilized from WEKA 3.8 to differentiate PWP from healthy case in the given dataset. The obtained results are presented in this section. Table 1 reports result from the third validation method.

As reported in Table 1, nearly 3 out of the 6 classifiers performed efficiently and classified the data with 100% accuracy. The MCC value for MLP, SGD and Bagging is 1; this indicated perfect prediction. IBK reported the worst performance by classifying only 84% of the data correctly. It has been applied with its default parameter in WEKA, *i.e.*, $k=1$.

Classifier results for 10-fold cross validation are presented in Table 2, where it can be obviously seen that nearly 3 out of 6 classifiers performed efficiently and classified the data with 100% accuracy. Again, IBK reported the worst performance by classifying only 87% of the data correctly. Classification results from LOO cross-validation are presented in Table 3. Also, here nearly 3 out of the 6 classifiers performed efficiently and classified the data with 100% accuracy. Again, IBK reported the worst performance by classifying only 87% of the data correctly.

Table 1. WEKA classification results.

-	Accuracy	Sensitivity	Specificity	MCC
MultiLayer Perceptron	1.00	1.00	1.00	1.00
SGD	1.00	1.00	1.00	1.00
SMO	0.92	0.85	1.00	0.85
Voted Perceptron	0.87	0.80	0.96	0.76
IBK	0.84	0.85	0.82	0.67
Bagging	1.00	1.00	1.00	1.00

Table 2. WEKA classification with 10-fold cross validation.

-	Accuracy	Sensitivity	Specificity	MCC
MultiLayer Perceptron	1.00	1.00	1.00	1.00
SGD	1.00	1.00	1.00	1.00
SMO	0.92	0.85	1.00	0.85
Voted Perceptron	0.88	0.83	0.94	0.77
IBK	0.87	0.88	0.86	0.74
Bagging	1.00	1.00	1.00	1.00

Table 3. WEKA classification with LOO cross validation.

-	Accuracy	Sensitivity	Specificity	MCC
MultiLayer Perceptron	1.00	1.00	1.00	1.00
SGD	1.00	1.00	1.00	1.00
SMO	0.92	0.85	1.00	0.85
Voted Perceptron	0.89	0.82	0.96	0.78
IBK	0.87	0.89	0.85	0.74
Bagging	1.00	1.00	1.00	1.00

5.2.1. Fuzzy Rough Results Analysis

Fuzzy rough classification was done using the fuzzy filter for classification in WEKA. Model validation and performance evaluation were conducted with the same methods and metrics. The results are presented in this section. The methods used under this category are summarized below:

- Discernibility Classifier
- FLR (Fuzzy Lattice Reasoning)
- FURIA (Fuzzy Unordered Rule Induction Algorithm)
- Fuzzy NN (Fuzzy Rough k nearest neighborhood)
- Ownership Nearest Neighborhood
- Fuzzy Rough NN
- NN

- OWANN (Ordering Weighted Averaging Nearest Neighbor)
- QSBA (Quantified Subsethood Based Approach)
- Quick Rules
- VQ NN (Vector Quantization K- nearest Neighborhood)
- VQ Rules (Vector Quantization Rules)

Fuzzy rough classification results for the third method are presented next in this section where it can be obviously seen that nearly 2 out of the 12 classifiers performed efficiently and classified the data with 100% accuracy. The QSBA classifier reported the worst performance by classifying only 65% of the data correctly. The best classifier under this validation method is FURIA. Table 4 reports Fuzzy rough classification results. The 10-fold cross-validated results from the fuzzy rough classifiers are presented in Table 5.

It can be easily seen from Table 5 that FURIA and FLR reported 100% accuracy. QSBA showed the worst performance by reporting 68% accuracy. Considering the 10-fold cross-validation results, the best classifier is FURIA. Finally, the results from the LOO validation model are presented in Table 6. Same as reported by the previous two validation methods, FURIA again outperformed the other classifiers and the worst classifier is QSBA.

5.2.2. Adaptive Neuro Fuzzy Classification (ANFC) Results Analysis

Adaptive neuro fuzzy classification was applied using MATLAB 2014b. The data used here is an ordered data which was shuffled to avoid feeding ordered class values to

the classifier. The first step in adaptive neuro fuzzy classification is generating a fuzzy inference system. MATLAB fuzzy logic toolbox has three types of ‘GENFIS’ functions which generate fuzzy inference systems. Genfis1 is used to generate fuzzy inference systems using grid partitioning on data. The fuzzy membership function used here is ‘PIMF’, and three membership functions have been generated. This function generates a huge number of rules which are not all necessary. Thus, applying genfis1 to the dataset with high dimensionality results in memory problems. To overcome this, attribute selection has been used to reduce the number of attributes in the dataset. Dataset dimensionality was reduced to three using the cfsSubsetEvaluation method in WEKA.

Table 4. Fuzzy rough classification results.

-	Accuracy	Sensitivity	Specificity	MCC
Discernibility Classifier	0.88	0.84	0.93	0.72
FLR	1.00	1.00	0.99	0.98
FURIA	1.00	1.00	1.00	1.00
FuzzyNN	0.79	0.71	0.88	0.53
Fuzzy Ownership NN	0.86	0.83	0.88	0.65
FuzzyRough NN	0.77	0.77	0.77	0.44
NN	0.86	0.83	0.89	0.66
OWANN	0.87	0.85	0.89	0.68
QSBA	0.65	0.81	0.45	0.17
QuickRules	0.93	0.92	0.95	0.82
VQNN	0.86	0.83	0.89	0.66
VQRules	0.93	0.89	0.97	0.85

Table 5. Fuzzy rough classification with 10-fold cross validation.

-	Accuracy	Sensitivity	Specificity	MCC
Discernibility Classifier	0.89	0.86	0.92	0.73
FLR	1.00	1.00	0.99	0.98
FURIA	1.00	1.00	1.00	1.00
FuzzyNN	0.82	0.80	0.85	0.56
Fuzzy Ownership NN	0.85	0.84	0.86	0.62
FuzzyRough NN	0.83	0.84	0.81	0.55
NN	0.88	0.87	0.89	0.68
OWANN	0.88	0.86	0.89	0.68
QSBA	0.68	0.79	0.57	0.23
QuickRules	0.94	0.92	0.96	0.85
VQNN	0.88	0.87	0.89	0.68
VQRules	0.94	0.92	0.95	0.83

Table 6. Fuzzy rough classification with LOO cross validation.

-	Accuracy	Sensitivity	Specificity	MCC
Discernibility Classifier	0.90	0.87	0.94	0.76
FLR	1.00	1.00	0.99	0.98
FURIA	1.00	1.00	1.00	1.00
FuzzyNN	0.83	0.80	0.86	0.58
Fuzzy Ownership NN	0.86	0.84	0.88	0.65
FuzzyRough NN	0.83	0.84	0.83	0.56
NN	0.89	0.88	0.90	0.71
OWANN	0.88	0.87	0.90	0.70
QSBA	0.69	0.80	0.57	0.24
QuickRules	0.94	0.92	0.95	0.83
VQNN	0.89	0.88	0.90	0.71
VQRules	0.93	0.93	0.92	0.95

Table 7. Results for ANFC using genfis2 and genfis3.

-	Accuracy	Sensitivity	Specificity	MCC
genfis2	0.97	0.98	0.96	0.90
genfis3	0.85	0.96	0.76	0.53

Table 8. ANFC results with reduced dimension.

-	Accuracy	Sensitivity	Specificity	MCC
genfis1	1.00	1.00	1.00	1.00
genfis2	1.00	1.00	1.00	1.00
genfis3	1.00	1.00	1.00	1.00

The second function used for fuzzy inference system generation is ‘Genfis2’. It utilizes subtractive clustering for fuzzy inference system generation. The membership function utilized here is ‘GAUSSMF’. The parameters needed for genfis2 are input and output matrices, and a vector specifying the range of influence for the cluster centers in each data dimension, and an optional matrix specifying the mapping of data from the input and the output into a unit hyperbox.

The third function used for fuzzy inference system generation is ‘GENFIS3’. It generates the fuzzy inference system using a fuzzy c-means clustering method.

After a Fuzzy Inference System (FIS) is generated, an adaptive neuro-fuzzy inference system is used to train the model based on neural networks. In ANFIS, a hybrid learning algorithm is utilized to tune the parameters of a Sugeno type Fuzzy Inference System (FIS). A combination of back propagation gradient descent and least squares methods is utilized to model a training dataset. Furthermore, the model is validated by ANFIS using a checking dataset to test overfitting in the training data.

To use k-fold cross-validation in MATLAB, the dataset was partitioned using the ‘CVPARTITION’ command which divides the input data into k equal partitions. The same validation methods were utilized, and the classifiers were evaluated based on the previously mentioned evaluators. The results from these three classifiers are presented next.

As it is obvious from the results reported in Table 7, the model with genfis2 outperformed the model with genfis3. The accuracy of the first model is 97%, while the second model classified 85% of the data correctly. To compare the performance of genfis1 with the other two classifiers, all of the three classifiers were run with the selected three features. The results are presented in Table 8. All the three models have classified the data with 100% accuracy. The cross validated model results are presented in Table 9. It can be easily seen from the reported in Table 9 that genfis2 outperformed genfis1. The former method reported 96% accuracy while the latter method reported 100% accuracy. In Table 10, the worst performance belongs to genfis3. All the evaluation metrics for genfis3 are lower their counterparts for the other two classifiers. Results of the LOO cross-validation method are presented in Tables 11 and 12. All of the classifiers reported 100% accuracy under the LOO cross validation.

Table 9. ANFC results with 10- fold cross validation.

-	Accuracy	Sensitivity	Specificity	MCC
genfis1	0.96	1.00	0.93	0.86
genfis2	1.00	1.00	1.00	1.00

Table 10. ANFC results with 10- fold cross validation and reduced dimension.

-	Accuracy	Sensitivity	Specificity	MCC
genfis1	1.00	1.00	1.00	1.00
genfis2	1.00	1.00	1.00	1.00
genfis3	0.95	1.00	0.91	0.81

Table 11. ANFC results with LOO cross validation.

-	Accuracy	Sensitivity	Specificity	MCC
genfis2	1.00	1.00	1.00	1.00
genfis3	1.00	1.00	1.00	1.00

Table 12. ANFC results with LOO cross validation and reduced dimension.

-	Accuracy	Sensitivity	Specificity	MCC
genfis1	1.00	1.00	1.00	1.00
genfis2	1.00	1.00	1.00	1.00
genfis3	1.00	1.00	1.00	1.00

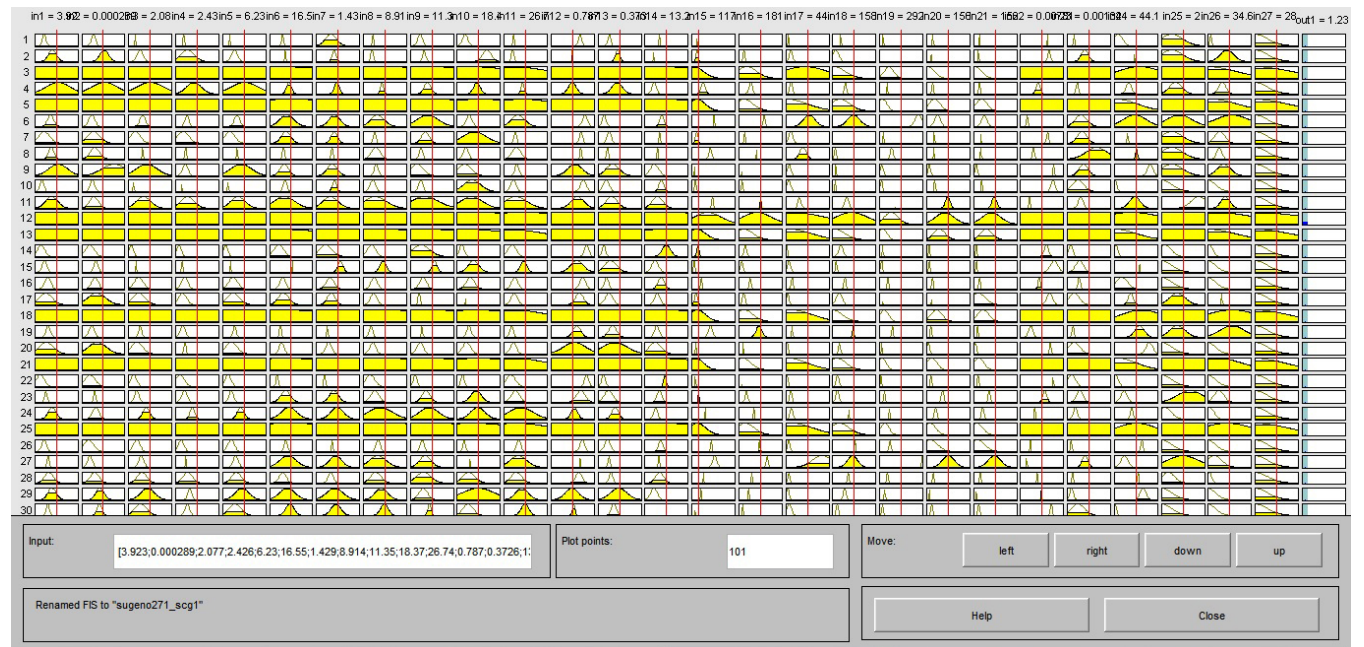


Fig. (3). Membership functions for scg1.

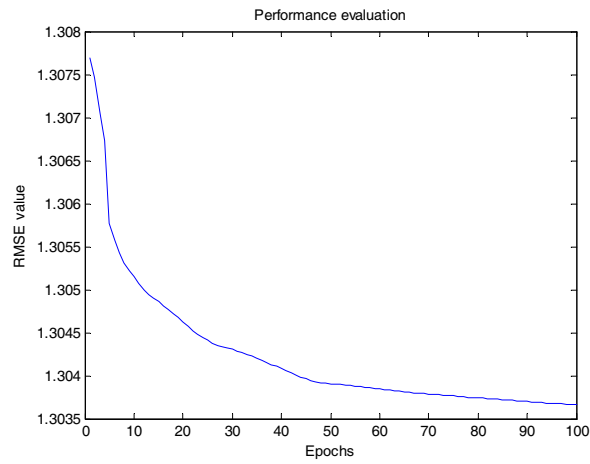


Fig. (4). Performance evaluation against RMSE for scg1.

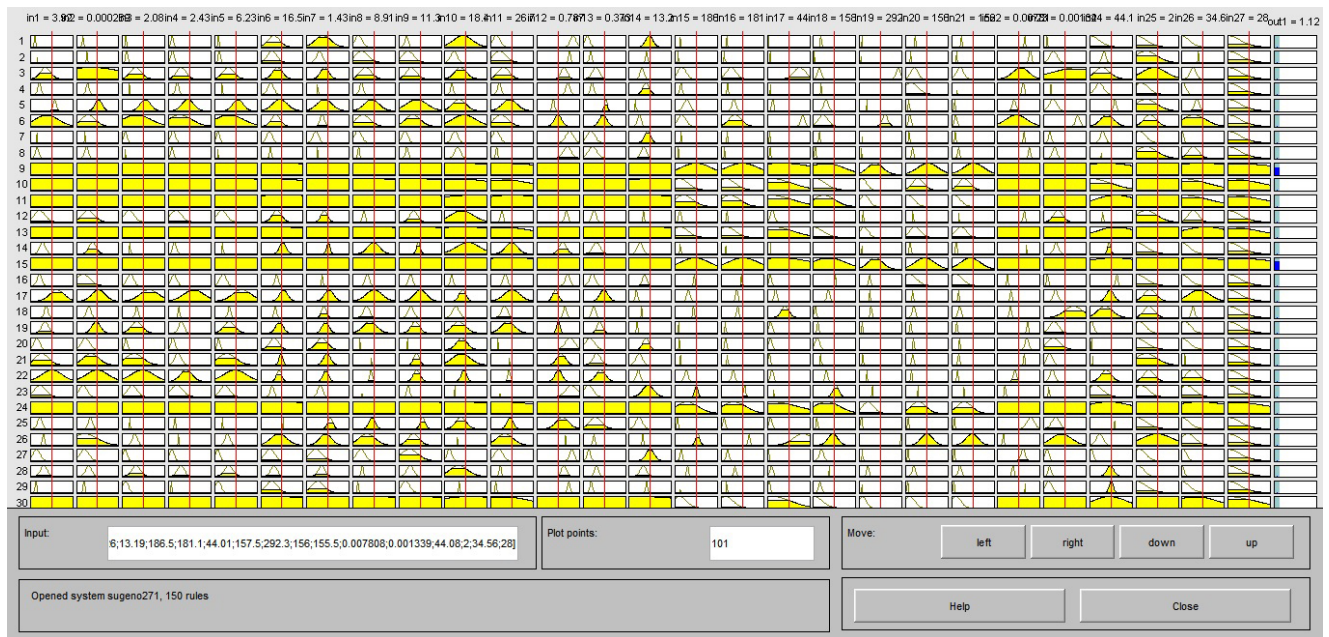


Fig. (5). Membership functions for scg2. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

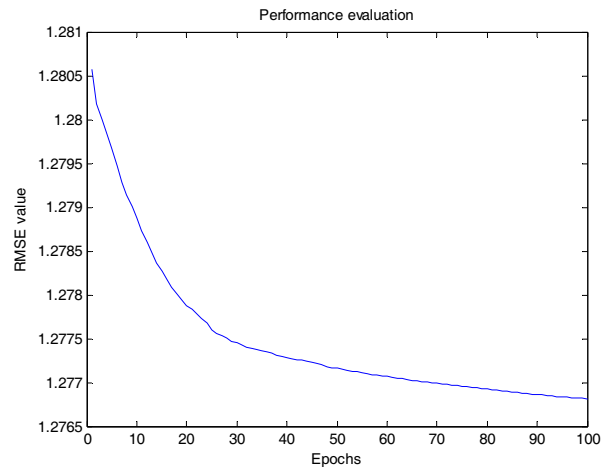


Fig. (6). Performance evaluation against RMSE for scg2.



Fig. (7). Membership functions for scg3.

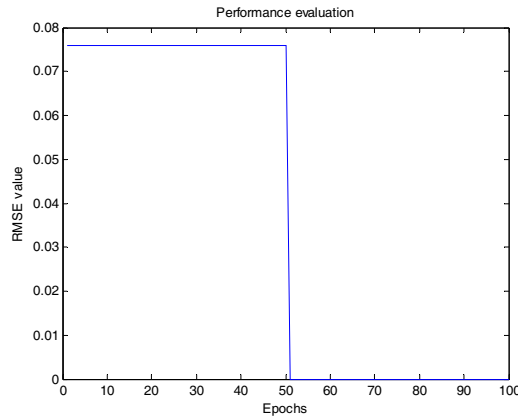


Fig. (8). Performance evaluation against RMSE for scg3.

5.2.3. SCG-NFC Method Results Analysis

As mentioned previously, three methods of scg- nfc have been applied to the dataset utilized in this study. The results are presented next in this section. The model evaluation has been done based on the accuracy measure and the model was validated by dividing the dataset based on a 0.7 to 0.3 ratio for the training and testing, respectively.

The results have been produced by considering 75 clusters using k-means clustering for each of scg1 and scg2, and 3 clusters for scg3. The number of clusters and epochs have been chosen based on a trial and error method. The membership functions (Figs. 3, 5, 7) and the performance evaluation against RMSE for each scg method (Figs. 4, 6, 8) are presented next in Table 13. The reported results clearly show that the worst classifier is the third one based on the accuracy measure.

5.2.4. IFC- NLR Results Analysis

The crisp input data was transformed into membership degrees using the IFC filter. The inductive fuzzy inference filter is a supervised filter available in WEKA. Here, the

inductive support is indicated by the membership degrees, and it is used for drawing the conclusion that the record belongs to the target class. To perform the operations mentioned above, the first step is the induction of the membership functions from the data. One optional decision is the possibility to display the induced membership functions. The next step is utilizing these induced functions to fuzzify the original attributes. In data mining, the IFC filter is used in two major ways, including prediction and visualization. Both prediction and visualization are based on the concepts of membership function induction and inductive attribute fuzzification.

The membership functions for three of the attributes are illustrated in (Figs. 9, 10 and 11).

Four fields are used to illustrate the numerical analytical variables (Fig. 11). Normalized likelihood ratios (NLRs) and the corresponding quantiles and average quantile values are shown as a table in the first field. In the second field, NLRs and their corresponding AVQs are displayed as a histogram. Membership functions for the analytical variables are

Table 13. Results from scg-nfc classifier.

	Accuracy
scg1	1.00
scg2	0.9959
scg3	0.7633

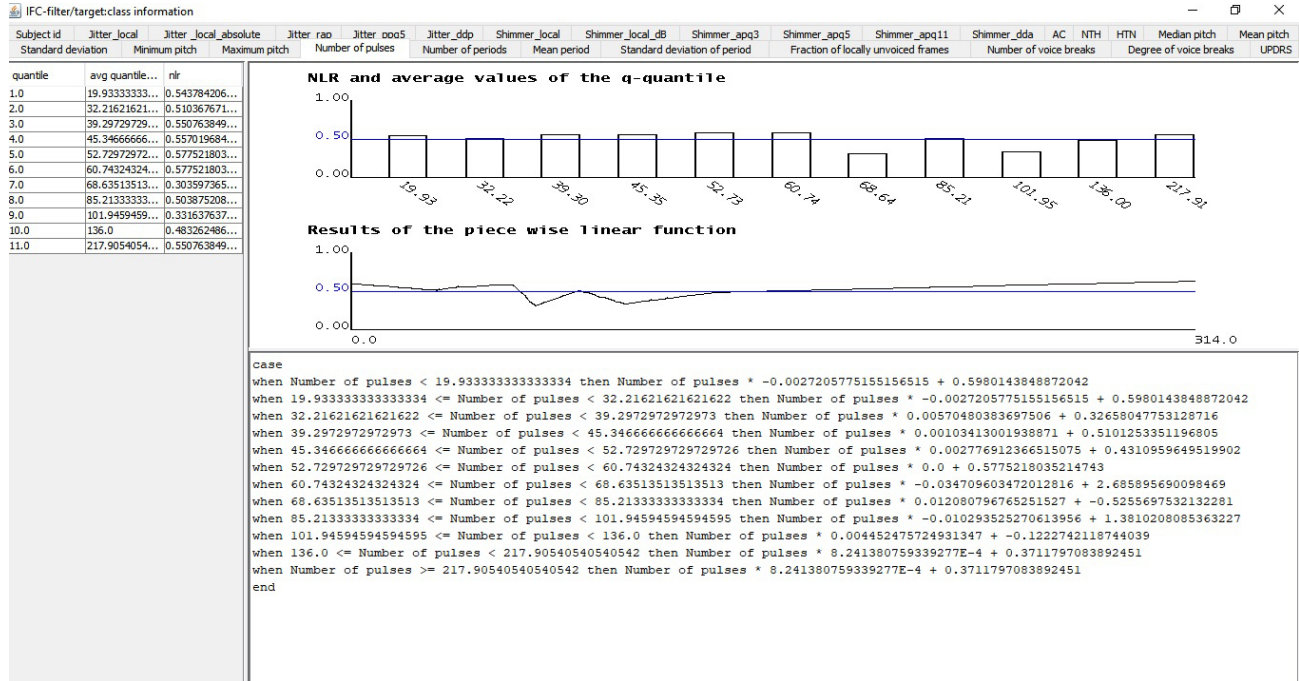


Fig. (9). Inductive membership function for jitter_local.

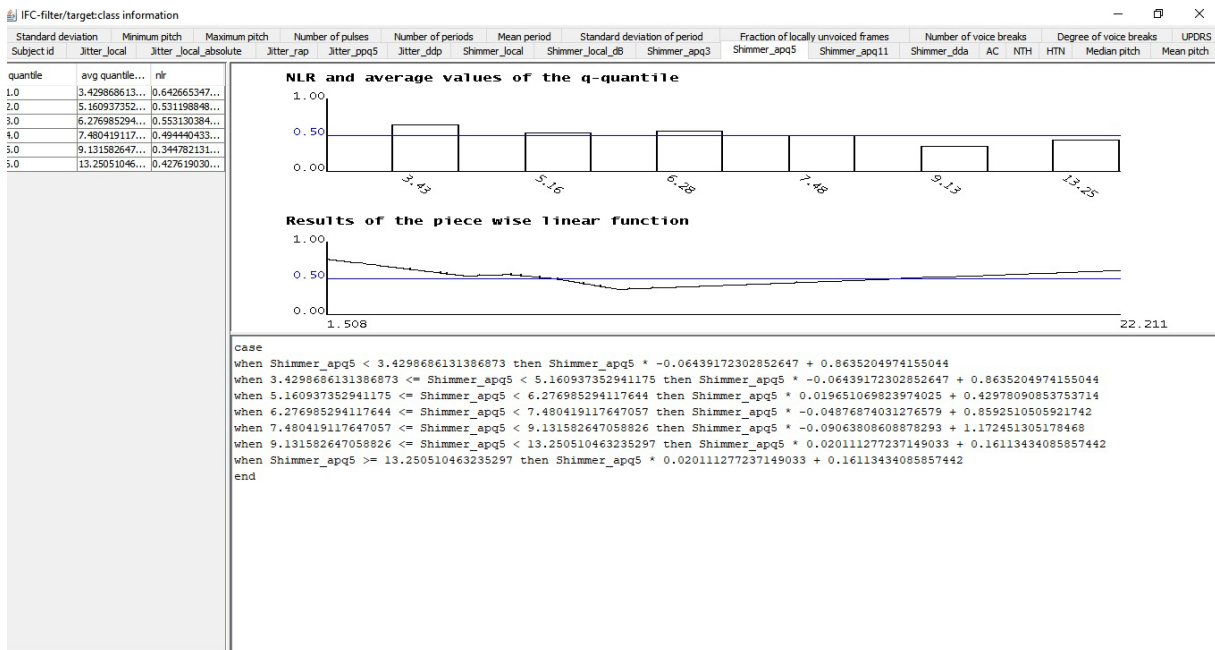


Fig. (10). Inductive membership function for number of pulses.

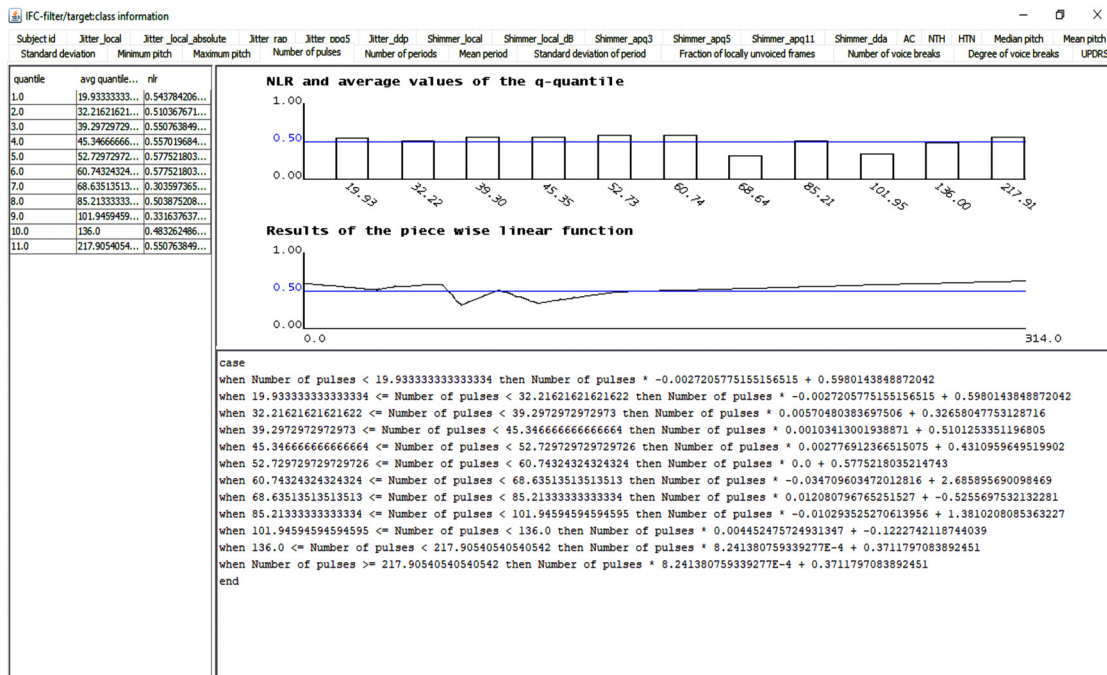


Fig. (11). Inductive membership function for shimmer aq5.

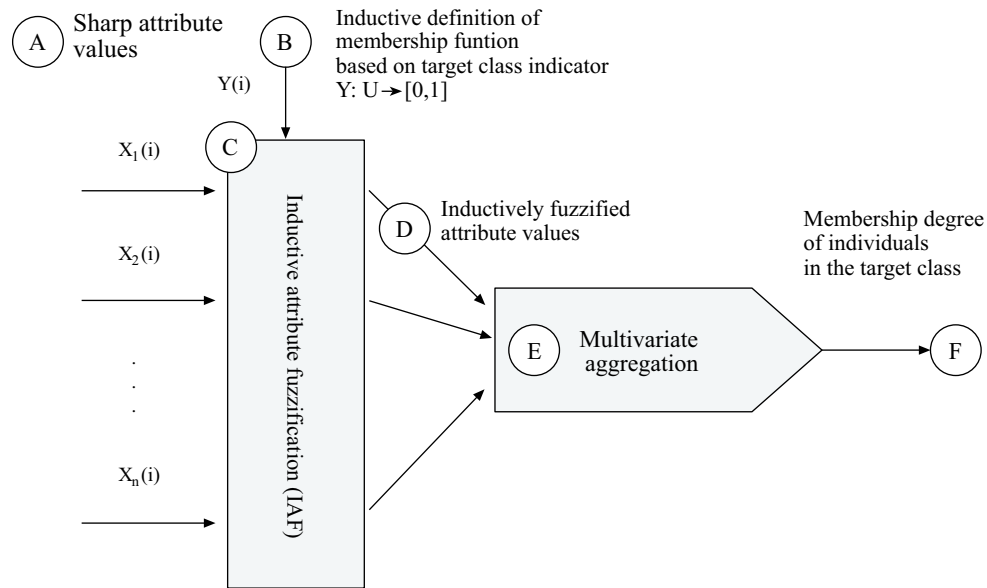


Fig. (12). Steps for the generation of an inductive membership function for an individual such as i [33].

illustrated using the third field. The membership function of the analytical variables is represented in the last field as SQL syntax which can be used for the fuzzy classification of the variables.

The nominal analytical variables are illustrated in a reduced form consisting of three fields. The first field includes a table representing normalized likelihood ratio and the quantile, and the average value of the quantile corresponding to NLR. The second field illustrates a histogram composed of the normalized likelihood ratios that correspond to the nominal values. The SQL syntax form of the membership

functions is presented in the last field. An additional tab is used to display the membership functions of all the analytic variables in the SQL syntax.

Making a multivariate inductive model for the target membership class is the basic idea of the inductive fuzzy classification method [33]. The proposed approach uses probability based IFC to forecast a univariate inductive fuzzification of analytic attributes before a multivariate conglomeration. This way, non-direct relationship between analytic attributes and the target membership can be demonstrated by a proper membership function. The steps presented in Fig.

(12) are connected with a specific end goal to determine an inductive membership degree of individual i in the forecast y' for class y :

- Step 1: The unprocessed data includes sharp valued attributes.
- Step 2: IFC-NLR is used to calculate an inductive definition for the membership function of attribute values in the predictive fuzzy class y' .
- Step 3: The membership functions derived in step B are utilized to fuzzify the attribute values. This step is a supervised univariate fuzzification of attribute values; it is called inductive attribute fuzzification.
- Step 4: In the next step, the dataset contains fuzzified attribute values in the range $[0,1]$ and indicates class memberships' inductive support.
- Step 5: The fuzzified analytic variables are congregated at a membership degree for individuals in the predictive class. It can be a straightforward disjunction or conjunction, a statistical model, or a set of fuzzy rules inferred by supervised machine learning methods, such as linear or logistic regression.
- Step 6: As a result of a multivariate aggregation, a multivariate membership function is generated. Outputs of the function generated are inductive membership degrees of individuals like i in a class like y . These outputs represent prediction resulted from these operations.

According to the suggestions of the previous researchers, the preprocessing step enhances the forecast exactness in case of using inductive fuzzy classification methods on analytical data. The performance of the existing data mining techniques can be improved by transforming the attributes used in data mining into inductive membership degrees in the so-called target class. The main idea in inductive fuzzy classification is a multivariate model of a target variable with a blend of attributes which are inductively fuzzified.

In this method, target class values are changed to numeric values. Thus, using regression for predictions of target class assignments is possible.

The model validation has been conducted by following the same validations applied for the previous classifiers. Classifier performance has been assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and average correlation coefficient of predicted classes and actual classes (CE). The results are presented next in this section.

As it is observed in Table 14, 8 out of 10 classifiers have classified the whole data correctly. The worst performance has been reported by RBFNetwork with an average correlation coefficient of 0.32.

Results of the 10-fold cross validation method are reported in Table 15. Here, the best results were obtained from the same 8 best classifiers reported in Table 14. Again, the worst performance belongs to RBFNetwork which has an average correlation coefficient of 0.57, though this is an improved value due to cross-validation. The results of the LOO cross validation method are presented in Table 16. The same top 8 classifiers reported in the previous two cases have the best performance here as well, and the worst classifier in this case is again the RBFNetwork method which has a correlation coefficient of 0.56.

6. MODEL VALIDATION AND COMPARATIVE ANALYSIS

To further investigate the classification methods suggested in this research, their results were compared with the results reported in [1]. This is a fair comparison because we used the same dataset produced by Sakar *et al.* [1]. They used two classifiers, namely, (1) K-NN classifier with k set to 4 different values 1,3,5 and 7, and the Euclidean distance metric, and (2) SVM with radial and linear basis function. The performance has been evaluated based on accuracy, sensitivity, specificity and MCC. Model validation is done using LOSO. The results from [1] are presented in Tables 17 and 18.

Table 14. Results from IFC-NLR.

	MAE	RMSE	CE
IsotonicRegression	0.00	0.00	1.00
LinearRegression	0.00	0.00	1.00
LogisticForIFC	0.00	0.00	1.00
AdditiveRegression	0.00	0.00	1.00
SimpleLinearRegression	0.00	0.00	1.00
LeastMedSq	0.46	0.68	0.00
SMOReg	0.00	0.00	1.00
RegressionByDiscretization	0.00	0.00	1.00
RBFNetwork	0.43	0.47	0.32
MultiLayerPerceptron	0.00	0.00	1.00

Table 15. Results from 10- fold cross validated IFC-NLR.

-	MAE	RMSE	CE
IsotonicRegression	0.00	0.00	1.00
LinearRegression	0.00	0.00	1.00
LogisticForIFC	0.00	0.00	1.00
AdditiveRegression	0.00	0.00	1.00
SimpleLinearRegression	0.00	0.00	1.00
LeastMedSq	0.48	0.69	0.00
SMOReg	0.00	0.00	1.00
RegressionByDiscretization	0.00	0.00	1.00
RBFNetwork	0.34	0.41	0.57
MultiLayerPerceptron	0.00	0.00	1.00

Table 16. Results from LOO cross validated IFC-NLR.

-	MAE	RMSE	CE
IsotonicRegression	0.00	0.00	1.00
LinearRegression	0.00	0.00	1.00
LogisticForIFC	0.00	0.00	1.00
AdditiveRegression	0.00	0.00	1.00
SimpleLinearRegression	0.00	0.00	1.00
LeastMedSq	0.48	0.69	0.00
SMOReg	0.00	0.00	1.00
RegressionByDiscretization	0.00	0.00	1.00
RBFNetwork	0.41	0.33	0.56
MultiLayerPerceptron	0.00	0.00	1.00

Table 17. K-NN results from [1].

K	Accuracy	Sensitivity	Specificity	MCC
1	0.5337	0.4962	0.5712	0.0007
3	0.5404	0.5327	0.5481	0.0008
5	0.5442	0.5365	0.5519	0.0008
7	0.5394	0.5404	0.5385	0.0008

Table 18. SVM results from [1].

Kernel	Accuracy	Sensitivity	Specificity	MCC
Linear	0.525	0.525	0.525	0.0006
RBF	0.55	0.6	0.5	0.1005

Table 19. Fuzzy rough nn results.

k	Accuracy	Sensitivity	Specificity	MCC
1	0.7616	0.3548	0.4068	0.450407
3	0.7616	0.3548	0.4068	0.450407
5	0.7616	0.3548	0.4068	0.450407
7	0.7616	0.3548	0.4068	0.450407

6.1. Fuzzy Rough Results

The results from the fuzzy rough classification methods are presented next. We have mainly applied these classifiers: fuzzy rough NN, FLR, FURIA and OWANN. The LOSO validation method has been used for all these classifiers.

For this classifier, parameter k was set to 1, 3, 5 and 7; the obtained results are presented in Table 19.

As it is obvious from Table 19, the results of the fuzzy rough NN outperformed those of k-NN and SVM reported in Table 17 and Table 18, respectively. The accuracy measure for this method did not vary for different values of parameter k; it is equal to 76.16%, which is higher than the same measure for k-NN and SVM presented in [1]. The MCC measures also outperforms the results presented in [1]. This shows the effectiveness and robustness of the classifiers applied in this paper. Finally, the results from the fuzzy lattice reasoning methods' application are presented below:

As it is obvious from the results reported in Table 20, FLR has an accuracy of 99.33%, which is better than those reported by k-NN and SVM. MCC measures also outperform the results reported in [1].

The results of FURIA are reported in Table 21 where it is obvious that this algorithm outperforms the ones presented in [1] in terms of accuracy and MCC measures. The high value

Table 20. FLR results.

-	Accuracy	Sensitivity	Specificity	MCC
FLR	0.9933	0.4981	0.0048	1

Table 21. FURIA results.

-	Accuracy	Sensitivity	Specificity	MCC
FURIA	1	0.5	0.5	1

Table 22. OWANN results.

k	Accuracy	Sensitivity	Specificity	MCC
1	0.7654	0.3414	0.4241	0.450407
3	0.7848	0.3376	0.4472	0.521715
5	0.7886	0.3405	0.4482	0.521715
7	0.8107	0.3664	0.4443	0.550123

of MCC indicates the effectiveness and robustness of this classifier.

The result of OWANN is reported in Table 22, where the values of k were taken as 1, 3, 5 and 7. The results from OWANN are presented in [1]. The best performance has been achieved for k= 7 with 81.07% accuracy.

6.2. Adaptive Neuro Fuzzy Classifier Results

The adaptive neuro-fuzzy classification method was implemented by employing subclustering and FCM methods; and the LOSO validation was utilized.

The subclustering method was performed using the genfis2 function in MATLAB. The radius parameter was set to 0.8 and the model was trained for 20 times. The results are presented in Table 23.

The accuracy of the adaptive neuro fuzzy classifier with subclustering is 97.21%; this is higher than the results reported in [1]. The MCC value is 0.8487, which confirms the effectiveness and sustainability of the method utilized in this paper.

In this method, a Sugeno type FIS is constructed using 3 rules and membership functions. The model was trained for 20 epochs and the results are reported in Table 24. The results from genfis3 outperform those reported in [1].

Table 23. Genfis2 results.

-	Accuracy	Sensitivity	Specificity	MCC
genfis2	0.9721	1	0.9286	0.8487

Table 24. Genfis3 results.

-	Accuracy	Sensitivity	Specificity	MCC
genfis3	0.8337	1	0.7647	0.4042

Table 25. Results from PCA.

-	Accuracy	Sensitivity	Specificity	MCC
genfis1	0.5683	0.5833	0.5714	0.0475
genfis2	0.6279	0.6429	0.6667	0.2374
genfis3	0.5577	0.5833	0.5714	0.0475

Table 26. Results from CFS subset evaluation.

-	Accuracy	Sensitivity	Specificity	MCC
genfis1	0.9942	1	1	1
genfis2	0.975	1	0.9286	0.8487
genfis3	0.8442	1	0.7647	0.4042

For further investigation, the adaptive neuro fuzzy classification method was applied to the dataset after attribute selection. The attribute selection methods used in this research are PCA and CFS subset evaluation. To force the CFS subset evaluation method to choose 3 attributes, we set variance to 0.6. The results are given in Table 25 and Table 26. A triangular membership function with 2 functions per input has been utilized in genfis1. The model has been trained for 40 times in PCA and 10 times in CFS subset evaluation. Parameters for the other functions are the same as before.

The results reported in Table 26 clearly demonstrate how the two attribute-selection methods outperform the methods described in [1]. Comparing the two attribute-selection methods, the CFS subset evaluation results are better than the PCA results.

CONCLUSION

This paper described fuzzy classification based approaches to distinguish PWP from healthy people. The study concentrated on analyzing a Parkinson speech dataset with multiple types of sound recordings gathered from 40 test subjects. From the results reported in this paper, it can be concluded that the fuzzy rough nearest neighborhood algorithm showed worst performance compared to the other classifiers. However, we compared the performances of the classifiers used in this paper to the performance of the classifiers

used in [1]. In general, the classifiers described in this paper outperformed those considered in [1]. As WEKA classifiers are concerned, MultiLayerPerceptron, Bagging and SGD reported 100% accuracy. The worst classifier was IBK with an accuracy measure of 0.87 for 10-fold and LOO cross validated models, and 0.84 for the model validated by data partitioning. Comparing the Fuzzy rough classifiers, the best performance belongs to FURIA with accuracy value equal to 1. The worst classifier in this category is QSBA with the accuracy of 0.68 for 10-fold cross validated model, 0.69 for LOO cross validated model, and 0.65 for the model which was validated by dataset partitioning. The best adaptive neuro fuzzy classifier is the one generated by genfis2, and the worst is the classifier generated using genfis3. Concerning the seg classifiers, scg1 classified the data in a completely correct manner and scg3 showed the least accuracy, namely 0.7633. The inductive fuzzy classification results showed that the best classifiers are the regression methods as offered in [33]. The worst classifier in this category was RBFNetwork.

AUTHORS' CONTRIBUTIONS

NKD wrote the programs, conducted the testing, and wrote the initial draft, OCS helped in testing, in analyzing the results and in manuscript, TO and RA developed the initial methodology and edited the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIAL

The data supporting the findings of the article is available in the UCI Machine Learning Repository at [https://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings].

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] Sakar BE, Isenkul ME, Sakar CO, *et al.* Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE J Biomed Health Inform* 2013; 17(4): 828-34. [http://dx.doi.org/10.1109/JBHI.2013.2245674] [PMID: 25055311]
- [2] de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. *Lancet Neurol* 2006; 5(6): 525-35. [http://dx.doi.org/10.1016/S1474-4422(06)70471-9] [PMID: 16713924]
- [3] Shahsavari MK, Rashidi H, Bakhsh HR. Efficient classification of Parkinson's disease using extreme learning machine and hybrid particle swarm optimization. 2016 4th International Conference on Control, Instrumentation, and Automation (ICCIA); 2016 June 2; Qazvin, Iran. [http://dx.doi.org/10.1109/ICCIAutom.2016.7483152]
- [4] Parkinson J. An essay on the shaking palsy. 1817. *J Neuropsychiatry Clin Neurosci* 2002; 14(2): 223-36. [http://dx.doi.org/10.1176/jnp.14.2.223] [PMID: 11983801]
- [5] Singh N, Pillay V, Choonara YE. Advances in the treatment of Parkinson's disease. *Prog Neurobiol* 2007; 81(1): 29-44. [http://dx.doi.org/10.1016/j.pneurobio.2006.11.009] [PMID: 17258379]
- [6] Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans Biomed Eng* 2009; 56(4): 1015-22. [http://dx.doi.org/10.1109/TBME.2008.2005954] [PMID: 21399744]
- [7] National Collaborating Centre for Chronic Conditions (UK). Parkinson's Disease: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care. London: Royal College of Physicians, UK, (2006). (NICE Clinical Guidelines No 35). 5, Diagnosing Parkinson's disease.
- [8] Cunningham L, Mason S, Nugent C, Moore G, Finlay D, Craig D. Home-based monitoring and assessment of Parkinson's disease. *IEEE Trans Inf Technol Biomed* 2011; 15(1): 47-53. [http://dx.doi.org/10.1109/TITB.2010.2091142] [PMID: 21062684]
- [9] Rigas G, Tzallas AT, Tsipouras MG, *et al.* Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors. *IEEE Trans Inf Technol Biomed* 2012; 16(3): 478-87. [http://dx.doi.org/10.1109/TITB.2011.2182616] [PMID: 22231198]
- [10] Marino S, Ciurleo R, Di Lorenzo G, *et al.* Magnetic resonance imaging markers for early diagnosis of Parkinson's disease. *Neural regeneration research* 2012; 7(8): 611.
- [11] Dastgheib ZA, Lithgow B, Moussavi Z. Diagnosis of Parkinson's disease using electrovestibulography. *Med Biol Eng Comput* 2012; 50(5): 483-91. [http://dx.doi.org/10.1007/s11517-012-0890-z] [PMID: 22399163]
- [12] Jeon H, Lee W, Park H, *et al.* Automatic classification of tremor severity in Parkinson's disease using a wearable device, *Sensors (Basel)* 2017; 17(9): E2067. [http://dx.doi.org/10.3390/s17092067]
- [13] Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Trans Biomed Eng* 2012; 59(5): 1264-71. [http://dx.doi.org/10.1109/TBME.2012.2183367] [PMID: 22249592]
- [14] Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *J R Soc Interface* 2011; 8(59): 842-55. [http://dx.doi.org/10.1098/rsif.2010.0456] [PMID: 21084338]
- [15] Chakraborty A, Chakraborty A, Mukherjee B. Detection of Parkinson's disease using fuzzy inference system intelligent systems technologies and applications. Springer 2016; pp. 79-90. [http://dx.doi.org/10.1007/978-3-319-23036-8_7]
- [16] Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng* 2010; 57(4): 884-93. [http://dx.doi.org/10.1109/TBME.2009.2036000] [PMID: 19932995]
- [17] Samà A, Pérez-López C, Rodríguez-Martín D, *et al.* Estimating bradykinesia severity in Parkinson's disease by analysing gait through a waist-worn sensor. *Comput Biol Med* 2017; 84: 114-23. [http://dx.doi.org/10.1016/j.combiomed.2017.03.020] [PMID: 28351715]
- [18] Parisi F, Ferrari G, Giuberti M, *et al.* Inertial BSN-based characterization and automatic UPDRS evaluation of the gait task of parkinsonians. *IEEE Trans Affect Comput* 2016; 7: 258-71. [http://dx.doi.org/10.1109/TAFFC.2016.2549533]
- [19] Murdoch, B., Whitehill, T., De Letter, M. and Jones, H. Communication impairments in parkinson's disease. *Parkinson's disease*, 2011 (2011).
- [20] Hlavica J, Prauzek M, Peterek T, Musilek P. Assessment of Parkinson's disease progression using neural network and ANFIS models. *Neural Netw World* 2016; 26: 2. [http://dx.doi.org/10.14311/NNW.2016.26.006]
- [21] Gürüler H. A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. *Neural Comput Appl* 2015; 1-10.
- [22] Martinez-Manzanera O, Roosma E, Beudel M, Borgemeester RW, van Laar T, Maurits NM. A method for automatic and objective scoring of bradykinesia using orientation sensors and classification algorithms. *IEEE Trans Biomed Eng* 2016; 63(5): 1016-24. [http://dx.doi.org/10.1109/TBME.2015.2480242] [PMID: 26394414]
- [23] Mousavi SJ, Ponnambalam K, Karray F. Inferring operating rules for reservoir operations using fuzzy regression and ANFIS. *Fuzzy Sets Syst* 2007; 158(10): 1064-82. [http://dx.doi.org/10.1016/j.fss.2006.10.024]
- [24] Fox CM, Morrison CE, Ramig LO, Sapir S. Current perspectives on the Lee Silverman Voice Treatment (LSVT) for individuals with idiopathic Parkinson disease. *Am J Speech Lang Pathol* 2002; 11: 2. [http://dx.doi.org/10.1044/1058-0360(2002)012]
- [25] Chen HL, Huang CC, Yu XG, *et al.* An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Syst Appl* 2013; 40(1): 263-71. [http://dx.doi.org/10.1016/j.eswa.2012.07.014]
- [26] Mamdani, Ebrahim H. Application of fuzzy algorithms for control of simple dynamic plant. *Proceedings of the institution of electrical engineers*. Vol. 121, No. 12, IET, 1974.

- [27] [http://dx.doi.org/10.1049/piee.1974.0328] El A, Edmonds J, Gonzalez J, Papa M. A framework for hybrid fuzzy logic intrusion detection systems. Proc of IEEE International Conference on Fuzzy Systems, (2005) .
- [28] Mitra S, Hayashi Y. Neuro-fuzzy rule generation: Survey in soft computing framework. IEEE Trans Neural Netw 2000; 11(3): 748-68. [http://dx.doi.org/10.1109/72.846746] [PMID: 18249802]
- [29] Hamdan H, Garibaldi JM. Adaptive neuro-fuzzy inference system (ANFIS) in modelling breast cancer survival. Proc of IEEE International Conference on Fuzzy Systems, 2010. [http://dx.doi.org/10.1109/FUZZY.2010.5583997]
- [30] Choi H, Yoo H, Jung H, Lim T, Lee K, Ahn K. An ANFIS-based energy management inference algorithm with scheduling technique for legacy device. international conference on artificial intelligence, Energy and Manufacturing Engineering (ICAEME'2015) Dubai, 2015.
- [31] Sun CT, Jang JS. A neuro-fuzzy classifier and its applications. Proceedings of IEEE International Conference on Fuzzy Systems; 1993 March 28- April 1; San Francisco, CA, USA. [http://dx.doi.org/10.1109/FUZZY.1993.327457]
- [32] Cetişli B, Barkana A. Speeding up the scaled conjugate gradient algorithm and its application in neuro-fuzzy classifier training. Soft Comput 2010; 14(4): 365-78. [http://dx.doi.org/10.1007/s00500-009-0410-8]
- [33] Kaufmann MA. Inductive fuzzy classification in marketing analytics. Bern: Ruf (2012) Kaufmann, M A Inductive fuzzy classification in marketing analytics. Switzerland: University of Fribourg 2012.
- [34] Graf C. Erweiterung des Data-Mining-Softwarepakets WEKA um induktive unscharfe Klassifikation 2010.
- [35] Pawlak Z. Rough sets: Theoretical aspects of reasoning about data. Springer Science & Business Media 2012.
- [36] Zadeh LA. Fuzzy sets. Inf Control 1965; 8(3): 338-53. [http://dx.doi.org/10.1016/S0019-9958(65)90241-X]