International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v11i2s.6143
***Article Received**: 26 November 2022 **Revised**: 22 December 2022 **Accepted**: 14 January 2023*

Improving Accuracy of Integrated Neuro-Fuzzy Classifier with FCM based Clustering for Diagnosis of Psychiatric Disorder

Chandrabhan Singh1, Dr. Mohit Gangwar2, Dr. Upendra Kumar3
[1]Research Scholar, Computer Science & Engineering
Dr. A.P.J. Abdul Kalam Technical University
Lucknow, UP, India.
[2]Professor, Department of Computer Science & Engineering
BNCET, Lucknow, UP, India.
Email: mohitgangwar@gmail.com
[3]Assistant Professor, Department of Computer Science & Engineering
IET, Lucknow, UP, India.
Corresponding Author: Dr. Mohit Gangwar (mohitgangwar@gmail.com)

Abstract— Parkinson's disease (PD) is a progressive neurodegenerative disorder. Autism spectrum disorder (ASD) is a neurodevelopment disorder. Clinical decision-making process is complex. Due to complex nature of disease sign and its symptoms clinical decision making may lead to misclassification. To deal with such complex medical problems methods or approaches of soft computing play an important role. This paper will focus on presenting an integrated Neuro-fuzzy model. This integrated model has the learning strength of neural network and knowledge representation ability of fuzzy logic. Modified Adaptive Neuro –Fuzzy inference system (M-ANFIS) is used here for classification and predication. Here Fuzzy C-mean (FCM) Clustering is used first to make classes of data before presenting in to ANFIS. This FCM based class will reduce the classifier computational overhead. Precision error and recall, F-measure and accuracy matrices are used to compare the experimental results with other classic methods.

Keyword – Parkinson's disorder, Autism spectrum disorder, Fuzzy C- mean clustering (FCM), Modified Adaptive Neuro-Fuzzy Inference System(M-ANFIS).

I. INTRODUCTION

Sign and symptoms of psychiatric disorders are not one to one mapped. They are complex in nature. Due to their overlapping nature the classification is not an easy task. Use of intelligent computing methods in medical health care domain increase in large scale. To classify between healthy and unhealthy i.e. with PD/ASD machine learning approaches play important role. The sign and symptoms of PD many times cause to classify it in other movement related disorders. The process of machine learning is defined in various phase. The output of each phase work as an input to the next layer. This research deal with integrated neuro-fuzzy classifier with Fuzzy C-mean clustering. Here intelligent feature selection approach is used to reduce the classifier computations time. As we know that ANFIS has high computational cost due to its complex structure as well as learning capability.

To overcome this the Fuzzy C-mean clustering is introduced by grouping the class of PD and ASD. Thus, grouping of class will reduce the complexity of computation. As we know that the computational cost and training cost of ANFIS architecture depends upon the size of numbers of parameters. To tune the parameters effective training approach is needed. Sugeno-Fuzzy inference approach is used in this work because it does not have output membership function as well as it is more flexible in designing system. In this paper section 2 will discuss the machine learning based methods used by various researchers. Section 3 will define the description about the dataset. Section 4 will elaborate the working of Modified ANFIS. Section 5 contain detail about experimental results followed by conclusion.

II. RELATE WORKS

In past decade lot of work related to medical computing imposing machine intelligence has been done by researcher. This section will discuss few of novel approaches applied previously.

Abiyev et al.[1] uses fuzzy neural based system for classification of PD from healthy class with very high accuracy. The diagnosis approach is combination of neural network along with fuzzy logic.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v11i2s.6143
***Article Received**: 26 November 2022 **Revised**: 22 December 2022 **Accepted**: 14 January 2023*

Aich et al. [2] uses support vector machine (SVM), random forest for the classification of Parkinson disease. The result of research shows good accuracy. Researcher has also compared the result with various other feature set based on genetic algorithm (GA) as well as principal component analysis (PCA)

Linear regression, Linear Discriminate Analysis (LDA), gaussian naïve bayes along with KNN and SVM was used by Ali et al. [3]. This research shows the SVM linear accuracy of approximate 70%.

Convolutional neural network (CNN) has been introduced by Gunduz [4] with leave one person out. Cross validation approach showing accuracy about 86.9%.

Classification and regression technique (CART) with SVM and Artificial Neural Network (ANN) for classification of PD is used by Karapinav [5] showing SVM accuracy about 93.8%.

Evolutionary wavelet neural network (EWNN) with training and testing ratio of 90:10 along with cross validation was used by Khan et al. [6] for the classification with high accuracy.

SVM with radial basis function Kernal (RBF) based classification proposed by Frid et al. [7] with accuracy of 81.8%. This research also shows the assess of PD severity.

Baby et al. [8] used ANN, SVM along with naïve bayes model for classifying PD from healthy class. In this cross-validation approach is also used. The ANN accuracy of this result is 86.75%.

ANN-SVM based modal was developed by Tahir et al [9]. In this research inter group and intra group-based normalization approach is used.

Wrappers based feature selection along with K nearest neighbor, SVM and random forest for classification has been introduced by Solana-Lavalle et al. [10].

Principal component analysis [PCA] along with multidimensional scaling based hierarchical clustering method is used by El Ansary et al [11] to show the difference between ASD and controlled person.

Goel et al [12] introduced modified grasshopper optimization algorithm (MGOA) which is a meta heuristic approach to detect ASD. The accuracy of this approach is very high in comparison to other approaches.

The related study clearly indicates that with specific reference to psychiatric disorder the integrated neuro-fuzzy based approach is not explored along with intelligent feature selection.

III. DESCRIPTION OF ASD AND PD DATASETS

Two bench marked dataset of PD and ASD was taken from UCI machine learning data repository [13] [14]. Parkinson data set has 240 instances where as ASD has 292 data instances. Number of attributes in PD dataset is approximate 46 where as in ASD it is 21. The dataset may have missing values also which will be handled during preprocessing phase. The ASD dataset contains screening data of children. Attribute A1 to A10 are questions whose value was classified with 0 or 1. Other attributes are like age, gender, ethnicity, country of res, used app before, relation, class/ASD, etc. The characteristics of this dataset is multivariate. PD dataset of UCI repository contains replications. The features were extracted from multiple voice recording of individuals.

IV. PROPOSED METHODOLOGY

Multiple levels of system are defined for the overall diagnosis system. This methodology is based on following phases: -

Table 1: Phase wise working methodology

Phase	Working	Approach used	Input	Output
Preprocessing	Cleaning/balancing	SMOTE approach	Row data	Preprocessed data
Feature selection	Selection of optimal feature sets	BOA+ELOA	All feature set	Optimal feature (subset)
Classification	First features go through FCM for categorization, then neural fuzzy classifier for classification	ANFIS classifier	Categorized feature set using FCM	Classified result for predication

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v11i2s.6143
***Article Received**: 26 November 2022 **Revised**: 22 December 2022 **Accepted**: 14 January 2023*

Table 1 represents the phase-wise working procedure of the integrated methodology. The datasets first go through with preprocessing phase to deal with cleaning, removal of missing values, normalization.

The proposed methodology starts with the initial stage of preprocessing where SMOTE approach is used. The Pseudo code for this phase is as follows: -

Algorithm 1: - Preprocessing (Advance SMOTE)
Input: - PD and ASD dataset
Output: - Normalized data
Steps: -

a. Missing and null value is handled.
b. Data set is divided between majority and minority class based on disease or healthy.
c. Using SMOTE technique [15] class imbalancing will be handled by creating synthetic instances.
d. Synthetic instances that are closer to the SMOTE boundary or majority class is deleted.

After the preprocessing feature selection will be performed as learning with complete data feature have lot of computational overhead and cost. The accuracy of classification system also degrades due to learning with all features. Attributes that play optimal and significant role in learning and predicting system are used whereas others are removed [16].

Meta heuristic-based optimization approaches play an important role in such classification. In this study nature inspired algorithm like Butterfly optimization (BOA) [17] and Lion based optimization algorithm [18] (LOA) is used to find the subset of suitable feature set. This hybrid optimization algorithm first incorporate BOA then LOA is clubbed with it. Due to the clubbed feature the shortcoming of BOA is addressed. The below figure clearly indicates that after feature selection, classification phase will be carried out.

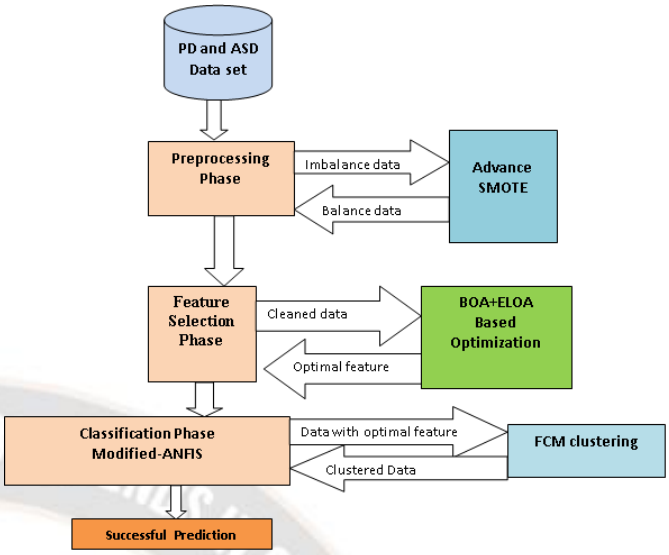

Figure 1- Overall approach of classification

For classification purpose first order sugeno fuzzy model is used. This modified Adaptive Neuro fuzzy inference system (M-ANFIS) first deal with FCM algorithm. The FCM based clustered ANFIS shows efficient result for complex problems [19]. The use of fuzzy C-Mean based clustering increases in last decade for diagnosis of complex medical diseases [20-22]. Using this the classes of PD and ASD are group together which helps in reduction of computational cost also.

The steps to perform the classification using M-ANFIS along with FCM based data clusters.

Algorithm 2- Classification
Input: - Data with optimal feature set
Output: - Classified result
Steps: -

I. Data with optimal feature set is clustered using fuzzy C-mean algorithm [20].
   a. During the clustering phase the PD and ASD data set is group in classes. Due to use of same distance for all iteration it will reduce the complexity.
   b. Clustering phase stops when it reaches to the max iteration or no improvement between objective function found.
   c. Data point with high membership grade from each cluster and cluster center is fetched.
II. First order sugeno fuzzy model is used on the input of step 1 to build fuzzy if then rules.
III. Five-layer ANFIS model is used for the classification using Gaussian membership function.
IV. Classified result as an output of step 3.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v11i2s.6143
***Article Received**: 26 November 2022 **Revised**: 22 December 2022 **Accepted**: 14 January 2023*

ANFIS model contains two type of nodes: fixed and adaptive.

Layer wise working of ANFIS model is as below-

Layer 1: - Gaussian membership function is applied on each node.

Layer 2: - Strength of rule generated in layer 1 is calculated.

Layer 3: - Normalized value is calculated.

Layer 4: - Consequent part of fuzzy rule is represented by each node.

Layer 5: - Summation operation perform defuzzification.

All steps perform the training in the learning system.

V. RESULT AND DISCUSSION

MATLAB toolbox are used for the experimental results figure 2 mentioned blow shows the working of advance SMOTE approach where synthetic examples are created to remove the imbalancing.

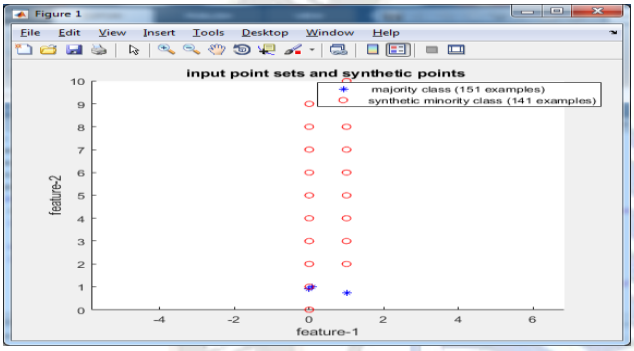

Figure 2: Working of Advance SMOTE

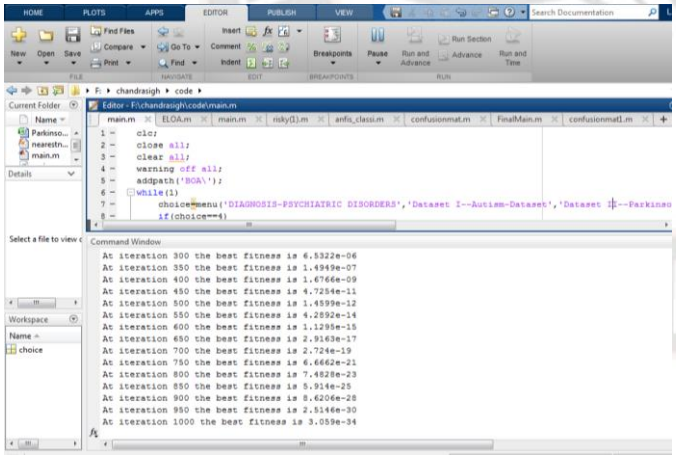

Figure 3: Working of Feature Selection

Figure 3 shows the working of integrated optimization approach to find the optimal feature set from the normalized dataset of PD and ASD. Figure 4 shows the classification step using ANFIS model. Here FCM clustering is applied which is then followed by ANFIS model for classification [23].

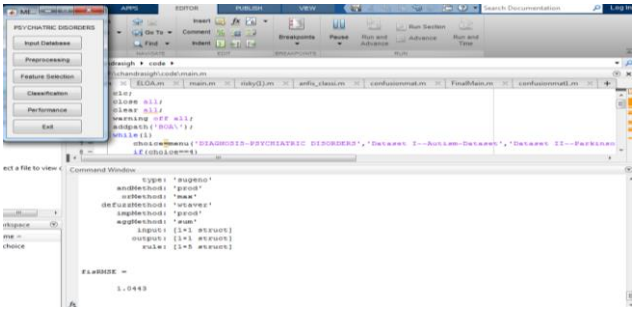

Figure 3: Modified-ANFIS for Classification

Table 2 shows the performance measure of the proposed approach using confusion matrices. Comparison is done between simple ANFIS and Modified approach for PD and ASD dataset. Various performance measures are calculated based on true positive, true negative, false positive, and false negative class instances.

Table 2: Comparison of performance Measure

Performance Measure	PD Dataset		Autism Dataset	
	ANFIS	Modified-ANFIS	ANFIS	Modified-ANFIS
Accuracy	89.84	92.86	89.93	92.80
Precision	88.65	92.90	89.23	92.19
Recall	90.14	92.87	90.45	92.35
F-measure	90.45	92.89	91.04	92.27
Error	9.20	7.14	10.12	7.14

The result of comparison clearly indicates that the FCM based ANFIS has batter result in form of fames like accuracy, precision, recall and F-measure value. Also this cluster based classification reduces the classifier computation overhead also.

VI. CONCLUSION

This study suggests a combined neuro-fuzzy classifier and method for intelligent feature selection based on optimization. The combination of BOA and ELOA carries out the intelligent feature selection (Enhanced Lion Optimization Algorithm). Opposition learning increases the ability of exploitation search. Prior to using these clusters as inputs to the ANFIS model for classification, features are first divided based on fuzzy C-mean clustering during the classification phase. The updated FCM clustering algorithm takes the same distance for all iteration. Due to this the complexity is majorly reduced. Thus, this research introduced an integrated model by combining neuro - fuzzy classifiers and feature selection methods for rapid and accurate detection of psychological disorders related to PD and ASD. A series of process are introduction for prediction of diseases. The proposed model contains new modified algorithms for achieving the classified result.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v11i2s.6143
***Article Received**: 26 November 2022 **Revised**: 22 December 2022 **Accepted**: 14 January 2023*

REFERENCES

[1]. Abiyev, R. H., and Abizade, S. (2016). Diagnosing Parkinson's diseases using fuzzy neural system. Comput. Mathe. Methods Med. 2016:1267919. doi: 10.1155/2016/1267919.

[2]. Aich, S., Kim, H., younga, K., Hui, K. L., Al-Absi, A. A., and Sain, M. (2019). "A supervised machine learning approach using different feature selection techniques on voice datasets for prediction of Parkinson's disease," in 2019 21st International Conference on Advanced Communication Technology (ICACT) (PyeongChang), 1116–1121. doi: 10.23919/ICACT.2019.8701961.

[3]. Ali, L., Khan, S. U., Arshad, M., Ali, S., and Anwar, M. (2019a). "A multi-model framework for evaluating type of speech samples having complementary information about Parkinson's disease," in 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (Swat), 1–5. doi: 10.1109/ICECCE47252.2019.8940696.

[4]. Gunduz, H. (2019). Deep learning-based parkinson's disease classification using vocal feature sets. IEEE Access 7, 115540–115551. doi: 10.1109/ACCESS.2019.2936564.

[5]. Karapinar Senturk, Z. (2020). Early diagnosis of Parkinson's disease using machine learning algorithms. Med. Hypoth. 138:109603. doi: 10.1016/j.mehy.2020.109603.

[6]. Khan, M. M., Mendes, A., and Chalup, S. K. (2018). Evolutionary Wavelet Neural Network ensembles for breast cancer and Parkinson's disease prediction. PLoS ONE 13:e0192192. doi: 10.1371/journal.pone.0192192.

[7]. Frid, A., Safra, E. J., Hazan, H., Lokey, L. L., Hilu, D., Manevitz, L., et al. (2014). "Computational diagnosis of Parkinson's disease directly from natural speech using machine learning techniques," in 2014 IEEE International Conference on Software Science, Technology and Engineering (Ramat Gan), 50–53. doi: 10.1109/SWSTE.2014.17.

[8]. Baby, M. S., Saji, A. J., and Kumar, C. S. (2017). "Parkinsons disease classification using wavelet transform based feature extraction of gait data," in 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT) (Kollam), 1–6. doi: 10.1109/ICCPCT.2017.8074230.

[9]. Tahir, N. M., & Manap, H. H. (2012). Parkinson disease gait classification based on machine learning approach. Journal of Applied Sciences (Faisalabad), 12(2), 180-185.

[10]. Solana-Lavalle, G., Galán-Hernández, J. C., & Rosas-Romero, R. (2020). Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features. Biocybernetics and Biomedical Engineering, 40(1), 505-516.

[11]. El-Ansary, A., Hassan, W. M., Daghestani, M., Al-Ayadhi, L., & Ben Bacha, A. (2020). Preliminary evaluation of a novel nine-biomarker profile for the prediction of autism spectrum disorder. PLoS One, 15(1), e0227626.

[12]. Goel, N., Grover, B., Gupta, D., Khanna, A., & Sharma, M. (2020). Modified grasshopper optimization algorithm for detection of autism spectrum disorder. Physical Communication, 41, 101115.

[13]. https://archive.ics.uci.edu/ml/datasets/Parkinson.

[14]. https://archive.ics.uci.edu/ml/datasets/Autistic_Spectrum_Disorder.

[15]. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

[16]. Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16-28.

[17]. Arora, S., & Singh, S. (2019). Butterfly optimization algorithm: a novel approach for global optimization. Soft Computing, 23(3), 715-734.

[18]. Yazdani, M., & Jolai, F. (2016). Lion optimization algorithm (LOA): a nature-inspired metaheuristic algorithm. Journal of computational design and engineering, 3(1), 24-36.

[19]. Benmouiza, K., Cheknane, A. (2019). Clustered ANFIS network using fuzzy c-means, subtractive clustering, and grid partitioning for hourly solar radiation forecasting. Theor Appl Climatol 137, 31–43. https://doi.org/10.1007/s00704-018-2576-4.

[20]. Pal, N. R., Pal, K., Keller, J. M., & Bezdek, J. C. (2005). A possibilistic fuzzy c-means clustering algorithm. IEEE transactions on fuzzy systems, 13(4), 517-530.

[21]. Wu Y, Duan H, Du S.(2015). Multiple fuzzy c-means clustering algorithm in medical diagnosis. Technol Health Care.;23 Suppl 2:S519-27. doi: 10.3233/THC-150989.

[22]. Chowdhary CL, Mittal M, P K, Pattanaik PA, Marszalek Z. (2020). An Efficient Segmentation and Classification System in Medical Images Using Intuitionist Possibilistic Fuzzy C-Mean Clustering and Fuzzy SVM Algorithm. Sensors (Basel);20(14):3903. doi: 10.3390/s20143903.

[23]. Wei, M., Bai, B., Sung, A. H., Liu, Q., Wang, J., & Cather, M. E. (2007). Predicting injection profiles using ANFIS. Information Sciences, 177(20), 4445-4461.