Attribute Optimization: Genetic Algorithms and Neural Network for Voice Analysis Classification of Parkinson's Disease

Yudi Ramdhani[1], Ade Mubarok[1], Syarif Hidayatulloh[1], Wildan Wiguna[2]
[1]*Universitas BSI, Bandung, Indonesia*
[2]*AMIK BSI Tasikmalaya, Tasikmalaya, Indonesia*

Keywords: Parkinson, Machine Learning, Data Mining, Feature Selection, Classification, Genetic Algorithm, Neural Network

Abstract: The Parkinson's disease is a degenerative disorder of the central nervous system that causes disturbances in the motor system, leading to impaired balance. Machine learning and data mining is able to detect this disorder in Parkinson's disease. Reviewing the phenomenon, the study aims to examine the genetic algorithm for feature selection and neural network algorithms for the classification of Parkinson's disease. Parkinson's diagnosis used a promising learning machine to be the solution as an early stage classification of Parkinson's disease. The research findings are submitted that in each calcification method through learning machine will get some obstacle in analyse medical data. One of the usual constraints on the neural network classification algorithm when the features contained in the dataset are not relevant to the classification. To reduce the irrelevant features used genetic algorithm selection feature to improve data analysis performance in better classification.

1 INTRODUCTION

Parkinson's disease (PD) is the second most common neurological disorder after Alzheimer's disease. It causes, during its course, a variety of symptoms. These include difficulty walking, talking, thinking or completing other simple tasks (Little, McSharry, & Hunter, 2009) (Ishihara & Brayne, 2006) (Jankovic, 2008). Approximately 90% of patients with Parkinson's disease have vocal disorders (O'Sullivan & Schmitz , 2007). With cur-rent prevalence rates, ranging from 10 to 800 people per 100,000, PD is one of the most common neurodegenerative disorders (Campenhausen, et al., 2005). PD is a movement disorder characterized by resting tremor, stiffness, slowing of movement, and loss of postural reflexes. Motor control disorder in PD involves motor processing planning, motor programming, motor sequencing, movement initiation and movement execution (Drotár, et al., 2016) (Contreras-Vidal & Stelmach, 1995). Vocal disorders do not appear suddenly. They are the result of a slow process whose initial stages may not be realized. For this reason, the development of early diagnosis and tele-monitoring systems with accurate, reliable and unbiased predictive models is very important for patients and research (Little, McSharry, Hunter, 2009) (Ruggiero, Sacile, & Giacomini, 1999). In the case of an assessment of speech disorders in Parkinson's patients, doctors and speech pathologists have adopted subjective methods based on acoustic cues to distinguish different disease states. To develop a more objective assessment, recent research uses sound quality measurements in time, spectral domains and cepstral to detect sound disturbances (Rani K & Holi, 2013) (Benba, Jilbab, Hammouch, 2014).

Data mining can be applied in the health sector for example diagnosing breast cancer, heart disease, diabetes and others (Larose, 2006). Genetic Algorithm is a better method for feature selection and parameter optimization. The best features selected for classification in the training dataset to classify cells (Mansoori, Suman, & Mishra, 2014). Genetic algorithm is one feature selection optimization algorithm. one of the selection processes is to take some of the best individuals. in addition, it can also be done with a proportional random sampling process, with proportions equal to the proportion of its quality (Sartono, 2010).

Neural Network is one of the many data mining analysis tools that can be used to make predictions of medical data (Karegowda, Manjunath, & Jayaram,

Ramadhani, Y., Mubarok, A., Hidayatullah, S. and Wiguna, W.
Attribute Optimization: Genetic Algorithms and Neural Network for Voice Analysis Classification of Parkinson's Disease.
DOI: 10.5220/0009947030743079
In *Proceedings of the 1st International Conference on Recent Innovations (ICRI 2018)*, pages 3074-3079
ISBN: 978-989-758-458-9

2011). Using neural network algorithms for cervical cancer cell classification shows that neural network algorithms have excellent performance for classification (Mariarputham & Stephen, 2015) (Ramdhani & Riana, 2017). Genetic algorithms are used to help classification algorithms in determining the attributes that must be used so as to increase the accuracy value. The use of Genetic Algorithms as a selection feature can improve predictive accuracy (Wahyuni, Sutojo, & Luthfiarta, 2014) (Ramdhani & Riana, 2017).

2 DATASET

Parkinson multiple sound recording dataset was performed with several types of voice recordings and several experiments were carried out with People with Parkinsonism through a doctor's examination process. During the collection of this dataset, 28 PD patients were asked to say "a" and "o" three times, each of which made a total of 168 records. The test group suffered from PD for 0 to 13 years and individual ages varied between 39 and 79. This dataset contains 1,040 records with 26 features in Table 1 manifested by all features contained in a dataset with two classes, namely healthy or unhealthy (Sakar, et al., 2013). This secondary data can be obtained from UCI Machine Learning which can be accessed via the page (https://archive.ics.uci.edu/ml/datasets).

3 METHODS

In this study the method used is to use Genetic Algorithm for selection features and Neural Network algorithm for classification.

3.1 Genetic Algorithm

Genetic algorithms (GAs) can be described as a heuristic search and optimisation technique that is inspired by natural evolution (McCall, 2005). Genetic Algorithms (GA) is heuristic method that is used to find the near-optimal solution in a large solution space.

Table 1: Features dataset of Parkinson's disease

Features	Group
Jitter (local) Jitter (local, absolute) (s) Jitter (rap) Jitter (ppq5) Jitter (ddp)	Frequency Parameters
Shimmer (local) Shimmer (local, dB) (dB) Shimmer (apq3) Shimmer (apq5) Shimmer (apq11) Shimmer (dda)	Amplitude Parameters
Mean autocorrelation (AC) Mean NHR Mean HNR	Hamronicity Parameters
Median pitch (Hz) Mean pitch (Hz) Standard deviation (Hz) Minimum pitch (Hz) Maximum pitch (Hz)	Pitch Parameters
Number of pulses Number of periods Mean period (s) Standard deviation of period (s)	Pulse Parameters
Fraction of locally unvoiced pitch frames Number of voice breaks Degree of voice breaks	Voicing Parameters

A population, i.e., a large number of chromosomes, is generated by some low computational approaches, such as random generation or greedy heuristics. In each iteration of the GA, the fitness values of all chromosomes in the population are evaluated, and the best chromosome is recorded. After a large number of iterations, the best chromosome in the population is translated as the selected solution (Qiu, Ming, Li, Gai,& Zong, 2015). Genetic Algorithms are algorithms used for search and optimization processes based on the principles of genetics and the process of natural selection. Genetic algorithms make a population consisting of many individuals who develop according to certain selection rules that have the optimization and provision of values.

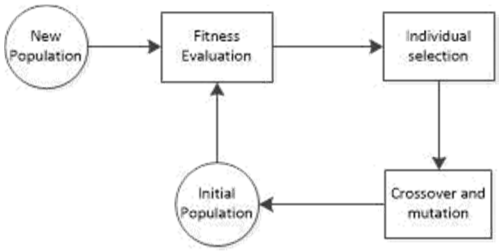

Figure 1. Genetic Algorithm Cycle (Imbar & Septiano, 2013)

Individuals state a possible solution. Individuals can be said to be the same as chromosomes, which are a collection of genes. Some important definitions that need to be considered in defining individuals to develop problem solving with genetic algorithms (Weise, 2009). In the genetic algorithm, there is a cycle that is performed to get the best optimization value or fitness value as described in Figure 1.

3.2 Neural Network

Neural networks or neural networks are a set of connected input / output units, where each connection has a weight. During the learning phase, the neural network adjusts the weight so that it can predict the correct class from the tupple (Han & Kamber, 2006). Information or input will be sent to neurons with a certain arrival weight, this input will be rated with the propagation function which will add up the values of all future weights. The sum result will be compared with the average threshold of some neuron functions (Kusumadewi, 2004). The learning process in backpropagation is done by adjusting the neuron weights with backward directions based on the error value in the learning process (Kusrini & Luthfi, 2009). To get an error, the forward propagation stage must be done first, when forward propagation, the neurons are activated by using an activation function that can be differentiated, such as a sigmoid function (Kusumadewi, 2004).

3.3 Proposed Method

In this study proposed a method for classification of Parkinson genetic algorithm as a selection feature and algorithm for neural network as a classification of Parkinson's disease. The proposed method can be seen in Figure 2.

The initial stage in this research is Normalization in the dataset with the aim of blocking data in a simple range using the z score transformation method.

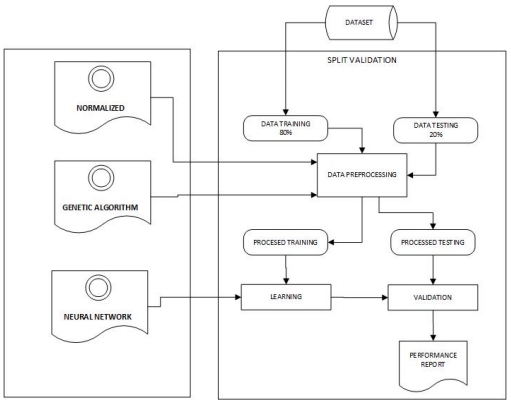

Figure 2: Proposed Method

The next step is done by separating training datasets and testing datasets using split validation method with data distribution of 80% training data and 20% testing datasets with data distribution on Table 1. Dataset training is applied to produce models from the Neural Network algorithm while for testing datasets to produce values accuracy.

Feature selection used in research using genetic algorithm. Genetic algorithm makes the population composed of many individuals selected with the most relevant values to classification so as to improve the performance of the accuracy value of classification of Parkinson's disease. Furthermore, features that have been selected by genetic algorithm are classified using the Neural Network algorithm. Classification results produce accuracy values and AUC (Area Under Curve).

In Figure 1, we describe the proposed method scheme for the study of Parkinson's disease classification. The results of the evaluation of the classification of Parkinson's disease with the proposed model have the maximum value results with feature optimization using genetic algorithms so as to influence the maximum results of calcifications carried out by the Neural Network algorithm. The multiple sound recordings dataset is classified into two classes, namely the healthy classification of Parkinson's and the class indicated by Parkinson's disease.

Table 2 Distribution of training data and testing data

Class Information	Dataset	Data Training	Data Testing
Sick PD	147	118	29
Healthy	48	39	9
Total	195	157	38

4 RESULT AND DISCUSSION

After the initial stage of the experiment was carried out classification using neural network algorithm obtained low results with accuracy results of 67.55% while for AUC values had a value of 0.74. Table 3 describes the classification results using other classification algorithms such as random forest algorithm, support vector machine, naïve bayes, and decision tree. It can be seen that the neural network algorithm has the highest accuracy value compared to other classification algorithms, because the neural network algorithm is suitable for classification of data that has a large record. While the algorithm that has the lowest classification is obtained by decision tree algorithm with an accuracy value of 54.09% and AUC value of 0.604, this is categorized as a poor classification value.

The classification results obtained by the Neural Network algorithm are still less than optimal so that feature optimization is performed using a genetic algorithm which is expected to increase the classification value. Before feature optimization there are 26 features contained in the dataset described in table 1, there may still be features that are less relevant from the dataset for classification.

The genetic algorithm produces features that are in accordance with the classification. Therefore, feature optimization was performed to increase the accuracy and AUC values for the classification of Parkinson's disease into two classes, namely healthy class or not identified Parkinson's disease and unhealthy Parkinson's disease. The selected features are described in Table 4, thus the features that were originally 26 attributes using the feature genetic algorithm selected into 13 selected attributes. Thus, the most relevant feature to classification is only 13 features. Next, the classification process of Parkinson's disease was carried out using a neural network algorithm by using the learning rate value of 0.5 and the momentum value of 0.5, then a feature optimization was performed using a genetic algorithm in the hope of increasing the classification results of Parkinson's disease.

Table 3 Classification Results

Algorithm	Accuracy	AUC
Neural Network	67.55%	0.74
Random Forest	58.41%	0.625
SVM	65.62%	0.723
Naïve Bayes	58.65%	0.604
Decision Tree	54.09%	0.549

The results obtained from the research that have been carried out have an accuracy value of 73.08% and AUC value of 0.794 thus there is an increase of the previous classification without using feature selection. Prior to optimizing the features, the accuracy value is 67.55% while the AUC value is 0.74 while after optimizing the feature selection optimization using genetic algorithms with neural network classification algorithm using learning rate value of 0.5 and momentum value 0.5 produces an accuracy value of 73.08% and AUC value of 0.74 thus a significant increase is seen after feature optimization.

Table 4 Selected features

No	Selected features
1	Jitter (ppq5)
2	Jitter (ddp)
3	Shimmer (local)
4	Shimmer (local, dB) (dB)
5	Shimmer (dda)
6	Mean autocorrelation (AC)
7	Mean NHR
8	Median pitch (Hz)
9	Mean pitch (Hz)
10	Standard deviation (Hz)
11	Minimum pitch (Hz)
12	Number of pulses
13	Mean period (s)

In Figure 3 it is explained about the increase in the accuracy value and the increasing value of AUC with a graphical form which shows a significant increase after it is done optimization.

Besides that, classification results are compared using random forest algorithms, support vector machine, naïve bayes, decision tree, with feature optimization using forward selection algorithm, backward elimination and greedy forward selection with an explanation in table 5. It is seen that neural network algorithms with algorithm feature selection genetics is superior to other algorithms. The highest value was obtained at 73.08% with an AUC value of 0.794 with a neural network classification algorithm with feature optimization using genetic algorithms. It can be seen that by using feature optimization of the classification algorithm has a significant effect with increasing the accuracy value of the classification results.

Table 5. Comparative results of classification of Parkinson's disease

Algorithm	Optimization Feature							
	Genetic Algorithm		Forward Selection		Backward Elimination		Greedy	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Neural Network	73.08%	0.794	67.31%	0.734	71.88%	0.778	65.87%	0.699
Random Forest	62.74%	0.666	60.82%	0.616	63.22%	0.644	58.89%	0.595
SVM	69.95%	0.753	63.46%	0.673	66.59%	0.712	63.46%	0.673
Naïve Bayes	64.66%	0.698	61.54%	0.618	62.98%	0.673	60.82%	0.648
Decision Tree	60.10%	0.603	54.33%	0.545	58.65%	0.609	54.33%	0.545

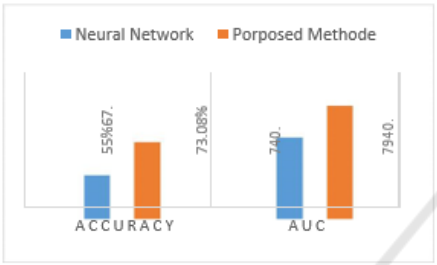

Figure 3. Comparison of Neural Network and Neural Network with genetic algorithm optimization

Whereas the lowest value of the classification of Parkinson's disease was produced by the Decision Tree algorithm with an accuracy value of 54.33% with an AUC value of 0.545 with a feature selection algorithm using the forward selection algorithm. Decision Tree classification algorithm has a very low value on the classification of Parkinson's disease despite feature optimization and before optimization. While other calcification algorithms after feature optimization are still not higher in classification values compared to the proposed algorithm.

5 CONCLUSION

The results obtained from the research that has been done, neural network algorithms have optimal values for classification of Parkinson's disease. After feature optimization using genetic algorithms, the accuracy of Parkinson's disease classification has a significant increase. Feature optimization has an important role in increasing the accuracy value. The results obtained from the classification values using genetic algorithms and feature optimization using neural network algorithms are still not optimal or are said to be good in the classification of Parkinson's disease. Therefore, it must be done to increase the accuracy value with classification algorithm optimization or with a hybrid model to increase the accuracy of Parkinson's disease so that the classification results can be used as an initial stage of classification to determine the classification of Parkinson's disease.

REFERENCES

Benba, A., Jilbab, A., & Hammouch, A. (2014). Voice analysis for detecting persons with Parkinson's disease using MFCC and VQ. *Recent Advances in Electrical Engineering and Computer Science*, 96-100.

Campenhausen, S. v., Bornschein, B., Wick, R., Botzel, K., Sampaio, C., Poewe, W., . . .

Dodel, R. (2005). Prevalence and incidence of Parkinson's disease in Europe. *European Neuropsychopharmacology* , 473 – 490.

Contreras-Vidal, J., & Stelmach, G. (1995). Effects of parkinsonism on motor control. *Life Sciences*, 165-176.

Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., & Faundez-Zanuy, M. (2016). Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artificial Intelligence in Medicine*, 39–46.

Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques Second Edition.* San Francisco: Morgan Kaufmann.

Imbar, R. V., & Septiano, K. (2013). Sistem HRD Perekrutan, Penggajian, dan Penjadwalan Menggunakan Algoritma Genetika pada Hotel Nirwana. *Jurnal Informatika*, 65-80.

Ishihara, L., & Brayne, C. (2006). A systematic review of depression and mental illness preceding Parkinson's disease. *Acta Neurologica Scandinavica*, 211-220.

Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 368-376.

Karegowda, A. G., Manjunath, A., & Jayaram, M. (2011). Application Of Genetic Algorithm Optimized Neural Network Connection Weights For Medical Diagnosis

Of Pima Indians Diabetes. *International Journal on Soft Computing (IJSC)*, 15-23.

Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining.* Yogyakarta: Penerbit Andi. Kusumadewi, S. (2004). *Membangun Jaringan Syaraf Tiruan Menggunakan Matlab & Excel Link.* Yogyakarta: Graha Ilmu.

Larose, D. (2006). *Data Mining Methods And Models.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Little, M. A., McSharry, P. E., & Hunter, E. J. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 1015 - 1022.

Mansoori, T. K., Suman, A., & Mishra, S. K. (2014). Feature Selection by Genetic Algorithm and SVM Classification for Cancer Detection. *International Journal of Advanced Research in Computer Science and Software Engineering Vol 4*, 357-365.

Mariarputham, E. J., & Stephen, A. (2015). Nominated Texture Based Cervical Cancer Classification. *Computational and Mathematical Methods in Medicine*, 1-10.

McCall, J. (2005). Genetic algorithms for modelling and optimisation. *Journal of Computational and Applied Mathematics*, 205-222.

O'Sullivan, S. B., & Schmitz , T. J. (2007). *Physical Rehabilitation 5th Edition.* Philadelphia: F. A. Davis Company.

Qiu, M., Ming, Z., Li, J., Gai, K., & Zong, Z. (2015). Phase-Change Memory Optimization for Green Cloud with Genetic Algorithm. *IEEE Transactions on Computers*, 3528 - 3540.

Ramdhani, Y., & Riana, D. (2017). Hierarchical Decision Approach Based on Neural Network and Genetic Algorithm Method for Single Image Classification of Pap Smear. *Informatic and Computing (ICIC).* Jayapura.

Rani K, U., & Holi, M. S. (2013). Automatic Detection of Neurological Disordered Voices Using Mel Cepstral Coefficients and Neural Networks. Point-of-Care Healthcare Technologies (PHT) (pp. 16-18). Bangalore, India: IEEE.

Ruggiero, C., Sacile, R., & Giacomini, M. (1999). Home telecare. Journal of Telemedicine and Telecare, 11-17.

Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., . . . Kursun, O. (2013). Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings. IEEE Journa*l of Biomedical and Health Informatics, 17*, 828-834.

Sartono, B. (2010). Pengenalan Algoritma Genetik Untuk Pemilihan Peubah Penjelas Dalam Model Regresi Menggunakan SAS/IML. *Forum Statistika dan Komputasi*, 10-15.

Wahyuni, D. T., Sutojo, T., & Luthfiarta, A. (2014). Prediksi Hasil Pemilu Legislatif DKI Jakarta Menggunakan Naïve Bayes Dengan Algoritma Genetika Sebagai Fitur Seleksi.

Weise, T. (2009). *Global Optimization Algorithms: Theory and Application, 2nd Edition.* Germany: Thomas Weise.