2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE) | 978-1-6654-0196-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICITISEE53823.2021.9655774

Feature Selection and Classification Algorithm Approaches for Parkinson's Disease Detection

1st Hendra Marcos
Doctoral Program Information System
Diponegoro University
Department Informatics
Universitas Amikom Purwokerto
Purwokerto, Indonesia
ORCID : 0000-0002-2284-2503

2nd Budi Warsito
Department of Statistics
Faculty of Science and Mathematics,
Diponegoro University
Semarang, Indonesia
email : budiwrst2@gmail.com

3rd Adi Wibowo
Department of Computer Science
Informatics Diponegoro University
Semarang, Indonesia
ORCID : 0000-0002-7966-1017

***Abstract*— Millions of people worldwide have developed Parkinson's disease (PD), which is more common by 50. Advances in information technology, especially in the field of machine learning, are being challenged for the development of early detection algorithms. This helps doctors to detect disease early and more accurately based on data. The available data has large dimensions and features so that the feature selection method can be used. This research paper aims to provide insight by comparing several algorithms for feature selection and PD classification. PD classification can help improve treatment efficiency and save time. The problem that arises is how the classification algorithm can show better and more efficient accuracy results. A large number of datasets is one of the challenges in the development of classification algorithms. The voice input dataset was obtained from the UCI machine learning repository. Bagging's algorithm out performed on average 100%, and MLP 96.58%, and RF 92.52%.**

Keywords—accuracy, algorithm, classification, detection, feature selection, parkinson's disease

I. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative illness that affects body movements, speech, and the neurological system. People living with Parkinson's may also experience mental and behavioral changes due to a brain chemical called dopamine. In 1817, Dr. James Parkinson discovered the ailment and used "shaking palsy" to describe it [1]. Other than Alzheimer's disease, brain cancer, degenerative neurological disease, and epilepsy, patients with Parkinson's disease had the second-highest risk of developing a neurodegenerative disease [2].

Impaired voice and motor function are the most acute symptoms whose incidence increases with the patient's age. Tremor, rigidity, and bradykinesia are the main symptoms of PD. Gait disturbances, burning in the eyes, dysphagia, signs of autonomic dysfunction, visual disturbances, pain, sensory complaints, and depression are other symptoms that affect a person [3]. The diagnosis of PD from symptoms at an advanced stage is easy and accurate, but requiring effective treatment is challenging. In addition, treatment started at an advanced stage is less effective at controlling the development of PD [4]. Therefore, early diagnosis is recommended for early treatment [5], [6].

Scientific research has developed rapidly in the last decade to diagnose several diseases, including PD disease early. Early diagnosis is developed using several machine learning-based algorithms. A suitable and robust machine learning model can assist clinicians in early detection. This goal can be achieved by many researchers working in this field. Prashanth et al. [7] demonstrated the efficacy of various classifiers for detecting PD, including SVM, RF, Naive Bayes, and Boosted Trees. SVM outperforms the other three classifiers in terms of performance. Naranjo et al. [8] suggested a PD classification model that minimizes processing time by utilizing Gibb's Sampling Algorithm and Bayesian Approach. However, the possible precision is not exceptionally high. Fayyazifar et al. [9] investigated ensemble modeling using AdaBoost and Bagging algorithms. They reduced the dimension of the feature collection using a genetic algorithm (GA). The authors of [10] used a speech input dataset to compare classifiers such as the MLP, SVM, KNN, and ANN with the Levenberg–Marquardt method.

Deep learning (DL) is a rapidly growing field for resolving health-related problems. Effective detection of PD has been demonstrated using DL-based techniques [11], [12]. Xiong et al. [13] used the adaptive Grey Wolf Optimization technique with a sparse autoencoder and an LDA model to boost detection performance. Ali et al. [14] used GA with NN and LDA to detect PD. The authors of [3] used grid-search and learning curve approaches to optimize the SVM, RF, and NN train parameters. The model's performance was evaluated using the Leave-One-Out cross-validation methodology. Finally, Almeida et al. [15] examined the effect of phonation on identifying Parkinson's disease.

The severity of Parkinson's disease (PD) determines the importance of automated diagnostic tools in recognizing it. If Parkinson's disease is discovered later, it might be fatal. Early diagnosis, on the other hand, dramatically improves the sick person's condition. Early diagnosis with the help of a computer requires data on the examination of PD patients. The dataset is obtained from UCI machine learning public data [16], where this dataset is the acoustic data of PD patients. The feature selection method that has been developed has its strengths and weaknesses. Reducing the data dimensions by selecting features can improve classification accuracy results [17]. Because of this, the focus of this work will be on feature selection using rough set theory, which processes multiple classification algorithms.

II. Review Of The Literature

A. *Parkinson's Disease (PD)*

Parkinson's Disease (PD, often referred to as idiopathic or primary parkinsonism, hypokinetic rigid syndrome/HRS, or paralysis agitans) is a degenerative neurological ailment that results in chronic movement disorders that deteriorate time. As illustrated in Figure 1, it is caused by the degeneration of dopamine-producing brain cells (neurons).

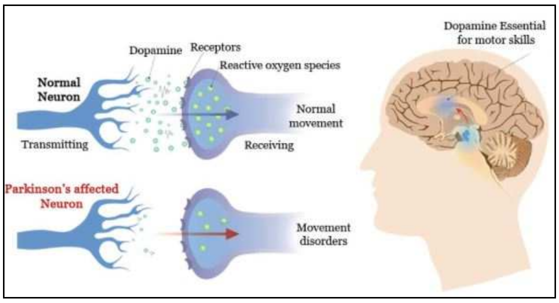

Fig. 1. Parkinson's disease [www.medindia.net]

As a result, the patient frequently suffers from uncontrollable shaking and tremors, delayed mobility, and difficulty in standing. Difficulty balancing and limb stiffness. It is primarily a problem for the middle-aged and elderly. The most frequent type of Parkinson's disease is idiopathic Parkinson's disease, which produces tremors, rigidity, and slowness of movement. Although there is no permanent treatment for this condition, medication and surgery can treat its symptoms. Dr. Richard Wade-Martins of the University of Oxford is researching on 'TRAPping Parkinson's disease-why do some cells die?' to determine why some cells die while others survive.

B. Rough set theory

The rough set theory (RST) is a popular mathematical tool for feature selection and the rule extraction [18], [19]. Let $I = (U, A)$ denote an information system, with U denoting a non-empty collection of finite objects known as the universe of discourse, and A denoting a non-empty collection of attributes. Every attribute with the letter A has a set of values (Va) associated with it. There is an associated equivalence relation IND (P) for a subset of attributes $P \subseteq A$, which is known as an indiscernability relation. Equation (1) can be used to define the relationship IND (P).

$$IND\,(P) = \{(x,y) \in U^2 \mid \forall\, a \in \mathrm{P},\, a\,(x) = a\,(y)\} \qquad (1)$$

If $(x,y) \in IND\ (P)$, then P attributes are unable to distinguish between x and y. $[x]p$ denotes the linear combinations of the P-indiscernibility connection. The RST's mathematical foundation is the indiscernibility relation. Within RST, the lower and upper approximations are two fundamental processes. For even a subset, $X \subseteq U$. X can be approached by building the P-lower approximation donated as $\underline{P}X$, which is also the set of all U items that can surely be categorized as X based on attribute set P. X as a P-upper approximate is indicated by $\overline{P}X$

$$\underline{P}X = \{X | [X]_p \subseteq \underline{P}X\}$$

$$\overline{P}X = \{X | [X]_p \cap X\, \underline{P} \neq \varphi\}$$

The RST chooses features with dependency of attributes and reduces the extraneous features picked by the RST are provided as input to the classification algorithm, where $\underline{P}X$ is portrayed as P-lower approximation and $\overline{P}X$ is portrayed as P-upper approximation [20].

C. Correlation-based feature

(CFS) [21] is a method which ranks the features with correlation-based criteria. CFS starts with an empty set and selects the features that are highly correlated with class labels (relevancy) and have low correlation with other selected features (redundancy). A well-known similarity metric between two attributes is a correlation. The correlation coefficient between two features is one if they are linearly dependent. The correlation coefficient is *0* if the features are uncorrelated. The correlation approach is used to determine the relationship between the features. The correlation between two random variables can be measured using two different methods. The first is based on linear correlation, whereas the second is based on information theory. The linear correlation coefficient is the most widely used of these two. According to conventional literature, the linear correlation coefficient 'r' for a pair of variables (X, Y) is calculated as follows :

$$r = \frac{\sum (X_i - \overline{X}_i)(Y_i - \overline{Y}_i)}{\sqrt{\sum (X_i - \overline{X}_i)^2}\sqrt{\sum (Y_i - \overline{Y}_i)^2}} \qquad (2)$$

D. Random Forest

Random forest [22] is really a popular supervised ensemble classification method that is also quite efficient. The ultimate result is the result of combining the results of a large group of decision tree classifiers. The feature set is randomly partitioned into subsets of features, each subset consists of features selected at random. For every individual decision tree, we choose a subset of cases and then use that subset to make a forecast. To come up with a random forest, you have to combine the trees, and majority vote is used to settle on the class label of the sample.

E. Artificial Neural Network (ANN)

ANN is a parallel design inspired by the way biological neural processing occurs. Although there are other ANN architectures, the MLP (multi-layer perceptron) is the most frequently employed (Fig. 2).

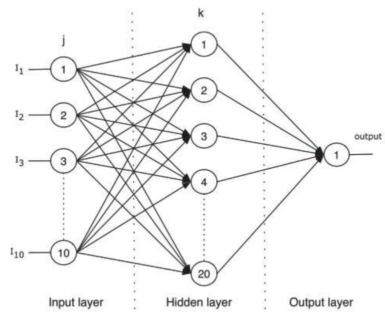

Fig. 2. Three-layer structure of a neuron [23]

MLP Networks alter weights using Rumelhart's 1986 backpropagation algorithm, which is a generalized delta rule. Levenberg–Marquardt, Neural network backpropagation gradient, and Resistant backpropagation are examples of backpropagation algorithms. The Levenberg–Marquardt algorithm is efficient and strongly recommended for neural network training on small and medium-sized networks, according to M.T. Hagan and M.Menhaj [24], hence the same methodology was used here.

It was initially envisaged that the computations of the backpropagation neural network would be performed using the so-called delta rule, which is also known as the steepest descendant training method, as the basis for its operation. The training sample data is transferred via an input layer, H hidden layers, and an output layer before returning to the training sample data in the traditional backpropagation methodology. An iterative backward pass of the sample is conducted to update the weight vector wki for all neurons I in layer k; as a result of this procedure, the term "backpropagation neural network" was coined. When all training samples have completed one forward and one backward pass before being submitted to the network, this is a single epoch. The parameter n represents the number of epochs that have been conducted. In a network with K layers, the output from layer k in the forward pass will be as follows: In a network with K layers, the output from layer k in the forward pass will be as follows: (Without taking into consideration the continuous bias term) [25].

Adam's optimizer is the approach we used to modify neuron weights iteratively in an epochal network. According to Wanjale et al. (2020) [26], Adam optimizer requires minor tuning and can also handle high variability data. The primary reason it was chosen above the others is that it uses virtually no resources to adapt to the neural network, which results in outstanding accuracy and sensitivity. Additionally, Adam optimizer considers the varied advantages of different optimizers and operates accordingly.

III. Material and Method

This section discusses the procedures and materials utilized to differentiate PD patients from healthy volunteers in this study.

A. Dataset

Naranjo et al. [8] conducted an experiment to duplicate voice recordings to distinguish PD people from healthy people. A total of 40 people with Parkinson's disease and 40 healthy people were analyzed. A total of 240 speech recordings were processed to extract 44 acoustic features, generating a 44-dimensional vector for each recording. The acquired features are classified into various categories according to whether they have a suitable formulation or not. This results in the following nine groups, four of which consist of a single characteristic, Table I is feature Information:

TABLE I. Feature Information

No	Feature	Description
1	ID	Subjects's identifier
2	Recording	Number of the recording
3	Status	Class 0=Healthy, 1=PD
4	Gender	0=Man, 1=Woman
5	Jitter	Relative (Jitter rel), absolute (Jitter abs), relative average perturbation (Jitter RAP), and pitch perturbation quotient (Jitter PPQ) are all pitch local perturbation measurements.
6	Shim	Local (Shim loc), shimmer in decibels (Shim dB), 3-point amplitude perturbation quotient (Shim APQ3), 5-point amplitude perturbation quotient (Shim APQ5), and 11-point amplitude perturbation quotient (Shim APQ11) are amplitude perturbation metrics.
7	HNR	Harmonic-to-noise ratio samples are done in the frequency ranges 0-500 Hz (HNR05), 0-1500 Hz (HNR15), 0-2500 Hz (HNR25), 0-3500 Hz (HNR35), and 0-3800 Hz (HNR38), respectively (HNR38)
8	MFCC	Order 0 to 12 Mel frequency cepstral coefficient-based spectrum measurements (MFCC0, MFCC1,..., MFCC12) and its variations are defined as follows: (Delta0, Delta1,..., Delta12)
9	RPDE	Recurrence period density entropy
10	DFA	Detrended fluctuation analysis
11	PPE	Pitch period entropy
12	GNE	Glottal-to-noise excitation ratio

B. Proposed Method

In this study, before classification, the dataset needs preprocessing first. The first step is the discretization of all numeric value attributes to be converted into categories or intervals. The discretization method used is boolean reasoning [27] because it is one of the static and supervised discretization algorithms to distinguish based on the boolean value of its class attribute.

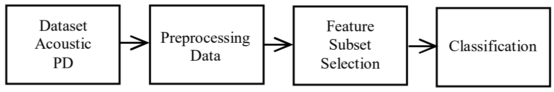

Fig. 3. Methods used in the classification of PD

In Figure 3, after preprocessing the data, the next step is to select features using rough set theory (RST) and Correlation-based feature (CFS). RST [18], [28] is a tool to compute the dependency between features (conditional attributes) and class labels (decision attributes). The dependency degree can measure relevance to rank features more relevant (dependent) with targets - the subset of features obtained using johnson's algorithm with complete reduct optimization. After the feature subset is obtained, the classification process is carried out using several classification algorithms by comparing the respective accuracy values. The supervised learning algorithm used to compare the classification results include Random Forest (RF), Artificial Neural Network (ANN) Multilayer Perceptron (MLP) model, k-Nearest Neighbor (k-NN), Bagging (BG), Naïve Bayes (NB).

IV. Result And Discussion

A. Result

Table I of the information system from the measurement dataset of 44 acoustic features of Parkinson's disease patients, where there is one decision feature (is "status"), 0 is not a disease (healthy) patient, and 1 is a patient with the disease (PD). The total results of measurements made are 240.

TABLE II. Dataset Voice Recordings To Distinguish PD

ID	Recording	Gender	Jitter_rel	Jitter_abs				Delta12	Status
CONT-01	1	1	0.255	0.000	.	.	.	1.355	0
PARK-39	2	0	0.576	0.000	.	.	.	1.109	1
PARK-39	3	0	0.233	0.000	.	.	.	1.142	1
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
PARK-40	1	0	0.269	0.000	.	.	.	1.456	1
PARK-40	2	0	0.454	0.000	.	.	.	1.307	1
PARK-40	3	0	0.347	0.000	.	.	.	1.423	1

The next process is that all data features with numeric values are discretized using a boolean reasoning algorithm. The numerical value features are Jitter, Shim, HNR, MFCC, Delta, RPDE, DFA, PPE, and GNE. The discretization results can be seen in Table III, where the value [*,0.54968) means Jitter_rel<0.54968, [0.54968,*) means Jitter_rel≥0.54968. For feature Delta_4 the value [*, 1.24596) means Delta_4<1.24596, [1.24596, 1.41213) means 1.24596≤Delta_4<1.41213, and [1.41213,*) means Delta_4≥1.41213, and so on for each result of the discretization of numeric features to be nominal. More details can be seen in Table IV.

TABLE III. DISCRETIZED DATASET

ID	Recording	Gender	Jitter_rel	.	.	.	Delta12	Status
CONT-01	1	1	[*, 0.54968)	.	.	.	[1.34710, *)	0
CONT-01	2	1	[*, 0.54968)	.	.	.	[*, 1.34710)	0
CONT-01	3	1	[*, 0.54968)	.	.	.	[*, 1.34710)	0
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
PARK-40	1	0	[*, 0.54968)	.	.	.	[1.34710, *)	1
PARK-40	2	0	[*, 0.54968)	.	.	.	[*, 1.34710)	1
PARK-40	3	0	[*, 0.54968)	.	.	.	[1.34710, *)	1

TABLE IV. RESULT VALUE FEATURE NUMERIK TO NOMINAL

No	Feature Numerik	Table of value	Count
1	Jitter_rel	[*,0.54968) [0.54968,*)	144 96
2	Jitter_abs	[*, 0.00004) [0.00004, *)	118 122
.	.	.	.
.	.	.	.
38	Delta_4	[*, 1.24596) [1.24596, 1.41213) [1.41213, *)	74 76 90
39	Delta_11	[*, 1.28537) [1.28537, *)	99 141
40	Delta_12	[*, 1.34710) [1.34710, *)	125 115

Next, Johnson's algorithm is reducted using rough set theory to get a feature subset with the number of selected features being 5. Finally, five feature subsets are tested to get classification results with several classification algorithms.

TABLE V. SUBSET FITUR SELECTION

No	Subset Fitur Selection	Method
1	{Jitter_abs, Shi_APQ11, HNR05, HNR25, GNE, MFCC1, MFCC2, MFCC3, MFCC5, Delta4}	RST_1
2	{Jitter_abs, Shi_APQ11, HNR25, DFA, GNE, MFCC1, MFCC4, MFCC6, Delta4, Delta5}	RST_2
3	{Jitter_abs, Shi_APQ11, HNR25, DFA, GNE, MFCC1, MFCC4, MFCC6, MFCC11, Delta4}	RST_3
4	{Jitter_rel, Shim_APQ5, Shi_APQ11, HNR25, DFA, GNE, MFCC1, MFCC5, MFCC6, Delta4}	RST_4
5	{Jitter_abs, HNR05, HNR25, HNR38, RPDE, GNE, MFCC1, MFCC3, MFCC5, Delta4}	RST_5
6	{HNR05, HNR35, Delta0, Delta2, Delta3, Delta5, Delta11, Delta12, MFCC3, MFCC6, MFCC4, MFCC10}	CFS-Subset Eval

When compared with other feature selection methods such as CFS-Subset Eval, the features obtained are 12 selected features {HNR05, HNR35, Delta0, Delta2, Delta3, Delta5, Delta11, Delta12, MFCC3, MFCC6, MFCC4, MFCC10}.

B. Discussion

Rough set theory (RST) feature selection resulted in 5 candidate feature subsets with ten features. In contrast, Correlation-based feature selection (CFS) resulted in 1 feature subset with 12 features out of a total of 40 initial features. Candidate features are also used to improve the learning model's predicted performance. The k-fold cross-validation procedure (k=10) was used for training and data evaluation. Accuracy, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were used to validate the model's results [27]. The formula for determining this metric is given below in the form of an equation. The necessity for several measures stems from the shortcomings of each one. In data with unbalanced classes, the accuracy metric fails to assess the model's performance accurately. The results of the classification algorithm model's performance are shown in Table VI.

TABLE VI. PERFORMANCE OF MODEL

No	Algorithm	Accuracy %				
		RST_1	*RST_2*	*RST_3*	*RST_4*	*RST_5*
1	MLP	97.08	96.25	98.58	98.87	97.18
2	NB	84.58	84.58	84.58	87.05	84.58
3	C45	74.17	74.17	78.33	79.15	80.42
4	RF	97.50	97.50	80.42	98.08	80.42
5	KNN	84.17	84.17	77.08	85.58	79.58
6	Bagging	100.00	100.00	100.00	100.00	100.00
7	AdaBoost	73.75	77.50	77.50	80.57	79.17
8	PART	74.58	78.33	78.33	78.35	77.91

TABLE VI. PERFORMANCE OF MODEL (EXT)

No	Algorithm	Accuracy %	
		CFS_Subset Eval	*PCA_rankker*
1	MLP	96.67	98.75
2	NB	86.25	94.58
3	C45	77.50	92.08
4	RF	97.08	100.00
5	KNN	84.58	100.00
6	Bagging	100.00	89.58
7	AdaBoost	79.17	92.58
8	PART	76.25	93.33

Algorithm with a high level of accuracy uses feature selection with the RST_4 algorithm, among the features are {Jitter_rel, Shim_APQ5, Shi_APQ11, HNR25, DFA, GNE, MFCC1, MFCC5, MFCC6, Delta4}, shown by meta-learning such as Bagging reaching 100%, then RF is 97.5%, and followed by MLP is 97.08%. Next, compared with other feature selection methods such as CFS-Subset Eval, Bagging results reach 100%, then RF is 97.08%, and followed by MLP is 96.67%. For PCA-ranker feature selection, each accuracy value is Bagging, reaching 89.58%, then RF at 100%, followed by MLP at 98.75%.

V. CONCLUSION

This experimental study aims to find discriminatory patterns of a sample of affected PD cases and control cases for accurate classification. The RST and CFS-Subset Eval feature selection methods are proposed because they have the mathematical roots of an exact decision table. The RST algorithm finds five optimal subsets with 11 practical features containing more visible factors and one feature subset from the CFS-Subset Eval algorithm. The proposed method proves its significance in the classification of PD to case controls. A significant accuracy value is influenced by the number of available features, the less the classification accuracy, the more significant the opportunity. The Bagging algorithm outperformed the average 100%, and MLP 96.58%, and RF 92.52%. Practical and in-depth computational models for diagnosing medical diseases are needed for early detection, and data availability is a significant factor. With more data, the learning model's performance will improve dramatically. In the future, data from a variety of PD patients with similar test criteria may help to build intelligent models for studying and understanding complex patterns.

ACKNOWLEDGMENT

This paper is an assignment for the Data Science for Business course supervised by Mr. Budi Warsito and Mr. Adi Wibowo. Thank you very much for the author's guidance so that this paper can be realized and Diponegoro University, where the author is studying in the Information Systems Doctoral program and especially to Amikom University Purwokerto, to the Chancellor and his staff. They have facilitated the author to continue doctoral studies.

REFERENCES

[1] J. Parkinson, "An essay on the shaking palsy," *J Neuropsychiatry Clin Neurosci*, vol. 14, no. 2. 2002, doi: 10.1007/s00702-017-1684-0.

[2] A. E. Lang and A. M. Lozano, "Parkinson's Disease," *N. Engl. J. Med.*, vol. 339, no. 16, pp. 1130–1143, Oct. 1998, doi: 10.1056/NEJM199810153391607.

[3] D. Braga, A. M. Madureira, L. Coelho, and R. Ajith, "Automatic detection of Parkinson's disease based on acoustic analysis of speech," *Eng. Appl. Artif. Intell.*, vol. 77, pp. 148–158, 2019, doi: https://doi.org/10.1016/j.engappai.2018.09.018.

[4] O. Yaman, F. Ertam, and T. Tuncer, "Automated Parkinson's disease recognition based on statistical pooling method using acoustic features," *Med. Hypotheses*, vol. 135, p. 109483, 2020, doi: 10.1016/j.mehy.2019.109483.

[5] G. Pahuja and T. N. Nagabhushan, "A Comparative Study of Existing Machine Learning Approaches for Parkinson's Disease Detection," *IETE J. Res.*, vol. 67, no. 1, pp. 4–14, 2021, doi: 10.1080/03772063.2018.1531730.

[6] R. Wahyudi, H. Marcos, U. Hasanah, B. P. Hartato, T. Astuti, and R. A. Prasetyo, "Algorithm Evaluation for Classification 'Phishing Website' Using Several Classification Algorithms," in *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, Nov. 2018, pp. 265–270, doi: 10.1109/ICITISEE.2018.8720975.

[7] R. Prashanth, S. Dutta Roy, P. K. Mandal, and S. Ghosh, "High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning.," *Int. J. Med. Inform.*, vol. 90, pp. 13–21, Jun. 2016, doi: 10.1016/j.ijmedinf.2016.03.001.

[8] L. Naranjo, C. J. Pérez, J. Martín, and Y. Campos-Roca, "A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications.," *Comput. Methods Programs Biomed.*, vol. 142, pp. 147–156, Apr. 2017, doi: 10.1016/j.cmpb.2017.02.019.

[9] N. Fayyazifar and N. Samadiani, "Parkinson's disease detection using ensemble techniques and genetic algorithm," in *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, Oct. 2017, pp. 162–165, doi: 10.1109/AISP.2017.8324074.

[10] G. Pahuja and T. N. Nagabhushan, "A Comparative Study of Existing Machine Learning Approaches for Parkinson's Disease Detection," *IETE Journal of Research*, vol. 67, no. 1. Taylor and Francis Ltd., pp. 4–14, 2021, doi: 10.1080/03772063.2018.1531730.

[11] V. J. Kadam and S. M. Jadhav, *Feature Ensemble Learning Based on Sparse Autoencoders for Diagnosis of Parkinson ' s Disease*. Springer Singapore.

[12] H. Gunduz, "Deep Learning-Based Parkinson's Disease Classification Using Vocal Feature Sets," *IEEE Access*, vol. 7, pp. 115540–115551, 2019, doi: 10.1109/ACCESS.2019.2936564.

[13] Y. Xiong and Y. Lu, "Deep Feature Extraction from the Vocal Vectors Using Sparse Autoencoders for Parkinson's Classification," *IEEE Access*, vol. 8, pp. 27821–27830, 2020, doi: 10.1109/ACCESS.2020.2968177.

[14] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated Detection of Parkinson's Disease Based on Multiple Types of Sustained Phonations Using Linear Discriminant Analysis and Genetically Optimized Neural Network," *IEEE J. Transl. Eng. Heal. Med.*, vol. 7, pp. 1–10, 2019, doi: 10.1109/JTEHM.2019.2940900.

[15] J. S. Almeida *et al.*, "Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques," *Pattern Recognit. Lett.*, vol. 125, pp. 55–62, 2019, doi: https://doi.org/10.1016/j.patrec.2019.04.005.

[16] L. Naranjo, C. J. PÃ©rez, Y. Campos-Roca, and J. MartÃ-n, "https://archive.ics.uci.edu/ml/datasets/Parkinson+Dataset+with+replicated+acoustic+features+." .

[17] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction : A review," *Egypt. Informatics J.*, vol. 19, no. 3, pp. 179–189, 2018, doi: 10.1016/j.eij.2018.03.002.

[18] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, 1982, doi: 10.1007/BF01001956.

[19] H. Marcos, N. Faisal, and H. A. Nugroho, "Rule Selection Based using Rough Set Theory for Diagnosis of Breast Cancer on Digital Mammogram Images," *ResearchGate*, no. October, 2015, [Online]. Available: https://www.researchgate.net/publication/319163459_Rule_Selection_Based_using_Rough_Set_Theory_for_Diagnosis_of_Breast_Cancer_on_Digital_Mammogram_Images.

[20] Z. Pawlak, "Rough set theory and its applications," *J. Telecommun. Inf. Technol.*, vol. nr 3, pp. 7–10, 2002.

[21] M. A. Hall, "Correlation-based Feature Selection for Machine Learning," no. April, 1999.

[22] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[23] F. Åström and R. Koker, "Expert Systems with Applications A parallel neural network approach to prediction of Parkinson ' s Disease," vol. 38, pp. 12470–12474, 2011, doi: 10.1016/j.eswa.2011.04.028.

[24] V. N. Vapnik, "The Nature of Statistical Learning Theory." Springer New York, 1995, doi: 10.1007/978-1-4757-2440-0.

[25] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. USA: Prentice Hall PTR, 1998.

[26] K. Wanjale, M. Nagapurkar, P. Kaldate, O. Kumbhar, and S. Bala, "Artificial Neural Network to Prescient the Severity of Parkinson's Disease," *Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020*, no. Iciccs, pp. 174–179, 2020, doi: 10.1109/ICICCS48265.2020.9121087.

[27] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. Cham: Springer International Publishing, 2015.

[28] H. Marcos and H. S. Utomo, "Perbandingan Kinerja Algoritme C.45 Dan Naive Bayes Mengklasifikasi Penyakit Diabete," *J. Inform. Darmajaya*, vol. 15, no. 2, pp. 141–148, 2015.