Journal of Natural Sciences and Mathematics Research
J. Nat. Scien. & Math. Res. Vol. 9 No. 1 (2023) 30-38
ISSN: 2460-4453 (Online); 2614-6487 (Printed)
Available online at http://journal.walisongo.ac.id/index.php/jnsmr

Early Detection Model of Parkinson's Disease Using Random Forest Method on Voice Frequency Data

Nurul Rifqah Fahira[1*], Armin Lawi[2], Masjidil Aqsha[3]
[1] Department of Biology, Hasanuddin University, Indonesia
[2] Information Systems Study Program, Hasanuddin University, Indonesia
[3] Department of Statistics, Hasanuddin University, Indonesia

* Corresponding author's e-mail: fahiranr.19h@student.unhas.ac.id

ABSTRACT

Parkinson's disease is the most common nervous system disease that affects all ethnicities, genders, and ages, with a higher prevalence in the elderly and men. Developing countries tend to have higher cases of Parkinson's. The prevalence of death due to Parkinson's in Indonesia reaches the fifth highest cases in Asia and 12th in the world. This neurodegenerative disease affects a person's ability to control movement. Currently, the diagnosis of Parkinson's disease is only based on observation of motor symptoms. Therefore, early detection of the disease cannot be done. His paper proposes an efficient way to detect Parkinson's disease symptoms by comparing the fundamental frequencies of patients' voices using the random forest method. Random forest is a Machine Learning method that applies the ensemble concept, which aims to improve the performance of the classification by combining several decision trees as a basis. Random forests have shown superior algorithm performance in numerous health studies. In this study, the dataset consisted of 20 patients with Parkinson's and 20 normal patients. Data for each patient was taken from 26 types of voice records, and thus, the total data was 1,040 observations. The obtained data is prepared by filtering and rescaling. Then, the data is split and modelled using the Random Forest Method. The random forest model obtained accuracy results of 72.50%, precision (normal) of 72.28%, precision (Parkinson's) of 72.73%, sensitivity (normal) of 73.00%, sensitivity (Parkinson's) of 72.00% and AUC is 80.70%. The built random forest model is quite good at Parkinson's disease detection.

Keywords:
Parkinson's Disease; Voice Frequency; Classification; Machine Learning; Random Forest.

Introduction

Parkinson's is the most common neuron deficiency disease after Alzheimer's (Solana-Lavalle et al., 2020). The disease can occur in different ethnicities, genders, and ages, with a higher prevalence in the elderly and men. Data from 1990-2016 showed increased mortality and disability rates in people with Parkinson's (WHO, 2022; Schiess et al., 2022). Parkinson's affects only 1-2 people out of 1000 but will double by 2040 (Dorsey & Bloem, 2018). Developing countries have higher cases of Parkinson's because the services and treatment provided to patients are still inadequate (Lestari et al., 2022). Previous research by Wikandikta et al. (2020) shows that the poor treatment of Parkinson's is predicted to be correlated to the high prevalence of deaths due to Parkinson's in Indonesia, with the fifth highest prevalence in Asia.

Parkinson's mainly occurs due to the progressive loss of cells in the substantia nigra, which is important in producing dopamine (Pyatha et al., 2022). The imbalance between dopamine and acetylcholine causes uncontrolled nerve activation. People with Parkinson's frequently complain about movement symptoms, such as resting tremors in which the limbs move uncontrollably (dyskinesia), postural instability, slowness of movement (bradykinesia), and muscle rigidity (Solana-Lavalle et al., 2020). Unfortunately, 60-80% of striatal dopamine neurons are lost when these motor symptoms are detected (Mantri et al., 2019). The degeneration of midbrain

mesencephalic dopamine neurons and the basal ganglia triggers cellular and synaptic alterations. These impulses are believed to be responsible for the motor symptoms of Parkinson's disease (Mallet et al., 2019).

Over the past few decades, technological developments have encouraged the digital storage of information in databases and increased exponentially, doubling monthly (Pynam et al., 2018). Machine learning applies artificial intelligence and statistics to process information from the past without the need for programming for each case. Machine learning is very promising in the effectiveness of data processing by minimizing human work (Faid et al., 2019). An integrative approach to systems biology and medicine will likely be a central computational strategy for generating new knowledge in biology and medicine (Zitnik et al., 2019). RapidMiner is a popular data mining software that can analyze data to be applied in research, education, and training. This software was developed by Ralf Klinkenberg, Ingo Mierswa, and Simon Fischer (Pynam et al., 2018). RapidMiner requires a shorter processing time than other machine learning software (Faid et al., 2019; László & Ghous, 2020).

One of the machine learning algorithms that shows superior performance in various research is the Random Forest proposed by Breiman (László & Ghous, 2020). This algorithm is a development of a decision tree (Hadiprakoso et al., 2022) that combines bootstrap aggregating (bagging) and random feature selection methods (Mantri et al., 2019). Within the random forest framework, multiple classifiers and regression trees are built using a randomly selected training dataset and a random subset of the predictor variables to model the results. The results from each tree are aggregated to predict each observation (Speiser et al., 2019).

Random forest algorithms have been widely used to detect diseases. Previous research by Triyono et al. (2021) showed the advantages of the Random Forest algorithm with Information Gain, with low error rates and high accuracy in large training datasets. The research of Fauzi et al. (2020) on breast cancer detection obtained the highest accuracy value produced by the Random Forest algorithm of 79.3103% with an AUC value of 0.843. Similar results were also obtained in the predictions of diabetes (Astuti et al., 2022), Alzheimer's (Akbar & Rahmaddeni, 2022), and risk classifications of babies born with low body weight (Yuliati & Sihombing, 2021). In this paper, the Random Forest method is tested to detect Parkinson's symptoms of vocal disorder based on a comparison of voice frequency data of normal people and people with Parkinson's. The dataset used is from the University of California Machine Learning Repository. This dataset consists of various biomedical voice measurements from 31 people: 23 people with Parkinson's disease and eight normal people.

Method

Workflow Diagram

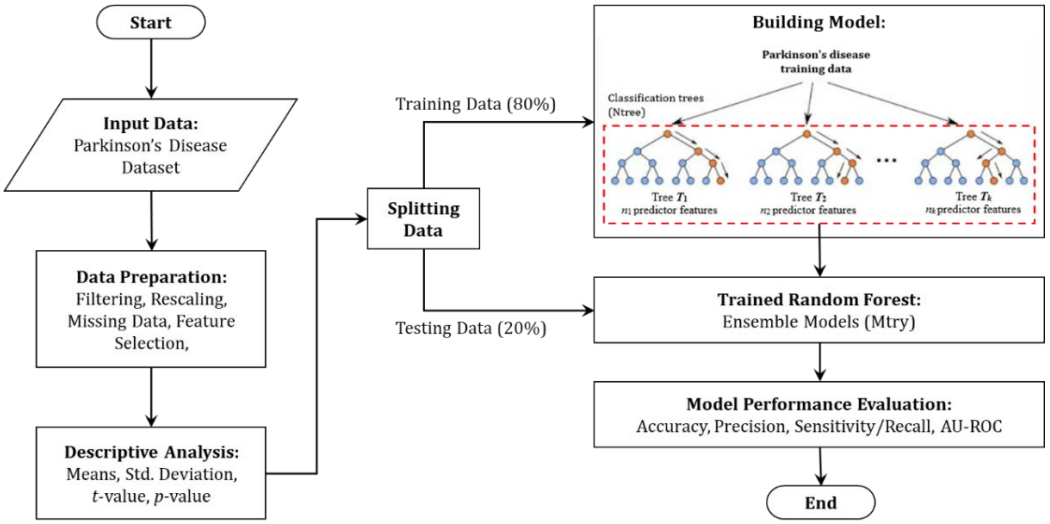

Figure 1. Workflow Diagram

Research workflow is conducted in 6 stages, as illustrated in Figure 1 as follows:

1. Input Data. The collection of voice data for people with Parkinson's disease is taken from the UCI Machine Learning Repository.
2. Data Preparation. This stage is carried out by filtering incomplete data and rescaling it to produce compatible data.
3. Descriptive Analysis.
4. Building Models. Training data is used for modelling the Random Forest. Model selection is done by considering the GridSearch and OOB-error values.
5. Trained Random Forest.
6. Performance Evaluation. Test data is used to test the performance of a model with a series of measurements, such as accuracy, precision, sensitivity, and ROC.

Data Source

The dataset used in this paper is secondary data from 40 patients, i.e., 20 normal and 20 patients with Parkinson's disease, respectively. Data were collected at the Department of Neurology, Cerrahpasa Medical Faculty, Istanbul University. The data was collected in different ways by recording each patient's voice, with 26 types found in 1,040 observations. The data has 26 predictors that are used to determine the patient's Parkinson's status carried out by expert doctors (Sakar et al., 2013). The data can be accessed on the UCI Machine Learning Repository website via the link https://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings. Table 1 describes the feature variables of Parkinson's disease in the dataset.

Table 1. Parkinson's Disease Dataset Description

Variable	Information	Data Type
$X_1 - X_5$	Jitter (local), Jitter (local, absolute), Jitter (rap), Jitter (ppq5), Jitter (ddp)	Numeric
$X_6 - X_{11}$	Shimmer (local), Shimmer (local, dB), Shimmer (apq3), Shimmer (apq5), Shimmer (apq11), Shimmer (dda)	Numeric
$X_{12} - X_{14}$	AC, NTH, HTN	Numeric
$X_{15} - X_{19}$	Median pitch, Mean pitch, Standard deviation, Minimum pitch, Maximum pitch	Numeric
$X_{20} - X_{23}$	Number of pulses, Number of periods, Mean period, Standard deviation of period	Numeric
$X_{24} - X_{26}$	Fraction of locally unvoiced frames, Number of voice breaks, Degree of voice breaks	Numeric
Y	Parkinson Status	Categoric

Data Preparation

Data preparation is carried out with the replace missing value operator, which can replace the loss of value with the maximum, minimum, or average value of the corresponding attribute so that the data is more compatible with the selected algorithm. In this stage, the data is filtered to remove data with incomplete information so that the data entered in the analysis process is only data that has complete information. In addition, rescaling was carried out so that the range of values in each variable was the same and variable selection was carried out using statistical t-tests to see the relationship between response variables (binary categorization) and numerical predictors.

Descriptive Analysis

Detection of anomalous or unusual variability of predictor data can be known through predictor dispersion analysis. Observation of standard deviation values is the most common and widely used analysis in determining the Degree of dispersal because it describes the spread of each unit of observation.

Building Models

The data is divided into two parts, including training data and test data, with a proportion of 80% training data and 20% test data. The training data constitutes 80% of 840 data, comprising 420 Parkinson's class data and 420 normal class data. The training data is used to build the model using the Random Forest method. The Random Forest method works by forming bootstrapped data that is randomly taken from a predetermined number of training data. Then, the bootstrapped data is used to build the classification trees that allow data classification through voting. However, the classification trees are built in a slightly different way from the classification trees in the community. In the Random Forest method, the classification tree is formed by only considering some randomly selected predictors. The Number of predictors selected randomly in building a classification tree can be set to obtain the best model.

Model Performance Measurements

The obtained random forest model was tested on test data by considering some measurement parameters, such as accuracy, precision, sensitivity, and ROC (Receiver Operating Curve) to measure the model's performance. Accuracy will show the accuracy of the Random Forest Algorithm in predicting instants. Precision is used to measure the correctness/accuracy of classifiers. The recall is used to measure the completeness or sensitivity of the classifier. At the same time, ROC is a representative of sensitivity and specificity, which measure the classification ability of Random Forest.

Following the approach of Zulfahmi et al. (2023), the evaluation of the model is performed by generating a Confusion Matrix table comprising the Number of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) data. By analyzing the confusion matrix table, it is possible to calculate the recall and precision values. The Recall value represents the ability of the model to identify the true positive class, which can be calculated as shown in equation (1)

$$\text{Recall} = \frac{\text{TP}}{(\text{TP}+\text{FN})} \tag{1}$$

The Precision value measures the model's correct positive predictive ability. The precision calculation is carried out using equation (2)

$$\text{Precision} = \frac{\text{TP}}{(\text{TP}+\text{FP})} \tag{2}$$

Accuracy can be calculated using equation (3)

$$\text{Accuracy} = \frac{\text{TP}+\text{FP}}{\text{N}} \tag{3}$$

Results and Discussions

Data Preparation Result

Parkinson's disease data consists of 1.040 observations and 26 predictors. After a thorough check, it has been confirmed that there is no incomplete data. Therefore, the filtering process is unnecessary. Then, the variables in the Parkinson's disease data were transformed to ensure that the value range remained consistent. The feature selection process was carried out using T-test statistics to see the relationship between predictors and the response variable. Predictors that do not have a significant relationship with the response variable are disregarded, as it was assumed that the predictors did not contribute enough to separate data in the response variable. The performance is evaluated by comparing the conditions of the first research object with the conditions of the object in the second study. The following are the results of the T-test statistics from predictors with response variables in the data of Parkinson's disease:

Table 2. Feature Selection using t-statistic test (α = 0.05)

Variable	*t*-value	*p*-value	Variable	*t*-value	*p*-value
X_1	-3.158	0.002	X_{14}	-1.306	0.192
X_2	-5.457	0.000	X_{15}	2.723	0.007
X_3	-3.603	0.000	X_{16}	3.345	0.001
X_4	-3.596	0.000	X_{17}	3.930	0.000
X_5	-3.603	0.000	X_{18}	1.908	0.057
X_6	-0.292	0.770	X_{19}	4.535	0.000
X_7	-0.876	0.381	X_{20}	-1.323	0.186
X_8	0.585	0.559	X_{21}	-1.519	0.129
X_9	0.728	0.467	X_{22}	-2.643	0.008
X_{10}	-4.766	0.000	X_{23}	1.897	0.058
X_{11}	0.585	0.559	X_{24}	3.962	0.000
X_{12}	-2.621	0.009	X_{25}	2.580	0.010
X_{13}	2.458	0.014	X_{26}	4.097	0.000

Based on Table 2, it can be seen that the predictors X_6, X_7, X_8, X_9, X_{11}, X_{14}, X_{18}, X_{20}, X_{21} and X_{23} do not have a significant relationship with the variable response (Parkinson's status). Therefore, these predictors were excluded from the analysis. Consequently, the analysis was performed using the remaining 16 predictors.

Descriptive Analysis Result

Analysis of the standard deviation data from each predictor was conducted to identify variability in specific symptoms associated with Parkinson's disease. A high standard deviation value indicates a wider distribution of data, which in turn suggests variation in Parkinson's disease symptoms. The following are the results of the descriptive analysis of Parkinson's disease predictors:

Based on Table 3, a very high standard deviation is obtained for predictors x15, x16, x19, x20, and x21. This indicates a significant risk of deviation, suggesting variations in symptoms or health parameters in the population. Previous research by Suphinnapong et al. (2021) has identified several voice disorder symptoms in individuals with Parkinson's disease. These symptoms include fundamental frequency standard deviation, jitter, fundamental frequency variations, PPQ (pitch perturbation quotient), sPPQ (smooth pitch perturbation quotient), shimmer, DUV (Degree of voiceless), NHR (noise to harmonic ratio), VTI (sound turbulence index), SPI (soft phonation index), and others. However, the role of an individual's average voice pitch as an acoustic parameter in Parkinson's disease is limited. Gender is another factor that causes variation in voice pitch, with women generally having a higher pitch than men. Therefore, measurements on an absolute frequency scale in Hz can experience deviations as a marker of Parkinson's disease.

Resulted in Random Forest Models

The determination of the optimal value of hyperparameters can be evaluated with GridSearch by considering various indices such as Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE) (Sakar et al., 2013). Grid search is applied to select the best parameters from the training sample for further testing.

In the Random Forest method, there are two main parameters used to build models, including the Number of classification trees (Ntree) and the number of variables used to build randomly selected models (Mtry). To minimize errors in random forests and validate the optimal number of Ntree, the calculation of the OOB error value is performed. The process of determining the out-of-bag (OOB) error value involves generating probability predictions for each training

observation i (where i = 1, ... , n) that is not included in the bootstrap sample. Subsequently, these predictions are compared to the corresponding actual values by calculating performance measures. This methodology proves to be an effective means of assessing the accuracy of a model without requiring a separate validation set (Probst & Boulesteix, 2018). Here is an out-of-bag (OOB) value error from the random forest model with different Ntree and Mtry value combinations.

Table 3. Descriptive Analysis of Parkinson's Disease Variables

Variable	Range	Mean	Standard Deviation
X_1	0.19 -14.376	2.68	1.765
X_2	6.495e-06 - 0.000776606	0	0
X_3	0.062 – 8.016	1.247	0.979
X_4	0.081 – 13.542	1.348	1.139
X_5	0.185 – 24.048	3.741	2.938
X_6	1.185 – 41.137	12.918	5.452
X_7	0.103 – 2.721	1.195	0.42
X_8	0.496 – 25.82	5.7	3.015
X_9	0.708 – 72.86	7.984	4.841
X_{10}	0.517 – 44.764	12.215	6.016
X_{11}	1.488 – 77.459	17.099	9.046
X_{12}	0.539566 - 0.997904	0.846	0.086
X_{13}	0.002106 - 0.869277	0.231	0.151
X_{14}	0.695 – 28.418	10	4.291
X_{15}	81.46 - 468.618	163.368	56.022
X_{16}	82.363 - 470.456	168.728	55.97
X_{17}	0.533 - 293.877	27.548	36.673
X_{18}	67.957 – 452.083	134.538	47.058
X_{19}	85.541 – 597.974	234.876	121.541
X_{20}	0 – 1490	109.744	150.028
X_{21}	0 – 1489	105.969	149.417
X_{22}	0.002038552 – 0.012070196	0. 007	0.002
X_{23}	5.5347e-05 - 0.006371201	0.001	0.001
X_{24}	0 – 88.158	27.683	20.975
X_{25}	0 -12	1.135	1.615
X_{26}	0 – 69.117	12.37	15.162
Y	1 – 55	13	15.895

Table 4. Parameter Tuning with GridSearch (OOB Error)

		Ntree					
		10	**25**	**50**	**100**	**250**	**500**
Mtry	3	0.4081	0.3923	0.3295	0.3534	0.3410	0.3367
	5	0.3924	0.3823	0.3532	0.3444	0.3337	0.3293
	7	0.3778	0.3726	0.3401	0.3469	0.3337	0.3192
	9	0.3725	0.3600	0.3476	0.3421	0.3337	0.3231
	11	0.3901	0.3684	0.3462	0.3409	0.3273	0.3224
	13	0.3854	0.3619	0.3535	0.3407	0.3394	0.3177
	15	0.4062	0.3576	0.3427	0.3396	0.3291	0.3180

Based on Table 4, it can be seen that the random forest model with Ntree = 500 and Mtry = 13 has the lowest out-of-bag error value. The model with the lowest OOB error value is selected

as the best data model to evaluate with some measure.

Table 5 displays that the random forest model acquired an accuracy of 72.50%, with precision (normal) of 72.28%, precision (Parkinson's) of 72.73%, sensitivity (normal) of 73.00%, and sensitivity (Parkinson) of 72.00%. The accuracy value indicates that the Random Forest model is quite efficient and has accurately predicted 72.50% of the test data. Additionally, the performance of the random forest model can be evaluated based on the following AUC graph:

Table 5. Confusion Matrix from Testing Data

		Reference		Precision
		Normal	**Parkinson**	
Prediction	**Normal**	73	28	**72.28%**
	Parkinson	27	72	**72.73%**
Recall/ Sensitivity		**73.00%**	**72.00%**	72.50% **(Accuracy)**

Performance Evaluation

The efficacy of the top-performing random forest model is evaluated by analyzing key performance indicators such as sensitivity, precision, and accuracy. The confusion matrix table below displays the evaluation results of the best model from test data:

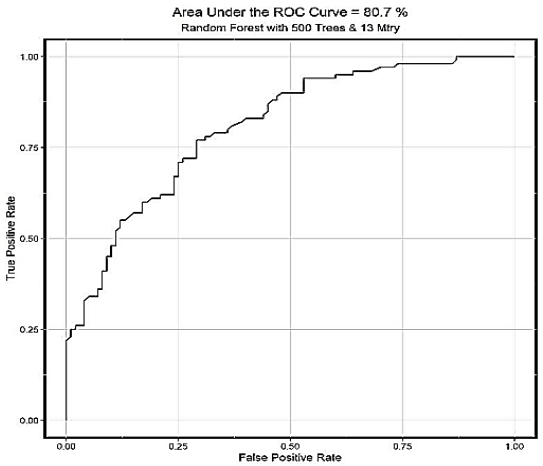

Figure 2. Receiver Operating Characteristic (ROC) Curve

According to Figure 2, the AUC value of the random forest model obtained is 80.70%. Based on previous research, an AUC value of 80.7% is a good criterion for diagnostic tests in general (Dahlan, 2009; Hastuti, 2012; Maskoen, et al., 2017; Alhabib, 2022). According to Janssens & Martens (2020), the rounded ROC curve represents a similar effect on disease risk on the variables contained in the prediction model. This demonstrates that the random forest model is quite effective at classifying or predicting Parkinson's disease. However, when dealing with smaller datasets, there is always a possibility of inaccuracies in predictions.

Conclusion

This paper represents the Machine Learning method for detecting Parkinson's disease based on sound frequency data using the Random Forest method. The process of preparing data shows that predictors x6, x7, x8, x9, x11, x14, x18, x20, x21, and x23 do not have a significant impact on the variable response (Parkinson's status) and are excluded from the analysis. By using the 16 remaining predictors, Parkinson's data is modelled using the Random Forest method. The random forest model obtained accuracy results of 72.50%, precision (normal) of 72.28%, precision (Parkinson's) of 72.73%, sensitivity (normal) of 73.00%, sensitivity (Parkinson's) of 72.00% and AUC of 80.70%. These results demonstrate that the random forest

model is quite effective at classifying Parkinson's disease based on sound frequency data. In future research, it is expected that different data preparation techniques and classification methods will be used to provide a clear picture of the Machine Learning method suitable for handling cases of disease classification based on sound frequency data, especially Parkinson's disease.

Acknowledgements

Hasanuddin University supported this work for its service and facilities.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

Akbar, F., & Rahmaddeni. (2022). Komparasi Algoritma Machine Learning Untuk Memprediksi Penyakit Alzheimer. *Jurnal Komputer Terapan*, *8*(2), 236–245.

Alhabib, I., Faqih, A., & Dikananda, F. (2022). Komparasi Metode Deep Learning, Naïve Bayes Dan Random Forest Untuk Prediksi Penyakit Jantung. *INFORMATICS FOR EDUCATORS AND PROFESSIONAL: Journal of Informatics*, *6*(2), 176–185. https://doi.org/10.51211/itbi.v6i2.1881

Astuti, Widya, L., Yulianti, E., & Dharmayanti, D. (2022). Feature Selection Menggunakan Binary Wheal Optimizaton Algorithm (Bwoa) Pada Klasifikasi Penyakit Diabetes. *Jurnal Ilmiah Informatika Global*, *13*(1), 5–12.

Dahlan, M. S. (2009). Penelitian Diagnostik. Jakarta: *Salemba Medika.*

Dorsey, E. R., & Bloem, B. R. (2018). The Parkinson Pandemic—A Call to Action. *JAMA Neurology*, *75*(1), 9–10. https://doi.org/10.1001/jamaneurol.2017.3299

Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, *8*(1), 11-16. https://doi.org/10.34148/teknika.v8i1.95

Fauzi, A., Supriyadi, R., & Maulidah, N. (2020). Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis Principal Component Analysis dan Random Forest. *Jurnal Infortech*, *2*(1), 96-101. https://doi.org/10.31294/infortech.v2i1.8079

Hadiprakoso, R. B., Aditya, W. R., & Pramitha, F. N. (2022). Analisis Statis Deteksi Malware Android Menggunakan Algoritma Supervised Machine Learning. *Cyber Security Dan Forensik Digital*, *5*(1), 1-5. https://doi.org/10.14421/csecurity.2022.5.1.3116

Hastuti, K. (2012). Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non Aktif. *Semantik*, *2*(1), 241-249.

Janssens, A. C. J. W., & Martens, F. K. (2020). Reflection on modern methods: Revisiting the area under the ROC Curve. *International Journal of Epidemiology*, *49*(4), 1397–1403. https://doi.org/10.1093/ije/dyz274

László, K., & Ghous, H. (2020). Efficiency comparison of Python and RapidMiner. *Multidiszciplináris Tudományok*, *10*(3), 212-220. https://doi.org/10.35925/j.multi.2020.3.26

Lestari, D. T., Harahap, H. S., Sahidu, M. G., Putri, S. A., Gunawan, S. E., Susilowati, N. A., & Hunaifi, I. (2022). Edukasi Deteksi Dini Penyakit Parkinson Pada Kader Puskesmas Dalam Rangka Hari Parkinson Sedunia. *Jurnal Abdi Insani*, *9*(3), 1012-1018. https://doi.org/10.29303/abdiinsani.v9i3.714

Mallet, N., Delgado, L., Chazalon, M., Miguelez, C., & Baufreton, J. (2019). Cellular and Synaptic Dysfunctions in Parkinson's Disease: Stepping Out of the Striatum. *Cells*, *8*(9), 1-29. https://doi.org/10.3390/cells8091005

Mantri, S., Morley, J. F., & Siderowf, A. D. (2019). The importance of preclinical diagnostics in Parkinson disease. *Parkinsonism & Related Disorders*, *64*, 20–28. https://doi.org/10.1016/j.parkreldis.2018.09.011

Maskoen, T. T., Masthura, A., & Suwarman. (2017). Nilai Area Under Curve dan Akurasi Neutrophil Gelatinase Associated Lipocalin untuk Diagnosis Acute Kidney Injury pada Pasien

Politrauma di Instalasi Gawat Darurat RSUP dr. Hasan Sadikin Bandung. *Anestesia dan Critical Care*, *35*(3), 158-164.

Probst, P., & Boulesteix, A.-L. (2018). To Tune or Not to Tune the Number of Trees in Random Forest. *Journal of Machine Learning Research*, *18*, 1–18.

Pyatha, S., Kim, H., Lee, D., & Kim, K. (2022). Association between Heavy Metal Exposure and Parkinson's Disease: A Review of the Mechanisms Related to Oxidative Stress. *Antioxidants*, *11*(2467), 1-18

Pynam, V., Spanadna, R., & Srikanth, K. (2018). An Extensive Study of Data Analysis Tools (Rapid Miner, Weka, R Tool, Knime, Orange). *International Journal of Computer Science and Engineering*, *5*(9), 4–11. https://doi.org/10.14445/23488387/IJCSE-V5I9P102

Sakar, B. E., Issenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., & Kursun, O. (2013). Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings. *IEEE Journal of Biomedical and Health Informatics*, *17*(4), 828–834. https://doi.org/10.1109/JBHI.2013.2245674

Schiess, N., Cataldi, R., Okun, M. S., Fothergill-Misbah, N., Dorsey, E. R., Bloem, B. R., Barretto, M., Bhidayasiri, R., Brown, R., Chishimba, L., Chowdhary, N., Coslov, M., Cubo, E., Di Rocco, A., Dolhun, R., Dowrick, C., Fung, V. S. C., Gershanik, O. S., Gifford, L., … Dua, T. (2022). Six Action Steps to Address Global Disparities in Parkinson Disease: A World Health Organization Priority. *JAMA Neurology*, *79*(9), 929–936. https://doi.org/10.1001/jamaneurol.2022.1783

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. https://doi.org/10.1016/j.bbe.2020.01.003

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, *134*, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028

Suphinnapong, P., Phokaewvarangkul, O., Thubthong, N., Teeramongkonrasmee, A., Mahattanasakul, P., Lorwattanapongsa, P., & Bhidayasiri, R. (2021). Objective vowel sound characteristics and their relationship with motor dysfunction in Asian Parkinson's disease patients. *Journal of the Neurological Sciences*, *426*, 1-8.. https://doi.org/10.1016/j.jns.2021.117487

Triyono, A., Trianto, R. B., & Arum, D. M. P. (2021). Early Detection Of Diabetes Mellitus Using Random Forest Algorithm. *Julia: Jurnal Ilmu Komputer An Nuur*, *1*(1), 25–31.

Wikandikta, I. P. G., Samatra, D. P. G. P., & Meidiary, A. A. A. (2020). Prevalensi gangguan tidur pada penderita parkinson di Poli Saraf RSUD Wangaya Denpasar tahun 2017. *Intisari Sains Medis*, *11*(3), 1085-1090. https://doi.org/10.15562/ism.v11i3.232

WHO. (2022). Parkinson disease: a public health approach: technical brief.

Yuliati, I. F., & Sihombing, P. R. (2021). Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *20*(2), 417–426. https://doi.org/10.30812/matrik.v20i2.1174

Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., & Hoffman, M. M. (2019). Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, *50*, 71–91. https://doi.org/10.1016/j.inffus.2018.09.012

Zulfahmi, I., Syahputra, H., Naibaho, S. I., Maulana, M. A., & Sinaga, E. P. (2023). Perbandingan Algoritma Support Vector Machine (SVM) dan Decision Tree Untuk Deteksi Tingkat Depresi Mahasiswa. *BINA INSANI ICT JOURNAL*, *10*(1), 238-247. https://doi.org/10.51211/biict.v10i1.2304