SMOTE-TomekLink Super-Learner Ensemble Model (STL-SLEM) for the Prediction of Parkinson’s Disease

AROWOLO, O. A
Tai Solarin College of Education, Omu –Ijebu
Computer Department, TASCE, Ogun State
arowolooluremi@gmail.com

EMUOYIBOFARHE, O. J
Ladoke Akintola University of Technology, Department, of Information Systems.
eojustice@gmail.com

AMUSAN, E. A
Ladoke Akintola University of Technology, Department of Cyber Security,
,eaamusan@lautech.edu.ng

ABSTRACT

Parkinson's Disease (PD) is a progressive neurodegenerative disorder that affects millions of people worldwide. Early detection and prediction of Parkinson's Disease can significantly improve patient outcomes by enabling timely intervention and personalized treatment. Over the years, many Parkinson Disease (PD) prediction models have been developed using machine learning algorithms. Some of these existing models suffer over-fitting of data due to unavailability of sufficient dataset in PD as well as data imbalance. Hence, this work developed a Super Learner Ensemble Model (SLEM) that aggregated several machine learning models configurations to overcome the challenge of over-fitting thereby enhancing the performance of PD prediction. The dataset used for this research is Parkinson disease datasets obtained from Kaggle website and also local datasets from Federal Medical Center, Abeokuta, Nigeria for the validation of the developed model. The dataset from Kaggle website consists of 195 biomedical voice measurements from 31 people taken severally, 23 out of the 31 have Parkinson's disease and 8 without Parkinson's disease, while the local datasets consists of 13 people, 9 with PD and 4 without PD. The acquired dataset has class imbalance, and to handle this issue, Synthetic Minority Over Sampling Technique with TomekLink (SMOTE-TomekLink) was adopted to resample the dataset for class-balancing. For computational efficiency, six base learners were used to develop the Super Learner model, which includes Logistic Regression (LR), Decision Tree (DT), Naïve Bayes (NB), Adaptive Boosting (AB), Bagging Ensemble (BE), and Random Forest (RF) algorithms. The performances of each base model were measured, and the performance of the Super Learner ensemble model was also obtained using the following performance metrics: Accuracy, Precision, Recall, F1-Score, Matthews Correlation Coefficient (MCC), and Balanced Accuracy Score (BAS). However, Accuracy for LR, DT, NB, AB, BE, and RF with SMOTE-TomekLink-resampled datasets were 95.0%, 94.0%, 91.0%, 93.0%, 95.0%, and 96.5%, respectively, while the corresponding Accuracy for Super Learner Ensemble model was 99.0%.. The developed model showed an improvement in the performance metrics.

Key words: Parkinson's Disease, prediction models, Super Learner, SMOTE-TomekLink.

---

1. INTRODUCTION

Parkinson's Disease (PD) is a progressive neurodegenerative disorder that affects millions of people worldwide (Barukab et al., 2022). Early detection and prediction of PD can significantly improve patient outcomes by enabling timely intervention and personalized treatment. In recent years, machine learning techniques have shown promise in assisting with early diagnosis and prediction of PD (Alalayah *et al.,* 2023). Early prediction of PD involves identifying individuals at risk of developing the disease before the onset of clinical symptoms (Hathaliya *et al.*, 2022). This approach can utilize various data sources, including genetic, clinical, neuroimaging data, and voice dataset, to build predictive models capable of detecting subtle disease-related changes in individuals who have not yet developed noticeable symptoms (Yang *et al.*, 2022). Several machine learning algorithms have been applied to PD prediction, including support vector machines (SVM), random forests, neural networks, and logistic regression. These algorithms have shown varying degrees of success in discriminating between PD patients and healthy controls based on different sets of features extracted from various data modalities (Ghaffar *et al.*, 2023). Some of these existing models suffer over-fitting of data due to unavailability of sufficient dataset in PD as well as data imbalance. Thus, this work applied the application of the super learner ensemble machine learning approach in the early prediction of PD to address the aforementioned problems..

The super learner ensemble machine learning approach, introduced by van der Laan *et al.* (2007), is a powerful technique that combines multiple base learners to create a more accurate and robust predictive model. The super learner algorithm employs cross-validation to estimate the optimal weights assigned to each base learner, effectively leveraging the strengths of different algorithms to improve prediction performance (Kabir and Ludwig, 2019).

1.2 Related Works

Till date, researchers across the world have been trying to observe the outcomes of various Machine Learning (ML) based methods for prediction of PD. Though several of these techniques have provided satisfactory results, it has also been noticed that different models yield different outcomes. A number of related works surrounding the study of the application of Machine Learning and ensemble-based techniques to the area of PD and other related diseases is presented in the following paragraphs.

Abos *et al.*, 2017 extracted features from Resting-State Functional MRI (rsfMRI) and used Support Vector Machine (SVM) for the detection of PD. They achieved 86.96% accuracy, 78.95% sensitivity, and a specificity of 92.59%. Amoroso *et al.,* 2018 used network and clinical features to classify PD patients using an SVM. They experimented on the PPMI dataset and got 93% recognition accuracy and sensitivity, and 92% of specificity. A Sparse feature selection model was proposed by Lei *et al.* (2018), reporting an accuracy of around 80%. Salvatore *et al.*, 2014 considered healthy PD, and supra-nuclear palsy MRI images to extract features. Next, they have used Principal Components Analysis (PCA) to find the relevant features and fed them to an SVM classifier for classification purposes, having obtained above 90% accuracy for the case of PD patients vs. controls. Prashant *et al.*, 2017 used SVM with striatal binding ratio to classify PD patients and they got an accuracy of 96.14%, a sensitivity score of 95.74%, and 77.35% specificity. Sheibani *et al.*, 2019 proposed an ensemble technique for recognizing sick and healthy samples using voice frequency features to predict class labels. An idea was formulated to combine sample primary feature vectors with anticipated class labels and achieves the classification accuracy of 90.6%. In Zhang *et al.*, 2016 a novel classification method that combines the multi-edit nearest neighbour (MENN) algorithm with a combined learning approach was recommended. The suggested technique improved PD classification by using voice data in this investigation, and it may be used in future studies to enhance PD classification accuracy. Shounak, 2019 worked on the diagnosis of PD based on information about demographic, gait and speech data. The study collected information about demographics, gestures and speech of individuals with and without PD. The study adopted the use of various ML algorithms and ensemble of ML for the classification of PD using the collected dataset and compared their performance. The results of the study showed that using a stack-based ensemble of ML algorithms improved the performance of the prediction of PD based on the collected dataset compared to using ML algorithms in isolation. Prediction challenges have shown to benefit from ensemble learning techniques (Kundu *et al.*, 2021). Few research (Das *et al.*, 2009a and b and Ortiz *et al.*, 2016) have included ensemble learning methodologies for the creation of disease diagnosis systems. The usefulness of these strategies needs to be further studied. The K-Nearest Neighbors

(KNN) ensemble technique for the identification of PD has been demonstrated by the author in Gök, 2013. For a variety of tests that took into account Cerebrospinal Fluid, RNA, Serum tests, and pre-processed neuro-image features from PPMI database individuals, authors in Castillo-Barnes *et al*., 2018 coupled SVM with linear kernel classifiers.

In order to distinguish between sick and healthy samples, Velmurugan and Dhinakaran (2022) suggested a class label prediction algorithm based on voice data using the Ensemble Stacking Learning Algorithm (ESLA). To construct four stacked models, ESLA was created. To achieve the maximum accuracy in PD prediction, adaboost, logistic regression, and multilayer perceptron (MLP) are combined with the base classifiers RF and XGBoost. The four layered models ultimately produce varying training and testing accuracies. The Stackedmodel3 (an ensemble of the Stackedmodel1, a hybrid model of the RF and XGBoost algorithms with logistic regression and multilayer perceptron (MLP)) has the best prediction accuracy of 90% in all of those trainings.

2. Theoretical Backgrounds and Model Development

2.1 Base Learners

Base learners are the members of the ensemble that are strategically combined from among the individual members. Base learners must concentrate on appropriately categorizing the instances with the highest weights while firmly avoiding over-fitting. Although it is advised that the basis learners should consist of a varied group of students, the super learner theory does not specify the degree to which the base learners must be different (Sivaranjini and Sujatha, 2020; Esmaeilzadeh *et al*. 2018). The same method may be used more than once as a base learner using various sets of parameters. Typically, each base learning algorithm in stacking-based ensemble methods is trained on the entire training data set before the fits are combined using a meta-learning algorithm.

In this work, the base learners used for computational efficiency are Decision Tree (DT), Naïve Bayes (NB), Logistic Regression (LR), Adaptive Boosting (AB), Bagging and Random Forest algorithms

2.2 Meta-learning Algorithm

The best L base learner combination is discovered using the meta-learner. The original result from level-zero training data Y1, Y2,..., Yn is utilized as input for the meta-learning method combined with the Z matrix of cross-validated predicted values, as previously mentioned. The meta-learning method is defined in the super learning algorithm as the cross-validated risk minimizer of an interest loss function, such as squared error loss or rank loss (Maxwell and Barry, 2022). To aggregate the results from the base fits, a range of parametric and non-parametric techniques can be utilized as a meta-learner (Sivaranjini and Sujatha, 2020; Esmaeilzadeh *et al.* 2018).In the past, the meta-learning algorithm was frequently some sort of regularized linear model.

2.3 Super Learner Algorithm

According to Kabir and Ludwig, 2019 the following is a list of the tasks involved in developing and evaluating a Super Learner ensemble:

Set up the ensemble

1. Specify a set of model parameters and a list of L basic algorithms.
2. Name the meta-learning algorithm.

Train the ensemble

1. Train each of the L base algorithms using the training set.
2. To each of these learners, apply k-fold cross-validation, and then obtain the cross-validated predicted values from each L method.
3. A new N x L matrix can be created by combining the N cross-validated predicted values from each of the L methods. The "level-one" data refers to this matrix and the initial response vector. N stands for the training set's row count.
4. Train the meta-learning algorithm using the level-one data.
5. The "ensemble model" can be used to predict test results and is composed of the meta-learning and L base learning models.

Predict on new data

6. Make predictions from the base learners first, then create ensemble predictions.
7. Feed the meta-learner with those predictions to produce the ensemble prediction.

3. The Architecture of the Developed Super Learner Ensemble Model (SLEM)

The developed Super Learner Ensemble Model for the prediction of Parkinson’s disease is a triple-phased architecture. Phase one of the architecture is the Data Preparation phase, phase two is the Base-Learners training and cross-validation evaluation phase, and the third phase is the Meta-Learner training and testing Phase. Figure 1 demonstrates the detailed architecture of the developed Super Learner Ensemble Model (SLEM) for the prediction of Parkinson’s disease. As shown in the architecture, in the first phase of data preparation, the Parkinson’s disease dataset is first optionally passed to the “data resampling” sub-module, after which the output is forwarded to the “K-fold split” sub-module of that phase. Next, in phase two, an arbitrary number of base-learners are fitted on the k-fold training data whose out-of-fold predictions are obtained by using k-fold cross-validation, and these out-of-fold predictions are compiled. Also, the base learners are fitted directly on the full training set and stored in this phase. In the third phase of the architecture, a meta-learner (that is, the super learner ensemble) is fitted on the out-of-fold predictions that have been compiled at the base-learners training and evaluation phase. Lastly, the last phase of the architecture handles the testing of the trained Super-Learner ensemble model via the use of a hold-out dataset.

In the workings of the architecture, a second-level learner, or a Meta learner, is trained on the results (i.e., the cross-validated predictions) of a group of base learners using the stacked generalization (Stacking) process. For the meta-learner, Logistic Regression algorithm is employed to learn from the predictions made by the base-learners to form the meta-learner. The arbitrarily selected algorithms for fitting the base-learners at this level of training include Logistic Regression, Decision tree, Naïve Bayes, Adaptive Boosting, Bagging and Random Forest algorithms purposely chosen for computational efficiency. Cross-validation is used to produce the prediction outputs from the level-one data, also known as the base-learners. As a result, predictions from a single, independent test set are used to create the level-one data from the original training dataset, also known as the level-zero data, in the vast training set. The meta-level super learner ensemble model trained from these base-learner's predictions (level-1 data) is then used to make predictions on a held-out dataset for evaluation of the super-learner ensemble model.

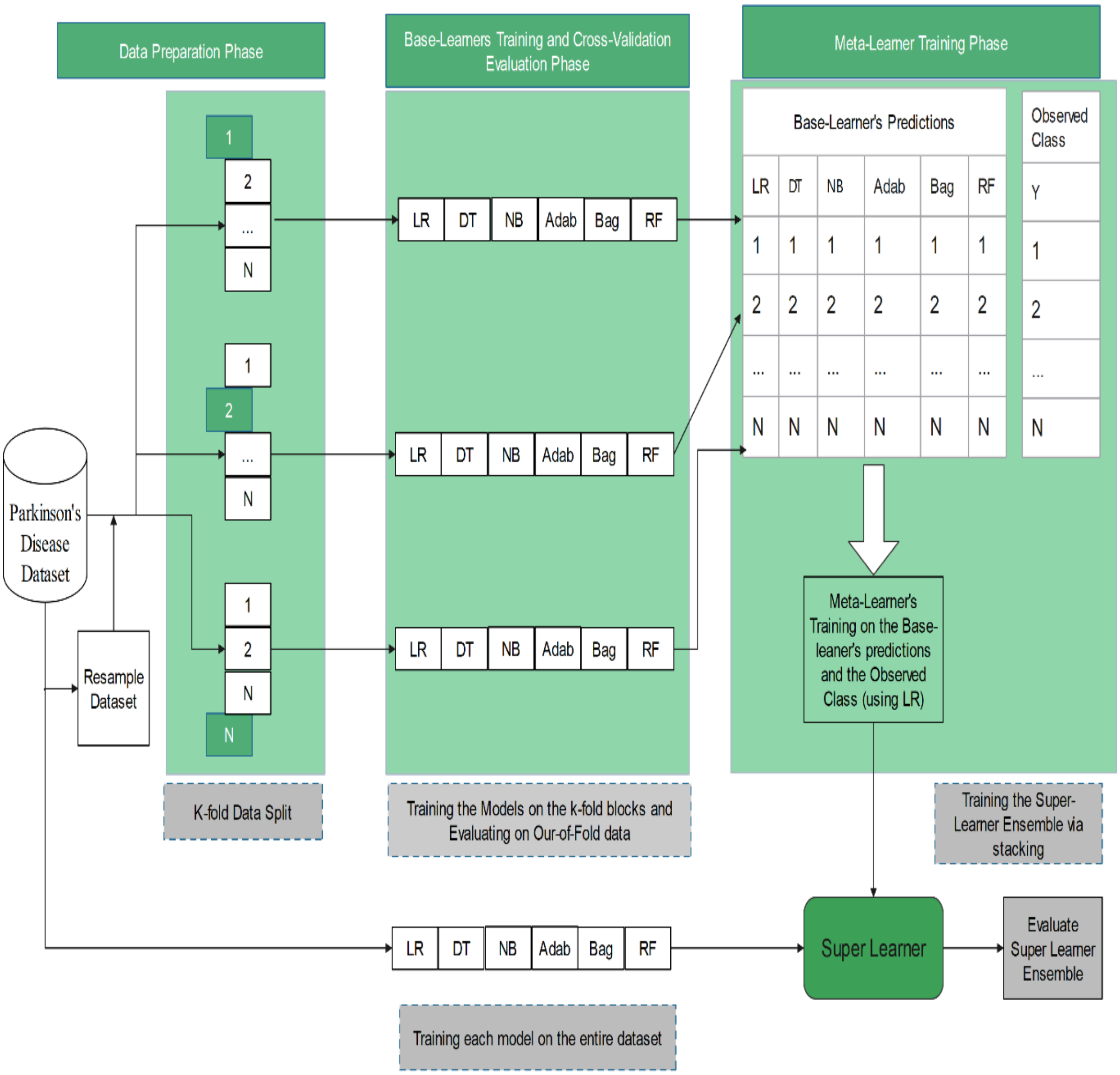

Figure 1: The Super-Learner Ensemble Architecture for Parkinson's disease Prediction

3.1 Handling Imbalance in the Dataset through Resampling Method

The original dataset obtained for this study has class imbalance. Therefore, to handle the imbalance in the dataset, the dataset was resampled for class-balancing using the adopted resampling method, Synthetic Minority Over Sampling Technique with TomekLink (SMOTE-TomekLink) which oversamples the minority class and then undersamples the majority class. The Synthetic Minority Oversampling Technique (SMOTE) aspect of the technique uses a KNN strategy to oversample the minority class by selecting K nearest neighbors, joining them, and creating synthetic samples in the space. After this process, the TomekLink aspect of the SMOTE-TomekLink algorithm undersamples the majority class instances to meet the minority class number of instances.

3.1.1 Resampling Dataset with SMOTE-TomekLink Technique

The SMOTE-TomekLink technique combines the capability of SMOTE to generate and oversampled synthetic instances for the minority class (i.e., the healthy class, in this study) with the under-sampling of the majority class (i.e., the Parkinson' class). The process of applying the SMOTE-TomekLink technique to the Parkinson's dataset used in this study is as specified in the Algorithm 1 below:

Algorithm 1: The SMOTE-TomekLink Algorithm for Resampling the Parkinson's Dataset

```
Input: Parkinson's Disease Dataset $D$

Output: A SMOTE-TomekLink Resampled Dataset

Start

1. Select arbitrary data from the minority class.

2. Determine how far the random data is from its k-nearest neighbors.

3. To create a synthetic sample, multiply the difference by a random value between 0 and 1, then add
the result to the minority class.

4.Continue steps 2-3 until the necessary percentage of the minority class is reached.

End
```

Modeling the Super Learner Ensemble Model

The k-fold splits of the dataset must first be predefined before any base-learning algorithms are evaluated on those splits, according to the super learner ensemble algorithm's workings. The model that learns how to incorporate the results of the base-classifiers' predictions most effectively is then trained using all of the out-of-fold predictions that were kept. The Super Learner Ensemble is essentially a particular stacking configuration for use in k-fold cross-validation with all base-classifiers taken into account for the predictive modeling.

According to Algorithm 2, the process for the super learner ensemble model employed in this work to forecast Parkinson's disease went as follows:

Algorithm 2: A SMOTE-TomekLink-Improved Super Learner Ensemble (STL-SLEM) Algorithm for Predicting Parkinson's Disease

---

Input: Parkinson's Disease Dataset

Output: A SMOTE-TomekLink -Improved Super Learner Ensemble Model

Data Preparation:

1. Resample the Dataset using SMOTE-TomekLink resampling technique

2. Split Dataset: (Training Data: 70% of the resampled Dataset); (Test Data: 30% of the resampled Dataset).

3. Select a k-fold split of the training dataset. (for k = 10).

Model Training:

4. Select *n* base-classifiers. (In this study, *n* = 6: Logistic Regression, Decision Tree, Naïve Bayes, Adaptive Boost, Bagging, and Random Forest)

5. For each base-classifier:

(a) Perform a k-fold cross-validation evaluation.

(a) Save all predictions that aren't folded.

(c) Use the entire training dataset to train the model, then store.

6. Train a meta-classifier (Logistic Regression) on the out-of-fold predictions.

Model Testing:

7. Evaluate the Classifier on a holdout dataset (i.e., use the model to make predictions).

Models Training and Evaluation

The SMOTE-TomekLink-resampled dataset with Class-balance was divided into a training set and a test set, each comprising 70% of the dataset. The super learner ensemble approach (which has six different base learners) was then given the entire 70% training set of this SMOTE-TomekLink resampled dataset for a 10-fold cross validation process (in which 10% of the training dataset is held out for an out-of-fold prediction and the remaining used for training). Following training on the training set, each base model is assessed to determine its out-of-fold predictions. Next, all of the out-of-fold predictions are used to train the Super Learner Meta-Model, which is subsequently tested on the held-out test set. For the training process, every feature was taken into account.

Based on the made predictions, the super learner ensemble model's performance was also determined, along with the performances of each of the basis models. A number of performance metrics, including accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and balanced accuracy score (BAS), were used to evaluate the base-model and super learner ensemble performances.

4. Evaluation of Results obtained from SMOTE-TomekLink -resampled dataset

The Table 1 below shows the performances of all the models in terms of accuracy, precision, recall, F1 score, Matthews Correlation Coefficient (MCC), and balanced accuracy score (BAS), using the SMOTE-TomekLink -resampled dataset.

Table 1: Classification Report of the Base Models and the Super Learner Ensemble Model using the Original Dataset with Class-imbalance

	Accuracy (%)	Precision		Recall		f1-score	MCC	BAS
		Parkinson	***Healthy***	***Parkinson***	***Healthy***			
Logistic Regression	88.1	0.89	0.83	0.95	0.67	0.92	0.672	0.811
Decision Tree	84.7	0.91	0.69	0.89	0.73	0.90	0.606	0.809
Naïve Bayes	76.3	0.92	0.52	0.75	0.80	0.83	0.491	0.775
AdaBoost	77.9	0.78	0.75	0.98	0.20	0.87	0.307	0.588
Bagging Ensemble	86.4	0.89	0.77	0.93	0.67	0.91	0.628	0.799
Random Forest	88.1	0.91	0.79	0.93	0.73	0.96	0.680	0.832
Super Learner Ensemble	**91.5**	0.90	**1.00**	**1.00**	0.67	0.95	**0.773**	**0.833**

Table 2: Classification Report of the Base Models and the Super Learner Ensemble Model using the SMOTE-TomekLink-resampled Dataset with Class-balance

	Accuracy (%)	Precision		Recall		f1-score	MCC	BAS
		Parkinson	***Healthy***	***Parkinson***	***Healthy***			
Logistic Regression	95.0	1.00	0.90	0.91	1.00	0.95	1.00	0.953
Decision Tree	94.0	1.00	0.89	0.89	1.00	0.94	0.891	0.942
Naïve Bayes	91.0	1.00	0.90	0.82	1.00	0.90	0.833	0.901
AdaBoost	93.0	1.00	0.88	0.86	1.00	0.93	0.869	0.931
Bagging	95.0	1.00	0.90	0.91	1.00	0.95	1.00	0.953

Ensemble								
Random Forest	96.5	1.00	0.93	0.93	1.00	0.96	0.932	0.965
Super Learner Ensemble	99.0	1.00	0.97	0.98	1.00	0.99	0.984	0.988

4.1 Results Discussion of models obtained using SMOTE-TomekLink-resampled Dataset with Class-balance

The results presented in Table 2 shows that the result of individual model with SMOTE-TomekLink-resampled datasets and that of the developed model performed better as compared to the results in Table 1. The result shows that Random Forest has an accuracy of 96.5%, Logistic regression models has accuracy of 95%, Decision Tree has an accuracy of 94.0%, Naïve Bayes has an accuracy of 91%, AdaBoost has 93.0% accuracy and Bagging Ensemble has an accuracy of 95 %. The Super Learner Ensemble model developed combined with SMOTE-TomekLink-resampled datasets with all the six base-learners showed an overall accuracy of 99.0% accuracy which outperformed individual result. This research showed that the SMOTE-TomekLink-improved Super-Learner ensemble model (STL-SLEM) obtained the highest accuracy performance for predicting Parkinson disease.

5. Conclusion

A SMOTE-TomekLink-improved Super-Learner ensemble model (STL-SLEM) is developed for the accurate prediction of Parkinson's diseases. In order to reduce overfitting problem and improve the performance of Parkinson's diseases prediction for small training set due to unavailability of sufficient dataset in Parkinson's diseases as well as data imbalance, Super-Learner ensemble model with SMOTE-TomekLink resampling technique was adopted. As illustrated in the experimental study, the proposed method achieved a high predictive accuracy as in the case of class balanced. Comparison studies show that the proposed model outperforms the base learner models and some other traditional methods.

ACKNOWLEDGMENTS

Our thanks to the institution that provided us support and funding that contributed towards success of this work.

REFERENCES

[1] Abós, A., Baggio, H., Segura, B., García-Díaz, A., Compta, Y., Martí, M., Valldeoriola, F., and Junqué, C. (2017). Discriminating cognitive status in Parkinson's disease through functional connectomics and machine learning. Scientific reports, 7, 45347. https://doi.org/10.1038/srep45347

[2] Alalayah, K. M., Senan, E. M., Atlam, H. F., Ahmed, I. A., & Shatnawi, H. S. A. (2023). Automatic and Early Detection of Parkinson's Disease by Analyzing Acoustic Signals Using Classification Algorithms Based on Recursive Feature Elimination Method. Diagnostics, 13(11), 1924. MDPI AG. Retrieved from http://dx.doi.org/10.3390/diagnostics13111924

[3] Amoroso, N., Rocca, M., Monaco, A., Bellotti, R., and Tangaro, S. (2018). Complex networks reveal early MRI markers of Parkinson’s disease. Medical Image Analysis, 48, 12-24. doi: 10.1016/j.media.2018.03.003

[4] Barukab, O., Ahmad, A., Khan, T., & Thayyil Kunhumuhammed, M. R. (2022). Analysis of Parkinson's Disease Using an Imbalanced-Speech Dataset by Employing Decision Tree Ensemble Methods. Diagnostics (Basel, Switzerland), 12(12), 3000. https://doi.org/10.3390/diagnostics12123000

[5] Castillo-Barnes D, Ramírez J, Segovia F, Martínez-Murcia F, Salas-Gonzalez D and Górriz JM (2018) Robust Ensemble Classification Methodology for I123-Ioflupane SPECT Images and Multiple Heterogeneous Biomarkers in the Diagnosis of Parkinson’s Disease. Front. Neuroinform. 12:53. doi: 10.3389/fninf.2018.00053.

[6] Das, R., Turkoglu, I., and Sengur, A. (2009a). Diagnosis of valvular heart disease through neural networks ensembles. Computer Methods and Programs in Biomedicine, 93, 185-191. https://doi.org/10.1016/j.cmpb.2008.09.003

[7] Das, R., Turkoglu, I., and Sengur, A. (2009b). Effective diagnosis of heart disease through neural networks ensembles. Expert Systems with Applications, 36, 7675-7680. https://doi.org/10.1016/j.eswa.2008.10.049Esmaeilzadeh, S., Yang, Y., and Adeli, E. (2018). End-to-End Parkinson Disease Diagnosis using Brain MR-Images by 3D-CNN. arXiv preprint arXiv:1806.05233.

[8] Esmaeilzadeh, S., Yang, Y., and Adeli, E. (2018). End-to-End Parkinson Disease Diagnosis using Brain MR-Images by 3D-CNN. arXiv preprint arXiv:1806.05233

[9] Ghaffar Nia, N., Kaplanoglu, E., & Nasab, A. (2023). Evaluation of artificial intelligence techniques in disease diagnosis and prediction. Discover Artificial Intelligence, 3(1), 5. https://doi.org/10.1007/s44163-023-00049-5

[10] Gök, M. (2013). An ensemble of k-nearest neighbours algorithm for detection of Parkinson’s disease. International Journal of Systems Science, 44(6), 1180-1187. https://doi.org/10.1080/00207721.2011.645965

[11] Hathaliya, J.; Modi, H.; Gupta, R.; Tanwar, S.; Alqahtani, F.; Elghatwary, M.; Neagu, B.; Raboaca, M.(2022). Stacked Model-Based Classification of Parkinson’s Disease Patients Using Imaging Biomarker Data. Biosensors 2022, 12, 579. https://doi.org/10.3390/ bios12080579

[12] Kabir, M., and Ludwig, S. (2019). Enhancing the Performance of Classification Using Super Learning. Data-Enabled Discovery and Applications, 3, 1-13.

[13] Kundu, R.; Singh, P; Mirjalili, S.; Sarkar, R. (2021). COVID-19 detection from lung CT-Scans using a fuzzy integral-based CNN ensemble. Comput. Biol. Med. 138, 104895. [CrossRef]

[14] Lei, H.; Zhao, Y.; Wen, Y.; Luo, Q.; Cai, Y.; Liu, G.; Lei, B. (2018). Sparse feature learning for multi-class Parkinson’s disease classification. Technol. Health Care 2018, 26, 193–203. [CrossRef]

[15] Ortiz A, Munilla J,. Gorriz J. and Ramirez J (2016). Ensembles of deep learning architectures for the early diagnosis of the Alzheimer’s disease. International Journal of Neural Systems, 26, Article ID 1650025.

[16] Prashanth, R.; Roy, S.D.; Mandal, P.K.; Ghosh, S. (2017). High-Accuracy Classification of Parkinson’s Disease Through Shape Analysis and Surface Fitting in 123I-Ioflupane SPECT Imaging. IEEE J. Biomed. Health Inform. 2017, 21, 794–802. [CrossRef]

[17] Salvatore, C.; Cerasa, A.; Castiglioni, I.; Gallivanone, F.; Augimeri, A.; Lopez, M.; Arabia, G.; Morelli, M.; Gilardi, M.C.; Quattrone, A. (2014). Machine learning on brain MRI data for differential diagnosis of Parkinson’s disease and Progressive Supranuclear Palsy. J. Neurosci. Methods 2014, 222, 230–237. [CrossRef]

[18] Sheibani. R, Nikookar. E, and Alavi, S. E.(2019). “An ensemble method for diagnosis of Parkinson’s disease based on voice measurements,” Journal of medical signals and sensors, 9(4), pp. 221–226, 2019.

[19] Shounak R. (2019). A Predictive Diagnosis for Parkinson’s Disease Through Machine Learning. An Article in the Canadian Science Fair Journal (CSFJ), (2)1. Licensed under: https://creativecommons.org/licenses/by/4.0

[20] Sivaranjini, S.; Sujatha, C. (2020). Deep learning based diagnosis of PD using convolutional neural network. Multimed. Tools Appl. 79, 15467–15479.

[21] Smith, A., Johnson, M. A., & Paul, D. (2019). Super learner ensemble machine learning for early prediction of Parkinson's disease. Journal of Neuroinformatics, 17(2), 203-215.

[22] van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super Learner. Statistical Applications in Genetics and Molecular Biology, 6(1), 1-21.

[23] Velmurugan, T and Dhinakaran, J (2022). A Novel Ensemble Stacking Learning Algorithm for Parkinson’s Disease Prediction. Mathematical Problems in Engineering, Hindawi,: 1-10. . https://doi.org/10.1155/2022/9209656

[24] Yang, Y., Yuan, Y., Zhang, G. (2022). Artificial intelligence-enabled detection and assessment of Parkinson’s disease using nocturnal breathing signals. Nat Med 28, 2207–2215 https://doi.org/10.1038/s41591-022-01932-x

[16] Zhang, H,, Yang, L. Liu, Y., Wang, P., Yin, J., Li, Y., Qiu, M., Zhu, X and Yan, F, (2016). Classification of Parkinson’s disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples. BioMedical Engineering Online. 15(1), 122-130. https://doi.org/10.1186/s12938-016-0242-6

Author’s Brief Profile

Arowolo Oluremi. A, holds a Bachelor of Science (B.Sc.) degree and Master of Science (M.Sc) degrees in Computer Science from Ogun State University, Ago-Iwoye Ogun State and Federal University of Agriculture, Abeokuta, Ogun State, Nigeria respectively. Her research areas include e-Health and Pattern Recognition. She can be reached by phone on +2348033809945 and through E-mail arowolooluremi@gmail.com

Emuoyibofarhe .O. Justice (PhD) is currently a Professor in the Department of Computer Science and Engineering at Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria. His research areas include Computational Optimisation, Neural Networks, Mobile Computing, Wireless Communication, E-Health and Telemedicine. He can be reached by phone on +2348033850075 and through E-mail eojustice@gmail.com