2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Towards Automated Classification of Parkinson's Disease: Comparison of Machine Learning Methods using MRI and Acoustic Data*

Noushath Shaffi
College of Computing and Inf.Sc
UTAS, Oman
noushath.shaffi@utas.edu.om

Vimbi Viswan
College of Computing & Inf.Sc
UTAS, Oman
vimbi.viswan@utas.edu.om

Mufti Mahmud
Department of Computer Science
Nottingham Trent University
Nottingham, United Kingdom
mufti.mahmud@ntu.ac.uk

Faizal Hajamohideen
College of Computing & Inf.Sc
UTAS, Oman
faizal.hajamohideen@utas.edu.om

Karthikeyan Subramanian
College of Computing & Inf.Sc
UTAS, Oman
karthikeyan.supramanian@utas.edu.om

***Abstract*—In this study, we focus on Parkinson's Disease (PD) classification and present a comparative analysis of prominent machine learning models using two distinct and independent modalities: Magnetic Resonance Imaging (MRI) and Acoustic data. Unlike many existing works that typically focus on a single modality, our research study provides performance evaluation on the performance of various algorithms on both MRI and Acoustic data. Through a detailed investigation, we provide an understanding of how different models perform when applied to each modality individually. Furthermore, our study extends beyond this comparative framework by introducing an ensemble approach aimed at enhancing the performance of machine learning models for PD classification using the acoustic data. Notably, our ensemble approach yields around a 12% increase in overall performance.**

***Index Terms*—Parkinson's Disease, Machine Learning, Ensemble, SVM, KNN, MRI, Accoustic.**

I. Introduction

Parkinson Disease is a complex neurodegenerative disorder with diverse manifestations. It causes tremors, stiffness, and difficulty walking. This disorder affects millions of people globally, which places a considerable financial strain on healthcare systems and society. Patients and caregivers frequently experience emotional distress because of the difficulties associated with managing the disease's progressive nature [1].

Hence, early prediction of PD becomes crucial. There are several benefits associated with the early prediction of PD: Firstly, it allows for timely interventions, potentially slowing down the disease progression and improving the quality of life for patients. Secondly, early detection enables targeted treatment approaches, making healthcare more effective and cost-efficient. Lastly, it opens up opportunities for further research into disease prevention and management strategies. Thus, automated systems can be utilized to enable real-time monitoring of patients with PD through continuous data collection which allows for the tracking of disease progression, treatment effectiveness, and personalized care adjustments.

In the automatic detection of PD, researchers and clinicians utilize various modalities to capture and analyze relevant data. These modalities include clinical data, neuroimaging, speech data, tremor analysis and specific biomarkers. Machine learning classifiers can analyze large and complex datasets, extracting subtle patterns and relationships that might not be apparent to human observers. As a result, automated classification achieves high accuracy and reliability in distinguishing between healthy individuals and those with PD [2].

Claas Ahlrichs et al. [3] carried out a work that reviews state-of-the-art works on recognizing motor symptoms of PD using machine learning algorithms. Another study [4] proposes an artificial intelligence-based voice analysis system for detecting Parkinson's Disease (PD) using voice recordings of patients and healthy subjects. The paper applies machine-learning techniques to analyze voice signals and identify PD-related patterns. The paper evaluates the performance and accuracy of the proposed system using different datasets and metrics and compares it with existing methods in the literature. Gunjan et al. [5] reviews the use of machine learning for Parkinson's disease diagnosis using voice data. The paper summarizes the main challenges and limitations of applying machine learning to PD diagnosis. This work provides an overview of the current state-of-the-art methods and their performance for PD diagnosis using machine learning, and compares them with conventional methods.Yet another study [6] proposes a machine learning algorithm for predicting

This work is funded by the Ministry of Higher Education, Research and Innovation (MoHERI) of the Sultanate of Oman under the Block Funding Program (Grant number: MoHERI/BFP/UoTAS/01/2021). Mufti Mahmud was supported by the British Embassy in Kuwait through the project titled 'Investigation of Emotional Deficits in Parkinson's Disease using Advanced Machine Learning Methods.'

Parkinson's Disease (PD) using voice data from patients and healthy subjects. The paper applies data mining techniques, such as data preprocessing, feature extraction, feature selection, and classification, to analyze the voice data and identify PD-related patterns. The study evaluates the performance and accuracy of the machine learning algorithm considering different classifiers, such as k-nearest neighbor, support vector machine, decision tree, and random forest. The paper shows that the proposed algorithm can achieve high prediction and low error rates for PD diagnosis using voice data. Shu et al. [7] proposed a study that aims to develop and validate a radiomics model based on whole-brain white matter and clinical features to predict the progression of PD. Radiomics is a technique that extracts quantitative features from medical images to reveal disease characteristics. The authors used conventional Magnetic Resonance Imagine (MRI) data from 100 PD patients and 100 healthy controls and extracted 1,024 radiomic features from each subject. They then applied machine learning algorithms such as support vector machine (SVM) and random forest (RF) to classify the subjects into PD or control groups and to predict the future change of Unified Parkinson's Disease Rating Scale (UPDRS) scores, which measure the severity of PD symptoms. The authors concluded that the radiomics model could be a valuable tool for predicting PD progression and providing insights into the pathophysiology of PD.

From this review of literature, we can understand that there exists several limitations such as small sample sizes and inconsistent evaluation metrics. To the best of our knowledge, the existing research on the comparative evaluation of machine learning algorithms for Parkinson's disease has primarily focused on single modality data. Surprisingly, despite the importance of considering multiple modalities to comprehensively understand the disease, no single paper has attempted to provide a performance evaluation of machine learning algorithms across different modalities. This gap in the literature presents an exciting opportunity for our research. By conducting a comparative evaluation on MRI and speech data modalities, we aim to address this critical limitation and gain a more comprehensive understanding of performance of machine learning algorithms in Parkinson's disease detection.

Evaluating multiple modalities provides a comprehensive description of the disease, capturing its complexity and allowing for a more nuanced understanding. In our research, we evaluated the performance of machine learning algorithms for classifying Parkinson's disease using MRI and acoustic data. We also highlighted the importance of having a standardized testing method. Our objective evaluation allows for a fair comparison of existing classifiers and helps to identify their limitations and strengths when evaluated under the same conditions. By establishing a common framework, we provide an unbiased analysis of various classifiers and address subjective claims found in the literature.

The paper is organized as follows: Overview of the relevant literature is provided in Section-II. Experimental setup and performance evaluations are presented in Section-III. Conclusions are drawn in Section-IV.

II. Review of Relevant Literature

Integrating supervised machine learning techniques with diverse data modalities has opened up innovative approaches to understand and address complex neurological disorders such as PD and its progression. In particular, applying supervised machine learning to PD on several data modalities, including MRI, acoustic data, and multimodal data fusion, has shown promise in enhancing PD diagnosis, prognosis, and treatment.

MRI is a crucial diagnostic tool for PD. Qualitative and quantitative analyses of MRI scans help understand PD and provide detailed brain insights for early detection, distinguishing PD from other disorders, identifying biomarkers, and tailoring treatments. The article by Solana-Lavalle et al. [8] focuses on enhancing PD diagnosis using quantitative analyses that involve voxel-based morphometry (VBM) and machine learning to classify 3D MRI scans. In the study, VBM identifies key brain regions based on first- and second-order statistics, followed by feature selection. Seven classifiers are employed separately for both genders. The approach shows impressive results of more than 96% accuracy for both male and female subjects. The article emphasizes the significance of VBM analysis and a gender-specific approach to considering neurobiological differences in enhancing the accuracy of PD diagnosis using MRI scans. The study by Salvatore et al. [9] uses machine learning to differentiate between PD and Progressive Supranuclear Palsy (PSP) through brain MRI data. ML techniques aid in the classification of these neurological disorders. The two neurological conditions present similar symptoms, though having different underlying pathologies. The approach is to train ML models on MRI data to differentiate between the two conditions effectively. The study underscores the application of ML techniques to brain MRI data in effectively distinguishing between PD and PSP. It showcases the potential for improved diagnostic accuracy in complex neurological conditions. Functional MRI (fMRI) and ML techniques can help predict optimal parameters for deep brain stimulation (DBS) in PD and enhance the efficacy of DBS treatment leading to the best outcomes for individual patients. Boutet et al. [10] employ fMRI data and machine learning algorithms to develop a predictive model for determining the optimal settings for DBS by analyzing brain activity patterns and neural responses from fMRI scans. The results emphasize using predicted parameters for improved treatment outcomes and personalized care.

Neurological diseases often cause subtle changes in speech patterns and vocal characteristics before physical symptoms appear. ML algorithms can detect these nuances, allowing for early identification and intervention. This helps clinicians understand the disease's trajectory and adapt treatment strategies. Diago et al. [11] used acoustic features from speech recordings in uncontrolled background conditions and integrated them with machine learning techniques to differentiate individuals with Parkinson's disease from those without. The study found that RF and SVM algorithms provide a reliable computational

method for estimating the presence of PD with high accuracy. In the works of Vikas et al. [12] and Kemal et al. [13], PD was diagnosed using acoustic data. In the study by Vikas et al. [12], a novel approach is used to partition and select features using Principal Component Analysis (PCA). Also, [13] uses a data sampling method known as one-against-all (OGA) to partition the dataset and extract features. In both studies, the partitioned dataset was validated using individual classifiers based on acoustic features, including weighted kNN, Logistic Regression (LR), and Medium Gaussian Kernel Support Vector Machine (MGSVM). They achieved a classification accuracy of over 74% and 77%, respectively.

Predicting PD with multimodal machine learning (ML) has shown significance in recent years due to its ability to combine diverse sources of information for more accurate, comprehensive, and early detection. Multimodal ML integrates data from various modalities, such as imaging, genetics, clinical observations, and acoustic-based data, to create a holistic and nuanced understanding of the disease. This approach offers several significant advantages, including enhanced accuracy, comprehensive biomarker discovery, and longitudinal monitoring. The authors in [14] aims to develop an automated ML package, GenoML, that uses ML and multimodal data to deliver accurate predictions and classifications of PD. They investigate top features, disease-relevant networks, and drug-gene interactions and perform automated ML on multimodal PD datasets to select and tune the best model. The significant results of the area under the curve (AUC) for the initial and tuned models demonstrate that combining data modalities outperforms the single biomarker paradigm. Studies have shown that there is a premotor stage preceding the onset of classic motor symptoms in PD diagnosis. This stage is characterized by a group of clinical features, including Rapid Eye Movement (REM), sleep behavior disorder (RBD), and olfactory loss. Prasanth et al. [15] used RBD alongside other biomarkers, such as cerebrospinal fluid measurements and dopaminergic image markers, to distinguish PD from normal subjects. The study employed Naive Bayes, SVM, Boosted trees, and RF for classification, with SVM showing the best performance. The study concludes that multimodal data aid in early detection. PD is characterized by difficulty in starting and stopping movements, besides several other motor symptoms. Juan et al. [16] proposed a methodology to model this difficulty by considering information from speech, handwriting, and gait. The transitions were used to train CNN to classify healthy and PD subjects. The authors also checked the robustness of the proposed model by considering speech signals in different languages. The results show that the fusion of information from the three modalities can accurately classify PD.

In essence, machine learning is shown to have transformative potential in utilizing MRI, acoustic, and multimodal data to advance the understanding of Parkinson's Disease. By merging these cutting-edge technologies, researchers can substantially contribute to early detection, personalized treatment, and improved quality of life for PD patients. As the field evolves, interdisciplinary collaborations and innovative solutions will pave the way for more effective PD management and care strategies.

III. Experimental Results

In this section, we detail a comprehensive series of experiments that leverage both MRI and acoustic data for the purpose of PD classification. The outcomes of these experiments are accompanied by appropriate analysis, providing valuable insights to facilitate informed decisions when selecting the most suitable model for PD classification.

A. Experimental Setup

In this section, we briefly describe the datasets used in our experiments, hyperparameters and performance metrices.

*1) **MRI Data**:* The NTUA dataset [17] encompasses binary labels denoting the severity of the neurodegenerative disorder, as quantified by the Clinical Dementia Rating (CDR) score. A CDR score of 0.0 signifies the absence of dementia, while a score of 1.0 indicates the presence of PD. This dataset comprises 4189 images categorized as CDR-0 and 4831 images categorized as CDR 1.0. These images possess dimensions of 128×128 pixels. These samples were then divided in the ration 70:30 for training and testing, respectively.

*2) **Replicated Acoustic Features Dataset**:* The dataset [18] encompasses 48 acoustic features extracted from three voice recording replications of sustained /a/ phonation for each of the 80 subjects, with 40 of them diagnosed with PD. It includes an array of vocal perturbation and spectral measures. These measures comprise pitch perturbation indicators like relative jitter and pitch perturbation quotient, amplitude perturbation metrics including local shimmer and amplitude perturbation quotients, harmonic-to-noise ratio in different frequency bands, Mel Frequency Cepstral Coefficients (MFCCs) and their derivatives, recurrence period density entropy (RPDE), detrended fluctuation analysis (DFA), pitch period entropy (PPE), and glottal-to-noise excitation ratio (GNE).

*3) **Hyperparameters**:* The hyperparameters for both the MRI and acoustic datasets were carefully selected using a 5-fold cross-validation technique. Specifically, distinct sets of hyperparameter values have been established for the MRI and acoustic datasets. The specific values chosen for these hyperparameters are provided in Table-I, ensuring a rigorous and unbiased approach to optimizing the model's performance.

*4) **Performance Metrics**:* The evaluation of models involved the utilisation of accuracy(A_c), sensitivity(S_e), and specificity(S_p) as performance metrics. Accuracy measures the ratio of correctly classified samples to the total number of samples. Specificity(S_p), or the true negative rate (TNR), quantifies the test's ability to correctly identify individuals without the specific ailment being tested. The false positive rate (FPR) can be derived as $1 - S_p$. Sensitivity(S_e), also referred to as recall or the true positive rate (TPR), evaluates the test's ability to accurately identify individuals with the targeted ailment. The false negative rate (FNR) can be calculated as $1 - S_e$. Furthermore, we took into account the computational runtime (in seconds) associated with the

TABLE I
HYPERPARAMETER VALUES OBTAINED FOR DIFFERENT MACHINE LEARNING MODELS USING 5-FOLD CROSS VALIDATION TECHNIQUE

ML Models	Hyperparameters (MRI)	Hyperparameters (Acoustic)
AdaBoost	learning rate=0.8, n_estimators=50, base_model used=DT	learning_rate=0.4 n_estimators=50 base_model=DT
GBoost	learning_rate=0.6 n_estimators=60	learning_rate=0.4 n_estimators=50
XGBoost	learning rate=0.3, n_estimators=100	learning_rate=1.0, n_estimators=10
RF	criterion=entropy, n_estimators=100	criterion=log_loss, n_estimators=60
DT	criterion=log_loss, splitter=best	criterion=gini, splitter=best
SVM	C=4.0, kernel=rbf	C=2.0, kernel=rbf
KNN	distance=minkowski n_neighbors=2	distance=minkowski, n_neighbors=4

Legend: AdaBoost – Adaptive Boosting; GBoost – Gradient Boosting; XGBoost – Extreme Gradient Boosting; RF – Random Forest; DT – Decision Trees; SVM – Support Vector Machines; KNN – K Nearest Neighbor; n_estimators – ; criterion – criterion for node split; n_neighbors – number of nearest neighbors; C – regularization parameter.

optimal hyperparameter values determined through a 5-fold cross-validation technique.

B. Results using MRI Data

This section presents experimental analysis of the machine learning model performance using MRI data. The models were trained on a set of 6314 samples and subsequently tested on 2706 samples. Results of this experimentation is presented in Table-II and Fig-1. Some observations from this experiment are outlined below:

TABLE II
COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS USING MRI DATA

ML Models	Acc	Spe	Sen	FNR	FPR	RT (secs)
AdaBoost	0.9301	0.9322	0.9282	0.0717	0.0677	8.2
GBoost	0.9531	0.9533	0.9432	0.0568	0.0467	12.5
XGBoost	0.9981	0.9992	0.9985	0.0014	0.0007	20.7
RF	0.9963	0.9992	0.9950	0.0049	0.0007	6.3
DT	0.9297	0.9236	0.9388	0.0661	0.0763	5.1
SVM	1.0	1.0	1.0	0.0	0.0	72
KNN	0.9988	0.9984	0.9992	0.0007	0.0015	2.8

Legend: AdaBoost – Adaptive Boosting; GBoost – Gradient Boosting; XGBoost – Extreme Gradient Boosting; KNN – K Nearest Neighbor; Acc – Accuracy; Spe – Specificity; Sen–Sensitivity; FNR – False Negative Rate; FPR – False Positive Rate; RT – Running Time.

- Table-II demonstrates the accuracy and running time of the algorithms. These results highlight the trade-off between accuracy and processing time for each algorithm. Among the tree-based algorithms, the XGBoost algorithm exhibited exceptional accuracy but required more processing time compared to other methods such as Decision Trees (DT) or RF. On the other hand, SVM achieved perfect accuracy but at the cost of increased processing time. K-Nearest Neighbor demonstrated impressive accuracy with efficient processing. This information provides valuable insights for researchers and practitioners seeking a suitable balance between accuracy and computational efficiency based on their specific application requirements.
- DT and AdaBoost have lower accuracy compared to some other algorithms due to their inherent characteristics and limitations. DTs can sometimes overfit the training data, meaning they create complex trees that capture noise and small fluctuations in the data. This can lead to reduced generalization to unseen test data, resulting in lower accuracy.
- AdaBoost is an ensemble technique that combines multiple weak learners to create a strong classifier. While AdaBoost can improve overall performance, it can be sensitive to noisy data and outliers, which may affect the quality of the base classifiers it relies on. In our case, the base classifier used was DT, which is prone to overfitting and impacts the AdaBoost's performance. Exploring a better base classifier might lead to improved accuracy.
- SVM achieved perfect accuracy, specificity, and sensitivity, with zero false negatives and false positives. This remarkable performance could be attributed to several inherent characteristics of SVM such as effective optimization of margin between classes, efficient handling of non-linear decision boundaries and resilience to outliers. In addition, as we utilized cross-validation to select optimal hyperparameters, the kernel function used to transform the data to higher-dimensional space may have allowed SVM to capture complex relationships between features.
- KNN on the other hand achieved high accuracy, specificity, and sensitivity, with extremely low false negatives and false positives. KNN makes decisions based on the classes of its nearest neighbors. When the neighbors share similar characteristics, KNN is likely to perform well.
- The algorithms tested obtained good sensitivity (True Positive Rate) and specificity (True Negative Rate) rates. The algorithms detected a large number of PD patients with sensitivity values of 92.82% to 99.92% proving the efficiency to detect people with PD. Similarly, Specificity scores varied from 92.36% to 100%, effectively identifying people without PD and reduce false positives. This combined achievement in sensitivity and specificity confirms the algorithms' clinical usability as accurate and reliable neurodegenerative illness diagnostic tools.

C. Results Using Accoustic Data

In this section, we present the results of ML models using the acoustic data. The experiment involved training ML models using a training dataset consisting of 56 records. Each record was characterized by 14 distinct acoustic features. These features served as the input variables for the models, enabling them to learn patterns and relationships within the data. Upon training, the models were put to test using a separate dataset of 24 records, each featuring the same 14 acoustic attributes.

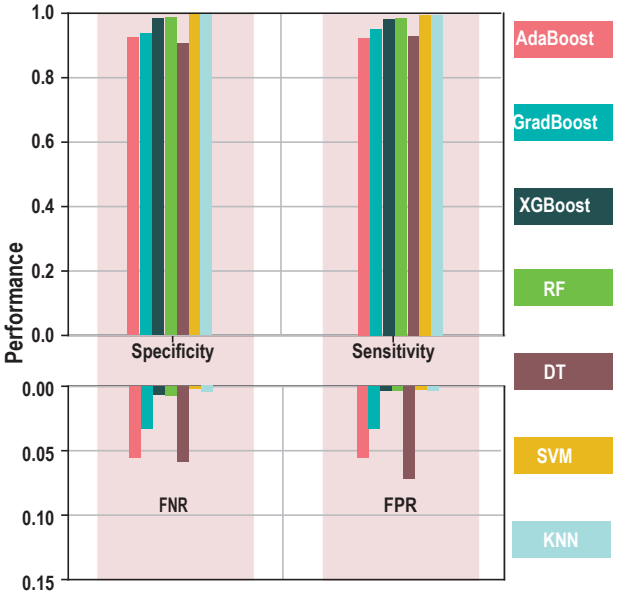

Fig. 1. Comparative Analysis of Diagnostic Performance Metrics: Specificity, Sensitivity, False Negative Rate (FNR), and False Positive Rate (FPR) in Binary Classification of PD using MRI Data.

TABLE III
COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS USING ACOUSTIC DATA

ML Models	Accuracy	Spe	Sen	FNR	FPR
AdaBoost	0.375	0.1	0.5714	0.4285	0.9
GBoost	0.4583	0.5	0.4285	0.5714	0.5
XGBoost	0.4166	0.5	0.3571	0.6428	0.5
RF	0.3333	0.5	0.2142	0.7857	0.5
DT	0.3333	0.1	0.5	0.5	0.9
SVM	0.4583	0.8	0.2142	0.7857	0.1999
KNN	0.625	1.0	0.3571	0.642	0.0
Proposed Ensemble **(SVM+KNN)**	**0.75**	**0.7**	**0.7857**	**0.2142**	**0.30**

Legend: Spe – Specificity; Sen – Sensitivity; FNR – False Negative Rate; FPR – False Positive Rate.

Based on the obtained experimental results (see Table-III, several observations can be made regarding the performance of different machine learning models in the classification of PD using acoustic data:

- Accuracy Variation: The accuracy of the models varies across the methods, ranging from 33.33% to 62.5%. This suggests differing levels of success in correctly classifying instances from both the PD and Healthy Classes (HC).
- Sensitivity and Specificity Trade-off: The models demonstrate a trade-off between sensitivity (ability to detect PD cases) and specificity (ability to correctly classify HCs). Some models achieve high specificity but relatively lower sensitivity, while others strike a balance between the two.
- Discrimination of PD Cases from HC: Certain models, such as K-Nearest Neighbor, demonstrate high specificity, suggesting their effectiveness in correctly identifying individuals with PD. This is essential for accurate disease detection and diagnosis.
- Among the methods, K-Nearest Neighbor stand out for achieving relatively balanced FPR value. It exhibits a reasonable trade-off between correctly identifying PD cases and minimizing false positives.
- RF and SVM have very high FNR values, indicating a substantial risk of missing actual PD cases. This could be problematic for a medical application where early detection is crucial.
- Potential for Further Improvement: Tree-based models (e.g., DT and RF) exhibit comparatively lower accuracy and sensitivity. This indicates that they might require additional feature engineering, or regularization to enhance their performance.

Overall, the choice of algorithm should consider the desired balance between correctly detecting PD cases and minimizing false positives, especially in the context of medical diagnosis. It is advisable to choose an algorithm that aligns with the specific requirements and priorities of the application.

Based on these results obtained from acoustic data we can conclude that Machine learning models require a diverse and representative set of training samples to learn meaningful patterns. Limited training samples as in this acoustic dataset might not adequately represent the underlying data distribution, leading to models that cannot generalize well to new data. Furthermore, as the acoustic dataset contains complex relationships and patterns, a small number of training samples might not be sufficient for the models to capture these nuances effectively. In addition, the limited feature dimension (14) can have an impact on the generalization ability of machine learning models.

D. Proposed Ensemble Model

In pursuit of achieving enhanced accuracy levels, we employed a simple ensemble approach by combining two of the best-performing models from our evaluations: Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) through soft-voting approach. The rationale behind this choice lies in the concept of algorithmic diversity, a crucial factor in successful ensemble strategies.

SVM and KNN were identified as strong candidates due to their consistent high performance across our assessments. Moreover, these two algorithms offer distinct characteristics that contribute to their diversity. KNN, a non-parametric method, excels at capturing local patterns within the data. In contrast, SVM, a parametric approach, excels at identifying non-linear decision boundaries within high-dimensional spaces. This inherent diversity in their functioning positions SVM and KNN to complement each other effectively. In addition, SVM often results in lower bias but higher variance, while KNN can lead to higher bias and lower variance. Ensembling these models helps strike a balance, potentially yielding a model with improved bias-variance trade-off. By fusing the strengths of SVM and KNN, we aimed to exploit their unique capabilities to collectively improve the overall classification performance. Our ensemble strategy leverages

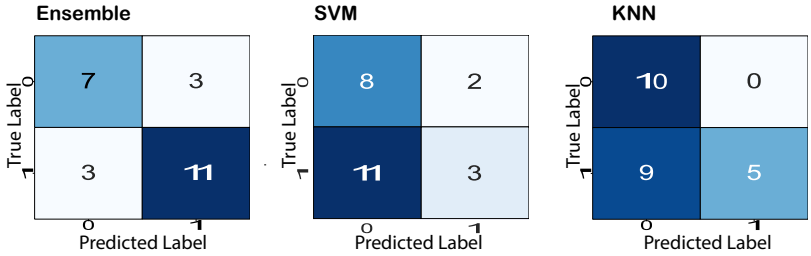

Fig. 2. Comparative Analysis of Ensemble and Individual Models

the local pattern recognition of KNN and the high-dimensional decision boundary identification of SVM. The collaborative nature of this ensemble has the potential to mitigate weaknesses and enhance accuracy, thereby showcasing the synergy achievable through a well-considered combination of diverse algorithms.

As can be seen in Table-III, the accuracy of the ensemble technique increased by 12% – a significant improvement considering the limitations of the dataset. In addition, the proposed ensemble demonstrates a balance between sensitivity and specificity, as both values are relatively high. This suggests that the model is capable of effectively identifying both positive and negative cases. Fig.2 presents the confusion matrices of ensemble, SVM and KNN methods.

IV. Conclusion

Our study addresses a critical aspect of PD classification by conducting a rigorous comparative analysis of prominent machine learning models using two distinct modalities: MRI and Acoustic data. By diverging from the common single-modality focus in existing works, our research uniquely contributes to the field's understanding of PD classification using MRI and acoustic data. Through meticulous evaluation, we unveil the performance nuances of various algorithms on both MRI and Acoustic data, uncovering the strengths and limitations of each modality. We hope that these insights empower researchers and practitioners to make well-informed decisions when selecting appropriate machine learning approaches for PD classification, based on the specific modality being considered.

Moreover, our study extends by introducing an ensemble approach that significantly enhances the performance of machine learning models for PD classification using the acoustic data. The ensemble model witnessed an overal performance increase of 12% in accuracy.

In essence, our work enriches the understanding of PD classification across distinct modalities, providing a comprehensive framework for selecting optimal algorithms and strategies. In future, we would like to work on providing comparative evaluation of machine learning and deep learning models using all possible modalities and biomarkers used in Parkinson's Disease Classification. In addition, we would like to investigate interpretable machine learning models that simultaneously delivers insights into the decision making process fostering trust in critical medical applications.

References

[1] F. Hajamohideen, N. Shaffi *et al.*, "Four-way classification of alzheimer's disease using deep siamese convolutional neural network with triplet-loss function," *Brain Inform.*, vol. 10, no. 1, pp. 1–13, 2023.

[2] N. Shaffi, V. Vimbi *et al.*, "Bagging the best: A hybrid svm-knn ensemble for accurate and early detection of alzheimer's and parkinson's diseases," in *Proc. Brain Informatics*, 2023, pp. 443–455.

[3] C. Ahlrichs and M. Lawo, "Parkinson's disease motor symptoms in machine learning: A review," *arXiv:1312.3825*.

[4] G. Costantini, V. Cesarini *et al.*, "Artificial intelligence-based voice assessment of patients with parkinson's disease off and on treatment: machine vs. deep-learning comparison," *Sensors*, vol. 23, no. 4, p. 2293, 2023.

[5] G. Pahuja and T. 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.

[6] R. Mathur, V. Pathak *et al.*, "Parkinson disease prediction using machine learning algorithm," in *Proc. ICETEAS 2018*, 2019, pp. 357–363.

[7] Z.-Y. Shu, S.-J. Cui *et al.*, "Predicting the progression of parkinson's disease using conventional mri and machine learning: An application of radiomic biomarkers in whole-brain white matter," *Magn. Reson. Med.*, vol. 85, no. 3, pp. 1611–1624, 2021.

[8] G. Solana-Lavalle and R. Rosas-Romero, "Classification of ppmi mri scans with voxel-based morphometry and machine learning to assist in the diagnosis of parkinson's disease," *Comput. Methods Programs Biomed.*, vol. 198, p. 105793, 2021.

[9] C. Salvatore, A. Cerasa *et al.*, "Machine learning on brain mri data for differential diagnosis of parkinson's disease and progressive supranuclear palsy," *J. neurosci. methods*, vol. 222, pp. 230–237, 2014.

[10] A. Boutet, R. Madhavan *et al.*, "Predicting optimal deep brain stimulation parameters for parkinson's disease using functional mri and machine learning," *Nat. commun.*, vol. 12, no. 1, p. 3043, 2021.

[11] D. Braga, A. M. Madureira *et al.*, "Automatic detection of parkinson's disease based on acoustic analysis of speech," *Eng. Appl. Artif. Intell.*, vol. 77, pp. 148–158, 2019.

[12] V. Mittal and R. Sharma, "Machine learning approach for classification of parkinson disease using acoustic features," *J. Reliab. Intell. Environ.*, vol. 7, no. 3, pp. 233–239, 2021.

[13] K. Polat and M. Nour, "Parkinson disease classification using one against all based data sampling with the acoustic features from the speech signals," *Med. Hypotheses*, vol. 140, p. 109678, 2020.

[14] M. B. Makarious, H. L. Leonard *et al.*, "Multi-modality machine learning predicting parkinson's disease," *npj Parkinson's Disease*, vol. 8, no. 1, p. 35, 2022.

[15] R. Prashanth, S. D. Roy *et al.*, "High-accuracy detection of early parkinson's disease through multimodal features and machine learning," *Int. j. med. inform.*, vol. 90, pp. 13–21, 2016.

[16] J. C. Vásquez-Correa, T. Arias-Vergara *et al.*, "Multimodal assessment of parkinson's disease: a deep learning approach," *IEEE j. biomed. health inform.*, vol. 23, no. 4, pp. 1618–1630, 2018.

[17] A. Tagaris, D. Kollias *et al.*, "Machine learning for neurodegenerative disorder diagnosis—survey of practices and launch of benchmark dataset," *Int. J. Artif. Intell. Tools*, vol. 27, no. 03, p. 1850011, 2018.

[18] L. Naranjo, C. J. Perez *et al.*, "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, 2017.