CONESCAPANHONDURAS2025paper29.pdf

Institute of Electrical and Electronics Engineers (IEEE)

Document Details

Submission ID
trn:oid:::14348:477758775

Submission Date
Jul 31, 2025, 11:52 PM CST

Download Date
Aug 12, 2025, 12:26 PM CST

File Name
CONESCAPANHONDURAS2025paper29.pdf

File Size
447.6 KB

6 Pages

3,885 Words

23,356 Characters

26% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

70 Not Cited or Quoted 20%
Matches with neither in-text citation nor quotation marks

10 Missing Quotations 2%
Matches that are still very similar to source material

9 Missing Citation 4%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

22% Internet sources

22% Publications

0% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

70 Not Cited or Quoted 20%
Matches with neither in-text citation nor quotation marks

10 Missing Quotations 2%
Matches that are still very similar to source material

9 Missing Citation 4%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

22% Internet sources
22% Publications
0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	arxiv.org 2%
2	Internet	doctorpenguin.com 1%
3	Internet	ouci.dntb.gov.ua <1%
4	Internet	eprints.whiterose.ac.uk <1%
5	Internet	www.medrxiv.org <1%
6	Internet	cinc.org <1%
7	Internet	hal.science <1%
8	Internet	www.frontiersin.org <1%
9	Publication	Rishit Singh, Yugnanda Malhotra, Jolly Parikh. "Integrative Approach for Early Det... <1%
10	Internet	iris.polito.it <1%

11 Internet
scyr.fei.tuke.sk <1%

12 Publication
H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in He... <1%

13 Internet
ijariie.com <1%

14 Publication
Saravanan Srinivasan, Parthasarathy Ramadass, Sandeep Kumar Mathivanan, Ka... <1%

15 Internet
d197for5662m48.cloudfront.net <1%

16 Internet
www.mdpi.com <1%

17 Internet
www.ncbi.nlm.nih.gov <1%

18 Internet
www.scielo.br <1%

19 Publication
Amina Adadi, Afaf Bouhoute. "Explainable Artificial Intelligence for Intelligent Tra... <1%

20 Internet
propulsiontechjournal.com <1%

21 Internet
9lib.net <1%

22 Publication
Daniela Xavier, Virginie Felizardo, Beatriz Ferreira, Henriques Zacarias et al. "Voic... <1%

23 Publication
Earl, Mark Paul. "Functional Group and Molecular Descriptor Aware Machine Lear... <1%

24 Publication
Laiba Zahid, Muazzam Maqsood, Mehr Yahya Durrani, Maheen Bakhtyar, Junaid B... <1%

25 Publication
Quoc Cuong Ngo, Mohammod Abdul Motin, Nemuel Daniel Pah, Peter Drotár, Pet... <1%

26 Publication
Machado, Diogo. "Advising Diabetes' Self-Management Supported by User Data i... <1%

27 Internet
prfastpower.weebly.com <1%

28 Publication
Dhiraj Baruah, Rizwan Rehman, Pranjal Kumar Bora, Priyakshi Mahanta, Kankana... <1%

29 Publication
Malay Vyas, Apurva. A. Mehta. "Chapter 21 Extracting the Pneumonia Signature: ... <1%

30 Publication
Wenhao Qi, Xiaohong Zhu, Bin Wang, Yankai Shi et al. "Alzheimer's disease digital... <1%

31 Publication
Keith Morrison. "Taming Randomized Controlled Trials in Education - Exploring K... <1%

32 Publication
Samsuddin Ahmed, Kun Ho Lee, Ho Yub Jung. "Robust Hippocampus Localization ... <1%

33 Internet
pmc.ncbi.nlm.nih.gov <1%

34 Internet
research.usq.edu.au <1%

35 Internet
revistamedica.com <1%

36 Internet
www.biorxiv.org <1%

37 Internet
ijeecs.iaescore.com <1%

38 Internet
isca-speech.org <1%

39 Internet
s3-eu-west-1.amazonaws.com <1%

40 Internet
sameni.info <1%

41 Internet
www.famemaine.com <1%

42 Internet
www.researchgate.net <1%

43 Publication
Jumanto Unjung, Rofik Rofik, Endang Sugiharti, Alamsyah Alamsyah, Riza Arifudi... <1%

44 Publication
Orhan Yaman, Fatih Ertam, Turker Tuncer. "Automated Parkinson's disease recog... <1%

45 Internet
assets.researchsquare.com <1%

46 Internet
gyan.iitg.ernet.in <1%

47 Internet
irosyadi.gitbook.io <1%

48 Internet
repositorio.uchile.cl <1%

49 Internet
vbn.aau.dk <1%

50 Internet
www.researchsquare.com <1%

51 Internet
www.rev-jec.org <1%

52 Internet
www.techethos.eu <1%

53 Publication
Birudala Venkatesh Reddy, Y V Krishna Reddy, Md. Abdur Razzak, Surender Reddy... <1%

54 Publication
Natália Soares Girão, Diogo Lugarini, Lúcia Valéria Ramos de Arruda, José Luís Fa... <1%

55 Publication
Recharla, Rugved. "AI-Powered Personalized Healthcare System for Disease Predi... <1%

56 Publication
Yang Wang, Tianying Liu, Jiaogen Zhou, Jihong Guan. "Video anomaly detection b... <1%

57 Internet
dokumen.pub <1%

58 Internet
uvidok.rcub.bg.ac.rs <1%

59 Internet
www.filesystems.org <1%

60 Internet
www.jeeemi.org <1%

61 Internet
www.journal.uestc.edu.cn <1%

62 Publication
Keshika Shrestha, H. M. Jabed Omur Rifat, Uzzal Biswas, Jun-Jiat Tiang, Abdullah-A... <1%

63 Publication
S. N. Kumar, Sherin Zafar, Sameena Naaz. "Computational Intelligence Algorithm... <1%

64 Publication
Siye Chen, Linghan Li, Shuyu Han, Wei Luo, Wenxia Wang, Yufan Yang, Xiaomeng ... <1%

65 Publication
Syed Nisar Hussain Bukhari, Kingsley A. Ogudo. "Ensemble Machine Learning App... <1%

66 Publication
Neny Sulistianingsih, Galih Hendro Martono. "Enhancing Predictive Models: An In... <1%

Parkinson's Disease Detection Through Vocal Features: Comparative Performance of Gradient Boosting Algorithms and a Stacked Ensemble Approach

1st Author
Department Name
Institution Name
City, Country

2nd Author
Department Name
Institution Name
City, Country

3rd Author
Department Name
Institution Name
City, Country

***Abstract*—Early detection of Parkinson's Disease (PD) remains a significant clinical challenge. The analysis of vocal features has emerged as a promising non-invasive biomarker; however, the robustness and reliability of predictive models are crucial for their clinical applicability. This work presents a comparative study of the performance of two gradient boosting algorithms, CatBoost and LightGBM, against a proposed Stacking ensemble architecture that combines Random Forest and XGBoost for PD classification. Using the publicly available UCI dataset, the models were evaluated through a rigorous protocol that included stratified splitting and cross-validation. The results show that while all models achieved high performance (F1-Score $>$ 0.92), the proposed model, despite a slightly lower overall accuracy, matched the best models in terms of Area Under the ROC Curve (AUC), with a value of 0.97. This finding suggests that the Stacking architecture provides greater reliability in class discrimination—an essential attribute for diagnostic support tools. It is concluded that the hybrid ensemble approach represents a more robust and clinically relevant solution for voice-based PD detection.**

***Index Terms*—Parkinson disease, Ensemble Learning, Vocal features , Stacking, Gradient Boosting**

I. Introduction

Parkinson's disease (PD) is a global public health emergency, with an estimated prevalence of 2–3% among individuals aged 65 and older, according to the latest epidemiological estimates [1]. As a multifactorial neurodegenerative disease, its pathology is characterized by the selective loss of dopaminergic neurons in the substantia nigra pars compacta, leading to the hallmark motor symptoms of resting tremor, bradykinesia, rigidity, and postural instability, along with a variety of non-motor manifestations [2]. Current clinical diagnosis, which relies mainly on motor criteria according to the Unified Parkinson's Disease Rating Scale (UPDRS), faces significant limitations in detecting the disease, typically identifying it in advanced stages when 60–80% of dopaminergic neurons have already degenerated [3].

In the development of biomarkers, early detection has become a high-priority research focus. Among the various approaches explored to date, computational analysis of vocal signals has proven to be particularly promising in the medical field, as up to 90% of patients develop speech impairments (hypokinetic dysarthria) during the course of the disease, often in its early stages [4]. Recent studies have measured these impairments using specific acoustic parameters: increases in fundamental frequency variability by approximately 30–50%, amplitude fluctuations by about 20–40%, and significant reductions in tonal range and articulation speed [5]. These subtle vocal anomalies, often imperceptible to the human ear, can be quantified using digital signal processing techniques, offering a unique window into early detection.

Machine learning systems have emerged as valuable tools for identifying these quantitative vocal patterns. Unlike traditional clinical methodologies that rely on subjective assessments, computational strategies enable the objective extraction of more than 300 distinct acoustic features from voice recordings [6]. However, their clinical implementation faces three main obstacles: first, limited model generalizability due to the widespread use of small (often under 200 samples) and demographically homogeneous datasets; second, the lack of standardization in acoustic feature extraction protocols and cross-validation methodologies; and third, the complex interaction of acoustic variability with early disease stages [7].

Despite significant advances in voice analysis for the early detection of Parkinson's disease, its translation into clinical practice still faces major challenges. According to E. Tolosa et al. [3], the lack of standardized protocols and the underestimation of early vocal symptoms are fundamental barriers. Additionally, the scarcity of healthcare professionals with specialized training in computational techniques has been highlighted by L. R. García et al. [4]. In light of these challenges related to standardization, clinical validation, and technology transfer, our research proposes the development of hybrid machine learning models designed to offer greater robustness and applicability in real-world clinical settings.

To address these challenges, this study conducts a comparative analysis of four machine learning architectures

particularly well-suited for vocal feature analysis: XGBoost (Extreme Gradient Boosting) [8], CatBoost (Categorical Boosting) [9], LightGBM (Light Gradient Boosting Machine) [10], and Random Forest [11]. Each of these algorithms offers distinct advantages: XGBoost excels in computational efficiency for tasks involving nonlinear relationships; CatBoost stands out for handling categorical variables without extensive preprocessing; LightGBM optimizes computational resource usage through histogram-based algorithms; and Random Forest contributes robustness against overfitting through the principle of ensemble averaging.

The methodological contribution of this work consists of designing a two-level stacking architecture using Random Forest and XGBoost in a hybrid model. This approach leverages the complementary strengths of Random Forest its ability to handle noisy data and outliers and XGBoost's precision in optimizing complex objective functions. Experimental validation was conducted using a protocol that included stratified data splitting (training and testing) and performance evaluation using metrics such as accuracy, F1-score, and AUC-ROC.

This paper is systematically organized to present these contributions. Section II details the materials and methods, including the dataset description and the processing pipeline. Section III and IV presents the experimental results, offering a comparative analysis of the evaluated algorithms and discussing the clinical implications of the findings, study limitations, and future research directions. Finally, Section V summarizes the main conclusions and their potential impact on the development of diagnostic tools.

II. MATERIALS AND METHODS

A. *Related literature*

This section provides a concise yet informative summary of previous research closely related to the present study, as detailed in Table I. These works represent a fundamental component for understanding the current research landscape, serving as both theoretical and contextual references. Through a critical analysis of these studies, it is possible to identify trends, patterns, and gaps in the existing knowledge, thereby highlighting the relevance of the present research.

For the identification of related literature, a systematic search was conducted in the Scopus database. The search strategy was designed to capture relevant studies at the intersection of Parkinson's disease, voice analysis, and machine learning techniques. Conceptually grouped keyword combinations were employed in the search process:

The first group, referring to the medical condition, included terms such as Parkinson's disease and Parkinson disease. The second group, focused on the analysis modality, included terms such as voice, speech, vocal features, acoustic signals, and acoustic parameters. Finally, a third group addressed methodological approaches and research objectives, using terms such as machine learning, deep learning, classification, classification algorithms, detection, detection methods, diagnosis, diagnostic system, feature selection, ensemble learning, ensemble methods, hybrid approach, hybrid machine learning, and gradient boosting.

Table I. The 8 most relevant scientific publications on this topic between 2012 and 2025, ranked by number of citations.

Authors (Year)	*Objective of the Article(Ref)*	*NC*
Tsanas A; Little MA (2012) [5]	Speech algorithms for accurate EP classification.	539
Gunduz H (2019) [12]	Deep learning for classifying EP with voice.	239
Alalayah KM et al. (2023) [13]	Early detection of PE: acoustic analysis and RFE.	49
Zhang L et al. (2020) [14]	Smart mobile system for voice-based EP diagnosis.	30
Suppa A et al. (2021) [15]	Objective diagnosis Essential Tremor: voice and ML.	21
Jain A et al. (2023) [16]	Differentiate dopaminergic response in PD with voice.	13
Hireš M et al. (2023) [7]	Inter-dataset generalization in voice EP detection.	11
García-Gutiérrez F et al. (2023) [2]	Voice to infer amyloid status in cognitive impairment.	7

In the research on Parkinson's disease (PD) detection through voice analysis, Tsanas and Little (2012) [5] established a milestone by achieving high accuracy with SVMs and novel vocal features, providing a benchmark and a relevant feature set for comparison with gradient boosting algorithms and hybrid ensembles. Subsequently, Gunduz (2019) [12] demonstrated the potential of deep learning (CNNs and LSTMs), outperforming traditional methods. This serves as an advanced reference for evaluating the effectiveness of gradient boosting and ensemble approaches, with its feature representation strategies potentially inspiring new designs. The importance of feature selection was highlighted by Alalayah et al. (2023) [13] through the use of RFE, a method particularly relevant for optimizing gradient boosting and hybrid ensemble models, while their comparison of base classifiers provides context for constructing such ensembles.

Moving toward practical application, Zhang et al. (2020) [14] validated a mobile system for PD diagnosis, emphasizing the relevance of real-world implementation for advanced algorithms such as gradient boosting and ensembles. Expanding the field, Suppa et al. (2021) [15] applied machine learning for the objective diagnosis of Essential Tremor, highlighting the potential of robust techniques like hybrid ensembles or gradient boosting to improve diagnostic objectivity in neurology. A more specific approach was proposed by Jain et al. (2023) [16], who used voice analysis to differentiate dopaminergic response in PD, suggesting that sophisticated algorithms may be valuable for disease monitoring.

Generalization across datasets a key challenge was investigated by Hireš et al. (2023) [7], and their findings are crucial for assessing the clinical robustness of gradient boosting and hybrid ensemble models. Finally, the versatility of acoustic analysis was demonstrated by García-Gutiérrez et al. (2023) [2].

B. Parkinson's Database

The Parkinson's dataset also known as the Oxford Parkinson's Disease Detection Dataset and available from the UCI Machine Learning Repository was contributed by Max Little in 2008. This dataset comprises 195 voice recordings from 31 participants, of whom 23 were diagnosed with Parkinson's disease (PD) and 8 were healthy individuals. For each recording, 22 numerical features were extracted from the voice signal. These features include various measurements of fundamental frequency (such as MDVP:Fo, Fhi, Flo, Jitter and its variants), amplitude variations (MDVP:Shimmer, Shimmer in dB, APQ), the ratio of noise to harmonic components (NHR, HNR), and nonlinear measures of dynamic complexity (RPDE, D2, DFA, spread1, spread2, PPE). The primary purpose of this dataset is binary classification to distinguish between healthy subjects and those with PD where a column labeled "status" indicates the diagnosis [17].

C. Modeling and Architecture

The experimental workflow, generally illustrated in Figure 1, it was designed to rigorously evaluate and compare different machine learning architectures

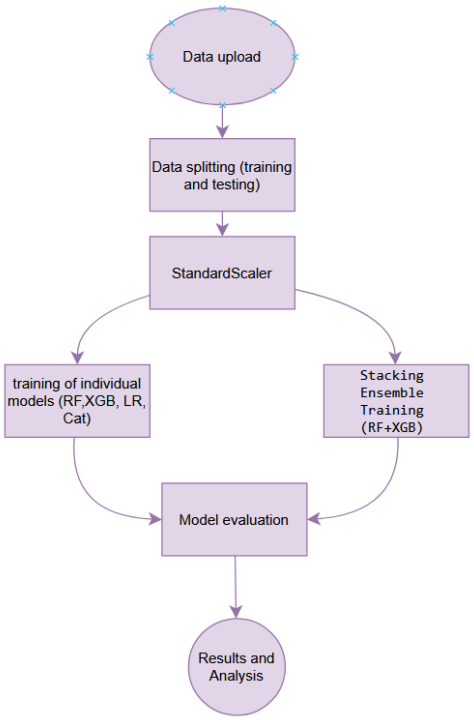

Fig. 1. General flowchart of the experimental pipeline, from data loading to final evaluation.

The process begins with the well-known "Parkinsons" dataset from the UCI repository. The predictive features (X) are separated from the target variable (y, the disease status), and non-informative columns are removed. The data is then split into 80% for training and 20% for testing using the train test split function, with stratified splitting to preserve the class distribution in both sets. Finally, before training, all features are normalized using StandardScaler to ensure a mean of zero and a standard deviation of one.

With the data prepared, two main approaches were evaluated. First, individual gradient boosting models (CatBoost and LightGBM) were trained as high-performance benchmarks. Second, a Stacking ensemble architecture was implemented, which constitutes the main contribution of this work

The architecture of the proposed model detailed in Figure 2, it specifically combines a Random Forest (RF) classifier and an XGBoost (XGB) classifier as base models (level 0), with a meta-classifier (level 1) that makes the final prediction.

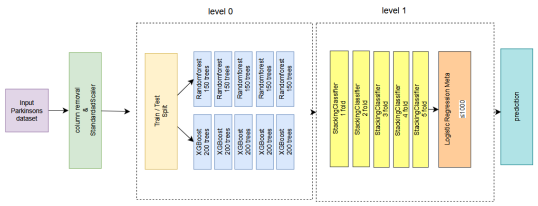

Fig. 2. Architecture of the proposed ensemble implemented with Random Forest and XGBoost as base models.

The core of this architecture lies in a StackingClassifier operating at two levels. At the first level (Level 0), to prevent data leakage, the base models (a RandomForestClassifier with 150 trees and an XGBClassifier with 200 trees) generate out-of-fold predictions using a 5-fold cross-validation procedure. These predictions are transformed into new meta-features.

At the second level (Level 1), these meta-features are concatenated with the original feature set an operation enabled by setting passthrough=True. This enriched dataset is then used to train the meta-classifier, which in this design is a LogisticRegression. Its role is to learn how to optimally weight and integrate all available information to produce the final prediction.

The performance of all trained models is ultimately evaluated on the test set, which remained isolated throughout the entire process, using metrics such as accuracy, precision, recall, and F1-score.

D. Evaluation Metrics

The evaluation relies on metrics derived from the confusion matrix (TP, TN, FP, FN), where TP and TN represent correct predictions, FP corresponds to Type I errors (false positives or false alarms), and FN represents Type II errors (false negatives or missed cases). Based on these values, the following key metrics are calculated:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

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

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5}$$

Model performance was evaluated through a multidimensional analysis. The overall ability to distinguish between classes was measured using the Area Under the ROC Curve (AUC), while the learning process was assessed by analyzing the training and validation curves. The training loss curve was used to verify the proper minimization of error, and the validation accuracy curve was employed to evaluate generalization capability. The simultaneous analysis of these two curves allowed for effective diagnosis of potential overfitting.

E. Hardware and software

The project was developed on a laptop equipped with an Intel Core i5 processor, 8 GB of RAM, and the Windows 10 Pro operating system. The code was implemented in Python 3.12 using Visual Studio Code, supported by libraries such as NumPy, Pandas, Scikit-learn, and XGBoost for data processing and modeling. Additionally, Matplotlib and Seaborn were used for visualization, and Joblib was employed to save the generated models.

III. Results

This section presents and analyzes the results obtained from the three implemented ensemble models. The evaluation is divided into a quantitative comparison of performance metrics and a detailed visual analysis of the models' behavior. Table II presents and analyzes the results obtained from the three implemented ensemble models. The evaluation is divided into a quantitative comparison of performance metrics and a detailed visual analysis of the models' behavior. Table X summarizes the key performance metrics obtained by each classifier on the test set.

It is important to emphasize that the entire final evaluation was conducted on the test set, consisting of 20% of the total data, which was not seen by the model during training. The complete dataset contains 195 voice recordings, of which 147 correspond to patients with Parkinson's disease and 48 to healthy subjects. Applying a stratified 80/20 split resulted in a test set composed of 29 samples from the Parkinson's class and 10 from the healthy class. This breakdown is essential for correctly interpreting the results of the confusion matrices presented in the visual analysis.

For a deeper understanding of the models' behavior, a detailed graphical analysis is presented below. The Figure 3 compares the performance of base models , meanwhile Figure 4 focuses exclusively on the proposed Stacking architecture.

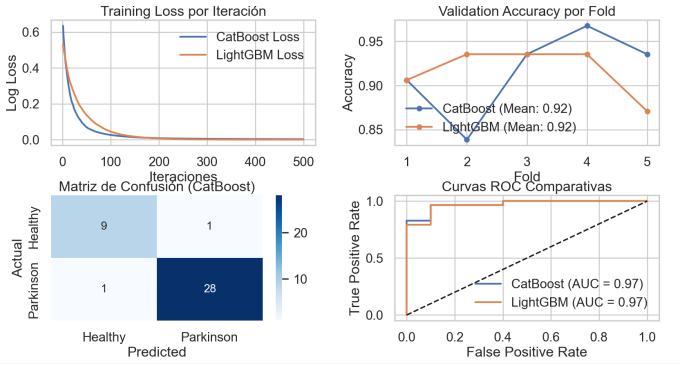

Fig. 3. Visualizing the performance of the CatBoost and LightGBM models.

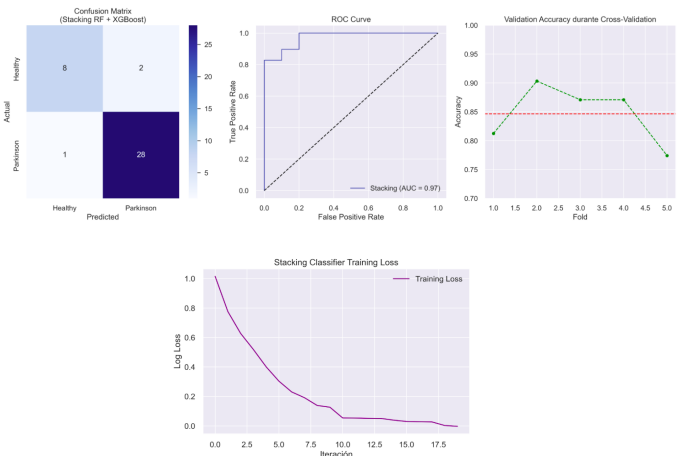

Fig. 4. Visual analysis of the proposed model (RF+XGB). The top row shows (from left to right): the confusion matrix, the ROC curve, and the accuracy during cross-validation. The bottom row shows the meta-classifier loss curve.

The Figure 4 provides a detailed view of the proposed model's performance. The confusion matrix reveals excellent performance on the majority class (Parkinson), with 28 out of 29 cases correctly classified. However, it shows 2 false positives and 1 false negative. Although the number of errors is low, minimizing false negatives (i.e., a patient with the disease being classified as healthy) is the most critical objective in a clinical diagnostic setting.

The ROC curve, with an Area Under the Curve (AUC) of 0.97, confirms the model's robust discriminative ability, matching the performance of the base models on this key metric and demonstrating its effectiveness in separating the classes. On the other hand, the validation accuracy plot shows some variability in performance across different folds, highlighting the importance of evaluating models through this method to obtain a reliable estimate of their generalization capacity.

Finally, the training loss curve of the meta-classifier exhibits a clear downward trend, converging smoothly. This indicates that the final model (Logistic Regression) successfully learned

Table II. Performance comparison between the ensemble models. Precision, Recall, and F1-Score metrics correspond to the weighted average

Modelo	Accuracy	Precision	Recall	F1-Score	ROC AUC
CatBoost [9]	0.95	0.95	0.95	0.95	0.97
LightGBM [10]	0.95	0.95	0.95	0.95	0.97
Proposed Stacking model (RF+XGB)	0.92	0.92	0.92	0.92	0.97

how to combine the predictions of the base models with the original features, consolidating the learning process of the ensemble.

IV. Discussion

The results obtained in this study demonstrate the high effectiveness of ensemble methods for Parkinson's disease classification based on vocal features. All three models implemented CatBoost, LightGBM, and the proposed Stacking ensemble—achieved outstanding performance, with precision and F1-scores exceeding 0.92 in all cases. This finding aligns with recent literature that positions decision tree-based algorithms as some of the most powerful solutions for this type of tabular classification problem.

The comparative analysis revealed an interesting trade-off between different aspects of performance. While CatBoost and LightGBM achieved slightly higher overall accuracy (0.95 vs. 0.92), the proposed Stacking architecture matched their performance on the most critical metric for clinical diagnostics: the Area Under the ROC Curve (AUC), reaching a value of 0.97. A high AUC indicates excellent discriminative ability, meaning the model's capacity to correctly distinguish between healthy and diseased individuals. It is hypothesized that the Stacking architecture, by combining predictions from heterogeneous models (Random Forest and XGBoost), constructs a more robust and generalizable decision boundary, which is reflected in the high AUC despite a slight trade-off in overall accuracy.

Computational efficiency also emerged as a key differentiating factor. LightGBM proved to be exceptionally fast, outperforming CatBoost by more than an order of magnitude without sacrificing predictive performance. This makes it a very attractive option for applications requiring rapid training or deployment in resource-constrained environments. On the other hand, the inherent complexity of the Stacking model, with its internal cross-validation process, makes it more computationally intensive, although its potential to improve diagnostic reliability may justify this additional cost.

It is also important to consider the limitations of this study. Firstly, the models were trained and validated using a single publicly available dataset. Although this dataset is widely used as a benchmark, validating the models' performance on larger, independent clinical cohorts is a necessary next step to confirm the generalizability of these findings.

V. Conclusion

This study investigated and compared the effectiveness of different ensemble learning architectures for the detection of Parkinson's disease using vocal features. Gradient boosting algorithms—CatBoost and LightGBM—were evaluated, and a hybrid Stacking model combining Random Forest and XGBoost was proposed, with the goal of developing a robust and accurate diagnostic tool.

The results demonstrated that all evaluated models exhibited high predictive performance, with F1-scores exceeding 0.92. The comparative analysis identified a clear trade-off between accuracy, discriminative robustness, and computational efficiency. LightGBM stood out for its exceptional training speed, offering a highly efficient solution without compromising accuracy. Meanwhile, the proposed Stacking architecture, despite a slight decrease in overall accuracy, matched the best performing models in terms of the Area Under the ROC Curve (AUC), reaching a value of 0.97. This result highlights its superior ability to reliably discriminate between healthy subjects and Parkinson's patients—an attribute of utmost importance in clinical applications.

It can be concluded that while gradient boosting algorithms like LightGBM provide a pragmatic and efficient solution, the Stacking ensemble architecture represents a significant advancement toward clinical reliability. By combining the strengths of heterogeneous models, the Stacking approach achieves greater robustness in class differentiation, which is essential for minimizing critical diagnostic errors, such as false negatives. This work reaffirms the potential of vocal analysis and machine learning as a non-invasive, accessible, and effective tool to support the early detection of Parkinson's disease.

For future work, it is recommended to validate these models on larger and more heterogeneous clinical datasets to confirm their generalizability. Additionally, the application of interpretability techniques, such as SHAP (SHapley Additive exPlanations), is suggested to help elucidate the acoustic factors that most influence the model's decisions, thereby increasing trust and transparency in the use of artificial intelligence tools in the medical field.

References

[1] Parkinson's Europe, "Parkinson's statistics," *Parkinson's Europe*, 2024. [Online]. Available: https://parkinsonseurope.org/facts-and-figures/statistics/

[2] F. García-Gutiérrez *et al.*, "Harnessing acoustic speech parameters to decipher amyloid status," *Front. Neurosci.*, vol. 17, pp. 1–12, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/37746151/

[3] E. Tolosa, A. Garrido, and S. W. Scholz, "Challenges in the diagnosis of parkinson's disease," *Lancet Neurol.*, vol. 20, no. 5, pp. 385–397, 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33894193/

[4] L. R. García, "Nuevos retos en la formación de logopedas en esta década," *Revista de Logopedia, Foniatría y Audiología*, vol. 42, no. 1, pp. 12–19, 2022. [Online]. Available: DOI:10.1016/j.rlfa.2022.07.001

[5] A. Tsanas *et al.*, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, 2012. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/22249592/

[6] G. Dimauro *et al.*, "Assessment of speech intelligibility in parkinson's disease using a speech-to-text system," *IEEE Access*, vol. 5, pp. 22 199–22 208, 2017. [Online]. Available: https://ieeexplore.ieee.org/document/8070308

[7] M. Hireš *et al.*, "On the inter-dataset generalization of machine learning approaches to parkinson's disease detection from voice," *International Journal of Medical Informatics*, vol. 179, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/37801807/

[8] T. Chen *et al.*, "Xgboost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016. [Online]. Available: https://arxiv.org/abs/1603.02754

[9] L. Prokhorenkova *et al.*, "Catboost: unbiased boosting with categorical features," *Advances in Neural Information Processing Systems*, vol. 31, 2018. [Online]. Available: https://arxiv.org/abs/1706.09516

[10] G. Ke *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Online]. Available: https://www.researchgate.net/publication/378480234_LightGBM_A_Highly_Efficient_Gradient_Boosting_Decision_Tree

[11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: https://link.springer.com/article/10.1023/A:1010933404324

[12] H. Gunduz, "Deep learning-based parkinson's disease classification using vocal feature sets," *IEEE Access*, vol. 7, pp. 115 540 – 115 551, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8807125

[13] K. M. Alalayah, "Automatic and early detection of parkinson's disease by analyzing acoustic signals using classification algorithms based on recursive feature elimination method," *Diagnostics*, 2023. [Online]. Available: https://www.mdpi.com/2075-4418/13/11/1924

[14] L. Zhang, "An intelligent mobile-enabled system for diagnosing parkinson disease: Development and validation of a speech impairment detection system," *JMIR Med*, 2020. [Online]. Available: https://www.sciencedirect.com/org/science/article/pii/S2291969420002653

[15] A. Suppa, "Voice analysis with machine learning: One step closer to an objective diagnosis of essential tremor," *Movement disorders*, 2021. [Online]. Available: https://doi.org/10.1002/mds.28508

[16] A. Jain, "Voice analysis to differentiate the dopaminergic response in parkinson's disease," *Frontiers*, 2023. [Online]. Available: https://doi.org/10.3389/fnhum.2021.667997

[17] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Oxford parkinson's disease detection dataset," 2008. [Online]. Available: https://doi.org/10.1109/TBME.2008.2005954