http://doi.org/10.35784/iapgos.5309 *received: 16.09.2023 | revised: 10.11.2023 | accepted: 26.11.2023 | available online: 20.12.2023*

COMPREHENSIVE MACHINE LEARNING AND DEEP LEARNING APPROACHES FOR PARKINSON'S DISEASE CLASSIFICATION AND SEVERITY ASSESSMENT

Oumaima Majdoubi, Achraf Benba, Ahmed Hammouch
Mohammed V University in Rabat, National School of Arts and Crafts, Electronic Systems Sensors and Nanobiotechnology, Rabat, Morocco

Abstract. *In this study, we aimed to adopt a comprehensive approach to categorize and assess the severity of Parkinson's disease by leveraging techniques from both machine learning and deep learning. We thoroughly evaluated the effectiveness of various models, including XGBoost, Random Forest, Multi-Layer Perceptron (MLP), and Recurrent Neural Network (RNN), utilizing classification metrics. We generated detailed reports to facilitate a comprehensive comparative analysis of these models. Notably, XGBoost demonstrated the highest precision at 97.4%. Additionally, we took a step further by developing a Gated Recurrent Unit (GRU) model with the purpose of combining predictions from alternative models. We assessed its ability to predict the severity of the ailment. To quantify the precision levels of the models in disease classification, we calculated severity percentages. Furthermore, we created a Receiver Operating Characteristic (ROC) curve for the GRU model, simplifying the evaluation of its capability to distinguish among various severity levels. This comprehensive approach contributes to a more accurate and detailed understanding of Parkinson's disease severity assessment.*

Keywords: Parkinson's disease, severity assessment, machine learning, XGBoost, Gated Recurrent Unit (GRU), comparative analysis

KOMPLEKSOWE METODY UCZENIA MASZYNOWEGO I UCZENIA GŁĘBOKIEGO DO KLASYFIKACJI CHOROBY PARKINSONA I OCENY JEJ NASILENIA

Streszczenie. *W tym badaniu naszym celem było przyjęcie kompleksowego podejścia do kategoryzacji i oceny ciężkości choroby Parkinsona poprzez wykorzystanie technik zarówno uczenia maszynowego, jak i głębokiego uczenia. Dokładnie oceniliśmy skuteczność różnych modeli, w tym XGBoost, Random Forest, Multi-Layer Perceptron (MLP) i Recurrent Neural Network (RNN), wykorzystując wskaźniki klasyfikacji. Wygenerowaliśmy szczegółowe raporty, aby ułatwić kompleksową analizę porównawczą tych modeli. Warto zauważyć, że XGBoost wykazał najwyższą precyzję na poziomie 97,4%. Ponadto poszliśmy o krok dalej, opracowując model Gated Recurrent Unit (GRU) w celu połączenia przewidywań z alternatywnych modeli. Oceniliśmy jego zdolność do przewidywania nasilenia dolegliwości. Aby określić ilościowo poziomy dokładności modeli w klasyfikacji chorób, obliczyliśmy wartości procentowe nasilenia. Ponadto stworzyliśmy krzywą charakterystyki operacyjnej odbiornika (ROC) dla modelu GRU, upraszczając ocenę jego zdolności do rozróżniania różnych poziomów nasilenia. To kompleksowe podejście przyczynia się do dokładniejszego i bardziej szczegółowego zrozumienia oceny ciężkości choroby Parkinsona.*

Słowa kluczowe: choroba Parkinsona, ocena ciężkości, uczenie maszynowe, XGBoost, Gated Recurrent Unit (GRU), analiza porównawcza

Introduction

Parkinson's disease (PD) is a progressive neurological condition that deteriorates as it progresses, primarily because of the untimely demise of dopaminergic neurons in the substantia nigra area [17]. Individuals are affected in their fundamental physiological systems, including respiration, balance, movement, and heart function [22]. In the early stages, their ability to speak fluently is hindered. Early detection of PD leads to extended patient lifespans, and accurate diagnosis requires robust health informatics tools. These solutions are created to assist healthcare professionals [7, 13] in assessing PD severity by utilizing diverse sensors. Symptoms of PD are linked to disturbances in dopaminergic pathways, where there is a deficiency of dopaminergic neurons, it results in a combination of motor and non-motor symptoms. Motor manifestations consist of trembling, rigidity, reduced mobility, and gait difficulties. On the other hand, non-motor symptoms encompass mood disorders, hallucinations, and incidents, urinary and reproductive issues, as well as sleep disturbances [5]. These symptoms emerge when approximately 60% of dopaminergic neurons remain, and they correlate with the effects of aging, ultimately diminishing quality of life [3]. This research supports healthcare professionals in assessing the severity of Parkinson's disease by employing various sensors. The approach involves utilizing diverse speech signal processing methods to capture vital attributes associated with Parkinson's disease. Subsequently, these attributes undergo analysis through machine learning algorithms, providing both dependable PD detection and an evaluation of disease severity. This information can facilitate early intervention and treatment strategies.

Recent years have witnessed a surge in the utilization of deep learning methods in medical research, showcasing their capacity to unravel intricate patterns from intricate data. This advancement enhances disease diagnosis accuracy and dependability. In this context, our research investigates the effectiveness of deep learning models. This entails utilizing contemporary neural network structures like MLP and RNN, in combination with established machine learning methods such as XGBoost and Random Forest, to improve the classification and severity evaluation of Parkinson's disease (PD).

Our research aims to bridge the gap between clinical expertise and computational prowess, presenting an encompassing framework. This framework not only precisely categorizes PD but also quantifies its severity. We tap into a well-curated dataset containing an array of patient features, executing a multifaceted strategy encompassing data pre-processing, exploratory analysis, model selection, regularization methods, and dimensionality reduction. By evaluating diverse models based on crucial metrics like precision, recall, and F1-score, we seek to comprehend the strengths and limitations of each approach concerning PD classification. Additionally, our study incorporates a GRU model to compute severity assessment percentages, enhancing the comprehensiveness of our analysis.

In this section, we will explore ongoing efforts in the classification of Parkinson's disease (PD). Using machine learning techniques and explore recent advancements in deep learning methods for the same purpose. Additionally, we will investigate the utilization of speech recordings not only for PD classification but also for evaluating disease severity. This entails using advanced computational models to categorize individuals based on PD presence and quantify the severity, offering crucial insights for personalized medical approaches.

The authors in [10] they established an advanced deep learning framework created to forecast Parkinson's disease using a dataset containing 42 voice recordings that have been subjected to pre-processing. Their study showcased improved accuracy compared to previous results. Nevertheless, it's noteworthy that the 81% accuracy achieved in 2018 is still regarded as relatively modest Within the scope of predicting Parkinson's disease.

The researchers Attained a 75% accuracy level and an 80% F1 score. In a research conducted by [20] they employed a 13-layer deep Convolutional Neural Network (CNN) model to detect Parkinson's disease from voice signals. Their experimentation involved a dataset comprising 20 patients. Despite achieving an 88% accuracy rate, their model made 361 incorrect predictions during the process.

In [21], introduced a novel classification approach for distinguishing between individuals with Parkinson's disease (PD) and those without, Utilizing dysphonia as a key factor, the researchers gathered data from 31 individuals, including 23 with PD and 8 healthy participants, who produced 195 sustained vowel sounds. Their method involved three crucial stages: feature computation, data preprocessing, feature selection, and the utilization of a linear kernel for classification. Impressively, the model achieved an accuracy of 91.4%.

In study [2], a novel deep learning framework, based on LSTM, is introduced to assess Parkinson's disease severity through gait pattern analysis. LSTM is utilized to capture temporal patterns in the data, eliminating the necessity for manual feature engineering and addressing the challenge of vanishing gradients. Furthermore, performance is enhanced through the incorporation of dropout, L2 regularization, and the application of the Adam optimizer. The results demonstrate outstanding performance with 98.6% accuracy in binary classification and 96.6% in multi-class classification, surpassing similar methods by 3.4%.

In [19] employed deep learning (DL) algorithms to detect Parkinson's disease (PD) and compared optimized and non-optimized methods. They found that K-fold cross-validation improved the accuracy of their approach. Another research [24] used artificial neural networks for PD detection.

In their study [23], researchers use machine learning algorithms on speech data to detect Parkinson's Disorder (PD) at an early stage, aiming to enable timely treatment and minimize its impact. They employ various models, including, Random Forest, Decision Tree , Naive Bayes, XGBoost, K Nearest Neighbor, and Support Vector Machine, along with Principal Component Analysis to reduce features. Their ensemble of the best-performing models achieves a 91% accuracy, improving disease recognition.

Our study investigates the application of advanced machine learning techniques, including XGBoost, Random Forest, MLP, and RNN, to enhance the classification and assessment of Parkinson's disease severity. Our dataset is meticulously curated, with features extracted from patient data. Our methodology encompasses a comprehensive data pre-processing phase that includes tasks such as data loading, managing missing values, and data normalization. Subsequently, we conduct exploratory data analysis, utilizing visual aids like histograms and box plots to gain insights into feature distribution and variability. These models undergo training and evaluation using performance metrics such as precision, recall, and F1-score. We enhance the effectiveness and ability to generalize of neural network models by integrating regularization methods such as L1 and L2.

Additionally, we utilize Principal Component Analysis (PCA) to simplify the intricacies of the dataset. The incorporation of GRU models enhances both the classification and severity assessment of PD. The integration of advanced ML methods provides insights into disease characterization. Notably, the analysis of severity percentages demonstrates the models' proficiency in evaluating severity. Remarkably, Random Forest achieves a severity analysis of 89.74%, with XGBoost closely following at 87.18%. These percentages bear clinical significance for well-informed medical decisions, emphasizing the importance of these models in Parkinson's disease analysis.

The results showcase promising achievements across all models, with XGBoost exhibiting an impressive precision of 97.4%. Random Forest, MLP, and RNN also demonstrate competitive precision rates. However, a more comprehensive analysis uncovers potential intricacies in accurately identifying negative instances. These insights underscore the significance of refining models to improve the detection of PD absence. Each model's performance is assessed using classification metrics, and detailed reports are generated to support in-depth comparisons.

The organization of this paper revolves around five distinct sections, ensuring a cohesive presentation of the content. Section 2 outlines the methodology employed in this study. Moving on to Section 3, we showcase the obtained results and engage in pertinent discussions. The conclusive insights drawn from this research are encapsulated in Section 4.

1. Methodology

1.1. Dataset

In this investigation, we make use of the dataset originally formulated by [15] with the principal objective of diagnosing PD by analyzing speech signals obtained from patients. The research encompasses voice samples collected from a total of 31 participants, among whom 23 exhibit Parkinson's disease while 8 are considered to be in a healthy condition. Within this dataset, there are 195 voice recordings, each characterized by 23 distinct attributes. The process of determining whether a particular voice recording pertains to an individual afflicted by Parkinson's disease or an individual in good health involves an assessment of 22 specific voice traits across the 195 biomedical voice recordings. This differentiation relies on the "status" column, where a value of 1 indicates the presence of Parkinson's disease and 0 indicates a healthy state. Further details concerning these 22 acoustic attributes can be found in table 1.

Fig. 1 depicts a histogram representing the acoustic properties of the dataset under analysis. Each chart shows the frequency of occurrences (on the y-axis) for different values (on the x-axis) associated with a particular attribute.

Table 1. Dataset description

Voice measure	Description
MDVP; Fo (Hz)	Average Vocal Fundamental Frequency
MDVP; Fhi (Hz	Maximum Vocal Fundamental Frequency
MDVP; Flo (Hz)	Minimum Vocal Fundamental Frequency
MDVP; Jitter (%)	Fundamental frequency perturbation (%)
MDVP; Jitter (Abs)	Absolute jitter in microseconds
MDVP; RAP	Relative Amplitude Perturbation
MDVP; PPQ	Five-point Period Perturbation Quotient
Jitter; DDP	Difference of differences between Cycles and period
MDVP; Shimmer	Shimmer Local amplitude perturbation
MDVP; Shimmer (dB)	Local amplitude perturbation (decibels)
Shimmer; APQ3	3-point Amplitude Perturbation Quotient
MDVP; APQ	11-point Amplitude Perturbation Quotient
Shimmer; APQ5	5-point Amplitude Perturbation Quotient
Shimmer; DDA	Average absolute difference between the amplitudes of consecutive periods
HNR, NHR	Harmonics-to-Noise Ratio, Noise-to-Harmonics Ratio
Status	Healthy (0) and Parkinson’s disease (1)
RPDE	Recurrence Period Density Entropy
DFA	Detrended fluctuation analysis
spread1, spread2, PPE	The fundamental frequency variation, pitch period entropy

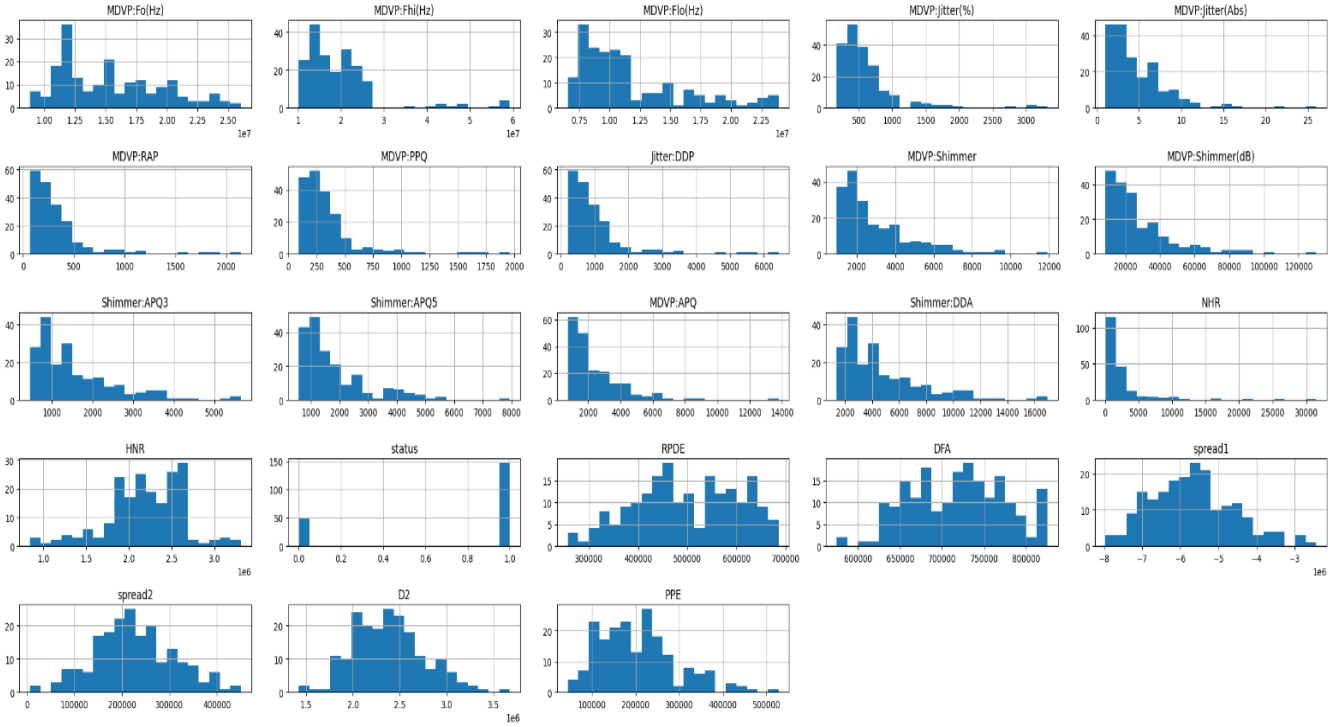

Fig. 1. Acoustic features histogram illustration

1.2. Feature extraction and advanced modeling techniques

This data analysis workflow commences by addressing missing data within the dataset. This process involves the removal of rows containing missing values and ensuring data conforms to the correct format by converting columns to numerical types while eliminating irrelevant columns. Following this, the dataset is partitioned into features (X) and the target variable (y). Subsequently, it is further divided into training and testing subsets. The data undergoes standardization using the "StandardScaler", and dimensionality reduction is executed through Principal Component Analysis (PCA).

Algorithm: Comprehensive Framework for PD Detection and Severity Assessment

Input: PD Dataset
Output: Disease Classification (PD or Healthy) and Severity Percentage

Commencement:
1. Load the PD dataset from a CSV file using Pandas.
2. Retrieve pertinent characteristics and transform them into an appropriate data format.
3. Preprocess the data:
   - Eliminate missing values through data cleansing.
   - Scale or standardize the features to the same range.
4. Split the dataset into training and testing subsets.
5. Perform PCA for dimensionality reduction:
   - Apply PCA to reduce feature dimensionality.

Training Phase:
6. Train a deep learning model:
   - Create a neural network with specified architecture.
   - Choose optimizer, activation functions, and regularization techniques.
   - Train the model on the training data using fit ().

Testing Phase:
7. Evaluate model performance on unseen test data:
   - Use the trained model to predict disease classification.
   - Compute metrics such as accuracy, precision, recall, and F1-score.
   - Create visual representations of the confusion matrix and ROC curve.
8. Calculate disease severity percentage using the trained model:
   - Utilize the trained model (GRU) to predict severity.
   - Apply the model to XGBoost, RandomForest, MLP, and RNN classifiers.
   - Calculate the percentage of severity based on different classifiers' predictions.

Conclusion.

When exploring models, a variety of neural network architectures are considered, commencing with a Sequential model that incorporates "Dense" and "Dropout" layers. This modeling approach utilizes the "Adam" optimizer and employs "sigmoid" activation for binary classification. Additionally, advanced ensemble techniques such as Random Forest are explored.

Furthermore, weight initialization in neural networks is fine-tuned, and advanced regularization methods, including L1 and L2 regularization, are applied to mitigate overfitting.

This comprehensive approach covers data pre-processing, feature extraction, model experimentation, and advanced techniques to ensure the development of robust and high-performing machine learning models.

The algorithm presents a thorough methodology crafted for the identification and evaluation of Parkinson's disease using voice signals. It commences by taking a Parkinson's disease dataset as input and systematically progresses through several stages. Initially, it divides the data, conducts sampling for training purposes, and then proceeds to train the model. During the training phase, a deep learning model is developed. Ultimately, the algorithm concludes by meticulously assessing the model's performance on unseen test data, providing valuable insights into its efficacy and suitability.

1.3. Data division

A traditional technique for splitting data entails a random division into two separate segments: one designated for training and the other for testing. The training subset is utilized to educate the model, while the test subset is employed to gauge its effectiveness. Typically, 80% of the data is assigned for training, leaving the remaining 20% for assessment. This methodology aligns well with your dataset.

1.4. Algorithms

XGBoost

XGBoost uses a convex loss function to quantify the difference between predicted and desired outcomes. This, combined with a penalty element to manage model complexity, seeks to minimize a regulated objective function involving L1 and L2 regularization. These functions essentially relate to regression trees. The iterative training process involves

integrating new trees that forecast residuals from prior trees. These additions are merged with existing trees to produce the final prediction. The term "gradient boosting" stems from employing gradient descent to minimize loss while incorporating new models [14]. Fig. 2 [11] shows the diagrammatic representation of XGBoost Algorithm.

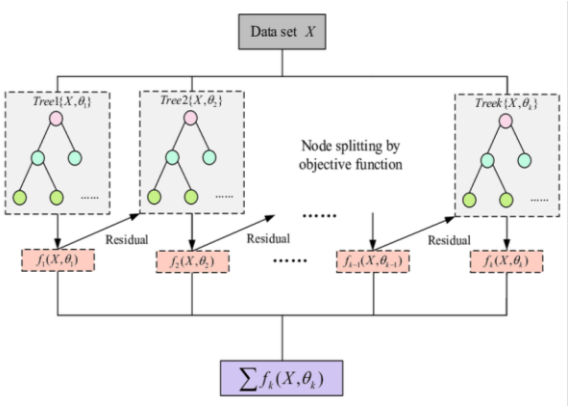

Fig. 2. Diagrammatic representation of XGBoost Algorithm

Random forest

A highly popular supervised machine learning algorithm adept at handling regression and classification tasks. It's based on ensemble learning, merging multiple classifiers to address complex challenges and bolster model performance. The "Random Forest" method employs multiple decision trees on diverse dataset subsets, combining their outcomes to boost dataset accuracy. Unlike relying on a single tree, it considers predictions from each tree to make its own based on majority consensus. Incorporating more trees heightens accuracy while guarding against overfitting [12].

MLP

The Multi-layer Perceptron, often referred to as MLP, represents a neural network architecture featuring densely connected layers that can adjust input dimensions to match desired output dimensions. It encompasses multiple layers of interconnected neurons, where the output of one neuron can serve as the input for another. An MLP typically comprises an input layer with one neuron for each input, an output layer with one neuron for each output, and it can include any number of hidden layers with varying node counts [6, 9]. In a standard MLP diagram, inputs are fed into the input layer, processed through the hidden layers, and ultimately yield outputs. All nodes shown in Fig. 3 within the MLP employ a sigmoid activation function to convert input values into a range spanning from 0 to 1 [4]. The sigmoid formula for this transformation is as follows: $\alpha(x) = 1 / (1 + \exp(-x))$.

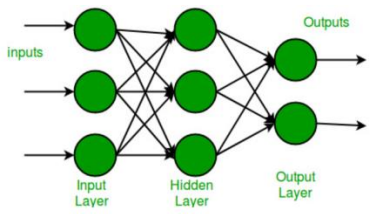

Fig. 3. A schematic diagram of a Multi-Layer Perceptron (MLP)

RNN

The network layers within an RNN establish cycles, which essentially involve utilizing the output of one layer as input for the next layer. Recurrent neural networks are commonly designed for tasks such as image description, automated translation, or processing natural language, as they assist in comprehending temporal or sequential data. RNNs find applications in tasks like automatic sleep apnea detection from nighttime ECG data [18] 16 and in the automated processing of speech [8].

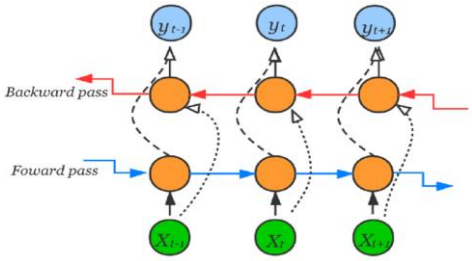

Fig. 4. The architecture of a recurrent network

2. Performance metrics

Performance metrics, as discussed [1], serve as essential tools for evaluating the effectiveness and precision of diverse models. These models rely on metrics such as accuracy, precision, recall, and the F1 score to make predictions based on provided data. In the assessment of these metrics, we consider four key components [16]:

- True Negatives (TN): in this context, we focus on correctly identifying negative cases.
- True Positives (TP): these represent instances where positive cases are correctly identified.
- False Positives (FP): conversely, false positives occur when the model incorrectly predicts a positive result for cases that are truly negative.
- False Negatives (FN): these arise when the model incorrectly predicts a negative outcome for cases that are actually positive.

Accuracy:

$$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN)$$

Precision:

$$\text{Precision} = TP / (TP+FP)$$

Recall:

$$\text{Recall} = TP / (TP+FN)$$

F1 score:

$$\text{F1Score} = (2\times\text{precision}\times\text{recall}) / (\text{precision} + \text{recall})$$

Specificity:

$$\text{Specificity} = TN / (TN+FP)$$

3. Results and discussion

3.1. Comparative analysis of machine learning models for classification performance

The role of performance comparison among models is to provide an objective and quantitative evaluation of their ability to minimize the classification task. Precision, measured individually for each model, helps quantify the accuracy of their positive predictions. In this specific case, we observe that XGBoost displays the highest precision, suggesting its capacity to minimize false positives. In comparison, the Random Forest exhibits slightly lower precision, indicating its ability to predict certain positive instances with slight uncertainty. MLP and RNN demonstrate similar precision performance, implying their capacity to maintain a good balance between true positives and false positives. This comparison sheds light on the respective strengths and weaknesses of the models, guiding the selection of the one that aligns best with the task's objectives and constraints.

The comparative analysis of machine learning model performances, as illustrated in table 2 for XGBoost, reveals significant differences. XGBoost stands out with the highest overall precision (0.974), showcasing its effectiveness in making accurate positive predictions. However, it is noteworthy that this high precision is not consistently maintained for the minority class, as highlighted in the table. While XGBoost excels in overall accuracy (97%), addressing concerns about its performance on the minority class could further enhance its effectiveness in diverse scenarios.

Table 2. Classification Report for XGBoost Model Performance

	precision	recall	F1-score	support
0	1.00	0.86	0.92	7
1	0.97	1.00	0.98	32
accuracy			0.97	39
macro avg	0.98	0.93	0.95	39
weighted avg	0.98	0.97	0.97	39

Table 3. Classification Report for Random Forest Model Performance

	precision	recall	F1-score	support
0	0.80	0.57	0.67	7
1	0.91	0.97	0.94	32
accuracy			0.90	39
macro avg	0.86	0.77	0.80	39
weighted avg	0.89	0.90	0.89	39

Table 4. Classification Report for MLP Model Performance

	precision	recall	F1-score	support
0	1.00	0.57	0.73	7
1	0.97	1.00	0.96	32
accuracy			0.92	39
macro avg	0.96	0.79	0.84	39
weighted avg	0.93	0.92	0.91	39

Table 5. Classification Report for RNN Model Performance

	precision	recall	F1-score	support
0	0.83	0.71	0.77	7
1	0.94	0.97	0.95	32
accuracy			0.92	39
macro avg	0.89	0.84	0.86	39
weighted avg	0.92	0.92	0.92	39

Turning to table 3, the Random Forest model exhibits slightly lower precision (0.897), indicating some challenges in maintaining precision, particularly for the minority class. This suggests that although Random Forest achieves an accuracy of 90%, further optimization may be beneficial for improving precision across both classes.

Meanwhile, as evident in tables 4 and 5, MLP and RNN demonstrate similar performances (0.923), with MLP showcasing a better balance between precision and recall. This balance is particularly crucial in scenarios where equal importance is placed on identifying both positive and negative instances. These findings emphasize the versatility of MLP in achieving a harmonious trade-off between precision and recall.

These insights underscore the need for a nuanced evaluation when determining the best-performing model. While XGBoost may excel in certain aspects, considering the broader context of precision, recall, and class-specific performance becomes imperative for making informed decisions. Additionally, exploring model-specific strengths and weaknesses, as outlined in the tables, allows for a more comprehensive understanding of how each model can be fine-tuned to achieve the best possible outcomes in diverse real-world applications.

In Fig. 5, the confusion matrices for different models offer a detailed look into the performance of the classification models. XGBoost exhibited a robust predictive ability, with only one misclassification out of 32 instances, showcasing its proficiency in accurately distinguishing between cases categorized as healthy (0) and unhealthy (1). The confusion matrices for MLP, RNN, and Random Forest models also displayed competitive outcomes, accurately predicting the majority of instances. However, a notable pattern emerged as these models encountered challenges in preventing false positives for the healthy class (0), as evident from the misclassifications.

These matrices provide a comprehensive understanding of the strengths and limitations of each model, streamlining their evaluation and selection for the classification task at hand, encompassing both healthy and unhealthy categories. By visually illustrating the actual and predicted class labels, the confusion matrices offer insights into the models' competence in correctly identifying instances from both classes. This information is crucial for informed decision-making about the appropriateness of each model for the classification task, aiding in the selection of the most suitable approach for the precise categorization of both healthy and unhealthy cases. The nuanced insights gained from the confusion matrices contribute to a more thorough assessment of each model's performance and guide the optimization process for enhancing their overall effectiveness in real-world applications.

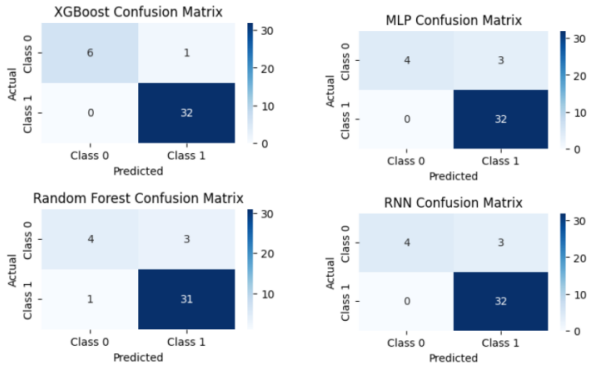

Fig. 5. Confusion Matrices for Different Models

3.2. Comparative Analysis of Machine Learning Models Using GRU for Disease Severity

The severity analysis percentages provide valuable insights into how effectively each machine learning model assesses the severity of the disease. Among the models examined, the XGBoost model achieved a severity analysis percentage of 87.18%, indicating its ability to accurately distinguish between severe and non-severe cases. The Random Forest model showed slightly better performance, with a severity analysis percentage of 89.74%. This suggests that the Random Forest model demonstrated a higher capability in correctly assessing the severity levels.

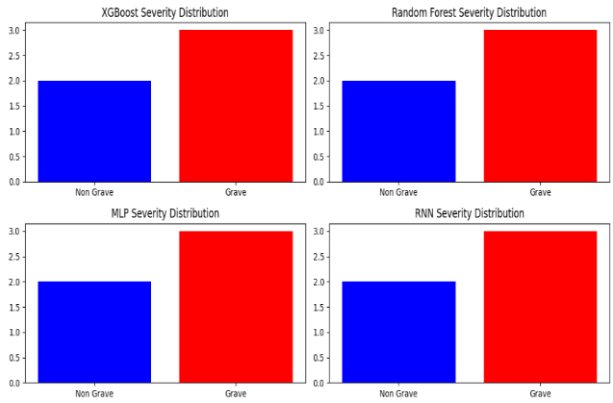

Fig. 6. Severity Distribution for Each Model

Additionally, in Fig. 6, the MLP and RNN models both exhibited a severity analysis percentage of 87.18%, aligning closely with XGBoost. This demonstrates that these models are consistent in their ability to analyze the severity levels, although they may face certain challenges, possibly in handling particular instances or class imbalances.

Regarding the GRU model, the provided code lacks its outcomes and severity analysis percentage. To comprehensively discuss its effectiveness, having the actual percentage calculated similarly to other models is crucial. Without this essential information, evaluating the GRU model's performance relative to others is challenging.

In summary, severity analysis percentages provide insight into each model's ability to evaluate disease severity. Notably, the Random Forest model stands out with the highest percentage. XGBoost, MLP, and RNN models consistently demonstrate their capabilities. To thoroughly evaluate the GRU model's efficiency, having its severity analysis percentage is essential for a comprehensive comparison.

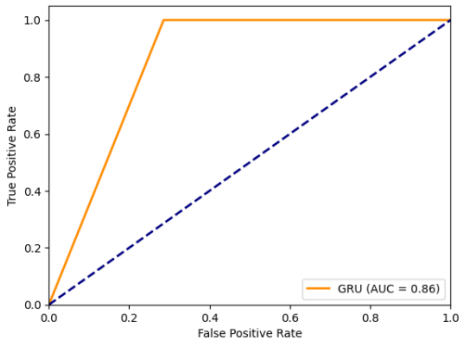

Fig. 7. Receiver Operating Characteristic (ROC) Curve for GRU Model

In Fig. 7, the achieved AUC of 0.86 for the GRU model demonstrates its effective performance in binary classification tasks. This robust performance signifies the model's reliability in distinguishing between positive and negative class samples, particularly in fields like healthcare, finance, and natural language processing. However, it's crucial to recognize that this performance should not be viewed in isolation. Future research endeavors should focus on enhancing model robustness, investigating the impact of class imbalances, and optimizing hyperparameters to maximize its utility in specific applications. In essence, this study contributes to advancing our understanding of GRU models in binary classification and paves the way for promising developments across various scientific and technological domains.

4. Conclusion

Our study harnessed advanced machine learning and deep learning techniques to classify Parkinson's disease and gauge its severity. Approaches like XGBoost, Random Forest, MLP, and RNN boosted disease categorization accuracy. Our work involved curated datasets, thorough preprocessing, and exploratory analysis to reveal crucial feature insights. Remarkably, our models achieved high precision, particularly XGBoost at 97.4%. However, challenges in identifying negative cases highlighted the need for model refinement. Furthermore, we introduced a GRU model for severity prediction, expanding our research horizon. By merging machine and deep learning, we contribute to disease diagnosis and prognosis advancement, facilitating potential personalized medical interventions. The findings emphasize computational methods' importance for timely interventions by healthcare professionals, ultimately enhancing the well-being of individuals grappling with this debilitating condition. In terms of future prospects, leveraging larger datasets and innovative approaches will likely enhance accuracy and personalized healthcare not only for Parkinson's disease but also for other neurological disorders.

References

[1] Abunadi I.: Deep and hybrid learning of MRI diagnosis for early detection of the progression stages in Alzheimer's disease. Connect. Sci. 34, 2022, 2395–2430.
[2] Balaji E. et al.: Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network. Applied Soft Computing 108, 2021, 107463.
[3] Benba A., Jilbab A., Et Hammouch A.: Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with Parkinson's disease and healthy people. International Journal of Speech Technology 19, 2016, 449-456.
[4] Bourdenx M. et al.: Identification of distinct pathological signatures induced by patient-derived ?-synuclein structures in nonhuman primates. Science advances 6(20), 2020, eaaz9165.
[5] Chaudhuri K. R., Schapira A. H.: Non-motor symptoms of Parkinson's disease: Dopaminergic pathophysiology and treatment. Lancet Neurol. 8, 2009, 464–474.
[6] El Bakali S., Ouadi H., Saad G.: Day-ahead seasonal solar radiation prediction, combining VMD and STACK algorithms. Clean Energy 7(4) (2023), 911–925.
[7] Erdogdu Sakar B., Serbes G., Sakar C. O.: Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease. PLoS ONE 12(8), 2017, e0182428.
[8] Gelly G.: Reseaux de neurones recurrents pour le traitement automatique de la parole. Ph.D. thesis, Université Paris Saclay (COmUE), Paris 2017.
[9] Gheouany S. et al.: Experimental validation of multi-stage optimal energy management for a smart microgrid system under forecasting uncertainties. Energy Conversion and Management 291, 2023, 117309.
[10] Grover S., Bhartia S., Yadav A., Seeja K.: Predicting severity of Parkinson's disease using deep learning. Procedia Comput. Sci. 132, 2018, 1788-1794.
[11] Guo R. et al.: Degradation state recognition of piston pump based on ICEEMDAN and XGBoost. Applied Sciences 10(18), 2020, 6593.
[12] Gupta I. et al.: PCA-RF: an efficient Parkinson's disease prediction model based on random forest classification. 2022, arXiv preprint arXiv:2203.11287.
[13] Gürüler H.: A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. Neural Computing & Applications 28(7), 2017, 1657-1666.
[14] Kumar A. et al.: A new Diagnosis using a Parkinson's Disease XGBoost and CNN-based classification model Using ML Techniques. International Conference on Advanced Computing Technologies and Applications – ICACTA. Coimbatore 2022, 1–6.
[15] Little M., McSharry P., Hunter E., Spielman J., Ramig L.: Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Nat. Preced. 2008.
[16] Majdoubi O., Benba A., Hammouch A.: Classification of Parkinson's disease and other neurological disorders using voice features extraction and reduction techniques. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOS 13(3), 2023, 16-22.
[17] Poewe W., Seppi K., Tanner C., Halliday G., Brundin P., Volkmann J., Schrag A., Lang A.: Parkinson disease. Nat. Rev. Dis. Prim. 3, 2017, 17013.
[18] Prakash P., Sebban M., Habrard A., Barthelemy J.-C., Roche F., Pichot V.: Détection automatique des apnées du sommeil sur l'ECG nocturne par un apprentissage profond en réseau de neurones récurrents (RNN). Médecine du Sommeil 18(1), 2021, 43-44.
[19] Quan C., Ren K., Luo Z., Chen Z., Ling Y.: End-to-end deep learning approach for Parkinson's disease detection from speech signals. Biocybern. Biomed. Eng. 42, 2022, 556-574.
[20] Rehman A. et al.: Parkinson's disease detection using hybrid lstm-gru deep learning model. Electronics 12(13), 2023, 2856.
[21] Sharanyaa S., Renjith P. N., Ramesh K.: An exploration on feature extraction and classification techniques for dysphonic speech disorder in Parkinson's Disease. Inventive Communication and Computational Technologies – ICICCT. Singapore, 2022.
[22] Sriram T. V. S., Rao M. V., Narayana G. V. S., Kaladhar D. S. V. G. K.: Diagnosis of Parkinson disease using machine learning and data mining systems from voice dataset. 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications – FICTA. Berlin, 2014, 151–157.
[23] Tallapureddy G., Radha D.: Analysis of Ensemble of Machine Learning Algorithms for Detection of Parkinson's Disease. International Conference on Applied Artificial Intelligence and Computing – ICAAIC. Salem, 2022, 354–361.
[24] Yasar A., Saritas I., Sahman M., Cinar A.: Classification of Parkinson disease data with artificial neural networks. IOP Conf. Ser. Mater. Sci. Eng. 675, 2019, 012031.

M.Sc. Oumaima Majdoubi
e-mail: oumaima_majdoubi@um5.ac.ma

Received the Master's degree in Electrical Engineering from the National School of Arts and Crafts of Rabat, Mohammed V University, Rabat, Morocco, in 2022. She is a research student of sciences and Technology of the Engineer in Ecole Nationale Supériure d'Informatique et d'Analyse des Systémes ENSIAS. She is part of E2SN research team at ENSAM-RABAT. She is interested in the detection of neurological diseases based on artificial intelligence and signal processing.

http://orcid.org/0009-0000-2968-7975

Prof. Achraf Benba
e-mail: achraf.benba@ensam.um5.ac.ma

Received the Ph.D. degree in electrical engineering from ENSAM, Mohammed V University, Rabat, Morocco, in 2017. He is a member of E2SN teams at ENSAM, Mohammed V University. His interests are in speech processing for detecting people with neurological disorders and Cardiac.

http://orcid.org/0000-0001-7939-0790

Prof. Ahmed Hammouch
e-mail: ah.hammouch@gmail.com

Received the M.S. degree and the Ph.D. degree in automatic, electrical, electronic by the Haute Alsace University, Mulhouse, France, in 1993, and the Ph.D. degree in signal and image processing by the Mohammed V University, Rabat, Morocco, in 2004. From 1993 to 2013, He is a member of E2SN teams at ENSAM, Mohammed V University.

http://orcid.org/0009-0005-8691-6662