

# Integrating ML Diagnosis and Multi-Criteria Decision Making for Activity Recommendations in Alzheimer's and Parkinson's Patients

1<sup>st</sup> Carlos Freitas

*Polytechnic School of Porto  
School of Engineering  
Porto, Portugal  
1240521@isep.ipp.pt*

2<sup>nd</sup> Daniel Dias

*Polytechnic School of Porto  
School of Engineering  
Porto, Portugal  
1240145@isep.ipp.pt*

3<sup>rd</sup> João Caseiro

*Polytechnic School of Porto  
School of Engineering  
Porto, Portugal  
1211334@isep.ipp.pt*

4<sup>th</sup> Pedro Rocha

*Polytechnic School of Porto  
School of Engineering  
Porto, Portugal  
1191689@isep.ipp.pt*

5<sup>th</sup> Vitor Castro

*Polytechnic School of Porto  
School of Engineering  
Porto, Portugal  
1140547@isep.ipp.pt*

**Abstract**—This paper reviews the state of the art for various machine learning techniques applied to Alzheimer's and Parkinson's disease diagnosis. We focus on XGBoost, Random Forests, Logistic Regression, Support Vector Machines (SVMs), and Decision Tree Classifiers, discussing their relevance in healthcare applications and their potential to support personalized activity suggestions for patients.

**Index Terms**—Alzheimer, Parkinson, Machine Learning, XGBoost, Personalized Activities, Healthcare

## I. INTRODUCTION

Neurodegenerative diseases such as Alzheimer's and Parkinson's pose significant challenges for healthcare systems worldwide. Both diseases are progressive and often lead to significant cognitive and physical impairments. Traditional diagnostic methods can be invasive, time-consuming, and expensive, necessitating innovative approaches to improve early detection.

Machine Learning (ML) has emerged as a powerful tool for diagnosing complex diseases by analyzing patterns within large datasets [1]. In this paper, we explore the state of the art of ML techniques, including XGBoost [1], Random Forests [2], Logistic Regression, SVMs [3], and Decision Trees [7], highlighting their applications in diagnosing Alzheimer's and Parkinson's. Additionally, we discuss how these predictive models can be extended to suggest personalized activities that support patients' cognitive and physical well-being.

## II. STATE OF THE ART

This section provides an overview of the theoretical foundations and existing technologies related to the application of Machine Learning (ML) in healthcare. It includes an introduction to Artificial Intelligence (AI) and ML, a description of learning paradigms, technologies and tools, ML algorithms, and a review of related work.

### A. Artificial Intelligence and Machine Learning

Artificial Intelligence (AI) enables machines to simulate intelligent human behavior, including learning, reasoning, and problem-solving. Machine Learning (ML), a subset of AI, focuses on algorithms capable of learning from data and improving performance over time [1]. These technologies are increasingly used in healthcare to identify patterns in medical data and support decision-making, especially in diagnosing neurodegenerative diseases such as Alzheimer's and Parkinson's.

### B. Learning Paradigms

ML encompasses various learning paradigms, each suited to different types of problems. This study focuses on supervised learning, a paradigm where models are trained with labeled data, associating inputs with desired outputs [2]. Supervised learning is highly effective for tasks like disease diagnosis, where labeled datasets are available. However, it requires large amounts of high-quality data, which can be costly and time-consuming to obtain.

### C. Technologies and Tools

The implementation of ML models for Alzheimer's and Parkinson's diagnosis relies on the following tools and frameworks:

- **Python:** A versatile programming language widely used for scientific computing and data analysis [3].
- **Scikit-learn:** A comprehensive Python library that provides a wide range of Machine Learning algorithms, including XGBoost, Random Forests, Logistic Regression, Support Vector Machines (SVMs), and Decision Trees, as well as utilities for model evaluation and validation [1], [2].

- **Pandas and NumPy:** Libraries for data manipulation and numerical computing, enabling efficient preprocessing and analysis of medical datasets [1].
- **Matplotlib:** A plotting library in Python used for visualizing the results of data analysis and model performance metrics, such as accuracy and ROC-AUC curves.
- **Jupyter Notebooks:** An interactive environment for developing and visualizing ML workflows, used extensively in the initial phases of this study [2].

#### D. Machine Learning Algorithms

Several ML algorithms have been applied to healthcare datasets for disease diagnosis. This subsection provides an overview of the main techniques used in this study:

1) *XGBoost*: XGBoost, a scalable and efficient implementation of gradient boosting, has demonstrated high accuracy in medical diagnosis [1]. Its ability to handle missing data and prioritize important features makes it particularly effective for healthcare datasets.

2) *Random Forests*: Random Forests are ensemble learning methods known for their robustness and interpretability. Their ability to reduce overfitting while maintaining high accuracy has been applied in various diagnostic tasks, including neurodegenerative diseases [2].

3) *Logistic Regression*: Logistic Regression is a classical statistical method often used as a baseline in medical studies. While simple, it is effective for binary classification problems such as the presence or absence of a disease [7].

4) *Support Vector Machines (SVMs)*: SVMs are widely used in high-dimensional datasets, making them suitable for analyzing medical data with numerous features. Their effectiveness in pattern recognition has been proven in imaging and biomarker studies [3].

5) *Decision Tree Classifiers*: Decision Trees offer intuitive decision-making paths, making them useful for explaining diagnostic outcomes to medical professionals. They are often used in conjunction with ensemble methods for better accuracy [7].

#### E. Related Work

Several studies have applied machine learning to neurodegenerative disease diagnosis. Here are some representative examples:

- **Alzheimer's Diagnosis:** Studies such as [4] and [5] explored the use of Random Forests and Support Vector Machines (SVMs) to analyze biomarkers and neuroimaging data. These methods achieved promising results for early detection of Alzheimer's disease, with accuracies exceeding 90.
- **Parkinson's Prediction:** XGBoost has been successfully applied to clinical and genetic data for identifying risk factors associated with Parkinson's disease, achieving high predictive accuracy and robustness in real-world datasets [6].
- **General Health Applications:** Logistic Regression and Decision Trees are widely used as baseline models in

healthcare applications due to their interpretability and simplicity. For example, [7] showed the effectiveness of Decision Trees in classifying patient risk levels in various clinical scenarios.

While these studies provide a foundation, our approach combines ML predictions with a multi-criteria decision-making framework to suggest personalized activities for patients, a novel contribution to the field.

### III. METHODOLOGY

The methodology for this project consists of the following steps:

#### A. Data Collection

Two datasets were used:

- **Alzheimer's Dataset:** Sourced from Kaggle [4], it includes biomarker and demographic information.
- **Parkinson's Dataset:** Sourced from Kaggle [6], it contains features such as vocal metrics, demographic details, and clinical data relevant for Parkinson's diagnosis.

#### B. Data Preprocessing

The datasets were preprocessed to ensure consistency and quality. Steps included:

- Handling missing values using imputation techniques.
- Normalizing features to ensure comparability across variables.
- Splitting datasets into training and testing subsets (80% training, 20% testing).

#### C. Algorithm Implementation

The following algorithms were implemented using Scikit-learn:

- **XGBoost:** Optimized for feature selection and gradient boosting [1].
- **Random Forests:** Used for ensemble learning and reducing overfitting [2].
- **Logistic Regression:** Served as a baseline model for binary classification [7].
- **Support Vector Machines (SVMs):** Applied to high-dimensional data for pattern recognition [3].
- **Decision Trees:** Used for intuitive and interpretable decision-making [7].

#### D. Evaluation Metrics

The models were evaluated using the following metrics:

- **Accuracy:** Proportion of correct predictions.
- **Precision, Recall, and F1-Score:** Metrics for evaluating classification performance.
- **ROC-AUC:** To assess the model's ability to distinguish between classes.

#### IV. EXPECTED RESULTS

The expected results include:

- **Model Performance:** High accuracy and ROC-AUC scores for detecting Alzheimer's and Parkinson's using XGBoost and Random Forests, with other algorithms serving as benchmarks.
- **Feature Importance:** Identification of key biomarkers and clinical features contributing to model predictions.
- **Activity Recommendations:** Integration of ML predictions with a multi-criteria decision-making framework to generate personalized activity suggestions for patients.

TABLE I  
ALGORITHMS AND EXPECTED BENEFITS

Algorithm	Expected Benefit
XGBoost	High accuracy, handles missing data
Random Forests	Robust, reduces overfitting
Logistic Regression	Simple, interpretable baseline
SVM	Effective for high-dimensional data
Decision Trees	Intuitive, easy to interpret

#### V. IMPACT DISCUSSION

This project has the potential to significantly impact the healthcare sector by:

- **Early Detection:** Enabling early diagnosis of Alzheimer's and Parkinson's through ML, potentially improving patient outcomes.
- **Personalized Care:** Tailoring activity recommendations to individual patient needs, promoting cognitive and physical engagement.
- **Scalability:** Demonstrating how ML models can be integrated into healthcare workflows for broader adoption in clinical practice.

Challenges include ensuring data privacy, model generalizability across populations, and acceptance among healthcare providers.

#### VI. DISCUSSION AND FUTURE DIRECTIONS

The results of this project demonstrate the potential of Machine Learning in improving the diagnosis of neurodegenerative diseases and supporting patient-centered care. However, several challenges remain, such as ensuring the generalizability of the models to diverse populations and validating the activity recommendations in real-world scenarios.

Future directions for this work include:

- Expanding the datasets to include more diverse populations for better generalizability.
- Testing the integration of ML models with healthcare decision-making systems in clinical settings.
- Exploring additional neurodegenerative diseases where similar methodologies can be applied.

#### VII. CONCLUSION

This study demonstrates the potential of Machine Learning to enhance the diagnosis of neurodegenerative diseases such as Alzheimer's and Parkinson's. By leveraging state-of-the-art algorithms like XGBoost and Random Forests, we aim to improve early detection and integrate the results into a multi-criteria decision-making framework for personalized activity recommendations.

Future work will focus on validating the activity recommendations through pilot studies and exploring scalability in diverse clinical settings. This project paves the way for innovative and patient-centered applications of AI in healthcare.

#### REFERENCES

- [1] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in \*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining\*, 2016.
- [2] L. Breiman, "Random Forests," in \*Machine Learning\*, 2001.
- [3] M. Scholkopf and A. J. Smola, "Learning with Kernels: Support Vector Machines," MIT Press, 2002.
- [4] H. Mohamed, "Alzheimer's Dataset," Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images>.
- [5] S. Ahmad et al., "Early Detection of Alzheimer's Disease Using Random Forest and SVM Models," \*Frontiers in Neurology\*, vol. 13, 2023.
- [6] R. El Kharoua, "Parkinson's Disease Dataset Analysis," Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/rabieelkharoua/parkinsons-disease-dataset-analysis>.
- [7] A. Brown et al., "Decision Tree Analysis in Medical Diagnostics," \*Journal of Clinical AI\*, vol. 10, no. 2, 2021.