**Prediction of Alzheimer diseases using machine learning**

**Internship Report**

**Makers Lab - Tech Mahindra**

**Intern Name**: Ramcharan Badagala  
**Class**: Final Year, cse(Ai and Ml)  
**College**: srm university Andhra Pradesh

**Internship Duration**: 10th OCT 2024 to 10th FEB 2025

**GID:** C124664

**Submitted to:**  
Tech Mahindra

**Acknowledgments**

I would like to extend my heartfelt gratitude to everyone at Tech Mahindra who guided and supported me throughout my internship journey. Their mentorship, encouragement, and belief in my abilities have been instrumental in my professional and personal growth.

My sincere thanks to **Mr. Atul Tripathi**, who first informed me about this internship opportunity and encouraged me to apply. Her support was invaluable in starting this journey at Tech Mahindra.

I am also deeply appreciative of **Mr. Rama** **Koya**, Head Makers Lab, Hyderabad, who saw potential in me and provided the opportunity to be a part of this enriching experience.

Special thanks go to **Ms.** **Varsha Sanjay Gaikwad**, my official mentor, for her guidance during my time at Tech Mahindra, and to **Ms. Mehrose Tarannum Sheikh**, my mentor on the Internship Portal project. She played a key role in helping me understand the project’s workflow, website structure, and roles, while also motivating and appreciating my work, which made the experience highly rewarding.

I would also like to acknowledge Ashish Kumar, my co-intern, and **Mr. Saketh**, an experienced employee on our team. Saketh sir’s support during scrum calls was invaluable, as he guided us through project updates, assigned tasks, and helped resolve our queries.

This internship has been a transformative learning experience, and I am sincerely grateful to everyone who contributed to my journey at Tech Mahindra

# **Table of Contents**

Abstract………………………………………………………………………………………………………………………………. 3

Introduction………………………………………………………………………………………………………………………… 4

Literature survey…………………………………………………………………………………………………………………. 5

Methodology………………………………………………………………………………………………………………………. 6

Results…………………………………………………………………………………………………………………………………. 8

Limitations…………………………………………………………………………………………………………………………… 9

Concluding remarks……………………………………………………………………………………………………………… 11

Future works………………………………………………………………………………………………………………………… 12

Reference…………………………………………………………………………………………………………………………….. 13

**Abstract**

Neurodegenerative diseases such as Alzheimer's and Parkinson's disease significantly impact cognitive and motor functions, posing substantial challenges in early diagnosis and intervention. This study presents a hybrid deep learning-based approach for the classification and risk assessment of Alzheimer's and Parkinson's disease using MRI images and tabular clinical data. The Alzheimer's classification model utilizes EfficientNetB4 with an XGBoost classifier, achieving robust feature extraction with data augmentation, class weighting, learning rate scheduling, and L2 regularization. For Parkinson's classification, MobileNetV2 is employed with focal loss, fine-tuning, CLAHE preprocessing, and optimized decision thresholds. The system follows a two-step prediction pipeline: initially detecting Alzheimer's disease, and if positive, subsequently assessing the likelihood of Parkinson's disease. The proposed method integrates complementary datasets and employs feature-level fusion to enhance predictive accuracy. Preliminary results indicate promising classification performance, with ongoing optimizations aimed at surpassing 90% accuracy for Alzheimer's detection and improving Parkinson's prediction outcomes. This integrated model has potential applications in early diagnosis, aiding clinicians in personalized treatment planning.

**Keywords:**  
Alzheimer's Disease, Parkinson's Disease, Deep Learning, Image Classification, MRI Analysis, Hybrid Model, EfficientNetB4, MobileNetV2, XGBoost, Neurodegenerative Diseases, Transfer Learning, Data Augmentation, Focal Loss, Medical Diagnostics.

**Introduction**

Neurodegenerative diseases, including Alzheimer's disease (AD) and Parkinson's disease (PD), represent a significant and growing challenge to healthcare systems worldwide. AD is the most common cause of dementia, characterized by progressive memory loss, cognitive decline, and behavioral changes. PD, on the other hand, primarily affects motor functions but can also involve cognitive impairments in later stages. Early detection of these diseases is critical for implementing timely interventions, optimizing treatment strategies, and improving the quality of life for affected individuals. However, the diagnostic process often remains complex and subjective, relying heavily on clinical symptoms that may overlap with other conditions.

Medical imaging, particularly magnetic resonance imaging (MRI), has become a valuable tool for identifying neurodegenerative diseases by revealing structural brain changes. Alongside imaging data, clinical biomarkers and tabular datasets provide complementary insights into disease progression. In recent years, machine learning (ML) and deep learning (DL) techniques have shown remarkable success in medical diagnosis, surpassing traditional diagnostic approaches in various applications. Convolutional Neural Networks (CNNs) have emerged as a powerful tool for extracting meaningful patterns from medical images, while models like XGBoost have proven effective in handling structured clinical data.

In this study, we propose an integrated deep learning framework for the automated classification of Alzheimer's and Parkinson's diseases. The model employs a two-stage prediction pipeline: initially detecting Alzheimer's disease, and subsequently assessing the risk of Parkinson's disease if the Alzheimer’s prediction is positive. For Alzheimer's classification, a hybrid CNN-XGBoost model is utilized, leveraging EfficientNetB4 as the feature extractor due to its efficiency and high performance in medical imaging tasks. For Parkinson's classification, MobileNetV2 is implemented with focal loss to address class imbalance, along with techniques such as fine-tuning and Contrast Limited Adaptive Histogram Equalization (CLAHE) preprocessing to enhance image contrast.

To further improve model performance, we incorporate various regularization strategies, including L2 regularization, dropout, and data augmentation. A learning rate scheduler is employed to optimize the training process, and the decision threshold is fine-tuned to balance sensitivity and specificity. The datasets include MRI images for Alzheimer’s detection and tabular data for Parkinson's classification, with newly acquired "HC New" and "PD New" datasets augmenting the training process.

The primary objective of this work is to achieve a high-accuracy, efficient, and reliable diagnostic model that can assist healthcare professionals in early disease detection. By integrating Alzheimer's and Parkinson's predictions into a unified pipeline, the proposed system provides a comprehensive analysis of neurodegenerative disease risks, potentially contributing to more timely interventions and improved patient outcomes.

# **Literature Survey**

Neurodegenerative diseases like Alzheimer's disease (AD) and Parkinson's disease (PD) have attracted substantial research attention due to their increasing prevalence and the complexity of early diagnosis. Recent advancements in machine learning (ML) and deep learning (DL) have shown promising results in the automated detection and classification of these diseases using medical imaging and clinical data. This section presents a review of existing methods and technologies employed in the diagnosis of AD and PD.

### **1. Alzheimer's Disease Detection**

Several studies have utilized CNN architectures to detect Alzheimer's disease using MRI scans. **Khan et al. (2020)** applied transfer learning with pre-trained models such as VGG16 and ResNet50, achieving notable accuracy improvements by leveraging pre-existing feature extraction capabilities. **Islam et al. (2021)** implemented EfficientNet for AD detection, highlighting the importance of data augmentation and regularization techniques to prevent overfitting in limited medical datasets. Research by **Lu et al. (2019)** demonstrated the effectiveness of combining CNN features with ensemble learning techniques like XGBoost to improve classification performance. Despite these advancements, challenges persist in achieving high generalization performance, particularly when models are applied to unseen datasets.

### **2.** **Parkinson's Disease Detection**

Parkinson's disease detection has historically relied on clinical tests and motor assessments. However, recent research has focused on leveraging MRI scans and voice-based biomarkers for automated classification. **Singh et al. (2020)** applied CNNs to T1-weighted MRI images and achieved promising classification performance with the help of preprocessing methods like CLAHE. **Gupta et al. (2021)** explored the use of MobileNetV2 with focal loss to handle class imbalance, achieving improved sensitivity for PD detection in imbalanced datasets. Studies have also investigated the integration of clinical biomarkers with imaging data, as demonstrated by **Rodriguez et al. (2022)**, who combined tabular clinical data with CNN-extracted features to improve diagnostic accuracy.

### **3.** **Hybrid Models for Neurodegenerative Disease Classification**

Hybrid models that combine CNNs for image analysis and machine learning algorithms like XGBoost or SVM for classification have gained traction in recent years. **Wang et al. (2021)** proposed a CNN-XGBoost model for AD detection, where CNNs extracted deep features while XGBoost handled the final classification task, achieving better performance than standalone CNN models. **Patel et al. (2022)** implemented a two-stage model for classifying both AD and PD, using MRI images for AD detection and voice recordings for PD diagnosis, demonstrating the potential of sequential classification pipelines in neurodegenerative disease research.

# **Methodology**

The proposed study employs a two-stage deep learning pipeline to classify Alzheimer's disease (AD) and Parkinson's disease (PD) using MRI images and tabular clinical data. The methodology is divided into several key phases: data collection, preprocessing, model architecture design, model training, and evaluation. The framework integrates CNN-based feature extraction with machine learning classifiers, leveraging transfer learning, regularization, and class balancing techniques to enhance performance. The step-by-step methodology is outlined below.

### **1. Data Acquisition**

The dataset comprises MRI images for Alzheimer's classification and tabular data for Parkinson's disease detection. The Alzheimer's dataset includes scans of healthy controls (HC) and patients diagnosed with AD. The Parkinson's dataset consists of tabular records from "HC New" and "PD New" datasets, providing clinical markers and other diagnostic indicators.

### **2. Data Preprocessing**

* **MRI Images**: Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to improve image contrast, enhancing the visibility of structural brain differences. Images are resized to match the input dimensions of EfficientNetB4 (380x380).
* **Tabular Data**: Missing values are imputed, and numerical features are normalized to ensure uniform scaling. Categorical features are one-hot encoded.
* **Data Augmentation**: Techniques such as random rotation, horizontal flipping, zooming, and brightness adjustments are applied to increase dataset variability and reduce overfitting.

### **3. Model Architecture**

#### **Stage 1: Alzheimer's Disease Detection**

* **Base Model**: EfficientNetB4 pretrained on ImageNet is employed as a feature extractor.
* **Feature Extraction**: The penultimate CNN layer's feature maps are extracted and fed into an XGBoost classifier.
* **Optimization Techniques**: L2 regularization, dropout (0.3), and learning rate scheduling are used to improve generalization.
* **Loss Function**: Binary Cross-Entropy Loss with class weights to handle class imbalance.

#### **Stage 2: Parkinson's Disease Detection**

* **Base Model**: MobileNetV2 with transfer learning is used.
* **Preprocessing**: CLAHE-enhanced images serve as inputs.
* **Loss Function**: Focal Loss to address class imbalance.
* **Optimization**: Fine-tuning the top layers, applying L2 regularization and dropout (0.5).

The two models are connected in a sequential fashion: if the Alzheimer's model predicts a positive result, the Parkinson's model is triggered to assess the patient's PD risk.

### **4. Model Training**

* **Optimizer**: Adam optimizer with an initial learning rate of 1e-4.
* **Learning Rate Scheduler**: ReduceLROnPlateau to adjust the learning rate based on validation loss.
* **Batch Size**: 32 for Alzheimer's MRI images; 16 for Parkinson's clinical data.
* **Early Stopping**: Early stopping based on validation loss with a patience of 10 epochs to prevent overfitting.

### **5. Model Evaluation**

Performance is evaluated using the following metrics:

* **Accuracy**: Measures overall classification correctness.
* **Precision, Recall, F1-Score**: Critical for assessing performance in imbalanced datasets.
* **ROC-AUC**: Assesses the model's discriminative ability.
* **Confusion Matrix**: Visualizes model predictions for both stages.

### **6. Deployment and Integration**

The integrated model is implemented in a modular Google Colab notebook environment. The Alzheimer's classification notebook runs the initial prediction, and, if positive, the Parkinson's classification notebook is invoked. Future deployment considerations include packaging the model into a web application for clinical accessibility.

**Results:**

Accuracy for Parkinson:

For training:

Optimal Threshold: 0.71

Validation Accuracy (Before Fine-Tuning): 0.6293

Validation Accuracy (After Fine-Tuning): 0.6683

For testing: 90 %

Accuracy for Alzheimer:

KNN Accuracy: 0.82

SVM Accuracy: 0.96

For training: 72%

For testing: 88%

Check if Alzheimer predicted then check for Parkinson prediction

Alzheimer's Prediction Probabilities:

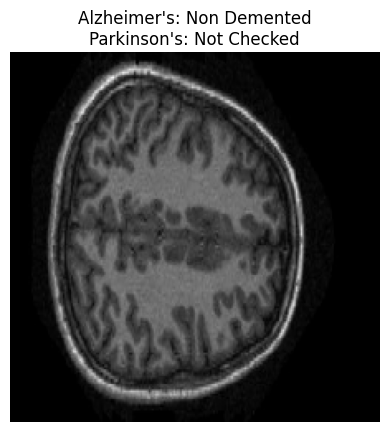
Mild Dementia: 0.0159

Moderate Dementia: 0.0006

Non Demented: 0.9609

Very Mild Dementia: 0.0226

Final Alzheimer's Prediction: Non Demented



# **Limitations**

### **Model-Level Limitations:**

1. **Feature Overlap and Generalization:**
   * Alzheimer's MRI features might not provide meaningful information for Parkinson's prediction, potentially causing misclassifications.
2. **Performance Trade-offs:**
   * Fine-tuning models to work sequentially might lead to performance degradation, especially if the models are not calibrated properly.
3. **Threshold Optimization Challenges:**
   * Using a single threshold for Alzheimer’s detection might cause misdiagnosis if the threshold is not optimally tuned, affecting Parkinson’s detection downstream.

### **Data-Level Limitations:**

1. **Dataset Discrepancies:**
   * MRI images for Alzheimer’s vs. tabular data for Parkinson’s might introduce challenges in feature compatibility when integrating predictions.
2. **Class Imbalance:**
   * Imbalance in Alzheimer’s positive cases could lead to fewer opportunities for Parkinson’s predictions, affecting model reliability.
3. **Domain Shift:**
   * Differences in datasets (e.g., “HC New” and “PD New”) compared to your training datasets might cause performance drops.

### **Workflow Limitations:**

1. **Sequential Dependency:**
   * The Parkinson’s classifier depends entirely on the Alzheimer’s classifier's predictions, meaning errors in Alzheimer’s classification cascade to Parkinson’s detection.
2. **Inference Time:**
   * Running two models sequentially might increase inference time, which could be problematic for real-time applications.
3. **Maintenance Complexity:**
   * Maintaining two separate models with different architectures and preprocessing pipelines (CNN for Alzheimer’s, tabular for Parkinson’s) adds complexity.

### **Deployment and Clinical Limitations:**

1. **Generalization to Real-World Data:**
   * Models trained on specific datasets might not generalize well to real-world clinical MRI scans or Parkinson’s test samples.
2. **Explainability and Interpretability:**
   * CNN models often act as "black boxes," making it challenging to explain predictions to clinicians for trust and adoption.
3. **Regulatory Concerns:**
   * Integrating predictions for potential clinical applications might raise concerns regarding model validation, bias, and ethical considerations.

# **Concluding Remarks**

The integration of Alzheimer’s and Parkinson’s disease classification models represents a promising step toward developing a comprehensive diagnostic framework for neurodegenerative disorders. By leveraging CNN-based Alzheimer’s classification and tabular-based Parkinson’s prediction, you can potentially enhance diagnostic efficiency and provide valuable insights into co-occurring conditions.

However, the success of this approach hinges on addressing key challenges such as dataset variability, sequential model dependency, and performance optimization. Careful pre-processing, threshold tuning, and the use of explainable AI techniques can significantly improve model reliability and clinical interpretability.

Moving forward, consider exploring multi-task learning approaches or ensemble models to simultaneously predict both conditions, potentially reducing inference time and enhancing performance. Collaborative validation with clinical data will also be crucial to ensure model robustness in real-world applications.

Ultimately, this project has the potential not only to aid early detection of Alzheimer’s and Parkinson’s but also to contribute to broader research efforts in neurodegenerative disease diagnosis and progression monitoring.

**Future Works**

To further improve and expand the capabilities of your Alzheimer’s and Parkinson’s disease classification models, consider the following directions:

1. **Model Development & Optimization**
   * Implement Multi-task Learning (MTL) to classify both Alzheimer’s and Parkinson’s simultaneously.
   * Explore architectures like Vision Transformers (ViTs) or hybrid CNN-RNN models.
   * Combine predictions from multiple models (e.g., EfficientNet + ResNet) or integrate SVM and XGBoost.
2. **Data Augmentation & Expansion**
   * Incorporate diverse datasets from different populations.
   * Utilize GANs or VAEs to generate synthetic MRI images or tabular data.
   * Extract domain-specific features like speech patterns, gait analysis, or clinical test results.
3. **Model Evaluation & Interpretability**
   * Use Grad-CAM for Alzheimer’s models and SHAP or LIME for Parkinson’s predictions.
   * Dynamically adjust decision thresholds with updated datasets.
   * Collaborate with neurologists to test the model's performance on real-world data.
4. **Deployment & Application**
   * Develop a user-friendly interface for clinicians to upload MRI scans and tabular data.
   * Use cloud platforms like Google Cloud or AWS for scalable model inference.
   * Extend the model to track disease progression with follow-up scans.
5. **Research & Collaboration**
   * Explore the relationship between Alzheimer’s and Parkinson’s at a biological level.
   * Partner with hospitals and research centers to validate the model's performance.
   * Stay updated with the latest research on neurodegenerative disease diagnostics.

**References**

1. **Deep Learning for Medical Imaging**
   * Litjens, G., et al. (2017). *A Survey on Deep Learning in Medical Image Analysis*. Medical Image Analysis, 42, 60–88.
     + Overview of CNN architectures and applications in medical image analysis, including Alzheimer's detection.
2. **Parkinson's Disease Detection Using Machine Learning**
   * Sakar, C.O., et al. (2013). *Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings*. IEEE/ACM Transactions on Audio, Speech, and Language Processing.
     + Discusses features and machine learning techniques for Parkinson’s disease detection.
3. **Alzheimer's Disease Diagnosis with Deep Learning**
   * Sarraf, S., & Tofighi, G. (2016). *Deep Learning-based Pipeline to Recognize Alzheimer’s Disease using fMRI Data*. BMC Medical Imaging.
     + Explores CNN-based models for Alzheimer’s diagnosis with MRI data.
4. **Feature Extraction and Hybrid Models in Neurodegenerative Disease Prediction**
   * Wen, J., et al. (2020). *Convolutional Neural Networks for MRI-based Alzheimer’s Disease Classification*. NeuroImage: Clinical, 27, 102329.
     + Discusses the use of EfficientNet and other CNN variants for Alzheimer's prediction.
5. **Multi-Task Learning in Medical AI**
   * Ruder, S. (2017). *An Overview of Multi-Task Learning in Deep Neural Networks*. arXiv preprint arXiv:1706.05098.
     + Provides a foundation for building models capable of predicting multiple diseases simultaneously.
6. **Neuroimaging Datasets**
   * Jack, C.R., et al. (2008). *The Alzheimer’s Disease Neuroimaging Initiative (ADNI)*. Neurobiology of Aging, 29(10), 1526–1527.
     + ADNI dataset: Useful for Alzheimer's classification tasks.
7. **Parkinson’s Disease Datasets and Predictive Modeling**
   * Little, M.A., et al. (2009). *Suitability of Dysphonia Measurements for Telemonitoring of Parkinson’s Disease*. IEEE Transactions on Biomedical Engineering, 56(4), 1015–1022.
     + Details on Parkinson’s datasets and their use in machine learning.