Multi-Class Classification of Neurodegenerative Diseases from Brain MRI Using Deep and Traditional Machine Learning Models

Lucas da Silveira Nascimento
Computer Engineering Undergraduate
CEFET-MG
Belo Horizonte, Minas Gerais
silveira0182@gmail.com

Rogério Martins Gomes
Department of Computing
CEFET-MG
Belo Horizonte, Minas Gerais
rogerio@cefetmg.br

***Abstract*—Neurodegenerative diseases are a growing global health concern, often leading to progressive cognitive and motor impairments. Accurate and early diagnosis remains a challenge due to the complexity and overlap of clinical symptoms. This study proposes an automated classification approach for Alzheimer's disease, Parkinson's disease, Lewy body dementia, and stroke using structural brain magnetic resonance imaging (MRI). We compared the performance of traditional machine learning algorithms—Random Forest, Support Vector Machines (SVM), and XGBoost—against a Convolutional Neural Network (CNN) in a multi-label classification framework. The models were trained and evaluated using over 18,000 MRI slices extracted from the OASIS-3 database. Our best model, a CNN architecture, achieved 95.07% accuracy and a Macro F1-score of 84.62% in a fixed-label prediction scenario. Additionally, we implemented a computer-aided diagnosis (CAD) system that enables real-time image visualization and probabilistic diagnostic output. The results highlight the potential of AI-based tools to support clinical decision-making in the early detection of neurodegenerative diseases.**

***Index Terms*—Machine learning, neurodegenerative disease, MRI, classification, CNN, computer-aided diagnosis**

I. Introduction

The human brain is the most complex organ in the body, and understanding its structure and function is fundamental to the diagnosis and treatment of neurodegenerative diseases [1]. These conditions are characterized by progressive cognitive and motor decline, affecting the quality of life of millions of individuals worldwide.

Although the precise etiology of neurodegenerative diseases remains unclear, multiple factors including aging, genetic predisposition, and environmental influences contribute to their development [2]. Clinical manifestations vary depending on the type of disease and the brain regions affected. Parkinson's disease (PD), for instance, typically presents with tremors, muscular rigidity, and speech impairments [2], while Alzheimer's disease (AD) is marked by memory loss, disorientation, and personality changes. Other conditions such as dementia with Lewy bodies (DLB) and Stroke [3] can exhibit overlapping or distinct symptoms. DLB shares features with Parkinson's disease but is distinguished by visual hallucinations and fluctuating cognition [4], [5]. Stroke, on the other hand, may lead to dementia depending on the affected brain areas, producing symptoms that can mimic those of AD [6].

In this context, early diagnosis plays an important role in managing neurodegenerative diseases. Brito et al. [7] emphasize the importance of neuroimaging techniques, such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT) for early detection and improved patient outcomes.

Among these, MRI has emerged as key tools in the non-invasive assessment of brain structure and pathology. MRI provides high-resolution, multi-planar views of soft tissues and allows the detection of subtle anatomical alterations such as hippocampal atrophy, white matter hyperintensities, and vascular lesions [3], [7]. A typical brain MRI scan may consist of 100 to 300 slices depending on slice thickness, which generally ranges from 1 to 3 mm. Thinner slices provide higher anatomical detail, while thicker slices can cover larger areas in less time. Additional MRI sequences, such as T1, T2, and FLAIR, further enhance diagnostic sensitivity by capturing various tissue contrasts and characteristics.

Given the vast volume and complexity of imaging data generated in MRI exams, artificial intelligence (AI) and machine learning (ML) techniques have emerged as powerful tools for identifying patterns associated with neurodegenerative diseases. These methods can automatically extract relevant features from brain images and detect subtle structural changes, such as hippocampal atrophy or amyloid plaque accumulation, thereby supporting diagnostic decisions. In particular, deep learning approaches like Convolutional Neural Networks (CNNs) [8] have demonstrated superior performance in classifying neurodegenerative conditions. When integrated into clinical workflows, such models have the potential to enable earlier and more accurate diagnoses, contributing to timely treatment and improved patient outcomes.

Despite growing interest in AI-based neuroimaging, most studies have focused predominantly on Alzheimer's disease, often through binary classification tasks (e.g., AD vs. controls).

This narrow focus limits generalizability and clinical applicability. Furthermore, disorders such as Lewy body dementia and stroke-related dementia remain underrepresented in the literature, highlighting the need for broader and more inclusive research efforts in this domain [9].

This work addresses these limitations by proposing a comprehensive machine learning (ML) framework for the automated classification of four neurodegenerative conditions from structural MRI. We explore the performance of both deep learning (CNN) and traditional models (Random Forest, SVM, and XGBoost) in multi-label and multi-class settings. In addition, we develop a Computer-Aided Diagnosis (CAD) interface that integrates model predictions with real-time MRI visualization, enhancing the interpretability and clinical usability of the system.

II. Related Works

In recent years, machine learning techniques have been increasingly applied to the classification of neurodegenerative diseases from neuroimaging data. Among these, deep learning approaches, particularly Convolutional Neural Networks (CNNs), have shown promising results in detecting Alzheimer's disease (AD), Parkinson's disease (PD), and other neurological disorders.

Ahmad et al. [10] proposed a convolutional neural network (CNN)-based model for detecting Alzheimer's disease using brain MRI scans. Their pipeline included preprocessing steps such as anatomical normalization and modulation to reduce inter-subject variability and emphasize structural differences. The model was trained on a dataset comprising 1,000 images—500 from patients with Alzheimer's disease and 500 from cognitively healthy individuals—and tested on a separate set of 500 images. The CNN achieved a classification accuracy of 92% on a binary dataset. The study also highlighted the limitations associated with small datasets and the critical importance of accurate labeling for effective training.

In a broader comparative study, Noor et al. [11] reviewed various deep learning approaches for detecting neurological disorders in brain MRI. Techniques assessed included CNNs, Deep Belief Networks (DBNs), Autoencoders (AEs), Recurrent Neural Networks (RNNs), Deep Neural Networks (DNNs), and Probabilistic Neural Networks (PNNs). The authors emphasized the importance of preprocessing steps, such as scaling, normalization, smoothing, and artifact correction, to enhance the quality of MRI data. Their findings indicated that CNNs outperformed other architectures for Alzheimer's and Parkinson's disease detection, while DNNs showed superior performance in classifying schizophrenia. These results underscore the need to select the most appropriate architecture for each specific neurological condition.

More recently, Zhang et al. [9] emphasized the need for multi-class classification models that go beyond Alzheimer's disease. They noted that most published works ignore less prevalent but clinically significant disorders such as dementia with Lewy bodies (DLB) and stroke-related cognitive impairment.

Salvatore et al. [12] approached the diagnostic challenge from a different perspective by developing a model to differentiate between Parkinson's disease (PD) and progressive supranuclear palsy (PSP), two conditions with overlapping clinical presentations. The researchers compiled MRI scans from multiple centers and applied standardized preprocessing procedures, including motion correction and spatial normalization. Principal Component Analysis (PCA) was used for feature extraction, followed by classification using a Support Vector Machine (SVM). The model achieved high classification accuracies: 90% for distinguishing PD from PSP, 95% for PD versus control, and 93% for PSP versus control. Key brain regions identified for discrimination included the caudate nucleus, putamen, thalamus, and prefrontal cortex.

While the literature presents encouraging results, four critical gaps remain:

1) The scarcity of large, well-annotated datasets and the complexity of accurately labeling clinical imaging data.
2) The overreliance on binary classification frameworks, limiting clinical applicability;
3) The scarcity of multi-disease models addressing coexisting or overlapping neurodegenerative conditions;
4) The lack of integrated diagnostic systems (CADs) that combine prediction with interpretability and usability.

This study aims to bridge these gaps by developing a unified framework that classifies four major neurodegenerative diseases: AD, PD, DLB, and stroke, from structural MRI using both deep and traditional ML models. Additionally, we propose a CAD platform for real-time prediction and visualization to support clinical decision-making.

III. Methodology

This study followed a structured machine learning (ML) development pipeline consisting of seven main stages: (i) database acquisition, (ii) data preprocessing, (iii) feature extraction, (iv) machine learning models, (v) implementation, (vi) results analysis, and (vii) integration into a computer-aided diagnosis (CAD) interface. The complete workflow is illustrated in Figure 1.

A. Database Acquisition

A preliminary survey of publicly available MRI repositories revealed a predominance of datasets centered on Alzheimer's disease. To ensure broader diagnostic coverage, we selected the OASIS-3 dataset (Open Access Series of Imaging Studies), a comprehensive longitudinal repository comprising over 30 years of neuroimaging, clinical, and cognitive data from 1,378 participants [13]. Among these, 755 individuals were cognitively healthy, while 622 presented varying degrees of cognitive impairment.

For this study, we focused on four neurodegenerative and cerebrovascular conditions: Alzheimer's disease, Parkinson's disease, Lewy body dementia, and stroke. Cognitively healthy individuals were included as the control group. Figure 2 illustrates the entity-relationship model employed to organize the dataset, highlighting the associations between patients,

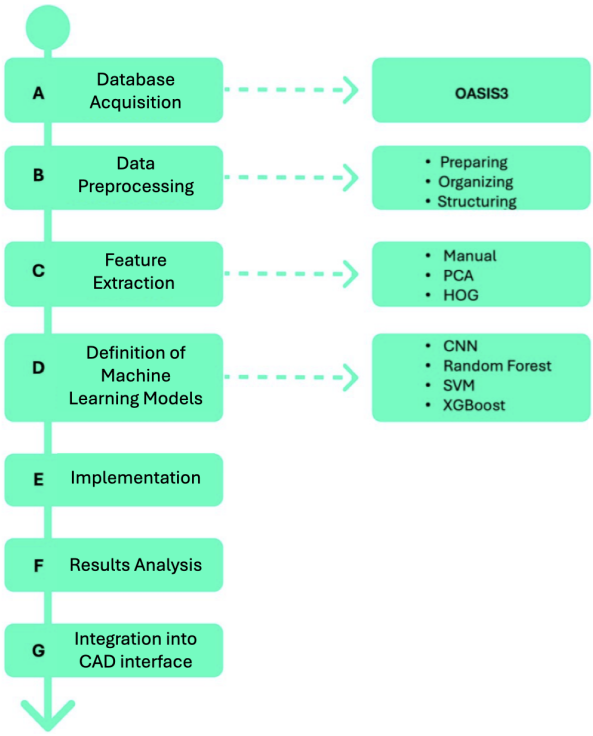

Figure 1: Workflow of the proposed classification model, including image preprocessing, feature extraction, model training, and integration into the diagnostic interface.

MRI scans, diagnoses, disease classifications, and medical evaluators.

Each MRI scan may be linked to multiple diagnoses and associated conditions. Conversely, each diagnosis is attributed to a single physician and may indicate either the absence of disease (control) or the presence of one or more pathological conditions.

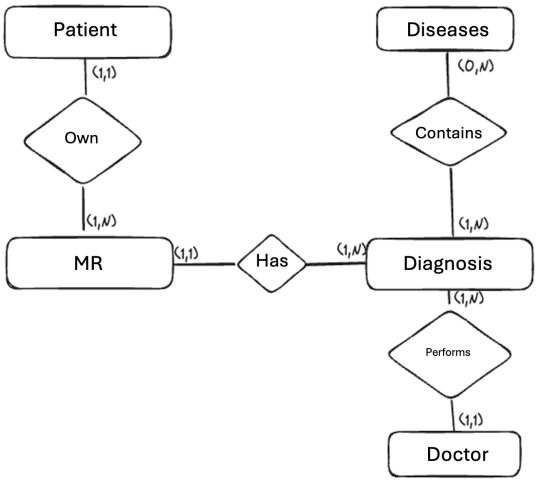

Figure 2: Entity-relationship diagram showing the structure of the dataset and relationships among patients, MRI scans, diagnoses, and diseases.

B. Data Preprocessing

The raw dataset underwent a structured preprocessing phase aimed at ensuring data integrity, internal consistency, and analytical suitability. Initially, duplicate entries and inconsistent records were identified and removed. MRI scans lacking diagnostic annotations were excluded, reducing the dataset from 2,400 to 1,700 valid volumes with corresponding clinical labels.

To reduce dimensionality while preserving diagnostic relevance, a targeted selection of metadata attributes was performed. Out of the original 84 fields, 8 key variables were retained based on their clinical significance and potential to contribute meaningfully to classification performance. These included four binary indicators corresponding to the presence of Alzheimer's, Parkinson's, Lewy body dementia, and stroke, as well as the MRI scan identifier, the diagnosis file, a subject-specific ID, and the control class label.

The selection was informed by domain knowledge and guided by the exclusion of variables deemed redundant, poorly correlated with diagnostic outcomes, or likely to introduce noise. By narrowing the feature space to a clinically coherent subset, the dataset became more interpretable and computationally efficient, supporting both model generalization and clinical relevance.

In preparation for image analysis, MRI data stored in NIFTI format were converted into 2D PNG slices, a common practice for deep learning model input due to their compatibility with standard image processing libraries and visual inspection. Figure 3 illustrates representative slices in axial, coronal, and sagittal orientations.

A consistent imaging protocol was established by selecting 11 consecutive axial slices per exam, centered around the midbrain. This included the central slice and five adjacent slices in both anterior and posterior directions. The axial plane was prioritized due to its ability to capture critical anatomical structures implicated in neurodegenerative diseases, such as the hippocampus, substantia nigra, and relevant cortical regions. The central axial region includes anatomical landmarks that are commonly affected across multiple conditions, including Alzheimer's disease, Parkinson's disease, Lewy body dementia, and stroke. By focusing on this region, the model is exposed to features that are not only structurally informative but also pathologically relevant across a diverse set of diagnoses. This strategy ensures anatomical consistency across samples and directs the model's attention toward the most diagnostically meaningful regions of the brain. While this fixed slice selection provides consistency, future work could explore the impact of varying slice numbers or alternative anatomical regions on model performance.

Finally, voxel intensity values were normalized to the 0-255 grayscale range using the nibabel Python library, a standard procedure to ensure consistent input scale for machine learning models. As the dataset had already undergone bias field correction and spatial smoothing during NIfTI generation, no additional correction steps were required.

C. Feature Extraction

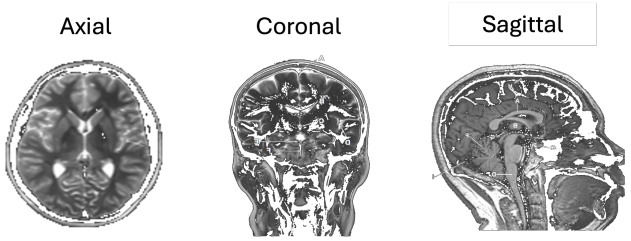

Figure 3: Example brain MRI slices in axial, coronal, and sagittal planes used for structural assessment and classification.

Feature extraction aimed to isolate informative descriptors from the MRI slices. We applied the Histogram of Oriented Gradients (HOG) method, which captures edge orientation patterns across local image regions [14].

As described by Chen et al. [15], the HOG feature extraction process begins with the application of square root gamma correction to the input color image. Subsequently, image gradients are computed for each of the three color channels using simple one-dimensional convolutional masks centered at [-1,0,1] in both horizontal and vertical directions. For each pixel, the gradient magnitude is determined from the channel exhibiting the highest gradient norm. The corresponding gradient orientations are then calculated and quantized into nine orientation bins: 0°– 20°, 20°– 40°, ..., 160°– 180°. A weighted vote for each edge orientation is cast based on the gradient magnitude, and these votes are accumulated in the respective orientation bins over local spatial regions, known as cells.

To ensure feature comparability, we applied z-score normalization using the StandardScaler method, assuming a normal distribution [16]. Dimensionality reduction was then performed using Principal Component Analysis (PCA) [17], which retains the components that preserved over 95% of the variance. In contrast, for the CNN model, raw image slices were used directly, allowing the network to learn features automatically through convolutional layers.

D. Definition of Machine Learning Models

Based on prior work and preliminary experiments, we selected the following classification algorithms for evaluation:

- Convolutional Neural Networks (CNNs): For end-to-end image feature learning [18].
- Random Forest: A robust ensemble method effective for tabular feature data [19].
- Support Vector Machines (SVM): Effective for small to mid-sized feature spaces [20].
- XGBoost: An efficient, gradient-boosted tree-based model optimized for structured data [21].

E. Implementation

Model training was conducted on 80% of the dataset, with 10% reserved for validation and 10% for testing. To ensure an unbiased evaluation and prevent data leakage, all records were randomly shuffled before this division. Given that MRI scans may include multiple diagnoses, we adapted the evaluation procedure to compare predicted disease sets with ground truth for each sample.

Class imbalance was observed in the dataset, with Alzheimer's and control samples outnumbering Parkinson's and Lewy body dementia cases. Figures 4 and 5 show the distribution of samples by disease and the number of diagnoses per scan, respectively. While this imbalance is a known challenge in medical datasets, no specific data augmentation or resampling techniques were applied in this study to mitigate its effects. This omission could potentially impact the model's robustness and generalizability, particularly for the less-represented disease categories, and will be considered in future work.

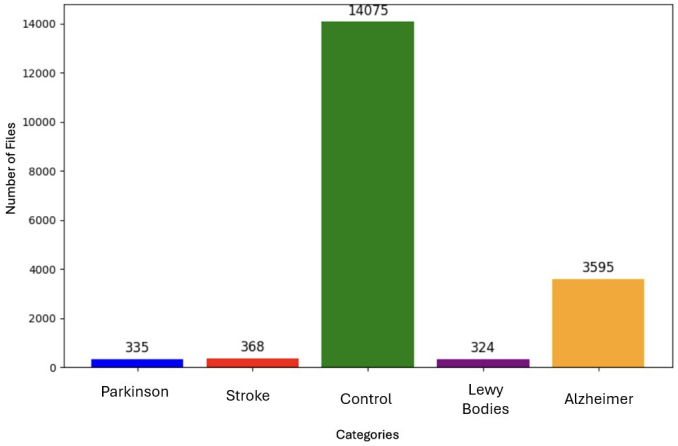

Figure 4: Distribution of MRI image samples by disease category.

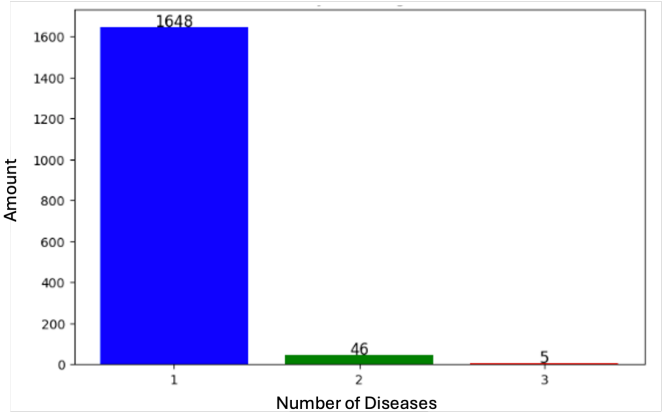

Figure 5: Frequency distribution of diagnoses per MRI scan. Most scans are associated with a single diagnosis.

To ensure robust evaluation, we applied 5-fold cross-validation using the cross_val_score function from Scikit-learn. For hyperparameter tuning, we used Optuna, a Bayesian optimization framework, conducting 100 iterations per model

[22]. Final metrics included accuracy, confusion matrices, and average training time.

Two experimental settings were considered:

1. Threshold-based multi-label prediction – a progressive probability threshold was applied to determine the set of predicted diseases.

2. Fixed-label prediction - for each exam, the model was constrained to predict exactly N labels, where N is the true number of diagnoses.

IV. Results and Analysis

This section presents the performance evaluation of the machine learning models in the task of classifying neurodegenerative diseases from MRI data. Two experimental scenarios were explored: (i) threshold-based multi-label classification and (ii) prediction based on the known number of diagnoses per exam. All models were optimized using the Optuna framework, and the results are presented using accuracy, precision, recall, F1-score, confusion matrices, and training time.

A. Hyperparameter Optimization

All models were tuned using 100 Optuna iterations for each model. Table I summarizes the best parameters obtained for each model.

Random Forest		SVM	
n_estimators	189	C	2.3374
max_depth	20	gamma	$8.67e^{-5}$
min_samples_split	2		
min_samples_leaf	1		
XGBoost		CNN	
n_estimators	171	num_conv_layers	2
max_depth	6	filters_conv_0	40
learning_rate	0.2616	filters_conv_1	30
subsample	0.9432	dropout	0.3117
colsample_bytree	0.9900	num_dense_layers	3
reg_alpha	0.1152	units_dense_0	171
reg_lambda	8.1302	units_dense_1	186
min_child_weight	5	units_dense_2	46
		learning_rate	0.0032

Table I: Optimized Hyperparameters of the Machine Learning Models

B. Scenario I: Threshold-Based Multi-Label Prediction

In this scenario, models were allowed to predict multiple diseases per scan by applying a variable probability threshold to sigmoid outputs. Predictions exceeding the threshold were retained and compared against the ground truth. Thresholds ranged from 1% to 100%, and the accuracy was calculated using the standard metric:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}} \quad (1)$$

Figure 6 shows the accuracy variation across thresholds for each model. The CNN outperformed all others at the optimal threshold, although its accuracy declined at higher thresholds. The SVM achieved rapid early gains but slightly dropped at higher levels. XGBoost performed consistently across a broad range, while Random Forest lagged behind all others. The experimental results revealed a strong dependence on the selected threshold values, with no single threshold performing optimally across all models. The final classification results, along with their corresponding optimal threshold values, are summarized in Table II. In addition, the confusion matrix had also been studied, showing relevant results in the multi-label classification task.

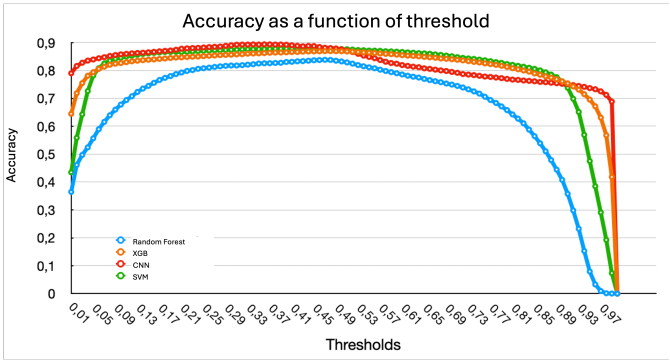

Figure 6: Accuracy variation across thresholds for each model.

To further analyze CNN performance, Figure 7 shows the confusion matrix under this setting. While Alzheimer's and control classes achieved high precision, some overlap was observed between PD and DLB.

Models	Threshold	Accuracy (%)	Macro-F1 (%)
Random Forest	0,46	83,82	27.88
SVM	0,40	87,89	66.04
XGBoost	0,47	86,95	71.00
CNN	0,35	89,37	74.75

Table II: Best Accuracy, Macro-F1 and Threshold per Model (Scenario I)

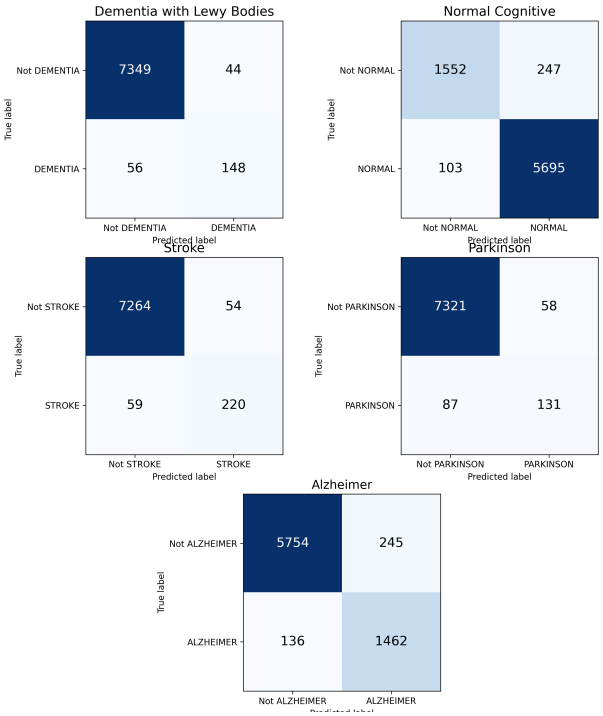

Figure 7: Confusion Matrix for each class

C. Scenario II: Prediction Based on Known Number of Diagnoses

In this setup, the model was constrained to predict exactly N classes per MRI, where N is the true number of clinical diagnoses for that exam. This scenario simulates a semi-supervised clinical setting where the number of conditions is known. Table III presents the accuracy and Macro F1-score results for this setting. The CNN once again delivered the

Models	Accuracy (%)	Macro-F1(%)
Random Forest	87,81	31.94
SVM	92,75	83.45
XGBoost	90,86	80.13
CNN	95,07	84.62

Table III: Model Metrics Under Scenario II

best performance, showing high separability for all classes. SVM and XGBoost also delivered competitive results, while Random Forest showed the lowest accuracy and Macro F1-score but offered faster training.

To provide a more comprehensive evaluation, Figure 8 and Table IV presents additional metrics for CNN in Scenario II. Results indicate high precision and recall across all disease categories. Given the observed class imbalance, the macro F1-score, which provides a more balanced evaluation across all classes, was also considered. The macro F1-score for the CNN model in Scenario II was 84.62%, further supporting its robust performance across both majority and minority classes.

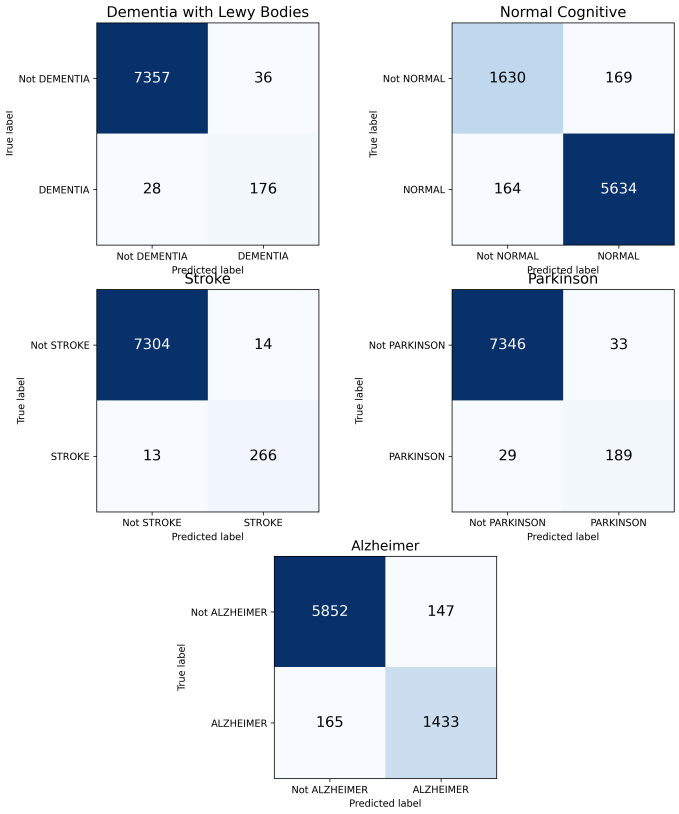

Figure 8: Confusion matrix for CNN model (Scenario II)

Class	Precision	Recall
Alzheimer's	0.95	0.94
Parkinson's	0.93	0.92
Lewy Bodies	0.89	0.90
Stroke	0.92	0.91
Control	0.97	0.96

Table IV: CNN Performance Metrics Per Class (Scenario II)

D. Training Time Comparison

Training times are summarized in Table V. While CNNs required longer training than tree-based models, their accuracy gain justified the computational cost.

Models	Execution Time
Random Forest	4 minutes
SVM	2 hours - 4 hours
XGBoost	7 minutes - 8 minutes
CNN	30 minutes

Table V: Average Training Time per Model (100 Trials)

E. One-CNN-per-Class Alternative Architecture

An additional experiment was conducted using four independent binary CNNs, each trained to detect a specific disease. Results are shown in Table VI.

Scenario	Accuracy(%)
Scenario I: Threshold-based (multi-label)	89,17
Scenario II: Fixed label count	93,49

Table VI: Performance of One-CNN-per-Class Architecture for both scenarios.

Although slightly less accurate than the unified CNN model, this modular architecture may be more interpretable and easier to update for specific diagnostic needs.

F. Diagnostic System and Interface

To bridge the gap between research and clinical practice, we implemented a Computer-Aided Diagnosis (CAD) platform that integrates the best-performing CNN model into an interactive interface for medical professionals (Figure 9).

Key Features of the System:

- Real-time visualization of MRI slices from NIfTI (.nii/.nii.gz) volumes;
- Dynamic probability display for each disease class upon file load;
- Interactive controls for contrast and brightness adjustment to assist in image inspection;
- Slice navigation tools to allow scrolling through the 11 axial views used by the classifier;
- Modular architecture, making it easy to update or replace the underlying CNN model;
- Lightweight deployment, capable of running on local workstations without GPU acceleration.

The CNN model embedded in the system is the one trained under the Scenario II (fixed number of labels), achieving the highest overall accuracy of 95.07%. Predictions are updated automatically as the user scrolls through different slices,

offering dynamic insights into how the classification evolves across brain regions.

The system was designed as a support tool, not a substitute for clinical expertise. Its purpose is to augment the diagnostic process by providing consistent, data-driven second opinions based on thousands of prior scans.

Figure 9: Screenshot of the CAD interface. MRI slice displayed on the left panel, with real-time probability scores shown on the right. Available in [23]

V. Conclusion

This study proposed a robust and scalable machine learning framework for the automated classification of neurodegenerative diseases using structural brain MRI. We evaluated and compared multiple models, including Convolutional Neural Networks (CNNs), Random Forest, Support Vector Machines (SVM), and XGBoost, across two experimental scenarios: one based on thresholded multi-label prediction and another based on the known number of diagnoses. Among the evaluated models, the CNN consistently outperformed the others, achieving a classification accuracy of 95.07% and a Macro F1-score of 84.62%, in the fixed-diagnosis scenario. Furthermore, the CNN maintained high precision and recall across all disease classes, including less commonly studied conditions such as Lewy body dementia and stroke-related cognitive impairment.

Beyond the predictive performance, the implementation of a Computer-Aided Diagnosis (CAD) interface represents a significant practical contribution. This system enables real-time MRI visualization, dynamic probability outputs, and user interaction, providing AI-driven support to assist rather than replace clinical decision-making.

Nevertheless, some limitations must be acknowledged. The dataset used (OASIS-3) reflects a specific demographic and MRI acquisition protocol, which may limit the generalizability of the results to other clinical settings. Additionally, handling cases with multiple co-occurring neurodegenerative conditions remains challenging for both multi-label and multiclass classifiers. The inherent clinical overlap between certain neurodegenerative diseases, such as Parkinson's disease and Lewy body dementia, presents a complex diagnostic challenge that our current models, while showing promising accuracy, still find difficult to fully disentangle. Further research is needed to develop more sophisticated approaches for accurately identifying and differentiating these concurrent pathologies. The current framework also does not incorporate temporal or longitudinal data, such as changes in cognitive function or MRI patterns over time.

Future work should aim to incorporate multimodal clinical data—including demographic, cognitive, and genetic information—to improve model robustness. Expanding the dataset to include more diverse populations and imaging protocols would enhance generalizability. Furthermore, integrating explainable AI (XAI) techniques, such as Grad-CAM or LIME, will be a priority to enhance model interpretability and build greater clinical trust by visualizing which MRI regions most influenced predictions. Finally, prospective validation of the CAD system in real-world clinical environments would provide important insights into its usability, reliability, and impact on patient care.

Overall, the integration of advanced machine learning techniques with real neuroimaging data demonstrates strong potential to enhance the early diagnosis and classification of neurodegenerative diseases. By combining high predictive performance with an interpretable and practical interface, this study contributes a valuable step toward the adoption of intelligent diagnostic tools in clinical neuroscience.

Among the evaluated models, the CNN consistently outperformed the others, achieving a maximum classification accuracy of 95.07% in the fixed-diagnosis scenario. Furthermore, the CNN maintained high precision and recall across all disease classes, including less commonly studied conditions such as Lewy body dementia and stroke-related cognitive impairment.

In addition to the predictive models, we developed a fully functional Computer-Aided Diagnosis (CAD) interface that integrates real-time image visualization with probabilistic diagnostic feedback. This tool is designed to support—not replace—clinical decision-making by providing interpretable AI-driven insights during MRI review.

Despite these promising results, certain limitations must be acknowledged:

- The dataset (OASIS-3) reflects a specific demographic and imaging protocol, which may limit generalization to other institutions;
- Cases involving multiple co-occurring pathologies remain a challenge for multiclass and multi-label models;
- The current system does not yet incorporate temporal data, such as cognitive decline over time or longitudinal MRI comparisons.

Future work will explore several directions:

- Incorporating multimodal clinical data (e.g., cognitive scores, genetic markers, laboratory values) to enhance model robustness;
- Applying explainable AI techniques (e.g., Grad-CAM, LIME) to increase transparency and trust in model decisions;
- Expanding the dataset to include a broader clinical population, including rare neurodegenerative conditions and more diverse MRI acquisition protocols;

- Conducting prospective evaluations of the CAD system in real clinical settings, in collaboration with neurologists and radiologists;

Overall, the integration of advanced ML techniques with real-world neuroimaging data demonstrates strong potential to assist clinicians in the early detection and classification of neurodegenerative diseases. As research and clinical partnerships evolve, such systems may significantly enhance diagnostic accuracy, optimize treatment planning, and ultimately improve patient outcomes.

ACKNOWLEDGMENT

The authors would like to thank the Centro Federal de Educação Tecnológica de Minas Gerais (CEFET- MG) for the institutional support and the conditions provided for this research.

REFERENCES

[1] National Institute of Neurological Disorders and Stroke (NINDS), "Brain basics," 2024. [Online]. Available: https://www.ninds.nih.gov/health-information/public-education/brain-basics

[2] D. J. Selkoe and J. Hardy, "The amyloid hypothesis of alzheimer's disease at 25 years," *EMBO Mol. Med.*, vol. 8, no. 6, pp. 595–608, 2016.

[3] R. G. González, "Clinical mri of acute ischemic stroke," *J. Magn. Reson. Imaging*, vol. 36, no. 2, pp. 259–271, 2012.

[4] S. B. Berman and C. Miller-Patterson, "Pd and dlb: Brain imaging in parkinson's disease and dementia with lewy bodies," *Prog. Mol. Biol. Transl. Sci.*, vol. 165, pp. 167–185, 2019.

[5] J. C. Hanson and C. F. Lippa, "Lewy body dementia," *Int. Rev. Neurobio.*, vol. 84, pp. 215–228, 2009.

[6] National Institute of Neurological Disorders and Stroke (NINDS), "Parkinson's disease information page," 2024. [Online]. Available: https://www.ninds.nih.gov/health-information/disorders/parkinsons-disease

[7] M. F. BRITO, A. C. de Araujo MATTOS, D. C. BERTOLIN, and W. SCRIGNOLLI, "Doenças neurodegenerativas importância do exame de imagem no diagnóstico precoce e no manejo: Uma visão geral," *Rev. Corpus Hippocraticum*, vol. 1, no. 1, 2023.

[8] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-Gonzalez, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, and O. Colliot, "Convolutional neural networks for classification of alzheimer's disease," *NeuroImage: Clin.*, vol. 25, p. 102203, 2020.

[9] Y. Zhang, W. Zhang, Y. Zhu, W. Wang, and D. Shen, "A review on multi-disease detection from mri using machine learning," *Comput. Biol. Med.*, vol. 158, p. 106749, 2023.

[10] M. F. Ahmad, S. Akbar, S. A. E. Hassan, A. Rehman, and N. Ayesha, "Deep learning approach to diagnose alzheimer's disease through magnetic resonance images," in *Proc. Int. Conf. Innovative Computing (ICIC)*. IEEE, 2021, pp. 1–6.

[11] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. A. Mamun, and M. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of alzheimer's disease, parkinson's disease and schizophrenia," *Brain Inform.*, vol. 7, pp. 1–21, 2020.

[12] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M. Gilardi, and A. Quattrone, "Machine learning on brain mri data for differential diagnosis of parkinson's disease and progressive supranuclear palsy," *J. Neurosci. Methods*, vol. 222, pp. 230–237, 2014.

[13] OASIS, "Oasis brains," 2024. [Online]. Available: https://www.oasis-brains.org/

[14] J. M. M. da Cruz and S. Blusseau, "Benchmark smil versus scikit-image (mathematical morphology features)," 2021.

[15] Y.-p. Chen, S.-z. Li, and X.-m. Lin, "Fast hog feature computation based on cuda," in *Proc. IEEE Int. Conf. Comput. Sci. Autom. Eng. (ICCSAE)*, vol. 4. IEEE, 2011, pp. 748–751.

[16] I. Izonin, B. Ilchyshyn, R. Tkachenko, M. Greguš, N. Shakhovska, and C. Strauss, "Towards data normalization task for the efficient mining of medical data," in *Proc. Int. Conf. Adv. Comput. Inf. Technol. (ACIT)*. IEEE, 2022, pp. 480–484.

[17] T. Kurita, "Principal component analysis (pca)," *Computer Vision: A Reference Guide*, pp. 1–4, 2019.

[18] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. Eng. Technol. (ICET)*. IEEE, 2017, pp. 1–6.

[19] A. Chaudhary, S. Kolhe, and R. Kamal, "An improved random forest classifier for multi-class classification," *Inf. Process. Agric.*, vol. 3, no. 4, pp. 215–222, 2016.

[20] W. S. Noble, "What is a support vector machine?" *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006.

[21] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794.

[22] Optuna, "Optuna," 2025. [Online]. Available: https://optuna.org/

[23] L. da Silveira Nascimento, "Web viewer for mri scans," 2025. [Online]. Available: https://viewer.gensokyo.shop/principal