EARLY DIAGNOSIS OF ALZHEIMER'S USING THE WEIGHTED PROBABILITY-BASED ENSEMBLE METHOD

HAMSE ELMI

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF SCIENCE IN COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE

DEPARTMENT OF
COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

STUDENT NUMBER

2023232

COMMITTEE

Sasha Kenjeeva
Dr. Eriko Fukuda

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

January 10, 2025

WORD COUNT

7134

ACKNOWLEDGMENTS

بِسْمِ اللهِ الرَحمٰنِ الرَحيمِ

I would like to start in the name of Allah (SWT) , the Most Compassionate and the Most Merciful. I want to express my sincere appreciation to all the loved ones who have helped me reach this point. May this research serve as a stepping stone in my journey toward knowledge and in helping others wherever I can, in sha Allah. A big thank you to my supervisor, Sasha, for guiding me and supporting me.

EARLY DIAGNOSIS OF ALZHEIMER'S USING THE WEIGHTED PROBABILITY-BASED ENSEMBLE METHOD

HAMSE ELMI

Abstract

This study explores the application of the Weighted Probability-Based Ensemble Method (WPBEM) for the classification of Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Normal Cognition (NC) using Amyloid PET imaging data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), specifically the ADNI3 phase dataset, which correlates to the 3rd edition of the study. The WPBEM integrates predictions from three convolutional neural networks (CNNs): DenseNet201, ResNet50, and VGG19. Making use of their complementary strengths to enhance diagnostic performance. The methodology involved pre-processing of Amyloid PET data, including slice selection, normalisation, and entropy-based prioritisation, followed by the training of individual CNN models and their integration within the WPBEM. The model's performance was evaluated using metrics such as accuracy, F1 score, area under the curve (AUC), sensitivity, and specificity. The results demonstrate the WPBEM's performance, achieving an accuracy of 98.48%, an F1 score of 0.985, and near-perfect sensitivity (0.991) and specificity (1.000). Comparative analysis revealed that the WPBEM outperformed individual CNN models, including DenseNet201 (96.34% accuracy), ResNet50 (97.56% accuracy), and VGG19 (81.71% accuracy), particularly in reducing misclassifications in MCI detection. These findings highlight the WPBEM's potential to advance early AD diagnosis by improving diagnostic accuracy and reliability. This study underscores the importance of ensemble learning in neuroinformatics and demonstrates the feasibility of adapting the WPBEM for functional imaging modalities like Amyloid PET.

1 DATA, CODE, ETHICS AND TECHNOLOGY STATEMENT

The dataset used in this thesis was sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (2004) and specifically corresponds to the ADNI3 collection. The data was obtained through an online request via the Image and Data Archive (IDA) (n.d.), a secure resource for archiving, exploring, and sharing neuroscience data. The dataset is anonymised, and the data owner has provided consent for its use in scientific research, adhering to ADNI's data use agreement. This thesis does not involve the collection of data from human participants or animals. Preprocessed PET-images were derived from the ADNI3 dataset and created as part of this thesis's workflow. Consent for using and preprocessing these images was granted during the initial data access request through the IDA. All code used for this thesis was written independently, with inspiration drawn from Fathi et al. (2024). The exception is the code highlighted in the provided GitHub repository Elmi (2024), which includes the Weighted Probability-Based Ensemble Method implementation. This specific portion of the code was shared by Fathi via email and is used with permission for non-commercial and scientific purposes. The generative language model ChatGPT 4o by OpenAI (2024) was employed to enhance the clarity of the thesis text, focusing on spell checking, and grammar improvements. Additionally, Grammarly (n.d.), which is a software writing tool, was used to identify and correct additional spelling and grammatical errors.

2 INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most common cause of dementia worldwide (Sperling et al., 2011). Characterised by cognitive decline, memory loss, and functional impairment, AD significantly affects patients and their caregivers, leading to emotional, physical, and financial burdens (Dubois et al., 2016). With the ageing global population, the prevalence of AD is projected to rise sharply, emphasising the need for early diagnosis and intervention. Timely detection of AD is crucial, as it enables the implementation of therapeutic interventions that can slow disease progression, improve patient outcomes, and reduce healthcare costs (Fathi et al., 2024) However, early diagnosis remains a formidable challenge due to the subtle nature of initial symptoms and the limitations of current diagnostic approaches (Sperling et al., 2011).

Neuroimaging is a cornerstone of AD research, offering valuable insights into both structural and functional brain changes associated with the disease. Ebrahimighahnavieh et al. (2020) mention how techniques like Magnetic Resonance Imaging (MRI) provide detailed information on anatomical changes such as cortical thinning and hippocampal atrophy, commonly seen in later stages of AD. However, these methods often struggle to detect early pathological changes. In contrast, functional imaging modalities, particularly Positron Emission Tomography (PET), enable the identification of early molecular abnormalities. For instance, Amyloid PET allows for the in vivo visualization of amyloid-beta plaques (Figure 1) long before clinical symptoms manifest. Amyloid-beta plaques are sticky clumps of misfolded protein fragments that accumulate in the spaces between neurons, disrupting cell communication (Scheltens et al., 2021). This capability highlights the potential of Amyloid PET in supporting early diagnosis and research into the progression of AD.

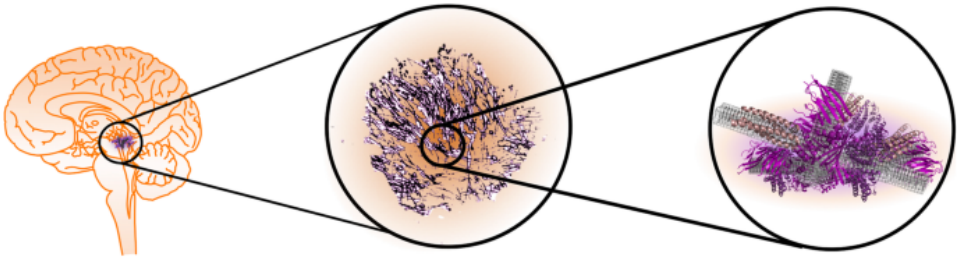

Figure 1: Illustration of the localisation of amyloid-beta plaques within the brain (left), the aggregation of amyloid-beta fibrils forming plaques (centre), and a molecular-level depiction of the amyloid-beta fibrils (right) (Rahman Lendel, 2021).

The integration of artificial intelligence (AI) and machine learning (ML) into neuroimaging has revolutionized the analysis of medical images. Convolutional Neural Networks (CNNs), a class of deep learning models, have demonstrated significant potential in classifying Alzheimer's disease and mild cognitive impairment (MCI) through automated analysis of neuroimaging data (Zhao et al., 2023). Although CNNs are powerful, their reliance on specific feature sets may lead to inconsistent performance across datasets. Ensemble learning, which combines predictions from multiple models, addresses this limitation by enhancing robustness and accuracy according to Logan et al. (2021). A recent framework, the Weighted Probability-Based Ensemble Method (WPBEM), has shown promising results in AD diagnostics using structural imaging modalities like MRI (Fathi et al., 2024). Its potential application to functional imaging, such as Amyloid PET, presents an exciting opportunity to advance early AD detection through AI-driven techniques.

This thesis aims to adapt the WPBEM framework to Amyloid PET neuroimaging data and evaluate its effectiveness compared to individual CNN models. By leveraging the unique strengths of Amyloid PET imaging and the robustness of ensemble learning, this research seeks to address a critical gap in early AD diagnostics. Specifically, the study will compare the performance of WPBEM to its constituent CNN models using key evaluation metrics, including accuracy, sensitivity, specificity, and the area under the curve (AUC).

The motivation for this research is both societal and scientific. Early diagnosis of AD can significantly enhance the quality of life for patients and their families, reducing the emotional and financial toll of late-stage care (Dubois et al., 2016). Additionally, it offers potential cost savings to healthcare systems by facilitating timely interventions and better disease management. From a scientific perspective, this research contributes to the growing field of neuroinformatics by advancing ensemble methods in functional imaging and exploring their utility in diagnosing neurodegenerative diseases at early stages. To address this challenge, the study poses the following research question:

> *What is the effectiveness of the Weighted Probability-Based Ensemble Method (WPBEM) compared to its individual convolutional neural network (CNN) models for the early diagnosis of Alzheimer's disease using Amyloid PET neuroimaging data?*

To achieve this goal, the thesis is organised as follows: the next section presents a review of related works (3) that provide the foundation for this research. This is followed by a detailed description of the methodology (4) employed to adapt and evaluate the WPBEM for Amyloid PET data. The

results (5) of the study are then analysed and discussed (6) in the context of existing research. Finally, the conclusion (7) highlights the implications of the findings for AD diagnostics and outlines potential future research directions.

Through this work, the study seeks to advance the field of early Alzheimer's diagnosis and demonstrate the potential of AI-driven ensemble methods in aiding medical imaging.

3 RELATED WORKS

3.1 *Imaging modalities*

Imaging modalities are indispensable in AD research, offering detailed insights into the structural and molecular changes in the brain. These tools are critical for identifying biomarkers, understanding disease progression, and enabling early diagnosis. Among the most extensively studied imaging techniques are Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET), each providing unique and complementary information about AD pathology (Wen et al., 2020).

3.1.1 *Magnetic Resonance Imaging (MRI)*

MRI is widely used to investigate the structural changes associated with AD. Key examples include hippocampal atrophy, where neuronal loss in the hippocampus, a region essential for memory, causes shrinkage, and cortical thinning, where the brain's outer layer becomes thinner due to cell loss (Ebrahimighahnavieh et al., 2020). These changes are strongly correlated with cognitive decline and are considered important biomarkers of disease progression. Advanced techniques, such as voxel-based morphometry, which quantifies brain tissue volume, and surface-based analysis, which examines cortical structures, have significantly enhanced MRI's diagnostic capabilities(Basher et al., 2021).

However, MRI has limitations. The earliest pathological changes in AD, such as the accumulation of amyloid-beta and tau proteins, occur years before structural damage is evident, making these biomarkers invisible to traditional MRI scans (Scheltens et al., 2021). Furthermore, age-related brain changes in older adults can mimic AD pathology, leading to potential misdiagnoses (Chapleau et al., 2022). These challenges necessitate the use of complementary imaging modalities that can detect molecular changes at earlier stages of the disease.

3.1.2 *Positron Emission Tomography (PET)*

PET imaging offers functional insights into AD pathology, focusing on two key markers: amyloid-beta plaques and tau tangles. Tau PET, for instance, visualises the accumulation of hyperphosphorylated tau protein, which is closely associated with cognitive decline and disease progression (Jo et al., 2020). Tau PET studies have demonstrated strong correlations between tau deposition and affected brain regions, such as the hippocampus, parahippocampus, and fusiform gyrus, which are critical to memory and cognitive functions (Odusami et al., 2023).

Amyloid PET imaging has gained prominence for its ability to detect amyloid-beta plaques, one of the earliest pathological hallmarks of AD, sometimes decades before clinical symptoms appear (Sperling et al., 2011). Using the radiolabeled tracer Florbetapir, Amyloid PET enables *in vivo* visualisation of amyloid deposition. Florbetapir, which is tagged with the radioactive isotope fluorine-18 (^{18}F), binds selectively to amyloid plaques, emitting positrons detectable by PET scanners (Fan et al., 2024). This technology provides both qualitative and quantitative insights into amyloid burden, making it a cornerstone for studying preclinical AD (Chapleau et al., 2022).

Amyloid PET's ability to capture plaque distribution has been instrumental in linking amyloid deposition patterns to disease progression. For example, Sperling et al. (2011) highlighted the detection of amyloid plaques 10–20 years before cognitive decline, offering a critical window for early diagnosis and therapeutic intervention (Dubois et al., 2016). However, its adoption in AD research has been limited by factors such as technological maturity, cost, accessibility, and regulatory approvals. The need for advancements in radioligand development and PET scanner capabilities has historically restricted the availability of Amyloid PET data (Kim et al., 2021).

3.2 *Advances in AD Classification*

Artificial intelligence (AI) and machine learning (ML) have demonstrated remarkable potential in automating and enhancing diagnostic accuracy in neuroimaging (Shanmugavadivel et al., 2023). Convolutional neural networks (CNNs), a subset of ML models designed for image analysis, have achieved promising results in neuroimaging studies (Alsubaie et al., 2024). CNNs have been employed to classify Alzheimer's disease (AD) and mild cognitive impairment (MCI), offering a powerful means to detect early neurodegeneration (Grueso Viejo-Sobera, 2021). However, individual CNN models can be limited by their reliance on specific feature sets and may yield inconsistent performance across diverse datasets (Odusami et al., 2023). Ensemble learning, which combines predictions from multiple models, offers a promising solution to this issue by aggregating the strengths of different models to improve robustness and accuracy (Logan et al., 2021).

In recent years, deep learning techniques have increasingly demonstrated their effectiveness in diagnosing AD at its earliest stages. Zhao et al. (2023) provided a comprehensive review of how deep neural networks have been employed in neuroimaging, particularly using MRI data, to classify AD and MCI, achieving high accuracy in detecting early neurodegeneration. Similarly, Qiu et al. (2020) developed a deep learning framework capable of

predicting the progression from MCI to AD, emphasising the importance of creating interpretable models for early-stage diagnosis. Odusami et al. (2023) demonstrated how a deep learning model using both PET and MRI data could provide an explainable diagnosis of AD by leveraging complementary information from both modalities.

Building on these efforts, Wen et al. (2020) utilised CNNs to analyse T1-weighted MRI (T1-MRI) scans, an MRI modality that highlights anatomical structures with high contrast between different tissues. This study focused on identifying structural biomarkers of AD, such as reduced brain volume and tissue degradation, by training CNNs to recognise subtle patterns in high-resolution images. Their efforts culminated in achieving an impressive diagnostic accuracy of 97.8%, demonstrating the potential of CNNs to detect nuanced but late-stage indicators of AD.

Additionally, Logan et al. (2021) proposed a groundbreaking approach that combined PET and MRI data with generative adversarial networks (GANs). By combining the structural insights from MRI with the functional data from PET, their study aimed to provide a holistic view of AD pathology. To address limited labelled neuroimaging datasets, the researchers employed GANs, a type of deep learning model capable of generating synthetic data. This innovative integration of multimodal imaging and synthetic data improved diagnostic accuracy to 90.1%, addressing the critical bottleneck of data scarcity in AD research.

AI applications in amyloid PET imaging are comparatively underexplored but have shown immense potential. For example, Fan et al. (2024) introduced AmyloidPETNet, which leverages PET imaging data to achieve high diagnostic accuracy (AUC >= 0.97), overcoming challenges such as differences between observers' measurements (inter-observer variability) and the computational demands of traditional image analysis methods. By precisely identifying amyloid plaques, these models offer significant advancements in early AD detection and staging.

3.3 *Ensemble Learning and WPBEM*

Ensemble learning has emerged as a solution to the variability and limitations of individual CNN models. By aggregating predictions from multiple models, ensemble methods improve robustness and accuracy. Hybrid models that combine CNNs with recurrent neural networks (RNNs) have also shown promise in analysing longitudinal imaging data, offering a way to track disease progression over time (Odusami et al., 2023).

One notable ensemble framework is the Weighted Probability-Based Ensemble Method (WPBEM), introduced by (Fathi et al., 2024). This novel deep learning framework assigns weights to individual model predictions

based on performance, enhancing reliability by combining the strengths of multiple models while mitigating weaknesses. The WPBEM has demonstrated success in analysing MRI data, achieving an accuracy of 98.57% in AD classification.

Despite its effectiveness in structural imaging, the application of WPBEM to functional imaging modalities like amyloid PET remain unexplored. Fathi et al. (2023) highlighted the potential for integrating PET data into ensemble frameworks, suggesting that such approaches could capture the molecular specificity of amyloid PET while leveraging the predictive power of ensemble learning. Amyloid PET, with its ability to detect amyloid plaques in vivo, presents a novel opportunity to extend the applicability of WPBEM and improve diagnostic capabilities in early AD detection (Dubois et al., 2016).

4 METHOD

4.1 *Data*

The dataset used in this study is sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (2004), specifically the ADNI3 phase. Established in 2004, ADNI is a landmark longitudinal study aimed at validating biomarkers for Alzheimer's disease (AD) and advancing diagnostic tools and treatments.

ADNI3, launched in August 2016, builds upon earlier phases with an emphasis on multi-modal neuroimaging and biomarker integration to better characterize AD progression (Weiner et al., 2017). This phase includes structural imaging modalities like Magnetic Resonance Imaging (MRI), functional imaging techniques such as Positron Emission Tomography (PET), and extensive clinical and cognitive assessments. Amyloid PET, introduced in later ADNI phases, employs tracers like Florbetapir and Florbetaben to visualize amyloid-beta plaques. Amyloid PET was not a core component in the early ADNI phases and was introduced in later stages as the technology matured and its utility in AD research became evident. This makes ADNI3 particularly valuable for studies like this one, which focus on the early detection of AD using Amyloid PET data.

The ADNI3 cohort comprises participants across a spectrum of cognitive states, including normal cognition (NC), mild cognitive impairment (MCI), and Alzheimer's disease (AD). Recruitment was facilitated through the Brain Health Registry, which uses web-based cognitive assessments to identify eligible participants (Weiner et al., 2018). Participants undergo regular follow-ups over 3–5 years, allowing for the longitudinal tracking of disease progression.

Key features of the dataset include high-resolution MRI scans with voxel sizes of approximately 1 mm^3. Amyloid PET provides direct measures of amyloid burden, while MRI captures structural and functional changes. Participants classified as NC include both clinically recruited individuals and community volunteers, with some selected based on amyloid positivity to explore preclinical AD stages (Weiner et al., 2017).

4.2 *Preprocessing and Proposed Model*

The proposed model builds upon the framework established by Fathi et al. (2024), whose work serves as a foundation for this study. Their WPBEM Python framework has been adapted to handle amyloid-PET data, with modifications tailored to meet the computational constraints of this research (Elmi, 2024). Specifically, the original six CNNs utilised in

Fathi's WPBEM have been streamlined to three key architectures: VGG19, DenseNet201, and ResNet50. These adjustments preserve the ensemble method's core strengths while ensuring its feasibility within the study's computational limitations.

This study employs a deep learning framework to classify AD, MCI, and Normal Cognition (NC) using Amyloid PET imaging data from the ADNI3 dataset. The model combines advanced preprocessing techniques with the WPBEM to capitalise on the complementary strengths of its constituent CNNs. The subsequent sections provide an explanation of the preprocessing pipeline and the model architecture as shown in Figure (2).

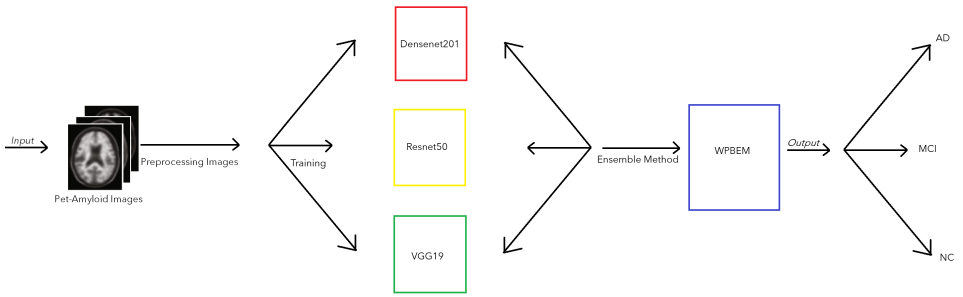

Figure 2: Workflow diagram of the WPBEM (Weighted Probability-Based Ensemble Method) for Alzheimer's diagnosis. The pipeline begins with preprocessed PET-amyloid images as input, which are then passed through individual CNN models (DenseNet201, ResNet50, and VGG19) for feature extraction. The ensemble method aggregates the predictions from these models to provide a final classification into Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), or Normal Cognition (NC).

4.2.1 *Preprocessing Strategy*

The preprocessing of raw Amyloid PET imaging data is a critical step to ensuring that the input data is both standardised and informative for model training. Given the complexity and variability of neuroimaging data as seen in Figure (3), a robust pipeline was designed to enhance the image quality while removing irrelevant or noisy data.
The preprocessing steps, implemented using `Python 3.7` (Python Software Foundation, 2018), libraries such as `NiBabel` (Brett et al., 2022), `Scikit-Image` (Van der Walt et al., 2014), and `NumPy` (Harris et al., 2020), were tailored to address the unique characteristics of PET scans.

1. **Conversion from DICOM to NIfTI:** Amyloid PET scans from the ADNI3 dataset were originally provided in DICOM format, which are raw files produced by Siemens' PET scanners (Graham et al.,

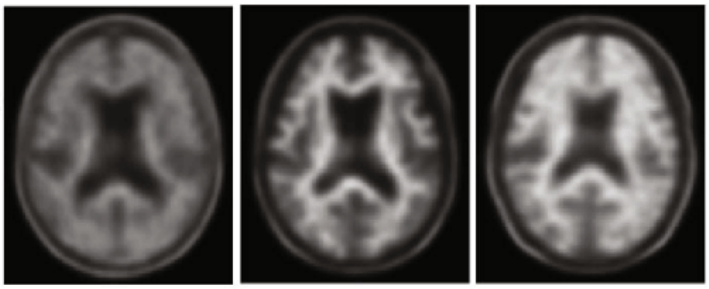

Figure 3: Images representing the three categories used for classification: NC (left), MCI (centre), and AD (right). These categories highlight the progression of cognitive decline observed in neuroimaging data.

2005). These files were converted to the NIfTI format using the `dcm2niix` tool (Li, 2016). The resulting NIfTI files were then handled using `NiBabel`, which provides functionality for reading, writing, and manipulating NIfTI data. This step facilitated subsequent operations such as slice selection and normalisation.

2. **Slice Selection:** PET scans comprise volumetric data with multiple slices along the z-axis with different thicknesses dependent on the settings chosen. These slices represent cross-sectional images that, when stacked together, form a 3D representation of the scanned region, allowing detailed visualisation of anatomical or pathological features at different depths. To optimise the data for 2D CNN architectures, slices were evaluated based on their variance, how much the intensity values (brightness or darkness of pixels) differ from their mean, to exclude non-informative regions. Using `NumPy`, the variance for each slice was calculated efficiently. Only slices with variance above the 20th percentile were retained as suggested by Fathi et al. (2024). This ensures that the selected slices reflect meaningful anatomical or pathological information. Due to variations in slice thickness among the images, the number of slices per category is unequal. The number of slices per sample and per diagnosis retrieved can be seen in Table (1).

Diagnosis	Number of Samples	Slices
AD	52	2329
MCI	59	2749
NC	62	2517

Table 1: Distribution of samples and slices across diagnostic categories.

3. **Normalisation and Standardisation:** Intensity normalisation was applied to each slice to scale pixel values to a range of $[0, 1]$, eliminating variations caused by scanner settings or patient-specific factors. This was implemented using `NumPy` for pixel-wise operations. Spatial normalisation, crucial for aligning all scans to a common template, was facilitated by tools compatible with `NiBabel` for image transformations and alignments, ensuring consistency in brain region positioning across participants.

4. **Resizing:** Each slice was resized to 256×256 pixels using `Scikit-Image`. The library's resizing functions preserved the anatomical fidelity of the images while adapting them to fixed input dimensions required by CNN models.

5. **Entropy-Based Slice Prioritisation:** Shannon entropy was calculated for each slice using `NumPy` to quantify its information content. Between 5 and 20 slices with the highest entropy values were selected for each scan based on how many slices were left over. This ensured that the dataset retained slices rich in diagnostic information, such as those capturing amyloid plaque distribution.

6. **Train-Validation-Test Split:** The pre-processed data was stratified and split into training (80%), validation (10%), and test (10%) sets. Stratification preserved the class distributions across splits, ensuring balanced datasets for each cognitive state.

4.2.2 *Proposed Model Architecture*

The diagnostic framework employed in this study integrates three advanced CNNs, VGG19, DenseNet201, and ResNet50, into a Weighted Probability-Based Ensemble Method (WPBEM) designed to improve classification accuracy for AD, MCI, and NC. The framework leverages the unique strengths of each model while addressing their individual limitations through ensemble learning. The framework is built using the deep learning library used for computer vision tasks called `torchvision` (TorchVision, 2016).

VGG19 is a widely used convolutional neural network (CNN) architecture that is both straightforward and powerful in its design, making it ideal for tasks requiring image analysis (Simonyan Zisserman, 2014). It is particularly well-known for its ability to extract hierarchical features, meaning it progressively identifies simple patterns like edges and textures in earlier layers and more complex structures, such as shapes and objects, in deeper layers. The architecture achieves this by using a series of convolutional layers arranged in a sequential and uniform manner, where each

layer applies filters of the same size (e.g., 3×3 filters) across the image. To adapt VGG19 for single-channel greyscale neuroimaging data, the first convolutional layer was modified to accept single-channel inputs by summing the original three-channel filters. The classifier was reconfigured to output predictions for three classes (AD, MCI, and NC).

DenseNet201 is a convolutional neural network (CNN) architecture designed to improve learning efficiency and accuracy by utilising densely connected layers (Huang et al., 2018). Unlike traditional CNNs, where each layer only connects to the next, DenseNet201 establishes direct connections between all layers. This means that each layer receives the feature maps from all preceding layers as inputs and passes its own outputs to all subsequent layers. This has the advantage that in deep networks, where gradients, the signals that update weights during training, don't diminish or vanish as they pass through many layers due to having multiple direct paths for gradients to follow. The input layer was modified to process single-channel inputs, and the classification head was adapted to predict three classes.

ResNet50 introduces residual connections, a mechanism that creates direct shortcut paths bypassing one or more layers in the network (He et al., 2015). These connections allow the gradient, the signal used to adjust model weights during training, to flow directly to earlier layers, even as the network becomes deeper. By bypassing layers, residual connections effectively address the vanishing gradient problem, where gradients diminish in magnitude as they propagate through many layers, making training unstable or ineffective. This design not only ensures that deeper networks like ResNet50 can train efficiently but also enhances their ability to learn complex patterns without being hindered by depth-related challenges. Similar to the other models, ResNet50 was adapted for single-channel inputs by modifying the initial convolutional layer. The output layer was replaced with a fully connected layer tailored to the three-class classification task. This architecture excels in capturing global features, patterns or characteristics capturing the overall structure or context rather than just fine details, complementing the local and dense features learned by VGG19 and DenseNet201.

The WPBEM integrates the predictive capabilities of DenseNet201, ResNet50, and VGG19, addressing inherent limitations of individual models, such as overfitting or bias toward specific features. By utilising a weighted aggregation mechanism, the WPBEM prioritises the contributions of models with a higher validation performance, resulting in a balanced and robust diagnostic ensemble.

4.3 *Training Phase*

Each CNN, DenseNet201, ResNet50, and VGG19, is independently trained using the pre-processed Amyloid PET dataset. The training employs cross-entropy loss, a widely used function in multi-class classification, which quantifies the divergence between predicted probabilities and ground truth labels (Zhang Sabuncu, 2018). To assess the optimal hyper-parameters, such as the number of epochs, learning rate, and batch size, we also employ hyper-parameter tuning to optimise training. This tuning involved testing various combinations to identify the configuration that yielded the best validation accuracy. The tuning process considered learning rates of 0.001 and 0.0001, with batch sizes of 32 and 64, and tested models over 100, 200, 300, and 350 epochs.

The validation accuracy of each CNN, obtained using this optimal configuration, was then utilised to determine its weight within the ensemble. The weighting mechanism assigns higher weights to models with superior validation accuracy, reflecting their greater reliability and confidence in predictions. Mathematically, the weight w_i for the i-th model is calculated as:

$$w_i = \frac{\text{Validation Accuracy of Model } i}{\sum_{j=1}^{N} \text{Validation Accuracies of All Models}}$$

where N is the total number of models in the ensemble. For instance, if DenseNet201 achieves a validation accuracy of 90%, ResNet50 88%, and VGG19 85%, DenseNet201 will receive a proportionally higher weight, signifying its greater influence on the final prediction.

This weighting mechanism ensures that the ensemble prioritises contributions from models that generalise better to unseen data. By leveraging the strengths of high-performing models and minimising the impact of less accurate ones, the ensemble achieves robust and reliable diagnostic performance.

4.4 *Ensemble Prediction Process*

Each trained CNN independently processes test samples, generating class probability distributions for AD, MCI, and NC. These probabilities reflect the confidence of each model in its classification. The class probabilities generated by the individual models are aggregated using a weighted averaging approach. For each class, the contribution of a model's probability is scaled by its pre-determined weight, ensuring that models with higher performance metrics have a more significant influence on the final ensemble output. However, all models contribute to the aggregation, capturing

a broader spectrum of insights. The ensemble then determines the final prediction by selecting the class with the highest weighted probability. This methodology ensures that the decision-making process integrates the strengths of all constituent models while mitigating individual weaknesses, thereby enhancing diagnostic accuracy and robustness.

4.5 *Evaluation Metrics*

The evaluation metrics used in this study are tailored to assess the performance of the Weighted Probability-Based Ensemble Method (WPBEM) and its constituent convolutional neural networks (CNNs) in classifying Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Normal Cognition (NC). **Accuracy** measures the overall proportion of correctly classified instances, calculated as:

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

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. While accuracy provides a straightforward evaluation of model performance, it may not fully reflect diagnostic effectiveness in datasets with imbalanced class distributions.

The **F1 score**, which is the harmonic mean of precision and recall, accounts for both false positives and false negatives. It is computed as:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}},$$

where precision is defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

and recall is given by:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

This metric is particularly valuable for evaluating the model's ability to distinguish MCI from other cognitive states in the presence of imbalanced data.

The **Area Under the Curve (AUC)** quantifies the model's ability to distinguish between classes across various decision thresholds. A higher AUC reflects stronger discriminative power, especially important for identifying early stages of cognitive decline from Normal Cognition (NC) to MCI and AD.

Sensitivity (recall or true positive rate) focuses on the proportion of actual positive cases correctly identified, emphasizing the model's ability to minimize missed diagnoses. It is expressed as:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

In contrast, **Specificity** (true negative rate) measures the proportion of true negative cases accurately classified, ensuring that individuals without cognitive impairment are not mistakenly flagged. It is calculated as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}.$$

Together, these metrics provide a comprehensive assessment of the model's diagnostic accuracy and reliability, addressing the challenges of early Alzheimer's detection by balancing the need for high sensitivity in identifying at-risk individuals with the need for high specificity to avoid false positives.

5 RESULTS

The study evaluated the classification performance of three CNNS, DenseNet201, ResNet50, and VGG19, alongside the WPBEM. Hyperparameter tuning resulted in the models being trained and validated using a learning rate of 0.001, batch size of 64, and 350 epochs. Key evaluation metrics included accuracy, F1 score, area under the curve (AUC), sensitivity, and specificity. The results demonstrated in Table 2 show the effectiveness of each model in classifying Alzheimer's disease (AD), mild cognitive impairment(MCI), and normal control (NC).

Table 2: Classification Performance of CNN Models and WPBEM

Model	**Accuracy (%)**	**F1 Score**	**AUC**	**Sensitivity**	**Specificity**
DenseNet201	96.34	0.963	0.998	0.991	0.990
ResNet50	97.56	0.976	0.998	1.000	0.980
VGG19	81.71	0.819	0.973	0.714	0.990
WPBEM	98.48	0.985	0.997	0.991	1.000

5.1 *Individual Model Performance*

5.1.1 *DenseNet201*

DenseNet201 achieved a high classification accuracy of 96.34%, with an F1 score of 0.963 and an AUC of 0.998.

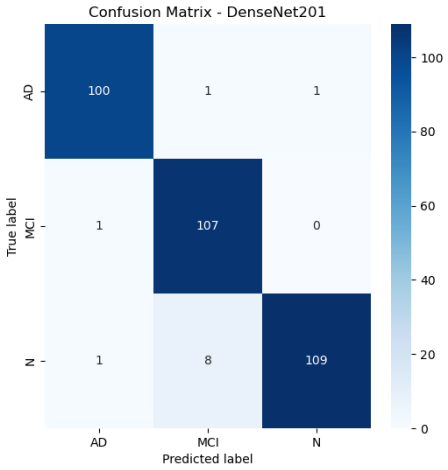

Figure 4: Confusion Matrix with classification results of DenseNet201. (N: Normal Cognition, MCI: Mild Cognitive Impairment, AD: Alzheimer's Disease)

Its sensitivity (0.991) and specificity (0.990) demonstrate its strong ability to correctly classify both positive and negative cases. However, the confusion matrix (Figure 4) shows that most misclassifications occurred between MCI and the NC group, highlighting areas for potential improvement.

5.1.2 *ResNet50*

ResNet50 outperformed DenseNet201 with an accuracy of 97.56%, an F1 score of 0.976, and an AUC of 0.998. Its perfect sensitivity (1.000) indicates it effectively identifies AD and MCI cases, while its slightly lower specificity (0.980) suggests minor trade-offs in false-positive rates. The confusion matrix (Figure 5) reveals the model's precision in distinguishing between the three cognitive states.

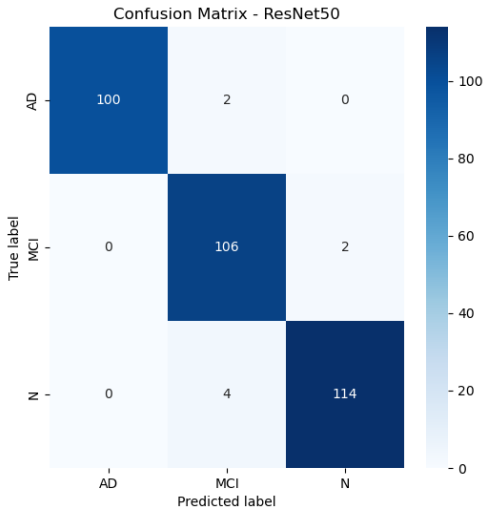

Figure 5: Confusion Matrix with classification results of ResNet50. (N: Normal Cognition, MCI: Mild Cognitive Impairment, AD: Alzheimer's Disease)

5.1.3 *VGG19*

VGG19 showed the lowest performance among the standalone models, with an accuracy of 81.71% and an F1 score of 0.819. Despite an AUC of 0.973 and specificity of 0.990, its sensitivity (0.714) highlights challenges in identifying AD and MCI cases. The confusion matrix (Figure 6) underscores significant misclassifications, particularly in the MCI class.

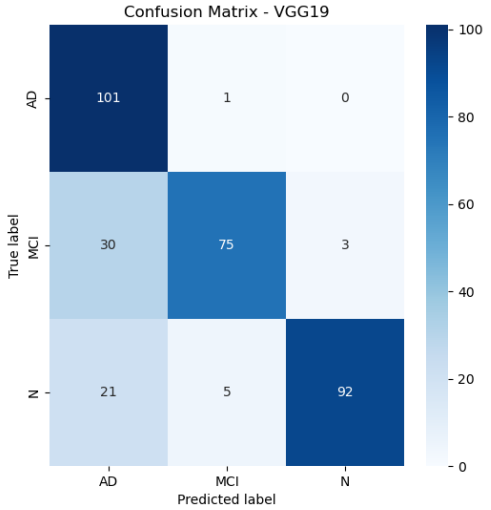

Figure 6: Confusion Matrix with classification results of VGG19. (N: Normal Cognition, MCI: Mild Cognitive Impairment, AD: Alzheimer's Disease)

5.2 *Ensemble Performance: WPBEM*

The WPBEM ensemble model outperformed all individual CNNs, achieving the highest accuracy (98.48%) and F1 score (0.985). Its AUC of 0.997 reflects exceptional discriminatory ability, and its balanced sensitivity (0.991) and specificity (1.000) demonstrate robust classification performance. The confusion matrix (Figure 7) highlights the model's ability to minimise errors across all classes, with no false positives recorded.

5.3 *Comparative Analysis*

The WPBEM demonstrated in Figure (8) its superior performance across multiple evaluation metrics compared to individual CNNs such as DenseNet201 and ResNet50. In terms of accuracy, the WPBEM achieved a 0.92% improvement over ResNet50, the best-performing standalone model, highlighting its enhanced ability to classify instances correctly. The F1 score of the WPBEM, calculated at 0.985, underscores its superior balance between precision and recall. This indicates that the model not only excelled in identifying true positives but also minimized false positives and false negatives, outperforming both DenseNet201 and ResNet50 in achieving a well-rounded classification performance.

The AUC of 0.997 further validated the WPBEM's strong discriminative power. Although all models achieved high AUC values exceeding

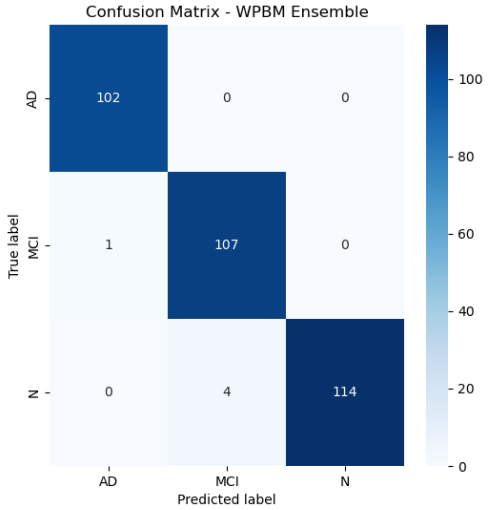

Figure 7: Confusion Matrix with classification results of the WPBEM. (N: Normal Cognition, MCI: Mild Cognitive Impairment, AD: Alzheimer's Disease)

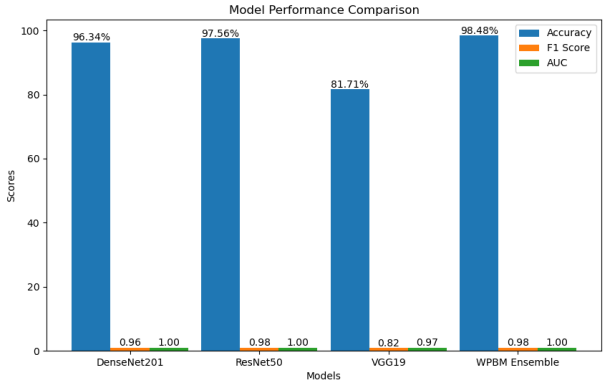

Figure 8: Comparison between individual CNNs and the WPBEM. The graph displays the following metrics: accuracy, F1 Score and the AUC score.

0.97, the WPBEM maintained a slight edge in distinguishing between cognitive states across varying decision thresholds. Moreover, the error reduction capabilities of WPBEM were evident from its confusion matrix, which revealed the fewest misclassifications, particularly in the challenging task of distinguishing MCI from NC. Collectively, these metrics emphasize the WPBEM's potential as a reliable and precise tool for advancing neuroimaging-based diagnostics.

6 DISCUSSION

6.1 *Interpretation*

The results of this study demonstrate the effectiveness of the WPBEM in classifying AD, MCI, and NC using Amyloid PET imaging data. The WPBEM consistently outperformed individual CNN models across all metrics, including accuracy, F1 score, and specificity. This improvement can be attributed to several factors inherent to this framework's implementation of ensemble learning. DenseNet201's dense connections allowed for effective feature reuse, enhancing the model's ability to detect subtle patterns in complex neuroimaging data. ResNet50's residual connections addressed the vanishing gradient problem, enabling deeper networks to train effectively and capture global features. And VGG19, while relatively simple, contributed through its hierarchical feature extraction capabilities, particularly for detecting local patterns.

The WPBEM's weighted combination of models allows it to aggregate their strengths while mitigating individual weaknesses, such as offsetting VGG19's lower sensitivity, which means it may miss positive cases, with ResNet50's and DenseNet201's superior performance in detecting those cases. Achieving nearly perfect sensitivity (0.991) and specificity (1.000), the WPBEM demonstrates a critical balance for clinical applications, adeptly identifying true positives while avoiding false positives, thereby addressing the significant consequences of both misdiagnosis which can allow the condition to progress unchecked while reducing the effectiveness of early treatments aimed at slowing cognitive decline (Dubois et al., 2016).

These findings align with previous studies in neuroinformatics, such as Logan et al. (2021) and Odusami et al. (2023), which demonstrated that ensemble learning frameworks improve diagnostic precision by combining multiple models' strengths. Moreover, the increased diagnostic performance of the WPBEM, particularly in distinguishing MCI from NC, complements prior findings that emphasize the challenges of early-stage diagnosis and the benefits of advanced neuroimaging modalities ((Wen et al., 2020); (Fathi et al., 2024)).

6.2 *Implications*

The findings of this study have important implications for both early Alzheimer's diagnosis and clinical practice. The WPBEM's ability to achieve high accuracy and specificity in detecting AD and MCI underscores its potential for early diagnosis, which is critical for improving patient outcomes. Identifying AD at the preclinical or early symptomatic

stages enables timely interventions that can slow disease progression and enhance quality of life. Similarly, the accurate identification of MCI is essential for monitoring individuals at risk of developing AD, offering opportunities for early therapeutic or lifestyle interventions.

These implications resonate with the broader context of advancements in neuroinformatics, as highlighted by Basher et al. (2021), where integrating ensemble methods into clinical workflows is seen as a transformative step. The WPBEM's high specificity minimises false positives, reducing unnecessary anxiety, invasive diagnostic procedures, and misallocated healthcare resources. By ensuring that only patients with a high likelihood of having AD or MCI are flagged for further evaluation, the WPBEM demonstrates its clinical applicability and potential to streamline diagnostic workflows. Furthermore, the WPBEM's performance highlights the importance of ensemble learning in addressing the variability inherent in neuroimaging data. As noted by Ebrahimighahnavieh et al. (2020), ensemble methods provide robustness against dataset-specific biases, enhancing the reliability of diagnostic tools across diverse populations. This is particularly relevant in the context of AD research, where early and accurate diagnosis can significantly influence treatment strategies and patient outcomes.

6.3 *Limitations*

Despite its promising results, this study faced several limitations that should be addressed in future research. The WPBEM relied on the ADNI3 dataset, which, while extensive, may not fully represent the diversity of the global population, as all participants are from North America. This limitation affects the generalisability of the findings to other populations. Additionally, the inherent scarcity of Amyloid PET data, driven by the high costs and logistical challenges of PET imaging, restricted the size of the training dataset, potentially impacting model performance. However, the upcoming development of ADNI4 offers hope for mitigating these limitations by providing more diverse and expansive datasets. Furthermore, adapting the WPBEM for PET data required significant preprocessing and computational resources due to the functional nature of PET imaging, which introduces challenges such as variability in tracer uptake and noise from low-resolution scans. The reliance on 2D slices rather than full 3D volumes, while computationally efficient, may result in the loss of critical spatial context needed to identify subtle patterns of amyloid plaque distribution.

The study also encountered limitations related to model architecture and design. While the WPBEM performed well, VGG19's lower sensitivity

and specificity highlight the challenges simpler architectures face when handling complex neuroimaging data. Although VGG19's inclusion enhanced the ensemble's diversity, it may have introduced minor biases that could affect results in nuanced cases. Moreover, the fixed weighting mechanism of the WPBEM, based on validation accuracy, effectively aggregated the models' contributions but might not fully account for their dynamic performance across different test cases.

6.4 *Future Work*

To address the limitations and build on the findings of this study, several avenues for future research are proposed. One key direction is the exploration of multi-modal integration, combining Amyloid PET with other imaging modalities such as structural MRI or Tau PET. This integration can provide complementary information, enabling the WPBEM to capture a broader range of pathological features and further improve diagnostic accuracy. Additionally, incorporating 3D convolutional neural network (CNN) architectures into WPBEM offers a promising approach to enhance the model's ability to analyse volumetric data. This adaptation would allow the preservation of spatial context and improve sensitivity to subtle patterns of amyloid deposition.

Another vital consideration for future research is improving population diversity within datasets. Including participants from diverse demographic and geographic backgrounds would enhance the generalisability of the findings. Collaborations with international consortia could provide access to larger, more representative datasets, making the WPBEM applicable across varied clinical settings. Alongside this, developing dynamic weighting mechanisms within WPBEM could optimise its performance further by adjusting weights based on class-specific performance or employing meta-learning techniques. These advancements could refine the model's robustness and diagnostic precision.

Lastly, enhancing explainability and clinical interpretability remains a critical challenge. While the WPBEM demonstrates high diagnostic accuracy, its limited interpretability hinders trust and usability in clinical settings. Future work should focus on developing explainable AI methods to identify the features or regions influencing model predictions. Additionally, optimising computational efficiency could facilitate real-time application in clinical workflows. These improvements would ensure that WPBEM is not only accurate but also transparent and practical for real-world deployment.

7 CONCLUSION

This study aimed to evaluate the effectiveness of the Weighted Probability-Based Ensemble Method (WPBEM) in classifying Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Normal Cognition (NC) using Amyloid PET imaging data from the ADNI3 dataset. The WPBEM framework combined the predictions of three CNNs—DenseNet201, ResNet50, and VGG19—leveraging their complementary strengths to improve diagnostic accuracy. The study addressed the critical need for accurate and early diagnosis of AD, which is essential for timely interventions and effective disease management.

The methodology involved preprocessing Amyloid PET data to standardise and optimise inputs for deep learning. Models were trained using hyper-parameters identified through hyper-tuning to achieve optimal configurations. Performance was evaluated using metrics such as accuracy, F1 score, AUC, sensitivity, and specificity. The WPBEM achieved superior results across all metrics.

Key findings highlight the WPBEM's effectiveness, achieving an accuracy of 98.48%, an F1 score of 0.985, and near-perfect sensitivity (0.991) and specificity (1.000). These results underscore the ensemble method's ability to mitigate the limitations of individual CNNs and provide robust and reliable classifications. In particular, the WPBEM demonstrated its strength in reducing misclassifications, especially in the challenging task of distinguishing MCI from NC and AD. The model's high specificity minimises false positives, a critical feature for clinical application to reduce unnecessary interventions and patient distress. The significance of these findings lies in the WPBEM's potential to advance early AD diagnosis. By enhancing diagnostic accuracy and reliability, frameworks like the WPBEM could support clinicians in identifying patients at risk of cognitive decline, enabling earlier interventions and improving patient outcomes. Furthermore, the ensemble framework serves as a generalisable approach that could be applied to other neurodegenerative diseases or multi-modal datasets.

In summary, the results of this study underscore the potential of ensemble learning in advancing the early diagnosis of Alzheimer's Disease. The WPBEM's performance demonstrates the value of combining complementary CNN architectures, while the study's limitations and future directions highlight opportunities for further innovation. By addressing these challenges and integrating additional data modalities, the WPBEM can be further refined to become a cornerstone in the evolving landscape of neuroinformatics and precision medicine for neurodegenerative diseases.

REFERENCES

Alsubaie, M. G., Luo, S., & Shaukat, K. (2024). Alzheimer's disease detection using deep learning on neuroimaging: A systematic review. *Machine Learning and Knowledge Extraction*, *6*(1), 464–505. Retrieved from https://doi.org/10.3390/make6010024 doi: 10.3390/make6010024

Alzheimer's Disease Neuroimaging Initiative (ADNI). (2004). *Adni | alzheimer's disease neuroimaging initiative.* Retrieved from http://adni.loni.usc.edu/

Basher, A., Kim, B. C., Lee, K. H., & Jung, H. Y. (2021). Volumetric feature-based alzheimer's disease diagnosis from smri data using a convolutional neural network and a deep neural network. *IEEE Access*, *9*, 29870–29882. Retrieved from https://doi.org/10.1109/ACCESS.2021.3059658 doi: 10.1109/ACCESS.2021.3059658

Brett, M., Markiewicz, C. J., Hanke, M., Cottaar, M., Cheng, C. P., Halchenko, Y. O., ... Ghosh, S. S. (2022). Nibabel: Access a cacophony of neuro-imaging file formats [Computer software manual]. Retrieved from https://nipy.org/nibabel/ (Neuroimaging data library)

Chapleau, M., Iaccarino, L., Soleimani-Meigooni, D., & Rabinovici, G. D. (2022). The role of amyloid pet in imaging neurodegenerative disorders: A review. *Journal of Nuclear Medicine*, *63*(Supplement 1), 13S–19S. Retrieved from https://doi.org/10.2967/jnumed.121.263195 doi: 10.2967/jnumed.121.263195

Dubois, B., Padovani, A., Scheltens, P., Rossi, A., Dell'Agnello, G., & Saykin, A. (2016). Timely diagnosis for alzheimer's disease: A literature review on benefits and challenges. *Journal of Alzheimer's Disease*, *49*(3), 617–631. Retrieved from https://doi.org/10.3233/JAD-150692 doi: 10.3233/JAD-150692

Ebrahimighahnavieh, M. A., Luo, S., & Chiong, R. (2020). Deep learning to detect alzheimer's disease from neuroimaging: A systematic literature review. *Computer Methods and Programs in Biomedicine*, *187*, 105242. Retrieved from https://doi.org/10.1016/j.cmpb.2019.105242 doi: 10.1016/j.cmpb.2019.105242

Elmi, H. (2024). *Csai bachelor thesis.* Retrieved from https://github.com/hamseelmi199/CSAI-Bachelor-Thesis2_HamseElmi (GitHub repository)

Fan, S., Ponisio, M. R., Xiao, P., Ha, S. M., Chakrabarty, S., Lee, J. J., & Flores, S. (2024). Amyloidpetnet: Classification of amyloid positivity in brain pet imaging using end-to-end deep learning. *Radiology*, *311*(3), e231442. Retrieved from https://doi.org/10.1148/radiol.231442

doi: 10.1148/radiol.231442

Fathi, S., Ahmadi, A., & Dehnad, A. (2024). A deep learning-based ensemble method for early diagnosis of alzheimer's disease using mri images. *Neuroinformatics*, *22*, 89–105. Retrieved from https://doi.org/10.1007/s12021-023-09646-2 doi: 10.1007/s12021-023-09646-2

Graham, R. N. J., Perriss, R. W., & Scarsbrook, A. F. (2005). Dicom demystified: A review of digital file formats and their use in radiological practice. *Clinical Radiology*, *60*(11), 1133–1140. Retrieved from https://doi.org/10.1016/j.crad.2005.07.003 doi: 10.1016/j.crad.2005.07.003

Grammarly. (n.d.). *Grammarly: Free online writing assistant.* Retrieved from https://www.grammarly.com/

Grueso, S., & Viejo-Sobera, R. (2021). Machine learning methods for predicting progression from mild cognitive impairment to alzheimer's disease dementia: A systematic review. *Alzheimer's Research Therapy*, *13*(1), 162. Retrieved from https://doi.org/10.1186/s13195-021-00900-w doi: 10.1186/s13195-021-00900-w

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020). Array programming with numpy. *Nature*, *585*, 357–362. Retrieved from https://numpy.org/ doi: 10.1038/s41586-020-2649-2

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*. Retrieved from https://arxiv.org/abs/1512.03385

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2018). *Densely connected convolutional networks.* Retrieved from https://arxiv.org/abs/1608.06993

Image and Data Archive (IDA). (n.d.). *Ida: Image and data archive.* Retrieved from https://ida.loni.usc.edu/

Jo, T., Nho, K., & Risacher, S. L. (2020). Deep learning detection of informative features in tau pet for alzheimer's disease classification. *BMC Bioinformatics*, *21(Suppl 21)*, 496. Retrieved from https://doi.org/10.1186/s12859-020-03848-0 doi: 10.1186/s12859-020-03848-0

Kim, S., Lee, P., Oh, K. T., Byun, M. S., Yi, D., Lee, J. H., & Kim, Y. K. (2021). Deep learning-based amyloid pet positivity classification model in the alzheimer's disease continuum by using 2-[18f]fdg pet. *EJNMMI Research*, *11*(1), 56. Retrieved from https://doi.org/10.1186/s13550-021-00798-3 doi: 10.1186/s13550-021-00798-3

Li, C. R. (2016). dcm2niix: Dicom to nifti conversion [Computer software manual]. Retrieved from https://github.com/rordenlab/dcm2niix

(Tool for converting DICOM to NIfTI format)

Logan, R., Williams, B. G., Ferreira da Silva, M., Indani, A., Schcolnicov, N., Ganguly, A., & Miller, S. J. (2021). Deep convolutional neural networks with ensemble learning and generative adversarial networks for alzheimer's disease image data classification. *Frontiers in Aging Neuroscience*, *13*. Retrieved from https://doi.org/10.3389/fnagi.2021.720226 doi: 10.3389/fnagi.2021.720226

Odusami, M., Maskeliūnas, R., Damaševičius, R., & Misra, S. (2023). Explainable deep-learning-based diagnosis of alzheimer's disease using multimodal input fusion of pet and mri images. *Journal of Medical and Biological Engineering*, *43*(3), 291–302. Retrieved from https://doi.org/10.1007/s40846-023-00801-3 doi: 10.1007/s40846-023-00801-3

OpenAI. (2024). *Chatgpt: A generative language model (gpt-4o).* Retrieved from https://openai.com/ (Version used: GPT-4o)

Python Software Foundation. (2018). Python 3.7 [Computer software manual]. Retrieved from https://www.python.org/ (Python programming language, version 3.7)

Qiu, S., Joshi, P. S., Miller, M. I., Xue, C., Zhou, X., Karjadi, C., & Chang, G. H. (2020). Development and validation of an interpretable deep learning framework for alzheimer's disease classification. *Brain*, *143*(6), 1920–1933. Retrieved from https://doi.org/10.1093/brain/awaa137 doi: 10.1093/brain/awaa137

Rahman, M., & Lendel, C. (2021). Extracellular protein components of amyloid plaques and their roles in alzheimer's disease pathology. *Molecular Neurodegeneration*, *16*, 59. Retrieved from https://doi.org/10.1186/s13024-021-00465-0 doi: 10.1186/s13024-021-00465-0

Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., ... van der Flier, W. M. (2021). Alzheimer's disease. *The Lancet*, *397*(10284), 1577–1590. Retrieved from https://doi.org/10.1016/S0140-6736(20)32205-4 doi: 10.1016/S0140-6736(20)32205-4

Shanmugavadivel, K., Sathishkumar, V. E., Cho, J., & Subramanian, M. (2023). Advancements in computer-assisted diagnosis of alzheimer's disease: A comprehensive survey of neuroimaging methods and ai techniques for early detection. *Ageing Research Reviews*, *91*, 102072. Retrieved from https://doi.org/10.1016/j.arr.2023.102072 doi: 10.1016/j.arr.2023.102072

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. In *Computer vision and pattern recognition.* Retrieved from http://export.arxiv.org/pdf/1409.1556

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S.,

Fagan, A. M., & Iwatsubo, T. (2011). Toward defining the preclinical stages of alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's Dementia*, *7*(3), 280–292. Retrieved from https://doi.org/10.1016/j.jalz.2011.03.003 doi: 10.1016/j.jalz.2011.03.003

TorchVision. (2016). *Torchvision: Pytorch's computer vision library.* https://github.com/pytorch/vision. GitHub.

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., ... Yu, T. (2014). scikit-image: Image processing in python. *PeerJ*, *2*, e453. Retrieved from https://scikit-image.org/ doi: 10.7717/peerj.453

Weiner, M. W., Nosheny, R., Camacho, M., Truran-Sacrey, D., Mackin, R. S., Flenniken, D., ... Veitch, D. (2018). The brain health registry: An internet-based platform for recruitment, assessment, and longitudinal monitoring of participants for neuroscience studies. *Alzheimer's Dementia*, *14*(8), 1063–1076. Retrieved from https://doi.org/10.1016/j.jalz.2018.02.021 doi: 10.1016/j.jalz.2018.02.021

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., ... Trojanowski, J. Q. (2017). The alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's Dementia*, *13*(5), 561–571. Retrieved from https://doi.org/10.1016/j.jalz.2016.10.006 doi: 10.1016/j.jalz.2016.10.006

Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., & Dormont, D. (2020). Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation. *Medical Image Analysis*, *63*, 101694. Retrieved from https://doi.org/10.1016/j.media.2020.101694 doi: 10.1016/j.media.2020.101694

Zhang, Z., & Sabuncu, M. R. (2018). *Generalized cross entropy loss for training deep neural networks with noisy labels.* Retrieved from https://arxiv.org/abs/1805.07836

Zhao, Z., Chuah, J. H., Lai, K. W., Chow, C.-O., Gochoo, M., Dhanalakshmi, S., ... Wu, X. (2023). Conventional machine learning and deep learning in alzheimer's disease diagnosis using neuroimaging: A review. *Frontiers in Computational Neuroscience*, *17*, 1038636. Retrieved from https://doi.org/10.3389/fncom.2023.1038636 doi: 10.3389/fncom.2023.1038636