RADIOMIC ANALYSIS OF ALZHEIMER'S DISEASE USING FDG-PET IMAGING

by

Ramin Rasi

B.S. in Computer Engineering, Azad University, 2009

M.S. in Computer Science, Tabriz University, 2014

Submitted to the Institute of Biomedical Engineering

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Biomedical Engineering

Boğaziçi University

2025

RADIOMIC ANALYSIS OF ALZHEIMER'S DISEASE USING FDG-PET IMAGING

APPROVED BY:

Prof. Dr. Albert Güveniş (Thesis Advisor)	.....
Prof. Dr. Cengizhan Öztürk	.....
Prof. Dr. Hale Saybaşılı	.....
Assist. Prof. Dr. İmran Göker	.....
Assist. Prof. Dr. Mehmet Kocatürk	.....

DATE OF APPROVAL: 16 April 2025

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Prof. Dr. Albert Güveniş, for his invaluable guidance, unwavering support, and insightful feedback throughout my research journey. His expertise and mentorship have been instrumental in shaping both my academic and professional development.

I am also profoundly grateful to my parents for their unconditional love, encouragement, and sacrifices that have made this achievement possible. Their endless support has been my greatest source of strength and motivation.

Finally, I sincerely appreciate the Institute of Biomedical Engineering at Boğaziçi University for fostering an intellectually enriching environment and providing the essential resources that enabled this research. The knowledge and experiences I have acquired during my time here have been profoundly impactful and transformative.

This research benefited significantly from the financial support provided by the Bogazici University Research Fund (BAP), project code 19774. Their contribution enabled the successful conduct of the experimental work and subsequent data analysis.

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, Ramin Rasi, hereby certify that I am aware of the Academic Ethics and Integrity Policy issued by the Council of Higher Education (YÖK) and I fully acknowledge all the consequences due to its violation by plagiarism or any other way.

Name :

Signature:

Date:

ABSTRACT

RADIOMIC ANALYSIS OF ALZHEIMER'S DISEASE USING FDG-PET IMAGING

Alzheimer's disease (AD) demands more advanced and personalized diagnostic methods. This dissertation proposes a radiomics-driven framework using FDG-PET to enhance the diagnosis, staging, and risk evaluation of AD. By integrating image-derived features and machine learning, we aim to facilitate early, non-invasive, and individualized assessment. We developed a fully automated brain radiomics platform capable of distinguishing between CN, MCI, and AD individuals. The hippocampus, entorhinal cortex, and amygdala emerged as key discriminative regions. Our simplified FDG-PET model achieved an AUC of 0.853 for predicting amyloid positivity, a critical pathological hallmark of AD, using features from the hippocampus, inferior parietal lobule, and isthmus cingulate. The platform was further extended for ApoE4 genotype prediction, achieving an AUC of 0.945 with features extracted from the hippocampus, amygdala, thalamus, and pars orbitalis. We also investigated hippocampus-amygdala connectivity, identifying robust biomarkers such as Shape Mesh Volume and GLDM Small Dependence Low Gray Level Emphasis (AUC = 0.88). Subregional analysis of hippocampal and amygdaloid structures revealed additional radiomic features, including GLRLM Long Run Emphasis and GLDM Small Dependence Emphasis, which differentiate AD from MCI and CN, indicating early microstructural and metabolic changes prior to visible atrophy. Our findings establish FDG-PET radiomics as a reliable, non-invasive imaging biomarker for diagnosing and monitoring AD. The proposed framework enables risk stratification and supports clinical decision-making, paving the way for preventive and personalized AD management.

Keywords: Alzheimer's disease (AD), FDG-PET, Radiomics, Machine Learning, Amyloid Beta, ApoE4, Mild Cognitive Impairment (MCI), Hippocampus, Amygdala, Clinical Decision Support, Personalized Medicine, Neuroimaging Biomarkers.

ÖZET

FDG-PET GÖRÜNTÜLEME KULLANARAK ALZHEIMER HASTALIĞININ RADYOMIK ANALIZI

Alzheimer hastalığı (AD), daha gelişmiş ve kişiselleştirilmiş tanı yöntemlerine ihtiyaç duymaktadır. Bu tez, AD'nin tanısı, evrelemesi ve risk değerlendirmesini geliştirmek amacıyla FDG-PET temelli bir radyomik yaklaşıma dayanan bir çerçeve önermektedir. Görüntüden türetinen özelliklerin makine öğrenmesi ile entegrasyonu sayesinde erken, invaziv olmayan ve bireyselleştirilmiş değerlendirme hedeflenmiştir. Tam otomatik bir beyin radyomik platformu geliştirerek bilişsel olarak CN, MCI ve AD olan bireyler arasında ayrım yapabildik. Hipokampus, entorhinal korteks ve amigdala temel ayırt edici bölgeler olarak öne çıkmıştır. Basitleştirilmiş FDG-PET modelimiz, hipokampus, inferior parietal lobül ve isthmus singulat bölgesinden elde edilen özelliklerle AD'nin kritik bir patolojik göstergesi olan amiloid pozitifliğini %85.3 AUC değeri ile tahmin etmiştir. Platform, AD için önemli bir genetik risk faktörü olan ApoE4 genotipini tahmin etmek amacıyla genişletilmiş ve hipokampus, amigdala, talamus ve pars orbitalis bölgelerinden elde edilen özelliklerle %94.5 AUC başarısı göstermiştir. Ayrıca, hipokampus–amigdala bağlantısı araştılarak *Shape Mesh Volume* ve *GLDM SDLGLE* gibi sağlam biyobelirteçler tanımlanmıştır (AUC = 0.88). Hipokampal ve amigdalaya ait alt bölgelerin analizi, AD'yi MCI ve CN'den ayıran *GLRLM Long Run Emphasis* ve *GLDM Small Dependence Emphasis* gibi erken mikroyapısal ve metabolik değişiklikleri ortaya koymuştur. Elde edilen bulgular, FDG-PET radyomiğini AD'nin tanı, izleme ve risk sınıflandırmasında güvenilir, invaziv olmayan bir görüntüleme biyobelirteci olarak konumlandırmakta; klinik karar destek süreçlerine katkı sağlayarak önleyici ve kişiselleştirilmiş tıp uygulamaları için zemin hazırlamaktadır.

Anahtar Sözcükler: Alzheimer hastalığı, FDG-PET, Radyomik, Makine Öğrenmesi, Amiloid Beta, ApoE4, Hafif Bilişsel Bozukluk, Hipokampus, Amigdala, Klinik Karar Destek, Kişiselleştirilmiş Tıp, Nörogörüntüleme Biyobelirteçleri.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

μ	Mean value of a distribution
σ	Standard deviation
ϵ	Small quantity / Epsilon in APOE4
p-value	Probability value in statistical tests
x_i	The i-th sample in a dataset
$f(x)$	Function of x used in modeling

LIST OF ABBREVIATIONS

AD	Alzheimer's Disease
MCI	Mild Cognitive Impairment
CN	Cognitively Normal
FDG-PET	Fluorodeoxyglucose Positron Emission Tomography
ROI	Region of Interest
MRI	Magnetic Resonance Imaging
SUVR	Standardized Uptake Value Ratio
SPM	Statistical Parametric Mapping
CSF	Cerebrospinal Fluid
APOE4	Apolipoprotein E $\epsilon 4$
ANTs	Advanced Normalization Tools
DKT	Desikan-Killiany-Tourville (Atlas)
IBSI	Image Biomarker Standardisation Initiative
PyRadiomics	Python Radiomics Extraction Toolkit
PCA	Principal Component Analysis
LASSO	Least Absolute Shrinkage and Selection Operator
RF	Random Forest
GB	Gradient Boosting
DT	Decision Tree
GNB	Gaussian NaÃ¯ve Bayes
GP	Gaussian Process
MLP	Multi-Layer Perceptron
AB	AdaBoost
QDA	Quadratic Discriminant Analysis
SHAP	SHapley Additive exPlanations
DCA	Decision Curve Analysis
ADNI	Alzheimer's Disease Neuroimaging Initiative

1. INTRODUCTION

1.1 Background and Significance

As Alzheimer's disease (AD) affects millions worldwide, understanding its complex mechanisms is critical for researchers and clinicians. Alzheimer's disease is the most common form of dementia, accounting for 60-70% of all dementia cases (55.2 million people worldwide), with nearly 10 million new cases diagnosed annually, equating to one new case every three seconds [1]. Progression in Alzheimer's includes complex neurobiological changes, including deposition of amyloid-beta plaques and tau tangles, along with pervasive inflammation. Each such factor plays a critical role in progressive cognitive decline [2]. Despite extensive research, the precise mechanisms driving the progression from mild cognitive impairment (MCI) to advanced dementia remain unclear, particularly the role of metabolic and structural changes [3]. Although several brain regions are associated with Alzheimer's progression, no single region has been conclusively identified as the primary driver [4]. Addressing these gaps necessitates a deeper exploration of understudied brain regions, network dysfunction, and metabolic alterations to identify novel biomarkers and therapeutic targets.

1.2 Importance of FDG PET in Alzheimer's Diagnosis

Diagnosis of AD by standard methods is usually dependent on a convergence of cognitive assessment, biomarker evaluation, and imaging [5], [6]. Early diagnosis, however, in the scenario of Mild cognitive impairment, is much more difficult. MCI involves apparent cognitive deterioration that might or might not actually develop into AD; hence, it is highly important to detect which of the cases is likely to progress. Imaging modalities allow us to explore the changes within the brain that accompany AD [7], [8]. Accordingly, Fluorine-18-deoxyglucose positron emission tomography (FDG-PET) imaging has been highly useful in tracing the changes in brain metabolism defining

neurodegeneration, and provides early indicators in the context of MCI and AD.

1.3 Current Methods for Diagnosing and Characterizing AD

Present diagnostic assessments of AD rely on assessment by a clinician, cognitive testing, evaluation of biomarkers, and imaging [9]. Cognitive batteries include the Mini-Mental State Examination (MMSE) as a major test of executive function and memory [10]. Supplementing the MMSE, other tests, the Rey Auditory Verbal Learning Test (RAVLT), also commonly quantify verbal learning and memory and permit a finer assessment of episodic function of the memory [11]. Similarly, the Montreal Cognitive Assessment (MoCA) provides a better evaluation of the cognitive areas of attention, visuospatial function, and linguistic function [12]. Tests are informative, yet remain subjective and less sensitive to detecting AD in its earliest stages when changes are merely molecular or metabolic, not yet structural [13], [14]. Used in conjunction with biomarker evaluation and imaging, the tests will provide better diagnosis at the stages of the illness. Neuroimaging has completely changed the diagnosis of AD, offering a window into brain function and structure. MRI is required to detect atrophy in critical regions, the hippocampus, so that it is possible to report objectively on neurodegeneration on an objective basis [15]. Functional imaging methods, i.e., FDG-PET, measure glucose metabolism and report on neuronal function [16]. Amyloid PET imaging offers a second line of assessment by imaging amyloid plaques, the hallmark of AD pathology. Biochemical markers measured by CSF analysis, i.e., concentrations of Aβ42 and tau protein, further refine diagnosis. Genetic APOE ϵ4 testing, however, offers the ability to detect those at higher genetic risk of late-onset AD. They are, however, each with major limitations, i.e., expense, invasiveness (in the context of CSF analysis), and varying sensitivity, especially in the context of AD at an earlier phase.

1.4 FDG PET Analysis of AD: Conventional Methods

FDG PET is an important imaging modality for assessing cerebral glucose metabolism, which indirectly reflects neuronal activity [17]. In Alzheimer's disease, FDG PET demonstrates typical patterns of hypometabolism, mainly involving the posterior cingulate cortex, precuneus, and parietotemporal regions, with relative sparing of primary motor, sensory, and visual cortices [18]. These findings have established FDG PET as an essential tool for both clinical diagnosis and research into AD progression. Different FDG-PET analysis methods vary in their strengths and limitations, ranging from traditional visual assessments to advanced radiomics-based approaches, as summarized in Table 1.1.

Table 1.1 Comparing Traditional and Radiomics-Based FDG-PET Analysis.

Method	**Strengths**	**Limitations**
Visual Analysis	Clinically intuitive; widely used	Subjective; low sensitivity
ROI-Based Analysis	Standardized; well-established	Predefined regions may miss subtle changes
Voxel-Based Analysis	Whole-brain approach; more sensitive than ROI-based	Requires large control datasets; statistical bias
Radiomics Analysis	High-dimensional feature extraction; objective; machine-learning compatible	Computationally intensive; requires validation

1.4.1 Visual Analysis

The most direct traditional approach is visual interpretation, based on qualitative evaluations by trained radiologists or clinicians. Areas of hypometabolism are determined by comparing the uptake of FDG in brain regions known to be typically affected by AD to those expected to be preserved. For instance, marked hypometabolism in the parietotemporal cortex and posterior cingulate relative to the sensorimotor or occipital cortex is indicative of AD.

However, the intuitive and wide use of visual analysis in clinical practice also entails some drawbacks. It is a subjective interpretation that may bring out interobserver

variability; this method is generally insensitive to the early or atypical presentation of AD. These limitations thus point to the need for quantitative methods to supplement visual assessments [19].

1.4.2 Quantitative Analysis: ROI-Based Methods

Quantitative analysis involves ROI-based methods that provide numerical measurements of metabolic activity [20]. FDG uptake in a specific ROI is normalized to a reference region presumed to be unaffected by the disease, yielding the Standardized Uptake Value Ratio (SUVR) (Eq. 1.1):

$$SUVR = \frac{SUVR_{target}}{SUVR_{reference}} \tag{1.1}$$

where:

- $SUVR_{target}$: The standardized uptake value in the target region (e.g., posterior cingulate cortex).
- $SUVR_{reference}$: The standardized uptake value in a reference region (e.g., cerebellum or pons).

The SUVR calculation enhances comparability across individuals and scans, mitigating variability caused by global differences in tracer uptake (Figure 1.1). This method has been widely used in both cross-sectional and longitudinal studies to measure metabolic decline in specific regions of interest [21].

However, ROI-based analysis has limitations. Predefining regions may overlook subtle or diffuse metabolic changes that fall outside the chosen boundaries. Further-

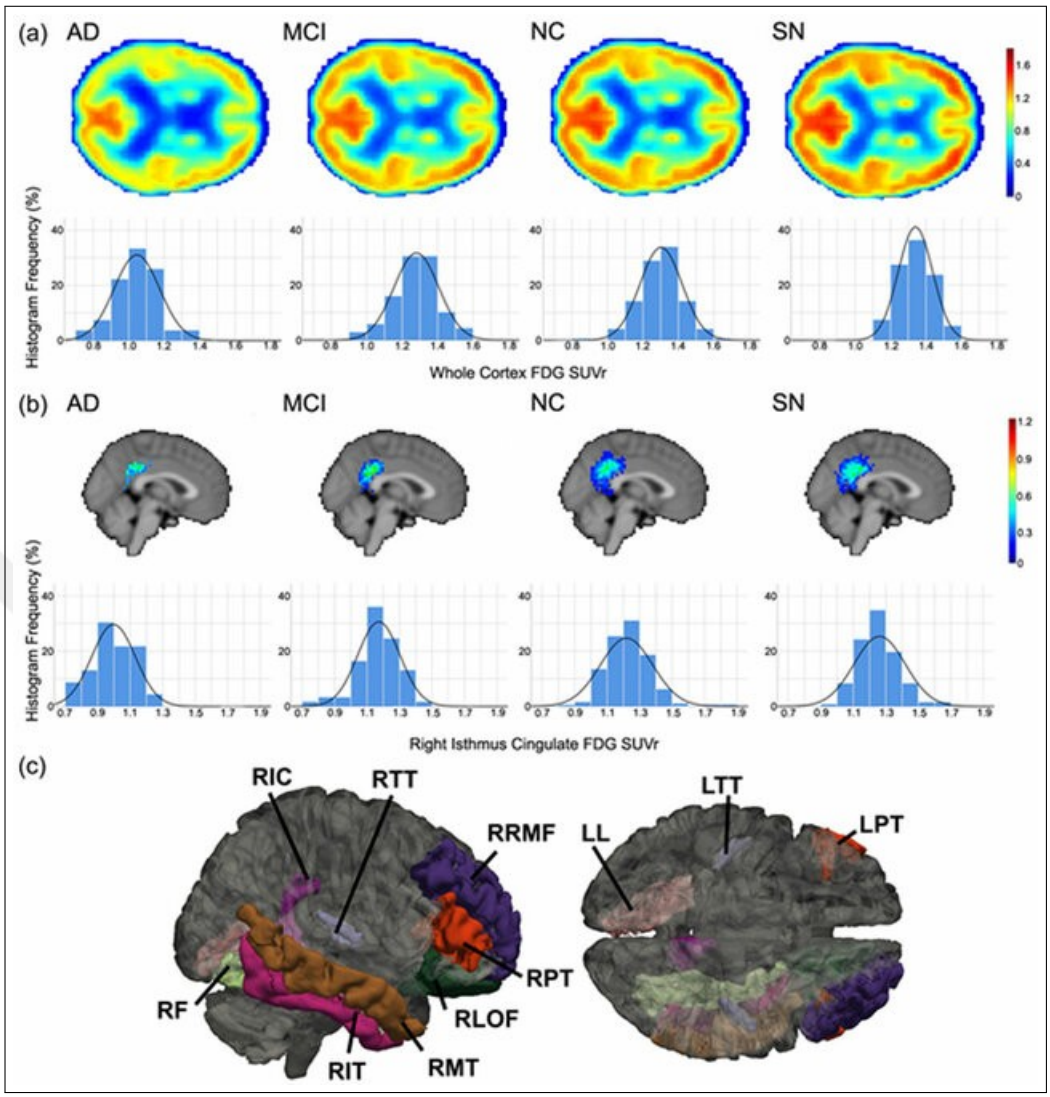

Figure 1.1 A schematic representation of key ROIs typically assessed in AD (e.g., posterior cingulate, precuneus, parietotemporal cortex) with their corresponding SUVR values.

more, the selection of the reference region can significantly impact results, introducing variability in interpretations [22].

1.4.3 Voxel-Based Analysis

Voxel-based methods, in turn, assess metabolic activity in the whole brain at a single voxel level. The common tool for this is statistical parametric mapping (SPM). Significant deviations in metabolism are identified voxel by voxel between the patient's

FDG PET and that of a control group [23]

Results are presented as statistical maps, showing areas with reduced glucose uptake, for example, $p < 0.05$ corrected for multiple comparisons. Voxel-based approaches are less dependent on pre-defined regions and provide a more comprehensive analysis but require robust statistical methods and access to control datasets.

1.5 Clinical Implications and Limitations

Conventional methods, including visual, ROI-based, and voxel-based analyses, have been pivotal in establishing FDG PET as a cornerstone for studying AD. They provide valuable insights into disease mechanisms and progression, particularly in identifying the characteristic hypometabolism patterns of AD. However, their limitations, such as reliance on subjective interpretation, predefined regions, or control datasets, highlight the need for advanced techniques.

Emerging machine-learning approaches and radiomics are poised to address these challenges by leveraging high-dimensional data from FDG PET to identify subtle and diffuse patterns of metabolic abnormalities. These methods complement conventional approaches, paving the way for more accurate and robust analyses in the future.

1.6 Shortcomings of Current Methods

Despite their contributions to AD research and diagnosis, current methods face several challenges:

- **Sensitivity Gaps:** Techniques like structural MRI detect physical brain changes that occur later in the disease, often missing the subtle molecular alterations in its early stages.

- **Invasive Testing:** CSF biomarker analysis, requiring lumbar puncture, introduces risks and discomfort, discouraging routine use in clinical practice.
- **Region-Specific Focus:** Traditional imaging approaches target specific brain areas, potentially overlooking diffuse or subtle abnormalities elsewhere in the brain.
- **Subjective Interpretation:** The reliance on expert interpretation introduces variability and increases the likelihood of diagnostic discrepancies.
- **High Costs:** Advanced imaging technologies, such as PET, and the need for specialized tracers impose significant financial burdens on healthcare systems.

These limitations emphasize the need for innovative approaches that offer higher sensitivity, non-invasiveness, and broader accessibility.

1.7 The potential role of radiomics in Advancing Image Analysis

Recent advancement in the field of radiomics, a computational algorithmic method of deriving numerical imaging data features, offers further insight into changes in the brain on the path from MCI to AD [25], [26]. Systematic textural, morphological, and intensity-derived feature analysis of brain structures is possible with radiomics, often detecting change on a level not visible to the naked eye [27]. The method is of special potential in PET and MRI imaging, where it enables fine-grained analysis of structures such as the hippocampus-amygdala complex, where AD's multifarious metabolic and structural change is underway. The window it affords is rich in which to explore brain imaging data in terms of imaging metrics other than atrophy, permitting visualization of region-by-region textural and connectivity sensitivity to AD progression [28], [29]. Texture-derived radiomic metrics, for instance, might highlight heterogeneity within a region, sensitive to cellular or metabolic heterogeneity of AD [30]. Textural and connectivity patterns in the hippocampus associated with AD have

been described in radiomics work, opening the possibility that these might be clinically useful markers of predictive cognitive decline [31]. Such analysis might assess specific subfields or nuclei within structures differently sensitive to AD change, however, which might be targets of critical importance to intervention.

1.8 Objectives of This Study

This study aims to leverage the comprehensive capabilities of radiomics to enhance the early detection and characterization of Alzheimer's disease through FDG-PET imaging. Specifically, the study will:

- Identifying Critical Features: Analyze FDG-PET radiomics features to pinpoint regions and characteristics linked to early AD pathology.
- Predicting Biomarkers: Develop predictive models for amyloid positivity and genetic risk factors like APOE4 using imaging data.
- Monitoring Disease Progression: Differentiate between cognitively normal individuals, those with mild cognitive impairment (MCI), and AD patients, and track disease progression over time.
- Validating Radiomics for Precision Medicine: Evaluate the clinical utility of radiomics for personalized diagnostics and treatment planning.

By achieving these goals, the study seeks to elevate the role of FDG-PET imaging in AD diagnostics, deepen our understanding of the disease, and contribute to the broader adoption of precision medicine in neurodegenerative disorders.

2. METHODOLOGY

In this study, we used radiomics techniques for different objectives, including amyloid positivity prediction, APOE4 status prediction, staging of Alzheimer's disease, and subregions analysis of the hippocampus and amygdala at scale. We utilized a comprehensive radiomics pipeline involving feature extraction, feature selection, classification, and SHAP and DCA-based interpretability analysis.

2.1 ADNI and Participants

Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD.

Due to the limited availability of data, our study utilized multiple participant groups sourced from the ADNI to address different study objectives, including predicting amyloid positivity, predicting APOE4 status, staging Alzheimer's disease, and conducting a deep analysis of the subregions of the hippocampus and amygdala (Table 2.1).

Table 2.1 Participant Clinical Characteristics for Each of the Different Studies.

Predicting Amyloid Status			
Group	**No. of Cases (Male/Female)**	**Age (Mean ± SD)**	**No. of Cases (Amyloid Positive/Negative)**
LMCI	28 / 41	72 ± 7.25	52 / 17
EMCI	51 / 40	71.5 ± 7.01	39 / 52
AD	33 / 36	73.9 ± 7.85	66 / 3
SMC	8 / 12	71.9 ± 4.23	7 / 13
CN	15 / 24	74.3 ± 6.44	14 / 25
MCI	11 / 2	75.3 ± 7.42	7 / 6
Total	146 / 155	73.2 ± 6.7	185 / 116
Staging Alzheimer's Disease (AD)			
Group	**No. of Cases (Male/Female)**	**Age (Mean ± SD)**	
AD	91 / 72	74.6 ± 8.12	
MCI	107 / 91	72.5 ± 8.07	
CN	91 / 97	73.6 ± 6.37	
Total	289 / 260	74.1 ± 7.02	
Investigating the Subregions of the Hippocampus and Amygdala			
Group	**No. of Cases (Male/Female)**	**Age (Mean ± SD)**	
AD	93 / 72	74.6 ± 8.1	
MCI	121 / 80	72.7 ± 7.4	
CN	97 / 92	73.7 ± 6.3	
Total	306 / 249	73.6 ± 7.3	
Predicting the APOE4 Status			
Non-e4 Carrier	33 / 23	71.34 ± 7.6	
Two e4 Carrier	31 / 25	69.35 ± 6.66	
Total	64 / 48	70.35 ± 7.2	
Investigating the Connectivity of the Hippocampus and Amygdala			
CN	174 (82 / 92)	73.5 ± 6.4	
EMCI	226 (126 / 100)	70.4 ± 7.2	
MCI	186 (109 / 77)	71.9 ± 8.9	
LMCI	123 (60 / 63)	72.1 ± 7.9	
AD	153 (84 / 69)	74.5 ± 8.1	

2.2 Diagnostic Grouping Criteria in ADNI

In the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, participant stages are defined using a combination of clinical assessments, cognitive testing, and sometimes biomarkers. Key criteria include the Mini-Mental State Examination (MMSE), Clinical Dementia Rating (CDR), and memory scores, often from the Logical Memory subscale of the Wechsler Memory Scale. It's important to note that ADNI criteria may evolve across different phases, and these classifications are for research purposes.

Table 2.2 Participant Clinical Characteristics for Each of the Different Studies.

Group	**CDR**	**MMSE**	**Logical Memory II Delayed Recall ($\geq$ 16 Years Education)**	**Other Criteria**
Cognitively Normal (CN)	0	24-30	$\geq$ 11	No memory complaints; Memory within normal limits; Functionally independent
Early MCI (EMCI)	0.5	24-30	9-11	Subjective memory complaints; Mild memory impairment; Preserved daily functioning
Late MCI (LMCI)	0.5	24-30	$\leq$ 8	Subjective memory complaints; Significant memory impairment; Functionally independent; No impairment in other cognitive domains
Mild Cognitive Impairment (MCI)	0.5	24-30	$\leq$ 11 or variable	General category used in some studies combining EMCI and LMCI
Alzheimer's Disease (AD)	$\geq$ 0.5	20-26	Severely impaired	Impairment in multiple cognitive domains; Functional decline; Meets NINCDS-ADRDA or NIA-AA criteria

2.3 FDG PET Acquisition

Our study used FDG PET scans that involved injecting participants with 185 MBq (5 mCi) of FDG and then performing a dynamic 3D scan consisting of six 5-minute frames acquired 30-60 minutes post-injection. ADNI provides four processed formats for FDG PET scans, ensuring compatibility between scanners and simplifying analysis. We utilized preprocessed FDG PET scans that underwent three stages of processing. Firstly, the averaged image from each subject's baseline FDG PET scan was co-registered. Secondly, this co-registered image was realigned into a standard 160x160x96 voxel image grid with 1.5 mm cubic voxels. Lastly, the image grid was aligned with the subject's anterior-posterior axis, ensuring it was parallel to the anterior and posterior commissure (AC-PC) line. All FDG PET scans, including the baseline and subsequent scans, were co-registered with this baseline image. An average image is created from the AC-PC co-registered frames, and intensity is normalized using a subject-specific mask [32].

2.4 MRI Acquisition

The ADNI project has collected a wide range of sMRI data across its phases (1/GO/2/3). While sMRI acquisition protocols have evolved, the initiative provides standardized datasets to address variations in sMRI scanners between sites and ensure consistency for research. ADNI 3 exclusively uses 3T scanners and offers two tiers ("Basic" and "Advanced") to manage this variability. We utilized standardized sMRI data, particularly T1-weighted MRI scans obtained using a sagittal MP-RAGE plane and 1.2mm slice thickness. These datasets were developed to promote data analysis consistency and facilitate researcher use [33].

2.5 Co-registration

We used Advanced Normalization Tools (ANTs) to co-register FDG PET and sMRI images. The tool ANTs is highly effective for image registration, correcting spatial discrepancies between scans [34]. Given the short interval (less than 90 days) between sMRI and FDG-PET acquisitions, rigid registration of the ANTs was employed. This fundamental image alignment method corrects for initial spatial discrepancies, including translations and rotations, without altering the original image shape or size. In this step, FDG-PET images were set as fixed images and sMRI images as moving images. This accurate alignment is crucial for subsequent analyses in the segmentation step. The tool ANTs uses a number of algorithms for registering two images to establish spatial correspondence: i) Symmetric Normalization: A cornerstone of ANTs, this method ensures unbiased registration by treating both images equally as moving and fixed ii) Diffeomorphic Transformations: These transformations guarantee smooth and invertible mappings, preserving anatomical structures iii) Multi-modal Registration: ANTs can handle images from different modalities (e.g., MRI, CT, PET) by incorporating appropriate similarity metrics.

2.6 Segmentation

2.6.1 Whole Brain

To segment the whole brain into regions of interest, our study utilized FastSurfer [35], a well-established neuroimaging tool, for a detailed volumetric analysis. FastSurfer, which replicates FreeSurfer's anatomical segmentation method with automated segmentation capabilities, allowed us to parcellate the entire brain into 95 distinct anatomical regions, as shown in Table 2.3, using the Desikan-Killiany-Tourville (DKT) atlas [36]. This standardized atlas ensures consistency and reliability in our measurements.

Table 2.3 IDs and corresponding regions segmented by FreeSurfer.

Left Hemisphere		Right Hemisphere	
ID	**Region (L)**	**ID**	**Region (R)**
2	Cortical white matter	41	Cortical white matter
4	Lateral Ventricle	43	Lateral Ventricle
5	Inferior Lateral Ventricle	44	Inferior Lateral Ventricle
7	Cerebellar White Matter	46	Cerebellar White Matter
8	Cerebellar Cortex	47	Cerebellar Cortex
10	Thalamus	49	Thalamus
11	Caudate	50	Caudate
12	Putamen	51	Putamen
13	Pallidum	52	Pallidum
17	Hippocampus	53	Hippocampus
18	Amygdala	54	Amygdala
26	Accumbens	58	Accumbens
28	Ventral DC	60	Ventral DC
31	Choroid Plexus	63	Choroid Plexus
1002	caudalanteriorcingulate	2002	caudalanteriorcingulate
1003	caudalmiddlefrontal	2003	caudalmiddlefrontal
1005	cuneus	2005	cuneus
1006	entorhinal	2006	entorhinal
1007	fusiform	2007	fusiform
1008	inferiorparietal	2008	inferiorparietal
1009	inferiortemporal	2009	inferiortemporal
1010	isthmuscingulate	2010	isthmuscingulate
1011	lateraloccipital	2011	lateraloccipital
1012	lateralorbitofrontal	2012	lateralorbitofrontal
1013	lingual	2013	lingual
1014	medialorbitofrontal	2014	medialorbitofrontal
1015	middletemporal	2015	middletemporal
1016	parahippocampal	2016	parahippocampal
1017	paracentral	2017	paracentral
1018	parsopercularis	2018	parsopercularis
1019	parsorbitalis	2019	parsorbitalis
1020	parstriangularis	2020	parstriangularis
1021	pericalcarine	2021	pericalcarine
1022	postcentral	2022	postcentral
1023	posteriorcingulate	2023	posteriorcingulate
1024	precentral	2024	precentral
1025	precuneus	2025	precuneus
1026	rostralanteriorcingulate	2026	rostralanteriorcingulate
1027	rostralmiddlefrontal	2027	rostralmiddlefrontal
1028	superiorfrontal	2028	superiorfrontal
1029	superiorparietal	2029	superiorparietal

Continued on next page

Table 2.3 (continued)

Left Hemisphere		Right Hemisphere	
ID	**Region (L)**	**ID**	**Region (R)**
1030	superiortemporal	2030	superiortemporal
1031	supramarginal	2031	supramarginal
1034	transversetemporal	2034	transversetemporal
1035	insula	2035	insula
14	3rd-Ventricle	–	3rd-Ventricle
15	4th-Ventricle	–	4th-Ventricle
16	Brain Stem	–	Brain Stem
24	CSF	–	CSF
77	WM-hypointensities	–	WM-hypointensities

2.6.2 Hippo-Amygdala

We performed an analysis of a collection of brain regions using its high-fidelity segmentation capability with FreeSurfer version 7.1. The brain was automatically split into 95 distinct regions by the software according to the DKT-Atlas [36]. Advanced hippocampal subfield and amygdala nuclei segmentations were also done by FS using a probabilistic atlas formed by combining high-resolution ex vivo scans and standard-resolution in vivo images. The process allocated voxels in the hippocampus to 19 subfields, fimbria, dentate gyrus, and to CA1. Amygdala segmentation used an atlas formed using high-resolution ex vivo data to split the amygdala into 9 subnuclei [37], [38] (Table 2.4).

Table 2.4 IDs and corresponding subregions of the hippocampus and nuclei of the amygdala as segmented by FreeSurfer, detailing the hierarchical organization and anatomical substructures.

Hippocampus		Amygdala	
FS ID	**Subfields**	**FS ID**	**Nuclei**
203	parasubiculum	7001	Lateral-nucleus
233	presubiculum-head	7003	Basal-nucleus
235	subiculum-head	7005	Central-nucleus
237	CA1-head	7006	Medial-nucleus
239	CA3-head	7007	Cortical-nucleus
241	CA4-head	7008	Accessory-Basal-nucleus
243	GC-ML-DG-head	7009	Corticoamygdaloid-transitio
245	molecular_layer_HP-head	7010	Anterior-amygdaloid-area-AAA

Continued on next page

Table 2.4 (Continued)

Hippocampus		Amygdala	
FS ID	**Subfields**	**FS ID**	**Nuclei**
211	HATA	7015	Paralaminar-nucleus
234	presubiculum-body		
236	subiculum-body		
238	CA1-body		
240	CA3-body		
242	CA4-body		
244	GC-ML-DG-body		
246	molecular_layer_HP-body		
212	fimbria		
226	Hippocampal_tail		
215	hippocampal-fissure		

2.6.3 Border region delineation

For the precise boundary area definition between adjacent segments within an image, we applied the concept of the distance transform. It is a process that assigns a pixel a distance label corresponding to the shortest distance to the boundary point. We applied the Euclidean distance transform [39], [40] in our approach specifically. The method has an added benefit over conventional dilation methods in that it computes the distance of every pixel to the two adjacent segment boundaries. We then specify a border region with user-devised thickness based on the computed distances (Eq. 2.1). The distance transform enables finer and controllable boundary area specification and is therefore particularly suitable where precise boundary characterization is required, e.g., segmentation verification and morphometric analysis.

$$D_x = \min_{y \in \partial\Omega} \|x - y\| \tag{2.1}$$

where:

- $D(x)$ is the distance transform at the point x.
- ∂S represents the border of the segment.

- $\|x - y\|$ denotes the Euclidean distance between points x and y.

2.7 Feature Extraction

To address the challenge of obtaining standard features in radiomics, we utilized PyRadiomics, an open-source platform that aligns with the International Biomarker Standardization Initiative (IBSI) [43]. This platform simplifies the extraction of a wide variety of features from medical images. PyRadiomics offers specialized tools for analyzing first-order statistics, shapes, and textures using different techniques [41]. Importantly, it allows feature extraction from both individual slices (2D) and entire volumes (3D) across various imaging modalities. This standardized workflow within PyRadiomics promotes consistency and reproducibility, enabling researchers to effectively compare findings across different studies [42]. The extracted features were classified into multiple categories:

- First order Order Statistics (19 features)
- Gray Level Dependence Matrix (14 features)
- Shape based (2D) (10 features)
- Gray level co-occurrence Matrix (24 features)
- Gray level run Length Matrix (16 features)
- Neighboring Gray Tone Difference Matrix (5 features)
- Gray Level Size Zone Matrix (16 features)
- Shape based (3D) (16 features)

2.8 Feature Selection

Radiomics analysis often produces redundant features, leading to overfitting. To mitigate this, we utilized eight feature selection methods from different categories to assess feature importance. This included filter methods like Analysis of Variance (ANOVA), Principal Component Analysis (PCA), and chi-square; embedded methods such as Least Absolute Shrinkage and Selection Operator (LASSO) and feature importance (FI); and wrapper methods like mutual information (MI), recursive feature elimination (RFE), and random forest algorithm (RFA). Before feature selection, preprocessing steps included removing constant, quasi-constant, and duplicated features to enhance computational efficiency. This process helped improve classification performance by reducing noise and redundancy. 2.10 Classification. The objective of classification methods is to analyze training data and formulate a hypothesis that enables the prediction of labels for both visible and invisible observations. It has been observed in machine learning studies that the performance of classification methods can vary depending on the specific problem being addressed[43]. Therefore, in radiomics studies, identifying the most suiTable machine learning method that aligns with the specific problem at hand poses a significant challenge. To address this challenge, it becomes necessary to explore and evaluate different methods in order to determine the optimal machine learning approach for the given radiomics study. To tackle this concern, we conducted a comprehensive analysis of nine classification methods derived from different classifier families. These methods include:

- K-Neighbors (KNN)
- Random Forest (RF)
- Gradient Boosting (GBM)
- Gaussian Process (GP)
- Decision Tree (DT)
- Multi-layer Perceptron (MLP)

- AdaBoost (AB)
- Gaussian Naive Bayes (GNB)
- Quadratic Discriminant Analysis (QDA)

2.9 Hyperparameter Tuning

It is actually quite critical to set the appropriate hyperparameters in machine learning so that the classifiers' accuracy is improved. Hyperparameter optimization could be quite impactful on the performance of the model. In our work, we took hyperparameter optimization very seriously and applied Randomized Search cross-validation. With a cross-validation number of 5 (CV=5), we were quite effective in searching over the hyperparameter space and keeping one good set of hyperparameters of the selected classifier [44], [45]. That involved searching over ranges of values of critical parameters like the learning rate, number of estimators, and maximum depth, just to mention a few, as described in Table 2.5.

Table 2.5 Hyperparameters Setting Values for Different Classifiers.

Classifier	**Parameter (Values)**
GB	learning_rate (0.01, 0.1, 0.2), n_estimators (100, 200, 300), max_depth (3, 5, 7)
HGB	learning_rate (0.01, 0.1, 0.2), max_depth (3, 5, 7), max_iter (100, 200, 300)
RF	n_estimators (100, 200, 300), max_depth (None, 10, 20)
DT	max_depth (3, 5, 10)
GNB	var_smoothing (1e-9, 1e-8, 1e-7)
MLP	hidden_layer_sizes: ((50,), (100,)), alpha ([0.0001, 0.001])
QDA	n_estimators (50, 100, 200), max_depth (None, 10, 20)
KNN	n_neighbors (3, 5, 7)

2.10 Evaluation Metrics

For comparing the performance of the nine classifiers in a composite fashion, a Receiver Operating Characteristic (ROC) curve analysis is first employed. Because correct medical diagnosis is so critical, we selected ROC AUC, a robust evaluation measure that measures a model's ability to differentiate between positive and negative samples, as our first-choice evaluation measure. This allowed for a comparison of the classifiers' performance without any influence of class skew. Having selected the best classifier, its further evaluation with other measures, Accuracy, Specificity, and Sensitivity, is performed. With the addition of these measures, we got a holistic perspective of the performance of the classifier in terms of overall accuracy as well as its accuracy in the classification of positive and negative samples.

2.11 Computational Environment

The computational infrastructure supporting our analysis consisted of:

- Linux Ubuntu 22.04 LTS
- Python 3.10.13
- ANTs for co-registration
- Freesurfer and FastSurfer for segmentation
- PyRadiomics for feature extraction
- CUDA-powered deep learning using PyTorch
- Hardware: Intel Core i7 Gen10th processor, 16GB DDR5 RAM, RTX2060 GPU (1920 CUDA cores, 240 Tensor Cores)

2.12 Computation Time

Given the complexity of the radiomics workflow, including multi-modal co-registration, segmentation, feature extraction, and classification, the computational demands of this study were significant. All experiments were conducted on a workstation equipped with an Intel Core i7 10th generation CPU, 16 GB DDR5 RAM, and an NVIDIA RTX 2060 GPU with 1920 CUDA cores and 240 Tensor cores, running Ubuntu 22.04 LTS.

2.12.1 Segmentation

For segmentation tasks, FastSurfer was employed due to its GPU-accelerated architecture. Processing a single T1-weighted MRI scan required approximately 2 to 3 minutes, a notable improvement over traditional FreeSurfer, which can take up to 8 hours on CPU-based systems.

2.12.2 Co-registration

Using ANTs, the rigid registration between FDG-PET and MRI scans typically required 5 to 7 minutes per subject. Despite being CPU-based, ANTs provided highly accurate spatial alignment, which is critical for subsequent feature localization.

2.12.3 Feature Extraction

Radiomics features were extracted using PyRadiomics, applied to both whole-brain and subregional masks. This step, implemented in Python 3.10, took about 2 to 3 minutes per subject when extracting approximately 11,400 features (120 features across 95 regions). Processing time varied slightly depending on the size of the masks and I/O latency.

3. RESULTS

Our research presents a comprehensive radiomics-based approach to analyzing FDG-PET images for Alzheimer's disease (AD) diagnosis, progression monitoring, and biomarker discovery. We conducted distinct but interrelated studies, each focusing on a crucial aspect of AD pathology. First, we introduced a radiomics platform capable of automatically identifying key brain regions and extracting high-dimensional features to optimize AD staging and classification. Second, we demonstrated the ability to predict amyloid positivity using metabolic imaging, providing a non-invasive alternative to amyloid PET scans. Third, we investigated the volumetric and textural properties of the hippocampus-amygdala connectivity area, revealing its potential as an imaging biomarker for tracking disease progression. Fourth, we explored the predictive power of radiomics in identifying the presence of the ApoE4 allele, an essential genetic risk factor for AD, further expanding the diagnostic utility of FDG-PET beyond conventional metabolic assessments. Lastly, we refined our radiomics approach by segmenting hippocampal and amygdala subregions, uncovering highly specific imaging biomarkers critical for early diagnosis and staging. Together, these studies contribute to a more interpreTable, explainable, and clinically applicable radiomics framework for AD research.

3.1 AD staging and classification

In this research, we conducted predictions in three distinct categories: AD and CN, MCI and CN, and MCI and AD. To perform these predictions, we independently evaluated each of these three predictions using a substantial number of features extracted during the initial stage, totaling 11,400 features for each image (120 x 95). It's crucial to note that having an excessive number of features can lead to overfitting, which is a common challenge in supervised machine learning[46]. To mitigate this issue, we employed feature reduction techniques aimed at reducing the feature count.

The selection of an appropriate algorithm for feature reduction is a critical decision in the field of machine learning, as highlighted in the reference[47]. Before applying the feature reduction methods to our extracted features, we first preprocessed these features. This preprocessing involved eliminating constant, quasi-constant, and duplicated features. This step is essential for ensuring the quality of the feature set. With our preprocessed feature set, consisting of 5,351 features, we embarked on a comprehensive exploration of various dimension reduction methods to identify the most effective approach for our problem. Subsequently, we calculated the accuracy of each subset of the most significant features using a rank-based method. Throughout this comparison, we adhered to the default parameters as defined in the Scikit-learn library for all classifiers. We refrained from hyperparameter tuning, as our primary focus was on the feature reduction aspect of the machine learning process.

3.1.1 AD vs CN

The outcomes displayed in Table 2.5 reveal that among the various classifiers and feature selection methods evaluated, the Random Forest (RF) classifier with LASSO exhibited the most impressive performance, achieving a good accuracy with an Area Under the Curve (AUC) value of 0.976. This accuracy surpassed the performance of all other combinations of classifiers and feature selection techniques, underscoring the superiority of the RF classifier when applied to features selected via the LASSO method. This finding highlights the effectiveness of this specific combination in achieving high-quality results in our study.

Based on the data presented in Table 3.1, it's evident that the Random Forest (RF) classifier outperformed other classifiers and feature selection methods in terms of accuracy, achieving an impressive AUC of 0.976. This level of accuracy was notably higher than what was attained with alternative combinations of classifiers and feature selection techniques. To reach this result, we followed a specific process.

First, we preprocessed all extracted features (n=11,400) by removing constant,

Table 3.1 Results of performing nine classifiers, GB, RF, DT, GNB, GP, MLP, QDA, AB, and KNN on the top 20 features selected by eight dimension-reduction methods, ANOVA, PCA, Chi-Square, LASSO, MI, RFA, FI, and RFE.

Feature Selection	GB	RF	DT	GNB	GP	MLP	QDA	AB	KNN
ANOVA	0.957	0.962	0.850	0.961	0.910	0.935	0.958	0.951	0.912
PCA	0.920	0.929	0.784	0.919	0.677	0.853	0.918	0.904	0.790
Chi Square	0.936	0.941	0.825	0.947	0.934	0.928	0.942	0.918	0.919
LASSO	0.971	0.976	0.874	0.971	0.914	0.919	0.972	0.967	0.909
MI	0.952	0.951	0.838	0.947	0.906	0.916	0.940	0.939	0.903
RFA	0.968	0.968	0.872	0.970	0.942	0.897	0.966	0.955	0.919
FI	0.949	0.955	0.838	0.955	0.920	0.941	0.952	0.933	0.921
RFE	0.959	0.964	0.853	0.965	0.899	0.939	0.955	0.943	0.927

quasi-constant, and duplicated features and reducing them to 5,351. Next, we applied the LASSO method to the preprocessed feature set. Subsequently, we utilized a rank-based method, which leveraged the coefficients of the features, to calculate the average accuracy across all subsets of the selected top 20 features by LASSO. This calculation was performed through 100 iterations for each subset, using 30% independent test data.

The RF classifier, when applied to features obtained via the LASSO algorithm and subjected to 100 iterations, demonstrated a high area under the curve (AUC) of 0.976, with a 95% confidence interval ranging from 0.95 to 0.98. This signifies the robustness and reliability of the RF classifier when working with these features of the LASSO method.

According to the ranked-based evaluation of all features, after including four features, there is minimal performance improvement. These four Regions of Interest (ROIs) are the isthmus cingulate (left hemisphere), inferior parietal (left hemisphere), hippocampus (left hemisphere), and entorhinal (right hemisphere). The selected features consist of first-order 90 Percentile, first-order Median, glrlm LongRunEmphasis, and gldm DependenceEntropy.

Furthermore, leveraging the four most crucial features, as depicted in Table 6, we evaluated accuracy with a tuned classifier. Once the RF classification method was chosen and its hyperparameters fine-tuned using randomized-search cross-validation (with a 5-fold cross-validation), we achieved an AUC of 0.961 with 100 iterations and a 95% confidence interval spanning from 0.938 to 0.982. These further underscore the efficacy of the RF classifier in producing consistent and high-quality results for our classification model.

3.1.2 AD vs MCI

In the context of distinguishing between AD and MCI, our analysis revealed that the Random Forest (RF) classifier, when applied to features selected through the LASSO method, delivered the highest level of accuracy. This achievement is shown in Table 3.2, where the classifier yielded an AUC of 0.917.

Further refinement of our analysis led us to identify the most critical features, six in total, based on their impact on accuracy. Even with this more concise feature set, the model exhibited commendable accuracy, resulting in an AUC of 0.862.

Table 3.2 Results of performing nine classifiers GB, RF, DT, GNB, GP, MLP, QDA, AB, and KNN on the top 50 features selected by the eight dimension-reduction methods ANOVA, PCA, Chi2, LASSO, MI, RFA, FI, and RFE to predict AD vs MCI.

Feature Selection	**GB**	**RF**	**DT**	**GNB**	**GP**	**MLP**	**QDA**	**AB**	**KNN**
ANOVA	0.815	0.815	0.683	0.843	0.772	0.806	0.818	0.787	0.781
PCA	0.758	0.782	0.625	0.796	0.615	0.744	0.777	0.751	0.693
Chi Square	0.781	0.813	0.668	0.830	0.836	0.812	0.804	0.768	0.796
LASSO	0.862	0.917	0.720	0.889	0.787	0.825	0.857	0.825	0.748
MI	0.843	0.837	0.698	0.828	0.793	0.798	0.815	0.826	0.793
RFA	0.872	0.862	0.724	0.871	0.791	0.799	0.845	0.851	0.782
FI	0.863	0.868	0.716	0.852	0.778	0.822	0.823	0.833	0.770
RFE	0.834	0.829	0.693	0.835	0.798	0.797	0.804	0.805	0.763

To carry out this prediction task for AD vs MCI, we initially applied LASSO to a preprocessed feature set containing 5,351 features. Subsequently, we employed a rank-based approach, focusing on the top 50 features and performing 100 iterations. Within this framework, the RF classifier consistently demonstrated accepTable performance, resulting in an average accuracy with an AUC of 0.917.

To improve the prediction accuracy, we pursued fine-tuning the RF classifier's hyperparameters, which had a positive effect on the results. With these enhancements, the AUC increased to 0.874 with 100 iterations, and we can express our confidence in this outcome with a 95% confidence interval spanning from 0.822 to 0.912. This demonstrates the robustness and reliability of our predictive model in discerning between AD and MCI, providing valuable insights for clinical applications.

3.1.3 CN vs MCI

As indicated by the findings presented in Table 3.3, the Random Forest (RF) classifier consistently demonstrated the highest level of accuracy when applied to features selected through the LASSO method. This achievement was particularly notable during 100 iterations, yielding an AUC of 0.879. Furthermore, we can express our confidence in this result with a 95% confidence interval ranging from 0.747 to 0.911.

In our quest to assess the average accuracy in distinguishing between CN and MCI, we adopted a rank-based approach focusing on the top 80 features, executing 100 iterations for each subset of features. Within this framework, the RF classifier consistently exhibited robust performance, achieving an average accuracy with an AUC of 0.879.

Our analysis further led us to identify the eight most influential features, as depicted in Table 3.4, based on their impact on accuracy. Even with this more refined feature set, the model maintained a commendable level of accuracy, resulting in an AUC of 0.778 using the final model.

Table 3.3 Results of performing nine classifiers GB, RF, DT, GNB, GP, MLP, QDA, AB, and KNN on the top 80 features selected by the eight dimension-reduction methods ANOVA, PCA, Chi2, LASSO, MI, RFA, FI, and RFE to predict CN vs MCI.

Feature Selection	GB	RF	DT	GNB	GP	MLP	QDA	AB	KNN
ANOVA	0.849	0.878	0.635	0.791	0.693	0.635	0.641	0.847	0.695
PCA	0.862	0.875	0.651	0.771	0.596	0.570	0.625	0.854	0.582
Chi Square	0.798	0.809	0.643	0.779	0.658	0.603	0.583	0.785	0.622
LASSO	0.861	0.879	0.631	0.754	0.692	0.684	0.645	0.858	0.695
MI	0.787	0.793	0.635	0.785	0.649	0.659	0.648	0.767	0.646
RFA	0.821	0.827	0.664	0.791	0.554	0.567	0.665	0.792	0.627
FI	0.802	0.814	0.635	0.775	0.600	0.597	0.676	0.786	0.646
RFE	0.787	0.799	0.638	0.791	0.547	0.631	0.681	0.746	0.680

We then embarked on the fine-tuning of the RF classifier's hyperparameters, specifically for the top features (eight in total). This fine-tuning effort yielded an average value for the area under the ROC curve (0.79) after 100 iterations. Our confidence in this result is substantiated by a 95% confidence interval spanning from 0.753 to 0.861. These findings underscore the model's robustness and reliability in distinguishing between CN and MCI, carrying significant implications for clinical applications and research.

To underscore the most efficient Regions of Interest (ROIs) as well as those commonly identified across various prediction scenarios, we present Figure 3.1, illustrating the Common ROIs consistently chosen by eight distinct feature selection techniques. Our examination reveals that the amygdala, entorhinal cortex, and hippocampus emerge as the most recurrently utilized regions among the foremost selected features by diverse feature selection methodologies.

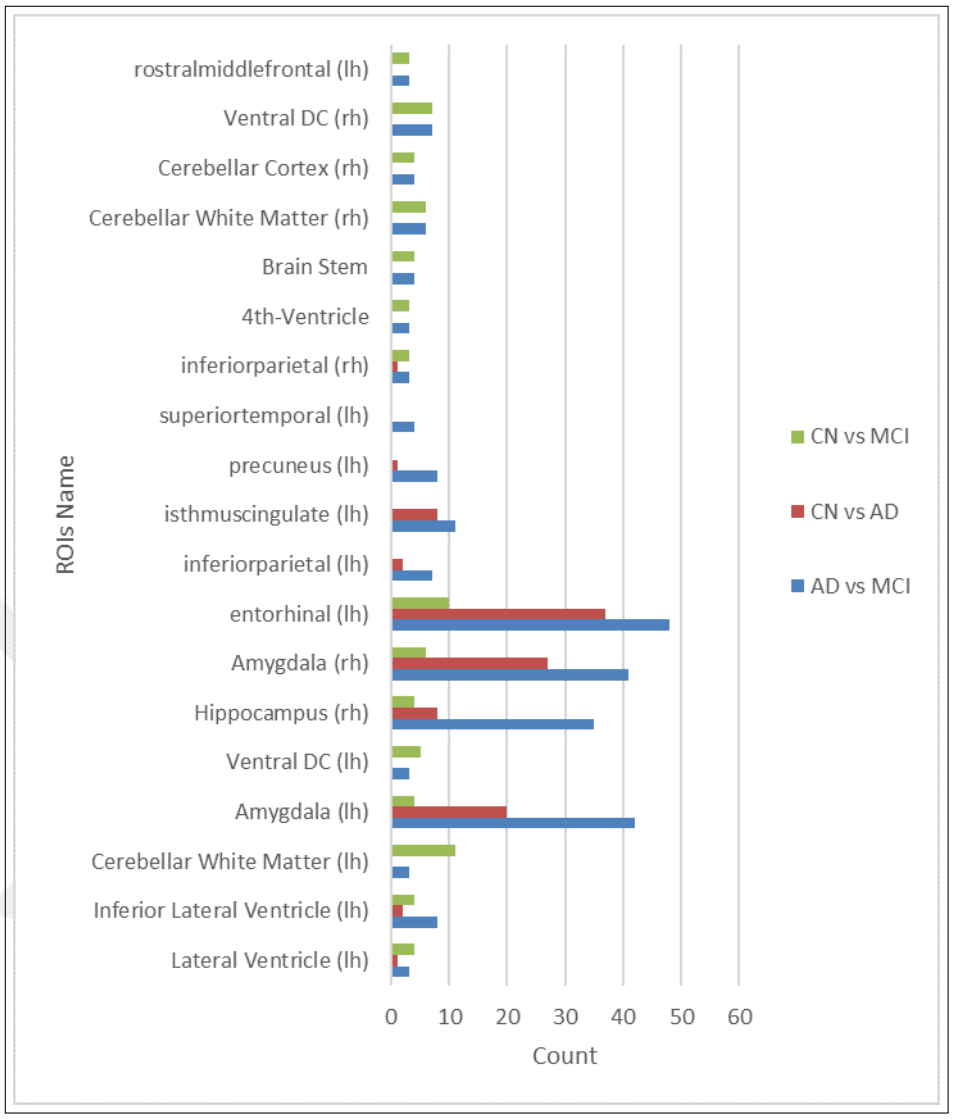

Figure 3.1 Common ROIs are frequently selected by eight different feature selection methods.

Table 3.4 Important Features and ROIs selected by the LASSO method.

Feature	ROI	Group 1		Group 2		P Value
		Mean	Std	Mean	Std	
CN vs AD						
firstorder 90Percentile 1010	isthmuscingulate	1.6166	0.095	1.4252	0.135	1.51E-38
firstorder Median 1008	inferiorparietal	1.2583	0.080	1.1193	0.140	6.77E-24
glrlm LongRunEmphasis 17	Hippocampus	51.153	4.959	39.967	7.042	4.51E-45

Continued on next page

(Continued from previous page)

Feature	ROI	Group 1		Group 2		P Value
		Mean	Std	Mean	Std	
gldm DependenceEntropy 2006	entorhinal	3.9792	0.089	4.1798	0.144	9.80E-39
AD vs MCI						
firstorder 90Percentile 1025	precuneus	1.4949	0.136	1.6106	0.108	8.02E-17
firstorder RootMeanSquared 13	Pallidum	1.1673	0.098	1.0814	0.091	3.29E-16
glrlm RunLengthNonUniformityNormalized 1009	inferiortemporal	0.1665	0.015	0.1529	0.012	1.59E-18
glrlm RunVariance 17	Hippocampus	12.474	2.071	14.372	2.249	1.37E-15
firstorder Median 2022	postcentral	1.2308	0.074	1.1870	0.062	4.66E-09
firstorder 90Percentile 2025	precuneus	1.5076	0.146	1.6092	0.108	1.42E-12
CN vs MCI						
firstorder Maximum 16	Brain Stem	1.4056	0.134	1.4988	0.177	9.51E-09
firstorder 90Percentile 2008	inferiorparietal	1.4771	0.079	1.4344	0.098	3.01E-06
gldm DependenceVariance 2012	lateralorbitofrontal	22.094	1.001	22.442	1.182	1.87E-03
glszm ZonePercentage 1022	postcentral	0.0004	3E-04	0.0004	3E-04	7.48E-01
firstorder Minimum 2035	insula	0.6296	0.075	0.5934	0.081	6.57E-06
shape Maximum 2DDiameter Column 2020	parstriangularis	29.712	2.342	30.471	2.811	3.97E-03
shape Sphericity 28	Ventral DC	0.5551	0.015	0.5571	0.016	2.08E-01
glrlm ShortRunEmphasis 1002	Thalamus	0.2481	0.022	0.2534	0.024	2.21E-02

3.2 Predicting amyloid positivity

In this study, the PyRadiomics tool was used to compute 120 radiomics features for each of the 95 segments extracted from FDG-PET scans. With this approach, a substantial number of 120*95 features were extracted from each scan, which was a high number considering the multiple regions of interest. However, before applying feature reduction methods, the features were preprocessed to improve the speed and quality of feature reduction methods. This was achieved by removing constant and quasi-constant features and identifying and eliminating duplicated features. As a result, the number of features was reduced to 5603, which is significantly lower than the original count of 11400 (120*95) features. This reduction in the number of features enabled faster feature selection methods, some of which are known to be time-consuming and computationally intensive.

Table 3.5 Accuracy (AUC) comparison of eight classifiers (GB, RF, KNN, GP, AB, GNB, MLP, and DT) applied to features selected by eight feature selection methods (RFA, RFE, LASSO, Chi-Square, ANOVA, FI, PCA, and MI).

Classifier	RFA	LASSO	RFE	Chi2	MI	ANOVA	PCA	FI
KNN	0.783	0.855	0.836	0.798	0.794	0.779	0.646	0.736
RF	0.889	0.916	0.862	0.850	0.873	0.828	0.732	0.852
MLP	0.687	0.788	0.842	0.834	0.834	0.761	0.594	0.761
GB	0.875	0.921	0.838	0.829	0.850	0.808	0.712	0.841
AB	0.846	0.907	0.811	0.812	0.840	0.796	0.689	0.799
DT	0.734	0.769	0.712	0.707	0.717	0.696	0.600	0.708
GP	0.796	0.877	0.840	0.839	0.836	0.828	0.684	0.792
GNB	0.876	0.924	0.845	0.819	0.866	0.849	0.776	0.846

Overfitting is a frequently encountered issue in machine learning, whereby the model fits the training data excessively well and exhibits poor performance when presented with new, unfamiliar data. To address this problem, several techniques can be employed, such as cross-validation, regularization, early stopping, ensemble methods, feature selection, and data augmentation. Combining these methods can effectively mitigate the impact of overfitting and enhance the model's ability to generalize to new data. Notably, cross-validation and feature selection are two techniques that can

be particularly effective in addressing the overfitting issues we used in our proposed method.

Selecting the right algorithm to decrease the number of features is crucial in machine learning. To find a high-performing algorithm, we applied eight feature selection techniques to the extracted features. These methods were chosen from various categories, such as wrapped (RFA and RFE), filtered (Chi-Square and MI), embedded (LASSO), and hybrid (FI+RF) techniques. We also utilized Principal Component Analysis (PCA), which is not strictly a feature selection technique but transforms the original features into principal components that capture the most important information in the data, leading to efficient dimensionality reduction while retaining most of the relevant information. Additionally, we employed ANOVA, a statistical method that serves as a filtered method for feature selection.

In Table 3.5, we present the results of using eight feature selection methods in combination with eight classification methods, which resulted in a total of 64 outcomes based on the AUC metric. To determine the best combination, we proceeded by applying each of the eight classification methods to the top 20 most important features identified by each of the eight feature selection techniques. We employed a ranked-based method to evaluate the entire subset of 20 features and identify the optimal set of features for our research objective. The entire process was repeated 100 times, with each replication conducted under completely random conditions for each subset of features. In each replication, 70% of the data was randomly selected for training and validation purposes, while the remaining 30% was reserved for testing. To maintain a balanced division based on amyloid status, we took care to ensure that the division was equal. Ultimately, we selected the highest AUC score obtained from the ranked-based method as the final result for the selected combination of classifier and feature selection technique.

The GNB classifier Eq. (3.1), as a straightforward and efficient method in conjunction with the LASSO method, exhibits the highest accuracy, with an average AUC value of 0.924, as shown in Table 10. While the default parameters defined in the

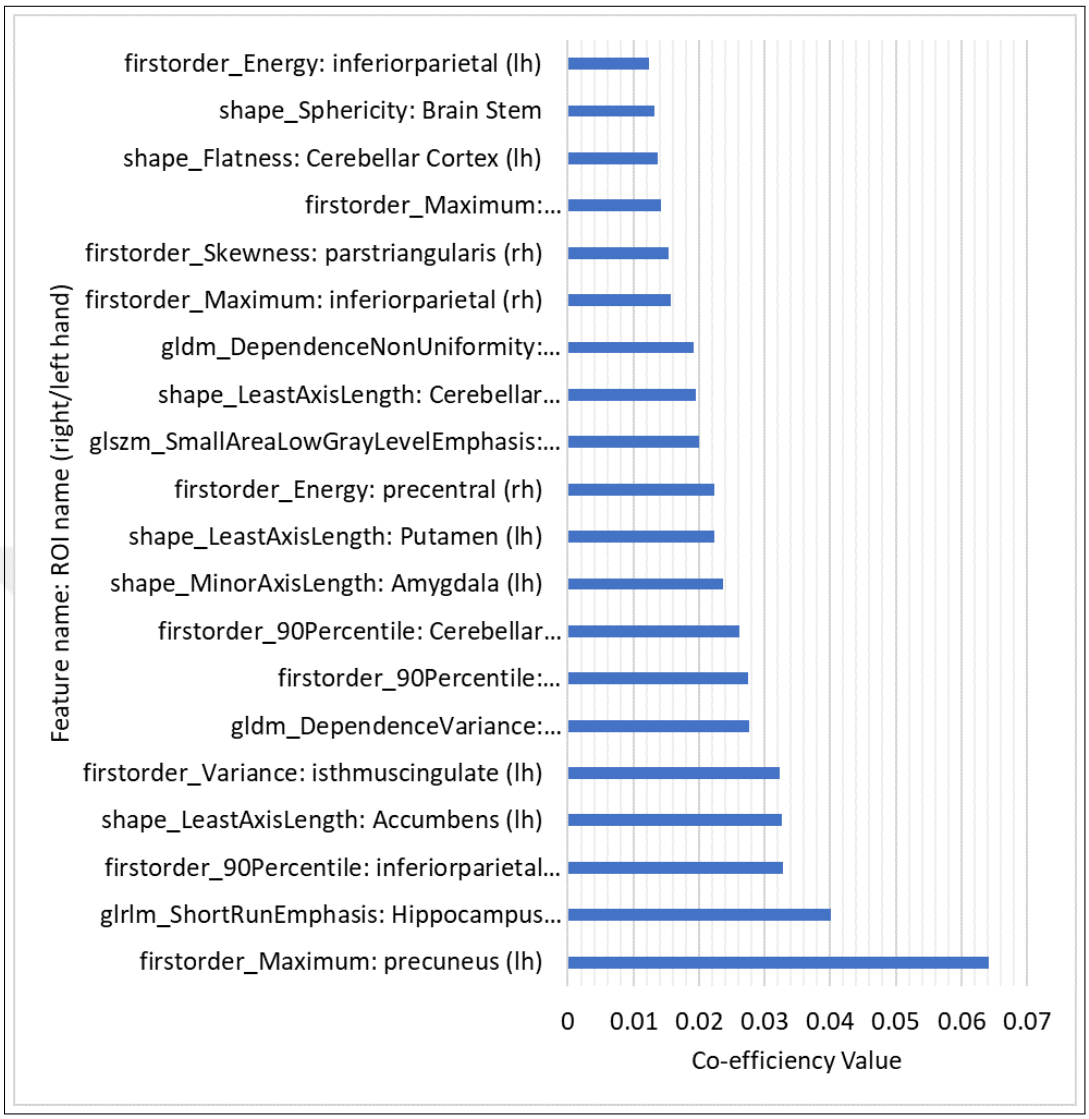

Figure 3.2 Selection of the highest coefficient values from the LASSO method to identify the top 20 features from regions of interest (ROIs).

Scikit-learn library were utilized for all classifiers without hyper-parameter optimization, it is important to note that tuning the classifier can potentially enhance accuracy. To evaluate the methods under the same conditions and determine the best combination of classifier and feature selection method, we performed the Gaussian Naive Bayes (GNB) classifier on the most important features (n=20) obtained through the LASSO method, as depicted in Figure 3.2. The feature selection was conducted using a rank-based method with 100 iterations for each step. We achieved an average AUC value of 0.924, with a 95% confidence interval ranging from 0.918 to 0.929, by utilizing the 18 most important features selected by the LASSO method based on their highest

coefficients.

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \tag{3.1}$$

To evaluate the accuracy of the Gaussian Naive Bayes (GNB) classifier, we utilized the rank-based method on sequentially chosen subsets of the top 20 features. These features were selected through the LASSO method based on their coefficient values. In order to set the hyperparameters of the GNB classifier, we employed Randomized Search cross-validation (CV=5) on training data (70%). The remaining 30% of the data was excluded and kept for testing purposes. After tuning the hyperparameters of the GNB classifier on LASSO features, we set the value of 1×10^{-9} for the var_smoothing parameter of the classifier.

To make the results more understandable and meaningful for medical experts and radiologists, we thoroughly examined the prominent features and regions of interest that were consistently identified by all feature selection methods. Throughout this analysis, we standardized the importance values of features from each method on a scale of 0 to 1. We then assessed their collective importance by aggregating these values. As a result, we labeled the chart exclusively with Regions of Interest (ROIs), omitting feature names, to depict the most crucial ROIs (Figure 3.3).

$$SRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j|\theta)}{i^2 j^2}}{N_r(\theta)} \tag{3.2}$$

$$rMAD = \frac{1}{N_{10-90}} \sum_{i=1}^{N_{10-90}} |X_{10-90}(i) - \bar{X}_{10-90}| \tag{3.3}$$

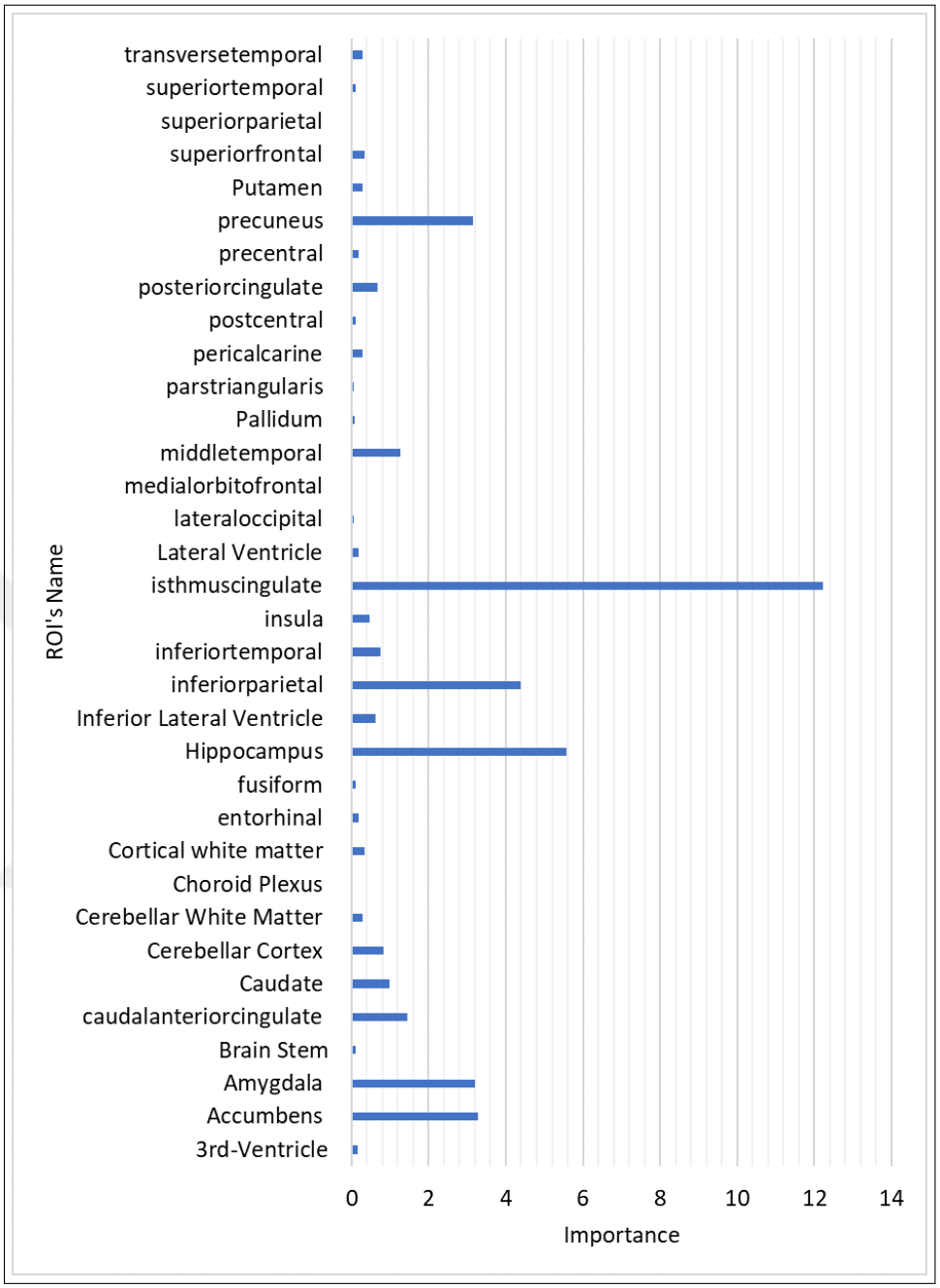

Figure 3.3 Commonly utilized regions of interest (ROIs) determined by eight distinct feature selection methods. The importance value represents the cumulative coefficient values for features within the selected ROIs across all feature selection methods.

Our analysis revealed three features of significance for achieving an acceptable level of accuracy. These features include SRLGLE Eq. (3.2), which quantifies the joint distribution of shorter run lengths with lower gray-level values. Another feature

is rMAD Eq. (3.3), which calculates the mean distance of all intensity values from the Mean Value obtained from a subset of the image array, specifically the gray levels between or equal to the 10th and 90th percentiles. The third feature, 90Percentile, is extracted from distinct regions of interest such as the Hippocampus, inferior parietal, and isthmus cingulate, representing the intensity value at the 90th percentile within each region. Subsequently, we employed a rank-based method and conducted the GNB classifier using the aforementioned three features. This process was repeated 100 times. For each iteration, we employed Randomized Search cross-validation (CV=5) on training and validation data (70%). The remaining 30% of the data was kept as independent test data. The results showed an average AUC value of 0.853, with a 95% confidence interval ranging from 0.850 to 0.863, using the three selected features.

Subsequently, we employed a rank-based method and conducted the GNB classifier using the aforementioned three features. This process was repeated 100 times. For each iteration, we employed Randomized Search cross-validation (CV=5) on training and validation data (70%). The remaining 30% of the data was kept as independent test data. The results showed an average AUC value of 0.853, with a 95% confidence interval ranging from 0.850 to 0.863, using the three selected features.

3.3 Hippocampus-amygdala connectivity

Our study focused on identifying key brain regions associated with AD. Initially, we segmented each brain FDG PET image into 95 regions based on the DKT atlas. Then, we extracted the 120 radiomic features for each of the 95 regions and applied eight different feature selection methods to identify the frequently selected features of the regions. These methods helped us identify the most informative regions that best differentiate between healthy and AD participants.

Analysis of the features revealed that the Hippocampus and Amygdala were frequently selected by the eight different feature selection methods, suggesting their significance in distinguishing AD from CN (Figure 3.4). This finding aligns well with

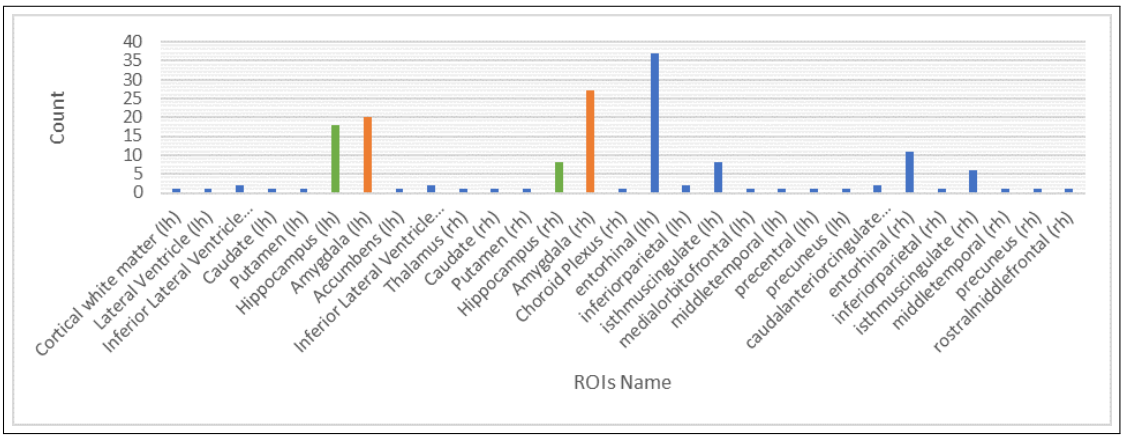

Figure 3.4 Overlapping Regions of Interest (ROIs) identified by eight feature selection algorithms.

established knowledge in AD research. Both the Hippocampus and Amygdala are critically involved in memory and emotional processing, and their atrophy is a well-documented characteristic of AD [48].

Interestingly, the Hippocampus and Amygdala are not only functionally important but also closely connected. They form a network with strong communication pathways, which allows them to work together in various cognitive processes [49]. The hippocampus, often visualized as resembling a seahorse, is a critical memory center in the brain. However, it's not a single, homogenous structure. It comprises distinct subfields, each with specialized functions. Studies have shown that the head, or anterior, portion of the hippocampus, specifically the CA1, CA3, and CA4 subfields, is particularly vulnerable in AD [50]. These subfields are involved in memory consolidation, a process where short-term memories are converted into long-term ones. Damage to these subfields disrupts this process, leading to memory problems that are characteristic of AD.

Furthermore, the amygdala, located deep within the temporal lobe, is often described as an almond-shaped structure. It plays a vital role in processing emotions, particularly fear and anxiety. However, the amygdala isn't a single unit either. It consists of multiple nuclei, each with distinct functions. Research suggests that specific nuclei within the amygdala, especially those involved in emotional processing and

memory formation, are more susceptible to AD pathology [51]. Damage to these nuclei can contribute to the emotional changes and behavioral disturbances sometimes observed in AD patients.

So, in our investigation of Alzheimer's disease diagnostic markers, we focused on the hippocampus and amygdala, regions critically implicated in the disease process. To capture potential information beyond the core structures, we defined a new region (CR) encompassing the parts of both the hippocampus and amygdala using the distance transform method as shown in Figure 3.5. This method calculates the distance of each voxel (3D pixel) within the brain image to the nearest boundary of the segmented hippocampus and amygdala. By setting a thickness of 5 voxels, we ensured the CR incorporated not only the parts of the core regions of the hippocampus and amygdala but also some surrounding tissue, potentially containing valuable information about tissue health and disease progression.

Following the CR definition, we extracted 120 features characterizing various tissue properties within the region. These features likely included measures of intensity, texture, and shape, providing a comprehensive profile of the tissue microstructure. To identify the most informative features for AD interpretation, we used several feature selection methods to identify the most informative features for our model.

$$SMVol = \sum_{i=1}^{N_f} \frac{a_i \cdot (Ob_i \times Oc_i)}{6} \tag{3.4}$$

We employed the ANOVA method to select the top 20 features with the strongest relationship to the target variable, used the Chi-square test to identify the top 20 features based on their independence from the target variable, and applied the MI method to rank features by their mutual dependency with the target variable. We also employed RFA with a Gradient Boosting Classifier to incrementally add features based on their contribution to model performance. Additionally, we assessed FI using

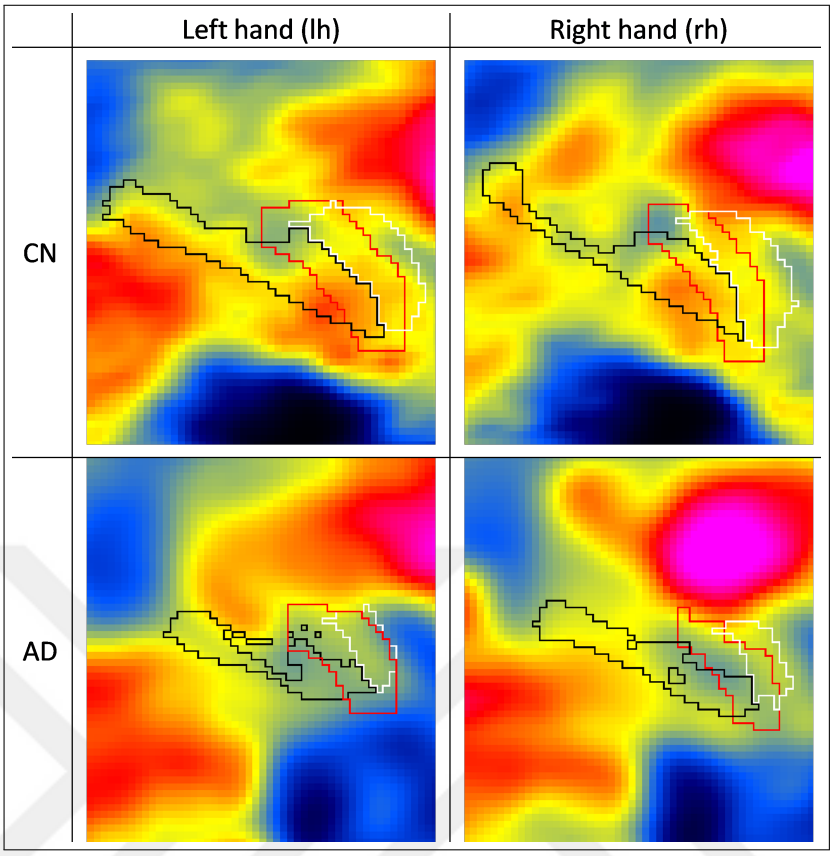

Figure 3.5 Illustration of the result of selecting the border region of the hippocampus (black) and the amygdala (white) using the distance transform method. The red region represents the CR between these structures.

a Random Forest Classifier to determine each feature's contribution. RFE recursively removed the least important features, also using a Random Forest Classifier. PCA was used to reduce dimensionality by transforming features into orthogonal components that capture the maximum variance. Finally, we utilized the LASSO method, applying a range of alpha values from 0.001 to 0.1 in increments of 0.001. Each method provided a unique perspective, ensuring a robust feature selection process.

$$SDLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{p(i,j)}{i^2 j^2}}{N_z} \tag{3.5}$$

Our analysis identified two pivotal features for differentiating between CN and

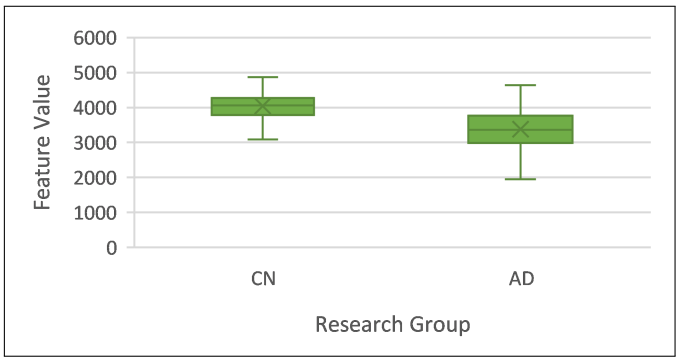

Figure 3.6 Analysis of SMVol Feature average value for two different research groups CN vs AD.

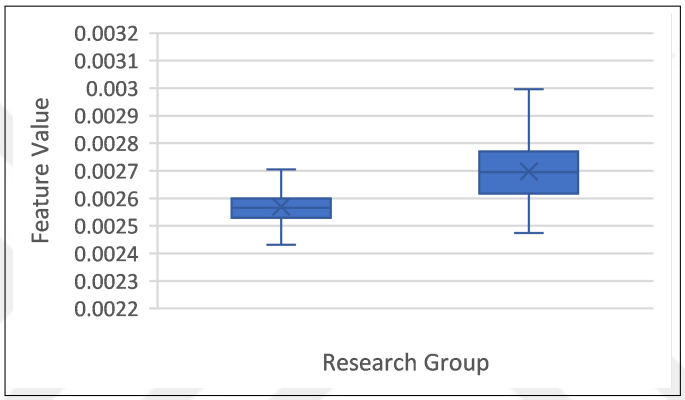

Figure 3.7 Analysis of gldm_SDLGLE Feature average value for two different research groups CN vs AD.

AD: gldm SDLGLE (gldm Small Dependence Low Gray Level Emphasis) (Eq. 3.5) and SMVol (shape MeshVolume) (Eq. 3.4). The gldm SDLGLE feature measures the emphasis on small dependencies with low gray levels within the tissue, indicating texture homogeneity and fine textures. Such texture features are essential for characterizing tissue heterogeneity, which is often linked to tissue behavior. The SMVol feature quantifies the tissue's volume, providing crucial information on its size and spatial extent. Volumetric features are significant in understanding conditions like atrophy in AD patients, as they reflect changes in tissue volume and structural integrity. These features likely capture specific characteristics of the CR region that undergo significant alterations in AD pathology, as shown in Figures 3.6 and 3.7.

When applying the HistGradientBoosting Classifier to a 30% independent test dataset, the analysis achieved an acceptable level of accuracy in differentiating between

CN brain tissue and AD using two features, SMVol and gldm SDLGLE. This is evident from the average ROC AUC of 0.88, indicating well-discriminative capability (Figure 3.8). The model also performed well in identifying both CN and AD cases, with an average sensitivity of 0.73 and an average specificity of 0.82 for CN versus AD. These results demonstrate the effectiveness of the classifier in distinguishing AD-related tissue changes based on the chosen features. Additionally, ROC curve analysis revealed consistent performance across all 100 iterations, with a mean AUC of 0.88 and a narrow confidence interval of 0.86-0.89, supporting the reliability of these findings. In addition, to assess the individual contribution of each feature, we calculated metrics (AUC, accuracy, sensitivity, and specificity) for SMVol and gldm SDLGLE separately.

Table 3.6 Influence of feature combination in distinguishing CN and AD patients using SMVol and gldm_SDLGLE.

Features	**ROC AUC**	**Accuracy**	**Specificity**	**Sensitivity**
SMVol	0.782	0.695	0.709	0.680
gldm_SDLGLE	0.787	0.688	0.730	0.642
SMVol and gldm_SDLGLE	0.878	0.777	0.814	0.736

The results, presented in Table 3.6, demonstrate how different feature combinations can influence the model's performance. Analyzing these results alongside the chosen features' characteristics (e.g., size and texture measurements) can provide valuable insights into the most informative characteristics for AD diagnosis.

We also analyzed the volume of the CR, finding variations between groups. This suggests potential atrophy or shrinkage in AD patients. Shape_MeshVolume, a more advanced measurement feature, can help quantify this shrinkage precisely. Atrophy in the hippocampus and amygdala is a well-established feature of AD and is likely due to the progressive neurodegeneration characteristic of the disease. By measuring the volume and shape changes of the CR, we can gain further insights into the severity and progression of AD.

Table 3.6 explores the predictive potential of the CR region in AD by analyzing brain metabolisms and volumes across distinct groups: CN, AD patients, and individuals at various stages of mild cognitive impairment (EMCI, MCI, LMCI). The findings

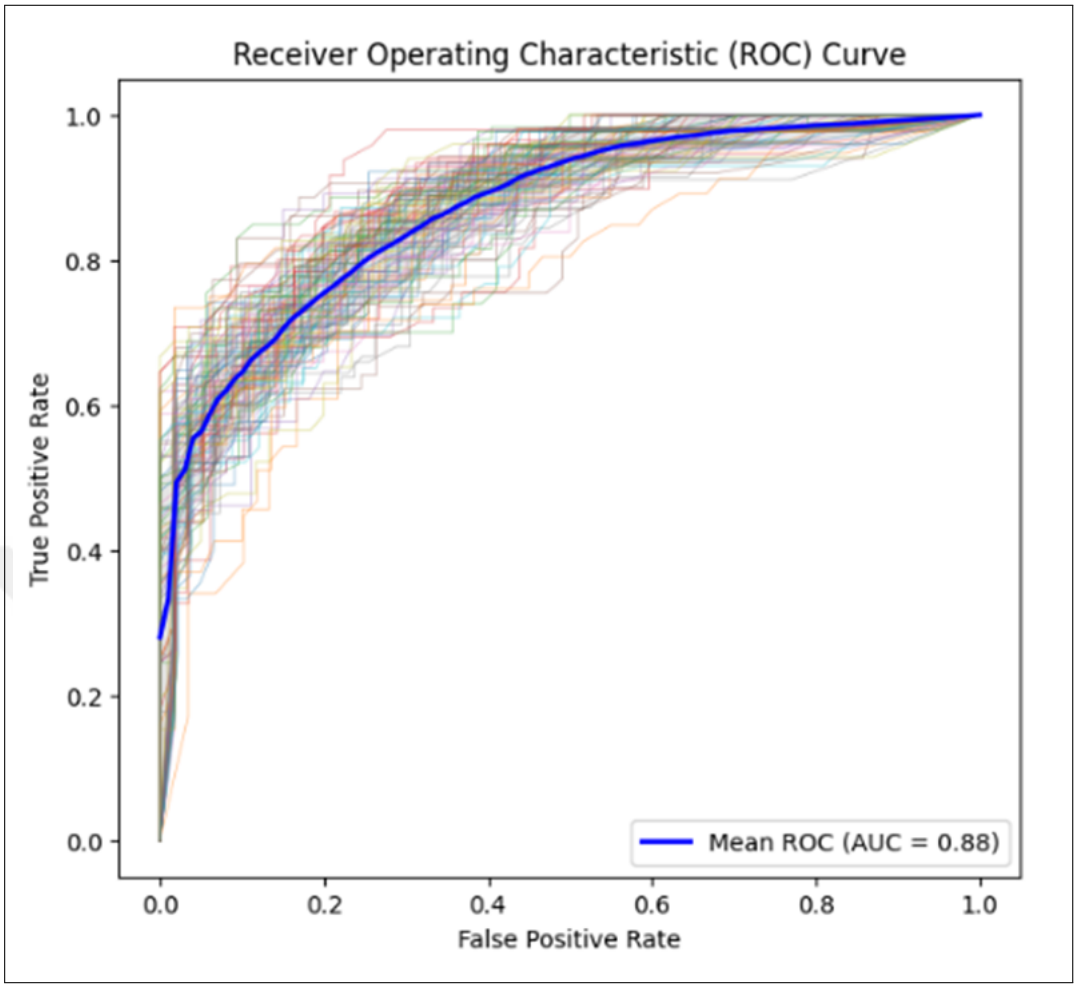

Figure 3.8 Performance of the Model in Distinguishing CN and AD Patients using SMVol and GLDM SDLGLE Features.

highlight a statistically significant decrease in CR region volume (SMVol) among AD patients compared to CN individuals (p-value = 1.03E-37), underscoring its potential relevance in AD pathogenesis. Moreover, Table 3.7 presents the result on SDLGLE, detailing variations in mean values and standard deviations across these groups. These observations suggest a possible association of SDLGLE with the progression of AD, prompting further in-depth investigation into its role as a biomarker or indicator of disease severity in addition to SMVol.

Table 3.7 Means and p-values for comparisons between groups based on SMVol and GLRLM features.

Groups	**SMVol**			**SDLGLE**		
	Mean	**SD**	**P-Value**	**Mean**	**SD**	**P-Value**
CN	4045.91	376.83	1.36×10^{-31}	0.002 569	0.000 056	1.03×10^{-37}

Continued on next page

(Continued)

Groups	SMVol			SDLGLE		
	Mean	SD	P-Value	Mean	SD	P-Value
AD	3376.12	544.80		0.002 697	0.000 098	
CN	4045.91	376.83	2.24×10^{-1}	0.002 569	0.000 056	1.07×10^{-2}
EMCI	3999.96	372.05		0.002 585	0.000 068	
CN	4045.91	376.83	5.31×10^{-3}	0.002 569	0.000 056	1.58×10^{-7}
MCI	3918.00	478.49		0.002 607	0.000 077	
CN	4045.91	376.83	3.47×10^{-9}	0.002 569	0.000 056	1.32×10^{-11}
LMCI	3736.80	497.18		0.002 632	0.000 097	
AD	3376.12	544.80	3.69×10^{-33}	0.002 697	0.000 098	1.21×10^{-32}
EMCI	3999.96	372.05		0.002 585	0.000 068	
AD	3376.12	544.80	6.22×10^{-20}	0.002 697	0.000 098	6.03×10^{-19}
MCI	3918.00	478.49		0.002 607	0.000 077	
AD	3376.12	544.80	3.40×10^{-8}	0.002 697	0.000 098	8.20×10^{-8}
LMCI	3736.80	497.18		0.002 632	0.000 097	
EMCI	3999.96	372.05	5.12×10^{-2}	0.002 585	0.000 068	2.22×10^{-3}
MCI	3918.00	478.486		0.002 607	0.000 077	
EMCI	3999.96	372.05	4.66×10^{-8}	0.002 585	0.000 068	2.58×10^{-7}
LMCI	3736.80	497.182		0.002 632	0.000 097	
MCI	3918.00	478.486	1.48×10^{-3}	0.002 607	0.000 077	1.33×10^{-2}
LMCI	3736.80	497.182		0.002 632	0.000 097	

3.4 Predicting the presence of the ApoE4 allele

To evaluate the impact of combining different feature selection methods and classifiers, we selected three methods, such as ANOVA, PCA, and LASSO, from distinct categories and applied them to the preprocessed features, which are more detailed in the next section. We then assessed the features selected by these methods using nine classifiers: AB, DT, GNB, GP, GB, HGB, MLP, QDA, and RF. After evaluating 27 different combinations, we found that the HGB classifier performed best when combined with LASSO, as shown in Table 3.8.

Table 3.8 Classification results for discriminating between MCI patients with and without APOE4 using various feature selection methods (ANOVA, PCA, and LASSO) in combination with nine different classifiers (AB, DT, GNB, GP, GB, HGB, MLP, QDA, RF).

Classifier		AB	DT	GNB	GP	GB	HGB	MLP	QDA	RF
ANOVA	**AUC**	0.77	0.72	0.81	0.79	0.78	0.77	0.87	0.80	0.81
	ACC	0.72	0.69	0.74	0.71	0.72	0.73	0.79	0.67	0.76
PCA	**AUC**	0.74	0.67	0.62	0.61	0.72	0.63	0.55	0.62	0.68
	ACC	0.63	0.65	0.55	0.62	0.66	0.56	0.51	0.51	0.62
LASSO	**AUC**	0.92	0.78	0.89	0.87	0.91	0.95	0.94	0.84	0.92
	ACC	0.87	0.76	0.83	0.85	0.88	0.88	0.87	0.73	0.86

3.4.1 Feature Extraction: Preprocessing with LASSO

To address the high-dimensional nature of the dataset, which included 14,760 features (95 regions of interest * 120 radiomic features for the whole brain, 19 hippocampal subfields * 120 features, and 9 amygdala subnuclei * 120 features), we implemented a multi-step feature selection process to reduce the risk of overfitting during the classification stage.

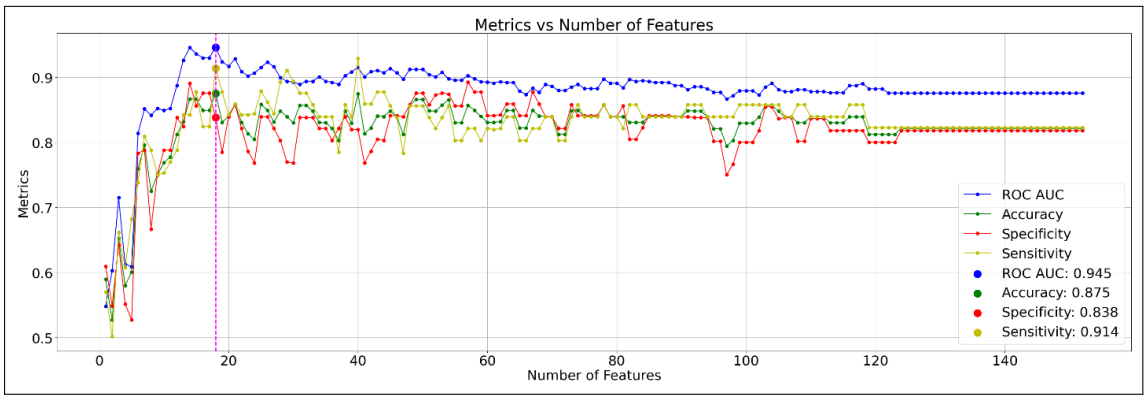

Figure 3.9 Rank-based performance of a classification model using features selected by LASSO. The plot illustrates the evolution of four key metrics (ROC AUC, Accuracy, Specificity, and Sensitivity) as the number of features increases, determined by the LASSO feature selection method. The optimal number of features is marked, corresponding to the point with the highest ROC AUC value.

The initial step involved data pre-processing to eliminate irrelevant features like constants, quasi-constants, and duplicates. This initial cleaning reduced the feature set to a more manageable size of 6,278. Following pre-processing, according to the obtained

results in Table 3.8, we employed the LASSO method for further feature selection. LASSO works by shrinking coefficient values toward zero, effectively removing features with minimal predictive power. This technique significantly reduced the feature set to a more focused set of 152 features likely to be the most informative for our classification task. The LASSO implementation involved data scaling, a technique to ensure all features are on a similar scale, and setting a high maximum number of iterations (10,000) to guarantee the model converges on a solution. Finally, we optimized the model parameters using grid search with a range of alpha values (a key parameter in LASSO) logarithmically spaced between 10^-4 and 10^2 and 10-fold cross-validation with ROC AUC scoring to evaluate model performance.

3.4.2 Classification

Building on the feature selection process, we used the HGB classifier for classification based on the results presented in Table 3.8. The HGB classifier is a powerful ensemble learning method that iteratively builds decision trees, focusing on improving classification for previously misclassified instances. This approach allows the model to learn from its mistakes and refine its predictions over time. The HGB classifier utilized the selected set of 152 features by LASSO for classification. Notably, we employed a rank-based scoring method that emphasized features with higher absolute LASSO coefficients. These features were likely identified as having a stronger association with the target variable (ApoE4 genotype) during the feature selection step. This approach of feature selection and HGB classification achieved excellent results in differentiating MCI patients with homozygous ApoE4 carriers and non-ϵ4 carriers. The model demonstrated excellent discrimination between the two patient groups, with an average Area Under the ROC Curve (AUC) of 0.945. The 95% confidence interval for the AUC ranged from 0.868 to 0.926, highlighting its strong predictive performance (Figure 3.9).

Additionally, the model achieved an average accuracy of 0.875, signifying a high proportion of patients were correctly classified. Furthermore, the model demonstrated

exceptional sensitivity (0.914) for detecting individuals with homozygous ApoE4, highlighting its effectiveness in identifying this specific patient group. The model also maintained good specificity (0.838) in correctly identifying MCI patients without the ApoE4 genotype. It's important to note that achieving this level of accuracy with only 18 features selected for their highest coefficients highlights the model's efficiency and its potential for clinical applications. This suggests that a smaller set of highly informative features can be sufficient for accurate classification, leading to a more streamlined and potentially more interpretable model.

We computed Pearson correlation coefficients to assess the linear relationships between these top features. A high correlation value (positive or negative) indicated a significant overlap in information between features. Utilizing a correlation matrix, we removed features with high correlations to avoid redundancy. We focused on the absolute value of the correlation coefficient to identify the strength of the relationship, regardless of direction. Self-correlations were filtered out, and only the pairs with the weakest correlations were retained.

This process led to the selection of seven features from the top 18 LASSO features, categorized into three distinct groups: glszm (SmallAreaEmphasis, ZoneVariance) related to Gray Level Size Zone Matrix, shape (SurfaceVolumeRatio, Sphericity, Maximum2DDiameterColumn, SurfaceArea) related to shape characteristics, and glrlm (RunEntropy) related to Gray Level Run Length Matrix.

These seven key features were identified in specific brain regions of the left hemisphere, including the hippocampus, entorhinal cortex, amygdala, thalamus, and pars orbitalis. This approach ensured that the final set of features was informative and non-redundant, providing a solid basis for further investigating the link between ApoE4 and MCI.

Table 3.9 Comparison of Features Between Homozygous ApoE4 Carriers and Non-ϵ4 Carriers

FS ID	Region (left/right hand)	Feature	homozygous ApoE4 carriers				non-E4 carriers				P_Value
			mean	SD	min	max	mean	SD	min	max	
238	CA1-body (Hippocampus) (lh)	glszm_SmallAreaEmphasis*	0.3000	1.2963	-0.3682	3.6735	-0.3000	0.4157	-0.3681	2.6631	0.0016
2017	paracentral (rh)	girlm_RunVariauce	-0.1289	0.9366	-2.3246	2.3972	0.1289	1.0608	-2.0647	3.1278	0.1756
1006	entorhinal (lh)	glszm_ZoneVariance*	0.2701	1.1090	-0.6391	2.7400	-0.2701	0.8111	-0.6391	3.4895	0.0040
1011	lateraloccipital (lh)	shape_Flatness	0.1819	1.0437	-1.8295	3.0743	-0.1819	0.9379	-2.7474	2.1383	0.0549
10	Thalamus (lh)	shape_SurfaceVolumeRatio*	0.2874	1.0451	-1.8291	2.9666	-0.2874	0.8809	-2.8339	2.2949	0.0021
245	molecular_layer_HP-head (Hippocampus) (lh)	shape_Flatness	-0.2439	0.7830	-1.8315	2.1238	0.2439	1.1411	-1.7182	5.4304	0.0097
53	Hippocampus (rh)	firstorder_90Percentile	0.2082	0.9508	-2.2811	2.8583	-0.2082	1.0218	-2.1932	2.3466	0.0276
244	GC-ML-DG-body (Hippocampus) (lh)	girlm_RunEntropy	0.1656	1.0613	-2.0553	2.6429	-0.1656	0.9242	-2.1335	1.5815	0.0811
235	subiculum-head (Hippocampus) (lh)	shape_Flatness	-0.1804	0.9948	-2.0349	2.6071	0.1804	0.9902	-1.8604	2.5275	0.0571
226	Hippocampal_tail (Hippocampus) (lh)	glszm_SizeZoneNonUniformity	0.1602	1.2777	-0.3627	6.7370	-0.1602	0.5980	-0.3627	1.6094	0.0926
5	Inferior Lateral Ventricle (lh)	gldm_DependenceVariance	0.2497	0.8670	-3.1855	1.8362	-0.2497	1.0760	-4.4154	1.2877	0.0080
10	Thalamus (lh)	firstorder_Energy	-0.1668	0.9803	-1.8968	2.1634	0.1668	1.0093	-3.4538	2.1225	0.0789
7005	Central-nucleus (Amygdala) (lh)	girlm_RunEntropy*	0.3171	0.9355	-1.8487	2.4870	-0.3171	0.9780	-2.3047	1.4931	0.0007
226	Hippocampal_tail (Hippocampus) (lh)	shape_Sphericity*	0.1124	1.0234	-2.0888	1.7519	-0.1124	0.9815	-2.4489	1.6912	0.2379
7008	Accessory-Basal-nucleus (Amygdala) (lh)	shape_Maximum2DDiameterColumn*	0.2723	1.0032	-1.8863	2.2889	-0.2723	0.9375	-1.8863	2.2889	0.0037
2029	superiorparietal (rh)	firstorder_Minimum	0.1898	0.9722	-1.7638	2.2018	-0.1898	1.0089	-1.9986	2.3255	0.0451
1019	parsorbitalis (lh)	shape_SurfaceArea*	-0.2675	0.9132	-2.3284	1.4640	0.2675	1.0280	-2.1684	2.6100	0.0044
46	Cerebellar White Matter (rh)	glszm_SizeZoneNonUniformityNormalized	-0.2310	0.8696	-1.4481	1.6651	0.2310	1.0824	-1.4481	1.6651	0.0144

3.4.3 Classification with Reduced Feature Set

Leveraging these seven selected features, we built a classification model using the HGB model. To ensure model reliability and generalizability, we employed a Stratified

five-fold cross-validation. Hyperparameter tuning with grid search was performed to identify the optimal configuration for the HGB model. This involved searching through a range of specified parameters, including learning rates of 0.01, 0.1, and 0.2; maximum tree depths of 3, 5, and 7; training iterations of 100, 200, and 300; and minimum samples per leaf node of 20, 50, and 100. The tuning aimed to maximize the model's performance based on the ROC AUC score, a key metric for evaluating binary classifiers in medical applications.

Using a grid search to determine the optimal hyperparameters, the HistGradientBoosting (HGB) classifier was evaluated with five-fold stratified cross-validation. The model achieved an average performance with an ROC AUC of 0.889, accuracy of 0.796, specificity of 0.768, and sensitivity of 0.821, with a 95% confidence interval between 0.868 and 0.926.

The proposed model, trained on a reduced feature set of seven variables (Table 14, *), demonstrated robust and balanced classification performance for the ApoE4 genotype ($p < 0.05$). Notably, this model effectively discriminated between individuals with and without the homozygous ApoE4 genotype despite the small number of input features.

3.4.4 Demographic and Clinical Characteristics

The statistical test results presented in Table 3.10 revealed that there are no significant differences between the two groups (APOE4 carriers and non-carriers) for the demographic and clinical characteristics tested.

The t-tests for Age, MMSE, CDRSB, ADAS13, and RAVLT Forgetting all resulted in p-values greater than 0.05, indicating that these continuous variables do not differ significantly between the two groups. The t-test for RAVLT Forgetting has a p-value of 0.0737, which is marginally close to significance but still above the commonly used threshold of 0.05. Overall, the analysis indicates that there are no substantial de-

Table 3.10 T-Test and Pearson's Chi-Square Test Results.

T-Test		
Variable	**t-statistic**	**p-value**
Age	1.469806	0.144469
MMSE	0.477482	0.633967
CDRSB	0.558488	0.577646
ADAS13	-1.191564	0.236020
RAVLT_forgetting	-1.805812	0.073708
Pearson's Chi-Square Test		
Variable	**Statistic**	**p-value**
Gender	0.036458	0.848572

mographic or clinical differences between the two groups based on the tested variables.

3.5 Subregional Biomarkers of hippocampus and amygdala

Given the interdependence between feature selection methods and classifier performance, we evaluated multiple combinations of feature selection techniques, including ANOVA, PCA, LASSO, Chi-square, and classifiers including GB, RF, AB, DT, GNB, GP, MLP, QDA, and KNN to optimize diagnostic accuracy. Specifically, we applied these feature selection methods to radiomic features from two diagnostic groups: CN vs. AD and MCI vs. AD. Following feature reduction, each feature subset was tested across nine distinct classifiers. Classifier optimization was performed using grid search (cv=3) to fine-tune the hyperparameters. We employed Stratified K-fold cross-validation (k=5) to split the dataset into balanced folds, ensuring each fold maintained the same proportion of AD, MCI, and CN participants as in the original dataset. Each classifier was trained and validated five times, once on each fold, with a fold held out for testing and the remaining four used for training, thus maximizing the stability and generalizability of model performance across the dataset.

We assessed model performance based on ROC AUC and accuracy, presented in Table 3.11, to determine the optimal feature selection and classifier combinations. For distinguishing CN from AD, combining the MLP classifier with the LASSO feature

selection yielded the best ROC AUC scores of 0.957 and an accuracy of 0.901. In distinguishing MCI from AD, the MLP classifier paired with LASSO achieved the best AUC score of 0.867 and an accuracy of 0.77.

We analyzed the top-performing models to investigate further the underlying radiomic features and subregions contributing to model performance. This analysis focused on the feature subsets and classifiers that were most effective for each comparison group (CN vs. AD and MCI vs. AD), enabling a detailed exploration of potential imaging biomarkers within hippocampal and amygdala subregions.

Table 3.11 Classification Performance of Different Feature Selection Methods and Classifiers for AD and MCI vs CN.

Group	Feature Selection	Metric	AB	DT	GNB	GP	GB	HGB	MLP	QDA	RF
CN vs AD	ANOVA	AUC	0.860	0.856	0.886	0.883	0.891	0.886	0.903	0.885	0.896
		ACC	0.771	0.788	0.811	0.797	0.802	0.780	0.828	0.802	0.808
	PCA	AUC	0.863	0.861	0.879	0.882	0.901	0.901	0.892	0.878	0.905
		ACC	0.794	0.791	0.797	0.791	0.825	0.831	0.811	0.800	0.831
	LASSO	AUC	0.939	0.851	0.950	0.906	0.931	0.940	0.957	0.911	0.937
		ACC	0.859	0.799	0.890	0.856	0.876	0.898	0.901	0.839	0.879
	Chi-square	AUC	0.869	0.864	0.889	0.868	0.893	0.890	0.889	0.883	0.891
		ACC	0.794	0.802	0.814	0.771	0.816	0.814	0.802	0.814	0.819
MCI vs AD	ANOVA	AUC	0.763	0.773	0.816	0.800	0.796	0.803	0.811	0.799	0.774
		ACC	0.691	0.727	0.754	0.746	0.730	0.741	0.751	0.743	0.708
	PCA	AUC	0.739	0.744	0.778	0.778	0.790	0.778	0.782	0.778	0.777
		ACC	0.686	0.664	0.700	0.694	0.719	0.708	0.716	0.700	0.727
	LASSO	AUC	0.827	0.774	0.871	0.817	0.856	0.860	0.867	0.812	0.861
		ACC	0.757	0.729	0.792	0.787	0.771	0.776	0.770	0.746	0.798
	Chi-square	AUC	0.754	0.787	0.812	0.763	0.806	0.795	0.793	0.807	0.791
		ACC	0.708	0.732	0.568	0.713	0.735	0.721	0.735	0.711	0.711
CN vs MCI	ANOVA	AUC	0.674	0.678	0.721	0.691	0.729	0.735	0.719	0.662	0.728
		ACC	0.615	0.651	0.659	0.659	0.682	0.669	0.664	0.621	0.667
	PCA	AUC	0.665	0.601	0.703	0.595	0.672	0.663	0.714	0.675	0.689
		ACC	0.633	0.585	0.641	0.582	0.638	0.610	0.656	0.605	0.638
	LASSO	AUC	0.762	0.680	0.792	0.751	0.788	0.787	0.803	0.746	0.796
		ACC	0.700	0.641	0.674	0.708	0.723	0.732	0.731	0.685	0.741
	Chi-square	AUC	0.631	0.639	0.693	0.652	0.684	0.682	0.698	0.637	0.692
		ACC	0.595	0.595	0.572	0.621	0.623	0.633	0.618	0.531	0.626

3.5.1 CN vs. AD

The comparative analysis of CN vs. AD classification was performed to identify the most predictive feature subset using the MLP classifier. A rank-based approach was applied, systematically evaluating all subsets of the top 50 features identified through the LASSO method. Features were added incrementally in each evaluation step, following their ranking based on coefficient values. For each subset, four metrics, ROC AUC, accuracy, specificity, and sensitivity, were computed. The aim was to determine the feature subset achieving the highest average of ROC AUC and accuracy.

The optimal subset comprised 36 features, where the MLP classifier demonstrated outstanding performance, achieving an AUC of 0.951, an accuracy of 0.907, a specificity of 0.916, and a sensitivity of 0.897. This configuration was selected as the optimal point based on the average of ROC AUC and accuracy, emphasizing the significance of these metrics in evaluating prediction models. These results demonstrate the effectiveness of this subset in differentiating between AD and CN subjects.

To refine the model and achieve a more interpretable feature set, we applied Pearson correlation analysis to all the top 36 selected features. This step aimed to identify and reduce redundancy among the features by selecting those with the least interdependence. By focusing on uncorrelated features, we ensure that the selected variables provide unique contributions to the model, avoiding overlap in the information they represent.

To identify a refined feature set with minimal redundancy and diverse representation, we selected the feature pairs with the lowest absolute correlation coefficients based on Pearson correlation analysis, indicating weak or negligible relationships. To validate these pairs, we further calculated p-values for their correlation coefficients using the Spearman rank correlation test, ensuring statistical confirmation of the correlations' strength and significance while ruling out random variation. The resulting Table 3.12 summarizes the selected uncorrelated feature pairs, along with their correlation coefficients and corresponding p-values, providing a robust foundation for feature

selection.

Table 3.12 Highly uncorrelated feature pairs for distinguishing CN vs. AD identified through the Pearson correlation test, showing minimal correlation and corresponding p-values.

Feature 1 (Name ID Hand)	Feature 2 (Name ID Hand)	Correlation	P-Value
original shape Flatness 7003 lh	original firstorder Skewness 7010 lh	0.001	8.64E-01
original firstorder Kurtosis 7008 lh	original gldm SmallDependenceEmphasis 233 lh	0.002	8.93E-01
original firstorder Skewness 7010 lh	original glrlm LongRunEmphasis 7008 lh	0.003	6.58E-01
original glrlm RunVariance 242 lh	original glrlm RunVariance 7005 lh	0.003	3.20E-02
original shape Flatness 7010 lh	original glszm ZoneVariance 7009 lh	0.004	6.81E-01
original gldm SmallDependenceEmphasis 7001 lh	original firstorder Skewness 7010 lh	0.005	6.92E-01
original gldm SmallDependenceEmphasis 233 lh	diagnostics Image-original Mean 7006 lh	0.006	3.96E-01

To enhance model interpretability and reduce redundancy, Pearson correlation analysis was conducted on the top 33 selected features to identify those with minimal interdependence. This approach prioritized uncorrelated features to ensure that each contributed unique information without overlapping. From this analysis, the feature pairs with the lowest absolute correlation coefficients were selected, reflecting weak or negligible relationships. Their statistical significance was further confirmed using p-values derived from the Spearman rank correlation test, validating the strength of these low correlations and eliminating the likelihood of random variation. The resulting Table 3.13 presents the selected feature pairs, their correlation coefficients, and corresponding p-values, forming a solid basis for an optimized feature set.

3.5.2 CN vs. MCI

The comparative analysis of CN vs. MCI classification was conducted to identify the most predictive feature subset using the MLP classifier. Similar to the CN vs. AD evaluation, a rank-based approach was applied, systematically evaluating all

Table 3.13 Highly uncorrelated feature pairs for distinguishing MCI vs AD.

Feature 1 (Name ID Hand)	Feature 2 (Name ID Hand)	Correlation	P-Value
original glrlm RunVariance 234 lh	original firstorder Skewness 236 lh	0.001	5.81E-01
original gldm SmallDependenceEmphasis 233 lh	original glrlm LongRunEmphasis 7008 lh	0.002	6.72E-01
original firstorder Range 244 lh	original glrlm LongRunEmphasis 233 lh	0.002	6.55E-01
original gldm DependenceNonUniformity 240 lh	original firstorder Minimum 7001 lh	0.003	7.80E-01
original firstorder MeanAbsoluteDeviation 234 rh	original glrlm LongRunEmphasis 233 lh	0.004	3.99E-01
original gldm SmallDependenceEmphasis 7003 lh	original firstorder Maximum 242 lh	0.005	8.00E-01
original glrlm LongRunEmphasis 7008 lh	original gldm DependenceVariance 7003 lh	0.006	7.80E-01
original glrlm LongRunEmphasis 7008 rh	original gldm DependenceVariance 7003 lh	0.007	7.80E-01
original firstorder Skewness 236 rh	original shape MinorAxisLength 234 lh	0.008	8.05E-01
original glrlm RunVariance 234 lh	original gldm SmallDependenceEmphasis 7003 lh	0.009	2.09E-01

subsets of the top features identified through the LASSO method. Features were added incrementally based on their ranking by co-efficiency values. For each subset, metrics such as ROC AUC, accuracy, specificity, and sensitivity were computed. The objective was to identify the feature subset that achieved the highest average of ROC AUC and accuracy. The optimal subset for CN vs. MCI classification comprised 35 features, achieving an ROC AUC of 0.803, an accuracy of 0.731, a specificity of 0.735, and a sensitivity of 0.726. This configuration was selected as the optimal point based on the average ROC AUC and accuracy, highlighting the effectiveness of the subset in distinguishing between CN and MCI subjects.

To further refine the model and achieve a more interpretable feature set, Pearson correlation analysis was applied to the top selected features. This step aimed to reduce redundancy and ensure the selected features provided unique contributions to the model. By focusing on uncorrelated features, we minimized overlap in the information

represented by the selected variables.

To identify the most diverse and representative subset, feature pairs with the lowest absolute correlation coefficients were selected based on Pearson correlation analysis. The statistical significance of these correlations was validated using the Spearman rank correlation test, ensuring the robustness of the feature selection process. The resulting Table 3.14 summarizes the selected uncorrelated feature pairs, along with their correlation coefficients and corresponding p-values.

Table 3.14 Highly uncorrelated feature pairs for distinguishing CN vs MCI.

Feature 1 (Name ID Hand)	**Feature 2 (Name ID Hand)**	**Correlation**	**P-Value**
original firstorder Skewness 236 lh	original gldm DependenceVariance 7001 lh	0	0.706
original firstorder Range 7007 lh	original shape Maximum2DDiameterColumn 7005 lh	0.001	0.7609
original firstorder InterquartileRange 211 rh	original shape Elongation 242 lh	0.002	0.7156
original firstorder InterquartileRange 211 lh	original shape Elongation 242 lh	0.002	0.7156
original firstorder Kurtosis 215 rh	original firstorder Range 7007 lh	0.002	0.5728
original gldm DependenceNonUniformityNormalized 236 lh	original firstorder RobustMeanAbsoluteDeviation 7008 lh	0.002	0.7097
original firstorder Minimum 241 lh	original firstorder Minimum 7001 lh	0.002	0.5246
diagnostics Image-original Minimum 7006 lh	original firstorder Variance 239 lh	0.002	0.8809
original firstorder InterquartileRange 211 rh	original glrlm RunVariance 226 lh	0.003	0.7854
original glrlm RunVariance 226 lh	original firstorder InterquartileRange 211 lh	0.003	0.7854
original glszm LargeAreaEmphasis 246 lh	original glszm SizeZoneNonUniformity 7005 lh	0.003	0.3579
original firstorder Variance 239 lh	original shape Maximum2DDiameterSlice 212 lh	0.004	0.5479

3.5.3 Analysis of Key Features and Regions

Our in-depth analysis of uncorrelated features for both diagnosis groups identified two key features as the common subset for both groups CN vs. AD and MCI vs. AD: Long Run Emphasis (GLRLM) in the left accessory basal nucleus and Small Dependence Emphasis (GLDM) in the left presubiculum-head. These features were found to be the most effective in differentiating between MCI, CN, and AD individuals among the highly uncorrelated feature pairs.

Table 3.15 Performance evaluation of the four most effective features, individually and in combination, for classifying CN vs. AD, MCI vs. AD, and CN vs. MCI using MLP.

Groups	**Uncorrelated Features**	**ROC AUC ± SD**	**Acc ± SD**	**Sens**	**Spec**
CN vs. AD	original gldm SmallDependenceEmphasis 233 lh	0.782 ± 0.008	0.723 ± 0.016	0.618	0.815
	original glrlm LongRunEmphasis 7008 lh	0.885 ± 0.020	0.797 ± 0.039	0.739	0.847
	original gldm SmallDependenceEmphasis 233 lh & original glrlm LongRunEmphasis 7008 lh	0.903 ± 0.019	0.822 ± 0.049	0.776	0.863
MCI vs. AD	original gldm SmallDependenceEmphasis 233 lh	0.748 ± 0.052	0.686 ± 0.051	0.558	0.791
	original glrlm LongRunEmphasis 7008 lh	0.798 ± 0.025	0.735 ± 0.035	0.691	0.771
	original gldm SmallDependenceEmphasis 233 lh & original glrlm LongRunEmphasis 7008 lh	0.824 ± 0.026	0.749 ± 0.040	0.715	0.776
CN vs. MCI	original firstorder RobustMeanAbsoluteDeviation 7008 lh	0.631 ± 0.072	0.620 ± 0.038	0.611	0.629
	original firstorder Minimum 241 lh	0.631 ± 0.032	0.613 ± 0.014	0.607	0.619
	original firstorder RobustMeanAbsoluteDeviation 7008 lh & original firstorder Minimum 241 lh	0.677 ± 0.059	0.636 ± 0.076	0.606	0.666

These features achieved remarkable performance in both classification tasks using the MLP classifier. For CN vs. AD, the MLP classifier yielded an ROC AUC of 0.903, an accuracy of 0.822, a sensitivity of 0.863, and a specificity of 0.776. In the MCI vs. AD classification, the same classifier achieved an ROC AUC of 0.824, accuracy of 0.749, sensitivity of 0.776, and specificity of 0.715, as presented in Table

3.15. Furthermore, for CN vs. MCI, Robust Mean Absolute Deviation (First order) and Minimum (First order) were the most effective uncorrelated features, achieving an ROC AUC of 0.677 and an accuracy of 0.636 when combined.

Table 3.16 shows the structural and textural changes of the hippocampal-amygdala complex across CN, MCI, and AD, underlining the importance of specific radiomic features as biomarkers. In the comparison between CN and AD, the feature Small Dependence Emphasis (GLDM) presents a remarkable growth in mean values, growing from 0.0839 in CN to 0.1636 in AD. This corresponds to a highly increased gray-level heterogeneity. On the other hand, the Long Run Emphasis (GLRLM) significantly decreases from 0.6429 to 0.3988, reflecting less structural uniformity. Both features present highly significant p-values, which point out their relevance for diagnosis.

A similar trend can be seen in MCI vs. AD, where Small Dependence Emphasis (GLDM) increases from 0.0929 in MCI to 0.1636 in AD, and glrlm LongRunEmphasis decreases from 0.5912 to 0.3988, reinforcing these features as able to tell AD from earlier stages. The changes observed in CN vs MCI are subtler. Robust Mean Absolute Deviation (First order), for example, increases from 0.2042 to 0.2453, indicative of higher variability in the texture in MCI, and the Minimum (First order) decreased from 0.4864 to 0.4563, indicating a loss of signal intensity. This trend is underlined by their respective p-values of 0.0001 and 0.0325.

3.5.4 Case Study: Longitudinal Analysis of Radiomic Features in AD Progression

As a case study, we examined the long-term data of a single patient from the ADNI database. Over eight years, the patient was subjected to various scans and clinical assessments, transitioning from CN to AD. We selected four FDG-PET scans that illustrate this progression. Additionally, clinical evaluations were recorded, and scores involving MMSE, ADAS11, ADAS13, and CDRSB scores were presented in Table 3.17.

Table 3.16 Comparison of Radiomic Features Across CN, MCI, and AD Groups, Highlighting Mean, Standard Deviation, and Statistical Significance.

Group	Features	Group1 Mean	Group2 Mean	Group1 Std.	Group2 Std.	P-Value
CN vs. AD	original gldm SmallDependenceEmphasis 233 lh	0.0839	0.1636	0.0423	0.1193	0.001
CN vs. AD	original glrlm LongRunEmphasis 7008 lh	0.6429	0.3988	0.1223	0.1561	0.001
MCI vs. AD	original gldm SmallDependenceEmphasis 233 lh	0.0929	0.1636	0.0536	0.1193	0.001
MCI vs. AD	original glrlm LongRunEmphasis 7008 lh	0.5912	0.3988	0.1659	0.1561	0.001
CN vs. MCI	original firstorder RobustMeanAbsoluteDeviation 7008 lh	0.2042	0.2453	0.0941	0.1134	0.001
CN vs. MCI	original firstorder Minimum 241 lh	0.4864	0.4563	0.1238	0.1511	0.0325

Table 3.17 Radiomic Features and Clinical Scores Across Study Dates.

Feature / Score	4/11/2006	4/8/2008	1/20/2012	1/28/2014
MMSE	29	27	22	18
ADAS11	8	13.33	17	33
ADAS13	11	18.33	27	46
CDRSB	0	2.5	7	10
original gldm SmallDepEmph 233 lh	0.077763	0.352112	0.569209	0.688984
original glrlm LongRunEmph 7008 lh	0.576802	0.266491	0.207079	0.134096
firstorder RobustMAD 7008 lh	0.086043	0.125388	0.178754	0.203053
firstorder Min 241 lh	0.415374	0.432891	0.456195	0.705742

Radiomic features were extracted from the three subregions, including the accessory basal nucleus, presubiculum head, and CA4-head, using PyRadiomics. Key features such as Small Dependence Emphasis (GLDM), Long Run Emphasis (GLRLM), Robust Mean Absolute Deviation (First order), and Minimum (First order) were computed and analyzed across the four scans. 3.5.5.1 Results of the case study The patient exhibited a steady decline in clinical scores (MMSE from 29 to 18 and CDRSB from 0 to 10), correlating with significant changes in radiomic features. The temporal evo-

lution of each radiomic feature is shown in Table 3.17, highlighting trends and abrupt shifts in metabolic patterns. The possible interpretations of observed changes in these computed radiomics features for the above-mentioned specific three sub-regions of the hippocampus-amygdala complex are given in the next section.

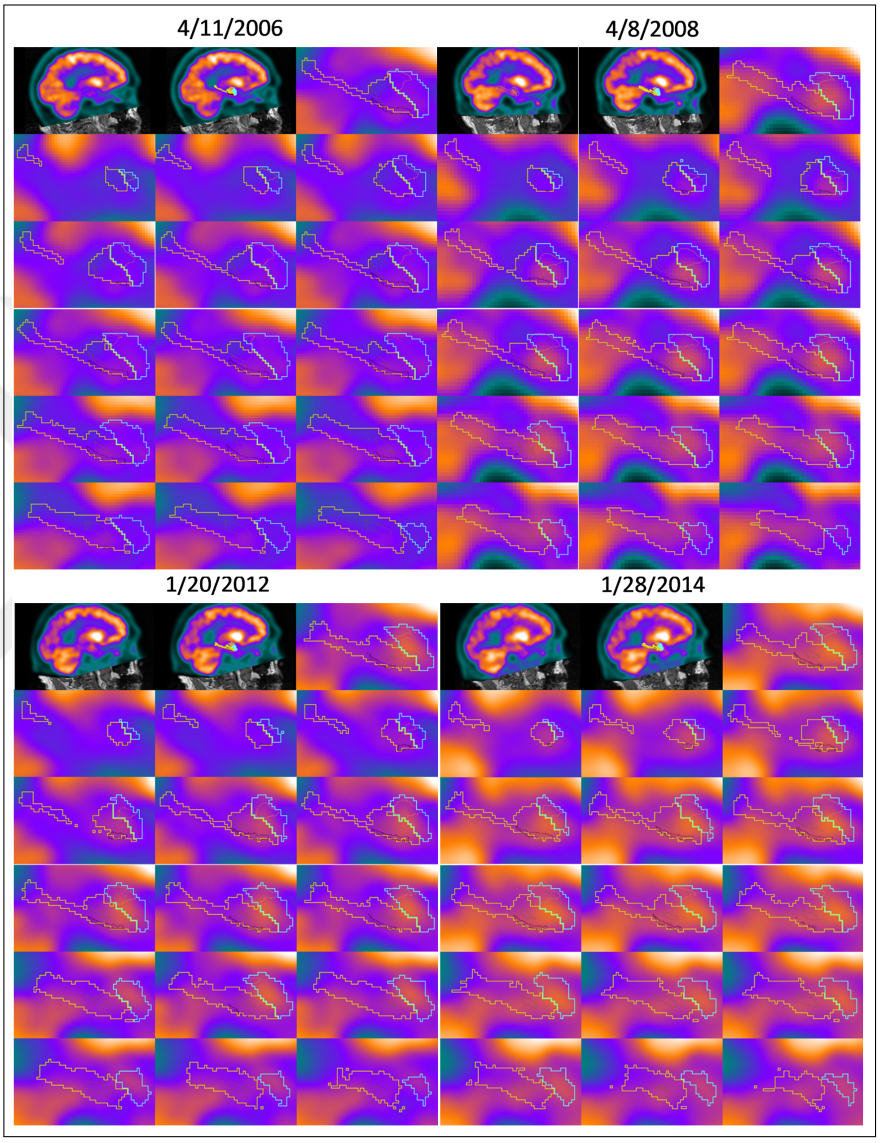

Figure 3.10 Longitudinal FDG-PET scans with segmented hippocampal and amygdala subregions (accessory basal nucleus, Presubiculum, CA4), demonstrating progressive changes in key areas.

3.6 Predicting tau positivity

We conducted a preliminary study to differentiate between individuals with TAU positivity and negativity. This study aimed to develop a classification model capable of distinguishing individuals across different research groups, Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer's Disease (AD), based on Tau protein levels, utilizing Fluorodeoxyglucose Positron Emission Tomography (FDG PET) scans. We analyzed FDG PET scans from 258 participants using the Desikan-Killiany Tourville (DKT) atlas in FreeSurfer, segmenting the brain into 95 distinct regions to extract a detailed feature set. Participants were categorized into two groups, tau-positive and tau-negative, based on a threshold of 375 pg/mL for tau levels (129 in each group). To reduce dimensionality and improve model interpretability, we applied Least Absolute Shrinkage and Selection Operator (LASSO) regression, selecting the top 10 most relevant features. A logistic regression classifier was then trained for binary classification based on the tau threshold. We assessed model performance using standard evaluation metrics. The final model, utilizing the three selected features (original shape VoxelVolume Hippocampus, original shape SurfaceVolumeRatio inferiorparietal, and original shape SurfaceVolumeRatio precuneus), achieved an Area Under the Receiver Operating Characteristic Curve (ROC AUC) of 0.76, indicating a good ability to differentiate between classes. The model's accuracy was 72%, demonstrating overall correctness in classifying participants. The selected features, hippocampal volume and parietal/precuneus surface volume ratios, align with known Alzheimer's Disease (AD) neuropathology, reflecting key regions affected by the disease. Further investigation into confounding factors and refined feature extraction is needed. This study demonstrates the feasibility of using FDG PET-derived features for Tau-based classification in AD. Future work should focus on larger datasets, multimodal data integration, and advanced machine learning to enhance performance and AD understanding.

4. DISCUSSION

Our study is a multi-objective investigation designed to extract different diagnostic and prognostic data from FDG-PET scans, offering a comprehensive view for radiologists. Alzheimer's disease (AD) is the most common type of dementia that affects the brain and slowly destroys the ability to think, remember, reason, and communicate. The neuropathological features of the disorder are the accumulation of amyloid-beta protein (AAβ), formation of tau protein tangles, neuroinflammation, and progressive neuronal death, which leads to brain atrophy and metabolic dysfunction. Current diagnosis is based on numerous biomarkers that include cerebrospinal fluid (CSF) analysis, FDG, amyloid, and tau positron emission tomography (PET) scanning, and structural magnetic resonance imaging (MRI), but such invasive, costly, or unresponsive tests are used. FDG-PET has been used extensively to assess glucose metabolism in brain tissue and thus neuronal function and to identify hypometabolic patterns in the development of AD. Therefore, based on the recent advancements in radiomics and machine learning algorithms, our work is aimed at exploring the potential of FDG-PET imaging to derive multimodal biomarkers of different aspects of AD pathology, including early identification, management, and risk assessment.

Traditional neuroimaging tests are primarily concerned with pre-specified brain regions or overall metabolism levels and, therefore, do not track more subtle structural and functional alterations that occur in the initial phases of AD. Radiomics, on the other hand, enables the assessment of high-dimensional features like first-order statistical features, texture-based features (e.g., Gray Level Co-occurrence Matrix and Gray Level Run Length Matrix), and shape features, which can identify subtle metabolic and morphological alterations. Our work integrates these radiomic signatures to accomplish several tasks: (1) prediction of amyloid positivity from FDG PET scans without the necessity of an additional amyloid PET scan; (2) identification of connectivity differences between key brain regions such as the hippocampus, and amygdala to monitor disease progression; (3) analysis of the interaction between genetic risk factors (e.g.,

ApoE4 genotype) and metabolic alterations; and (4) development of maximally accurate classification models to distinguish between cognitively normal (CN), mild cognitive impairment (MCI), and AD patients. By integrating findings from these analyses, we aim to establish a comprehensive understanding of how radiomic features contribute to characterizing AD. With the development of an end-to-end non-invasive pipeline, our research accelerates clinical decision-making, reduces uncertainty in diagnosis, and enables precision medicine in neurodegenerative disease.

4.1 Discussion of Individual Feature Objectives

4.1.1 Features for Staging and Diagnosis of AD

Accurate staging of AD is essential for effective treatment planning and disease management. Analysis of this study identified four frequently employed features (namely, gldm_DependenceEntropy, shape_SurfaceVolumeRatio, glrlm_RunPercentage, glrlm_LongRunEmphasis) by eight distinct feature selection methods. These features were consistently recognized as the most critical in predicting AD within three specific Regions of Interest (ROIs): the amygdala, entorhinal cortex, and hippocampus. GLDM measures the degree of randomness or non-uniformity in an image.

gldm_DependenceEntropy: This feature measures texture complexity, potentially reflecting alterations in tissue microstructure due to neurodegeneration in AD and MCI. In other words, a larger value of a textural feature extracted from GLDM corresponds to a heterogeneous and irregular spatial distribution, with a series of complex correlation structures between the voxels[52].

Textural features such as glrlm_RunPercentage and glrlm_LongRunEmphasis were employed to detect structural and morphological changes within brain regions, aiding in the diagnosis and comprehension of cognitive and behavioral impairments in individuals with AD and MCI [53], [54]. These features play a pivotal role in delineating subtle alterations in brain tissue, contributing valuable insights into the progression of

neurodegenerative conditions and facilitating more targeted interventions for affected patients.

Shape_SurfaceVolumeRatio: This feature acts as a sensitive gauge for morphological alterations within brain regions, particularly those susceptible to neurodegeneration in AD and MCI, like the hippocampus. An elevated ratio signifies a shrinking tissue volume relative to the remaining surface area, potentially reflecting cellular loss and atrophy. This phenomenon is most commonly observed in the hippocampus, a key structure for memory and navigation, and often worsens with the progression of AD.

4.1.2 Features for Predicting Amyloid Positivity

Our study demonstrated that FDG-PET-derived features such as gray-level zone size matrix (GLSZM) and neighboring gray-tone difference matrix (NGTDM) were highly predictive of amyloid positivity. These features, extracted from regions including the hippocampus, inferior parietal lobe, and isthmus cingulate, namely Short Run Low Grey Level Emphasis (SRLGLE), root Mean Absolute Deviation (rMAD), and 90th percentile, may be numerically associated with tissue properties in the context of amyloid pathology.

SRLGLE is a feature that measures the distribution of short runs of low grey-level values in an image. This feature could potentially reflect a heterogeneous texture or granularity of amyloid deposits. Compared to features like Long Run High Gray Level Emphasis (LRHGLE), which capture broader intensity variations, SRLGLE might be more specific to distinguishing amyloid deposits from other types of lesions that have a more uniform or smoother signal distribution. A higher value indicates finer structural textures and a greater concentration of low gray-level values in the image.

rMAD is a feature that is calculated as the average absolute deviation of the standardized uptake value (SUV) from the mean SUV of the region of interest. rMAD

may offer insights into cellular density and metabolic heterogeneity. 90th percentile is calculated as the SUV value below which 90% of the voxels in the region of interest fall. This feature may indicate regional metabolic heterogeneity and high-activity foci.

Although these individual features may potentially point to certain tissue characteristics, the complex patterns of multiple features can predict AB+ more accurately. More research is needed to confirm these associations in order to improve interpretability. The identified ROIs, hippocampus, inferior parietal, and isthmus cingulate, are known to play roles in memory, spatial processing, self-referential thinking, and emotional regulation. These brain regions are known to be particularly relevant to the study of AD and mild cognitive impairment (MCI)[55], [56]. It goes without saying that using the patterns of a higher number of features (18) is advantageous in terms of accuracy (AUC increases from 0.853 to 0.924) while presenting disadvantages associated with complexity.

4.1.3 Features for Tracking AD Progression via Connectivity Analysis

The neurodegenerative process in AD is not limited to isolated regions but extends to connectivity disruptions between critical brain structures. Our study identified key shape-based and texture-based features that characterize these connectivity changes, particularly between the hippocampus and amygdala. The gldm_SDLGLE quantifies the frequency of neighboring voxels with similar intensities within a small spatial distance. In healthy brain tissue, neuronal organization leads to a relatively uniform distribution of gray levels. Therefore, an increase in gldm_SDLGLE within the CR, as measured in this study, might indicate a disruption in the underlying neuronal microstructure in AD patients [57]. This disruption could be due to factors such as amyloid plaque deposition, neurofibrillary tangles, or neuronal loss, all of which are hallmarks of AD pathology. These pathological changes can alter the local tissue composition and connectivity, leading to a less uniform texture as detected by gldm_SDLGLE. According to the results, analyzing the volume and texture changes of the CR allows us to gain deeper insights into the severity and progression of AD. 4.1.4

Features for Predicting ApoE4 Genetic Risk Genetic predisposition, particularly the presence of the ApoE4 allele, is a well-established risk factor for AD. Our study explored whether metabolic imaging features from FDG-PET could serve as surrogate markers for genetic susceptibility. Our findings identified seven significant radiomic features, four of which are shape-based. In light of previous research, structural alterations in the thalamus, hippocampus, and amygdala have been observed in individuals with homozygous APOE4. Our shape-based features align with these earlier findings, further supporting the evidence of such changes in these brain regions [58], [59]. Specifically, the shape_SurfaceVolumeRatio in the thalamus suggests a change in the region's overall shape, potentially indicating atrophy or volume loss. The shape_Sphericity and shape_Maximum2DDiameter_Column in the hippocampus and amygdala highlight deviations from a normal, spherical shape, suggesting irregularities in these structures. The shape_SurfaceArea in the parsorbitalis region suggests alterations in the size or shape of this structure, potentially indicating atrophy, volume loss, or other structural abnormalities.

The remaining features, glszm_SmallAreaEmphasis, and glszm_ZoneVariance, are gray-level zone matrix features that provide insights into the distribution of gray levels within specific image regions. glszm_SmallAreaEmphasis focuses on the frequency of small, homogeneous areas, potentially reflecting changes in tissue texture or homogeneity. glszm_ZoneVariance quantifies the variability of gray levels within zones, suggesting alterations in the heterogeneity or complexity of the tissue.

Finally, glrlm_RunEntropy is a gray-level run-length matrix feature that measures the complexity or randomness of gray-level transitions within an image. It can be indicative of changes in tissue organization or microstructure. Collectively, these features offer a comprehensive assessment of structural and textural abnormalities in the brain regions associated with homozygous APOE4. By combining shape-based, gray-level zone matrix, and gray-level run-length matrix features, our model provides a more nuanced understanding of the underlying pathological processes associated with this genetic risk factor. Associations between cognitive capabilities and the 18 features have also been explored. Analysis of scatter plots in the supplementary file revealed

generally weak associations for MCI patients. However, these results may be different if a larger cohort is considered including healthy and AD patients.

4.1.4 Features for Subregional Biomarker Analysis in AD

Subregional metabolic pattern analysis of hippocampus and amygdala generated new diagnostic markers of AD. Four prominent radiomic features of FDG-PET imaging discriminating MCI, AD, and CN states, supporting minute gray-level changes in brain textural patterns, associated with Alzheimer's pathology, were our discoveries. Long Run Emphasis (GLRLM), associated with the left accessory basal nucleus, one of the center areas of the consolidation and emotion function, indicates potential functional change as a result of neurodegeneration with higher values.

Dependence on small (GLDM), associated with left head of presubiculum involved with navigation and memory, assesses smoothness of textural and detects earlier metabolic or micro- rather than later anatomical changes, possibly before visible atrophy. Furthermore, Robust Mean Absolute Deviation (First order) and Minimum (First order) are important to use to distinguish CN from MCI. Minimum (First order), which maps onto the CA4-head region, indicates the lowest intensity of voxels, which marks minute changes in the metabolic function and neuronal structure that might reflect beginning neurodegenerative processes significant in encoding of the memory. Concurrently, Robust Mean Absolute Deviation (First order) measures the intensity spread of voxels, with information regarding local heterogeneity of the brain tissue. In combination, these characteristics offer an effective biomarker for the early identification and characterization of Alzheimer's disease progression.

Figure 3.10 and Table 3.17 illustrate the potential of radiomic features to detect minor metabolic alterations associated with the worsening of the clinical status in the progression of AD:

- Elevated Small Dependence Emphasis might suggest the brain area has a less het-

erogeneous pattern of glucose metabolism, with less variance in intensity values between adjacent voxels. This might be associated with: Greater Degree Neuronal Dysfunction: Greater decrease in glucose metabolism within the area, indicating higher neuronal damage. Reduced Heterogeneity: Less complex metabolic pattern, possibly indicating a less dynamic or less active brain area.

- Long Run Emphasis (GLRLM) In AD FDG PET imaging, a decrease in the "Long Run Emphasis" property of Gray Level Run Length Matrix (GLRLM) might reflect higher heterogeneity of glucose metabolism within the regions of the brain relative. Long Run Emphasis: It is a parameter that indicates the frequency of long sequences of the identical intensity pixels. Smaller values reflect less frequent long sequences and frequent intensity value switching.

- Glucose Long Run Emphasis Measure GLRLM) decrease might reflect a less uniform pattern of glucose metabolism within the brain region, with increased switching between regions of higher and regions of lower glucose uptake. This could be secondary to a mixed pattern of neurodegeneration, i.e., regions of higher neuronal degeneration with coexisting regions of less severe degeneration. Or coexisting inflammatory processes or other metabolic processes possibly underlying higher glucose uptake in certain regions.

- Robust Mean Absolute Deviation (order 1) One less sensitive than standard deviation. Greater values suggest greater variability in the data. Possible Interpretations: 1) Greater Heterogeneity: Greater Robust Mean Absolute Deviation (order 1) might be suggesting greater variability of glucose metabolism of the region of the brain with areas of higher and other areas of low uptake. This might be on account of i) Mixed Patterns of Neurodegeneration: Co-occurrence of areas with greater and smaller degrees of neuronal loss, ii) Inflammation or other Processes: Co-occurrence of inflammatory processes or other metabolic processes which might be responsible for higher uptake in some regions. 2) Minimum (order 1) had increasing values rather than the expected decrease in metabolic function with the progression of the disease, possibly suggesting inflammation. Confirmation is, however, required.

4.2 Comprehensive Discussion of All Features

Through the integration of these feature-specific findings, we highlight the potential of radiomics to facilitate AD diagnosis as well as monitoring. While each feature provides specific information on different aspects of the disease, when taken together, it represents a multivariate approach to AD characterization.

Among the key conclusions is that first-order statistics perform well to detect overall metabolic loss, while texture features provide higher sensitivity in detecting the heterogeneity of the disease. Shape features, particularly those based on connectivity analysis, provide a new method of assessing shrinkage over time. Lastly, the potential of not having to perform the risk factors with a genetic origin in an invasive manner emphasizes the potential of radiomics in precision medicine.

One other important point is the interaction of numerous machine learning models and feature selection. All feature selection with Random Forest improved both robustness and interpretation simultaneously, while classifiers such as Random Forest, Multi-Layer Perceptron, and GNB provided high predictive accuracy across different study goals.

In summary, our research emphasizes the imperatives of a multi-objective approach to AD research. Leveraging the power of heterogeneous radiomic features by FDG-PET imaging not only improves diagnostic accuracy it also offers new information on mechanisms of disease, risk stratification, and monitoring of the disease. Future initiatives need to focus on external validation on independent data, longitudinal investigations to assess the progression of the disease, and the establishment of multimodal imaging paradigms to further develop AD characterization and increase utility in the clinic.

4.3 Advantages of Our Study for Interpretability

Our study provides several advantages in terms of interpretability, improving the transparency and clinical relevance of radiomics-based analysis for Alzheimer's disease (AD). By leveraging FDG-PET imaging biomarkers, we ensure that the extracted features, brain regions, and classification results are explainable and traceable to underlying neurobiological mechanisms. Below, we outline the primary advantages.

i. Feature Selection for Enhanced Interpretability Our approach carefully selects and reduces imaging features to improve interpretability. Instead of relying on black-box deep learning models, we use feature selection techniques such as LASSO, ANOVA, and PCA, ensuring that only the most relevant and biologically meaningful features are retained. As demonstrated in our study, reducing the feature set improved model interpretability while maintaining high predictive accuracy.

ii. Identification of Key Brain Regions Without Predefined Assumptions Unlike many traditional approaches that rely on predefined ROIs, our study automatically identifies critical brain regions using a data-driven framework. Regions like the hippocampus, amygdala, inferior parietal, entorhinal, and isthmus cingulate consistently emerged as crucial areas. These regions are well-established in the literature as being highly relevant to AD progression, reinforcing the interpretability and validity of our findings.

iii. Linking Radiomic Biomarkers to Disease Mechanisms Radiomics enables us to extract quantitative features that reflect underlying neurobiological changes. Our study demonstrates how specific texture and shape-based radiomic features correspond to metabolic abnormalities detected in FDG-PET studies. By establishing these associations, we improve model interpretability and provide a basis for clinical translation.

iv. Transparent Model Development and Validation Our study follows a structured pipeline, covering all aspects of the radiomics process: image preprocessing, seg-

mentation, feature extraction, selection, and classifier validation. Unlike deep learning models, which often lack transparency, our approach allows clinicians and researchers to follow each step of the analysis process, ensuring reproducibility and trust in the results.

v. Clinical Utility and Decision Support It is the explainability of our method that renders it suitable for clinical application. We avoid computational overhead by choosing a small subset of meaningful brain regions and features and maintaining high predictive power. It is even more promising to deploy it in decision-making in the clinic to help radiologists and neurologists in earlier diagnosis and customized planning of therapy.

vi. Comparison with Deep Learning Methodologies While deep learning is highly accurate, it comes with an issue of interpretability as it is based on high-level features. Unlike it, our radiomics approach is based on the fact that the features extracted map to clinically relevant biomarkers, and this translates into higher transparency in AD diagnosis. This renders our work closer to actual medical application.

vii. Future Relevance and Generalization We do not restrict our method to AD, and it is possible to generalize our method to other neurodegenerative disorders and imaging methods like Lewy body, Frontal temporal dementia, or Parkinson's. With an interpretable framework, future work has the ability to apply our method to multimodal imaging data. This makes our work relevant because advancing medical imaging technologies render our method adaptable.

4.4 Limitations and Future Directions

While this research demonstrates the potential of radiomics-based FDG-PET analysis to increase the detection, staging, and heredity-driven risk assessment of AD, several limitations should be acknowledged. These limitations should be addressed to make our findings reproducible, generalizable, and clinically applicable.

4.4.1 Study Limitations

Our work is based on data provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI), which is a well-structured and well-standardized dataset. ADNI subjects, however, do not represent the entire AD population worldwide because the population is primarily comprised of North Americans and Europeans with minimal demographic and ethnic heterogeneity. ADNI subjects also have greater education and better reserves, resulting in selection biases in machine learning models [11]. To enhance the generalizability of radiomics-based biomarkers, future research should incorporate multicenter datasets with diverse populations, including underrepresented groups with different genetic backgrounds, socioeconomic statuses, and comorbidities. This approach would ensure that the developed models remain robust across heterogeneous clinical settings.

Cross-Sectional Design: The Need for Longitudinal Studies

This study utilizes cross-sectional FDG-PET scans, which provide valuable metabolic snapshots of AD pathology at a single time point. However, AD is a progressive disorder, and neurodegenerative changes evolve over several years. Cross-sectional studies cannot fully capture disease progression or distinguish rapid vs. slow cognitive decline trajectories [28]. Future studies should focus on longitudinal FDG-PET data, tracking patients from preclinical to advanced AD stages. A longitudinal approach would enable the identification of:

- Early-stage metabolic changes that precede clinical symptoms.
- Patterns of progression that predict rapid vs. slow disease progression.
- The impact of ApoE4 and other genetic risk factors on disease evolution.
- Biomarker stability over time for personalized disease monitoring.

Additionally, incorporating repeated imaging at different disease stages would help re-

fine predictive models by distinguishing individual variability from true disease-related changes.

Standardization Challenges in FDG-PET Imaging

A major challenge in radiomics research is the lack of standardized imaging protocols across different scanners, acquisition sites, and preprocessing methods. FDG-PET studies are often performed on different PET scanner models, each with varying spatial resolution, reconstruction algorithms, and noise levels, leading to potential variability in radiomic feature extraction. For radiomics-based biomarkers to be clinically useful, there is a critical need for standardization:

- Harmonization of FDG-PET acquisition parameters across imaging centers.
- Development of standardized preprocessing pipelines to ensure consistency in feature extraction.
- Implement scanner normalization techniques, such as ComBat harmonization, to correct for site-specific differences.
- Consensus guidelines for radiomics studies, including quality control measures to ensure reproducibility.

Efforts such as the Image Biomarker Standardization Initiative (IBSI) have begun addressing these challenges, but further work is needed to ensure cross-site validation and regulatory approval for clinical applications. Model Interpretability and Clinical Integration Although machine learning models, particularly deep learning and radiomics-based approaches, have shown high classification accuracy, interpretability remains a major barrier to clinical adoption. Clinicians require transparent and explainable AI models that provide clear reasoning for predictions rather than "black-box" outputs. To improve interpretability:

- Feature selection techniques (e.g., LASSO, SHAP values) should be employed to

highlight the most biologically meaningful radiomic features.

- Visualization methods, such as heatmaps and saliency maps, can help clinicians understand which brain regions contribute most to model predictions.
- Comparisons with traditional biomarkers, such as CSF amyloid/tau levels, can help validate the reliability of radiomics-based biomarkers.

Developing AI models that integrate radiomics with traditional clinical and imaging markers would increase trust and adoption among neurologists and radiologists.

4.4.2 Future Research Directions

Future research on radiomics-based FDG-PET imaging in AD should be focused on the creation of better predictive models, method standardization, and incorporation into the clinic of AI tools. The incorporation of longitudinal FDG-PET data will also provide disease progression monitoring, detection of earlier markers, risk stratification, and tailored outcome prediction. Time-series modeling will also provide better disease staging as well as monitoring of responses to treatment, which will provide optimal AD treatment planning. Continuations of challenges include standardization and reproducibility with the need for harmonized methods of feature extraction, stable preprocessing, and validation across heterogeneous data. Standardized radiomics pipelines and higher-level AI-driven segmentation methods will provide higher reliability and applicability in the clinic. Greater incorporation of FDG-PET radiomics and EHR and clinic data will build predictive models of imaging markers mixed with cognitive testing and histories to detect and quantify cognitive impairment. For the integration of radiomics into the daily work of hospitals, decision-making systems driven by AI to automatically extract features, perform online analysis, and provide user-friendly interfaces to clinicians should be implemented. Cloud systems and federated learning will enable the realization of data security and multicenter trials. Novel AD-subtyping marker research employing radiomics, i.e., metabolic heterogeneity and inflammatory markers, have the ability to further tailor classification and treatment choice. Bringing

this into practical application, FDG-PET radiomics can be an essential component of precision neurology.

4.5 Conclusion

The emphasis of the present work was to explore radiomics-based FDG-PET imaging of AD with certain principal goals: (1) definition of radiomic features for proficient staging and diagnosis of AD, (2) biomarkers obtained from FDG-PET identifying amyloid positivity, (3) disruption in connection in salient brain areas for tracking progression of AD, (4) predictability of radiomic features regarding ApoE4 gene risk investigation, and (5) metabolic subregional differences as AD classification biomarkers research. With an integration of machine learning models and radiomic evaluation in high dimensions, our work introduces a new non-invasive method of capturing clinically useful biomarkers to advance AD diagnosis performance and interpretability. In order to develop diagnosis and staging toward the intent, we identified four key radiomic features, gldm_DependenceEntropy, shape_SurfaceVolumeRatio, glrlm_RunPercentage, and glrlm_LongRunEmphasis, with strong predictive AD status within important regions such as the amygdala, entorhinal cortex, and hippocampus.

These features were able to detect subtle structural and metabolic changes, also evidencing their value in the characterization of cognitive disease stage and impairment. Moreover, machine learning classifiers and radiomics-based utilization enhanced diagnostic performance compared to traditional imaging methods, providing an effective platform for staging and early detection. In the prediction of amyloid positivity, from our evidence, certain FDG-PET radiomic features like Short Run Low Grey Level Emphasis (SRLGLE) and root Mean Absolute Deviation (rMAD) are surrogate markers of amyloid burden.

These radiomic features, extracted on voxels of interest like the hippocampus, inferior parietal lobe, and isthmus cingulate, were discovered to be strongly predictable, thus preventing unneeded amyloid PET scanning.

This result cements the potential of FDG-PET as an independent imaging modality in AD screening and risk stratification. Monitoring of AD progression was also explored in this work through connectivity analysis and proved that texture- and shape-based radiomic features, especially gldm_SDLGLE, can effectively differentiate metabolic and microstructural alterations in the key brain areas, including the hippocampus and amygdala. The results confirm established patterns of neurodegeneration, demonstrating the potential of radiomics for quantitative, dynamic evaluation of the disease process. For genetic risk factor analysis, a number of radiomic features we found were linked to ApoE4 status, including thalamus.shape_SurfaceVolumeRatio, hippocampus.shape_Sphericity, and texture features glszm_SmallAreaEmphasis and glrlm_RunEntropy. These biomarkers also present non-invasive quantification of genetic susceptibility, revealing further insight into the influence of genetic predisposition on metabolic change. Lastly, our subregional metabolic biomarker study of the amygdala and hippocampus identified new radiomic features, including Long Run Emphasis (GLRLM) and Small Dependence Emphasis (GLDM), to differentiate cognitively normal controls, MCI patients, and AD patients. These features provide an improved understanding of disease pathology to enable early diagnosis and precision medicine strategies by disease trajectory on a case-by-case basis.

Despite our study highlighting the diagnostic ability of radiomics in AD, there are several challenges that remain, including the standardization of imaging protocols, independent dataset validation, and clinical adoption of AI models. Recent studies must overcome the multi-modal imaging data fusion, following disease progression in the long term, and develop explainable AI models that can be used for clinical decision-making. With additional development of machine learning and radiomics, FDG-PET imaging will presumably greatly enhance early diagnosis, risk stratification, and treatment monitoring of AD, and ultimately the practice of precision neurology.

REFERENCES

1. Zhang, J., Y. Zhang, J. Wang, and et al., "Recent advances in alzheimer's disease: mechanisms, clinical trials and new drug development strategies," *Signal Transduction and Targeted Therapy*, Vol. 9, p. 211, 8 2024.

2. Counts, S. E., M. D. Ikonomovic, N. Mercado, and et al., "Biomarkers for the early detection and progression of alzheimer's disease," *Neurotherapeutics*, Vol. 14, pp. 35–53, 1 2017.

3. Siddappaji, K. K., and S. Gopal, "Molecular mechanisms in alzheimer's disease and the impact of physical exercise with advancements in therapeutic approaches," *AIMS Neuroscience*, Vol. 8, pp. 357–389, 2021.

4. Planche, V., J. V. Manjon, B. Mansencal, and et al., "Structural progression of alzheimer's disease over decades: the mri staging scheme," *Brain Communications*, Vol. 4, 5 2022.

5. van Dyck, C., "The development of neuroimaging methodologies to assist in the diagnosis and treatment of alzheimer's disease," *Biological Psychiatry*, Vol. 95, p. S25, 5 2024.

6. Lista, S., M. Mapstone, F. Caraci, and et al., "A critical appraisal of blood-based biomarkers for alzheimer's disease," *Ageing Research Reviews*, Vol. 96, p. 102290, 4 2024.

7. Hnilicova, P., E. Kantorova, S. Sutovsky, and et al., "Imaging methods applicable in the diagnostics of alzheimer's disease, considering the involvement of insulin resistance," *International Journal of Molecular Sciences*, Vol. 24, p. 3325, 2 2023.

8. Dang, C., Y. Wang, Q. Li, and et al., "Neuroimaging modalities in the detection of alzheimer's disease-associated biomarkers," *Psychoradiology*, Vol. 3, 3 2023.

9. Guiot, J., A. Vaidyanathan, L. Deprez, and et al., "A review in radiomics: Making personalized medicine a reality via routine imaging," *Medicinal Research Reviews*, Vol. 42, pp. 426–440, 1 2022.

10. Chen, Z., S. Yan, S. Bi, and et al., "Early diagnosis of alzheimer's disease using multiparametric hippocampal signatures with 18f-fdg pet/mr radiomics," 2023.

11. Hu, J., Y. Wang, D. Guo, and et al., "Diagnostic performance of magnetic resonance imaging-based machine learning in alzheimer's disease detection: a meta-analysis," *Neuroradiology*, Vol. 65, pp. 513–527, 3 2023.

12. Kim, D. W., H. Y. Jang, K. W. Kim, and et al., "Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers," *Korean Journal of Radiology*, Vol. 20, p. 405, 2019.

13. Gao, Y., J. Liu, J. Wang, and et al., "Proteomic analysis of human hippocampal subfields provides new insights into the pathogenesis of alzheimer's disease and the role of glial cells," *Brain Pathology*, Vol. 32, 7 2022.

14. Al-Ani, L., A. Tao, J. Dyke, and et al., "Amygdala atrophy as an early manifestation of alzheimer's disease (s39.005)," *Neurology*, Vol. 100, 4 2023.

15. Hansson, O., "Biomarkers for neurodegenerative diseases," *Nature Medicine*, Vol. 27, pp. 954–963, 6 2021.

16. Dewanjee, S., P. Chakraborty, H. Bhattacharya, and et al., "Altered glucose metabolism in alzheimer's disease: Role of mitochondrial dysfunction and oxidative stress," *Free Radical Biology and Medicine*, Vol. 193, pp. 134–157, 11 2022.

17. Limberger, C., L. S. Machado, J. P. Ferrari Souza, and et al., "The relationship between 18F-FDG-PET signal and key players of brain glucose metabolism," *Alzheimer's & Dementia*, Vol. 19, 12 2023.

18. Kato, T., Y. Inui, A. Nakamura, and et al., "Brain fluorodeoxyglucose (fdg) pet in dementia," *Ageing Research Reviews*, Vol. 30, pp. 73–84, 9 2016.

19. Ringman, J. M., W. Pope, and N. Salamon, "Insensitivity of visual assessment of hippocampal atrophy in familial alzheimer's disease," *Journal of Neurology*, Vol. 257, pp. 839–842, 5 2010.

20. Bailly, M., C. Destrieux, C. Hommet, and et al., "Precuneus and cingulate cortex atrophy and hypometabolism in patients with alzheimer's disease and mild cognitive impairment: Mri and 18f-fdg pet quantitative analysis using freesurfer," *BioMed Research International*, Vol. 2015, pp. 1–8, 2015.

21. Chen, Y. J., and I. M. Nasrallah, "Brain amyloid pet interpretation approaches: from visual assessment in the clinic to quantitative pharmacokinetic modeling," *Clinical and Translational Imaging*, Vol. 5, pp. 561–573, 12 2017.

22. Astrakas, L. G., and M. I. Argyropoulou, "Shifting from region of interest (roi) to voxel-based analysis in human brain mapping," *Pediatric Radiology*, Vol. 40, pp. 1857–1867, 12 2010.

23. Mevel, K., B. Desgranges, J.-C. Baron, and et al., "Which spm method should be used to extract hippocampal measures in early alzheimer's disease?," *Journal of Neuroimaging*, Vol. 21, pp. 310–316, 10 2011.

24. Ziolko, S. K., L. A. Weissfeld, W. E. Klunk, and et al., "Evaluation of voxel-based methods for the statistical analysis of pib pet amyloid imaging studies in alzheimer's disease," *NeuroImage*, Vol. 33, pp. 94–102, 10 2006.

25. Chen, F., "Pet radiomics of white matter, can be employed as a biomarker to identify the progression of mild cognitive impairment to alzheimer's disease," *Academic Radiology*, Vol. 30, pp. 1885–1886, 9 2023.

26. Song, R., X. Wu, H. Liu, and et al., "Prediction of cognitive progression in individuals with mild cognitive impairment using radiomics as an improvement of the atn system: A five-year follow-up study," *Korean Journal of Radiology*, Vol. 23, p. 89, 2022.

27. Fusco, R., V. Granata, G. Grazzini, and et al., "Radiomics in medical imaging: pitfalls and challenges in clinical management," *Japanese Journal of Radiology*, Vol. 40, pp. 919–929, 9 2022.

28. Guiot, J., A. Vaidyanathan, L. Deprez, and et al., "A review in radiomics: Making personalized medicine a reality via routine imaging," *Medicinal Research Reviews*, Vol. 42, pp. 426–440, 1 2022.

29. Salvatore, C., I. Castiglioni, and A. Cerasa, "Radiomics approach in the neurodegenerative brain," *Aging Clinical and Experimental Research*, Vol. 33, pp. 1709–1711, 6 2021.

30. Rasi, R., and A. Guvenis, "Platform for the radiomics analysis of brain regions: The case of alzheimer's disease and metabolic imaging," *Brain Disorders*, Vol. 16, p. 100168, 12 2024.

31. Du, Y., S. Zhang, Q. Qiu, and et al., "The effect of hippocampal radiomic features and functional connectivity on the relationship between hippocampal volume and cognitive function in alzheimer's disease," *Journal of Psychiatric Research*, Vol. 158, pp. 382–391, 2 2023.

32. Jagust, W. J., S. M. Landau, R. A. Koeppe, and et al., "The alzheimer's disease neuroimaging initiative 2 pet core: 2015," *Alzheimer's & Dementia*, Vol. 11, pp. 757–771, 10 2015.

33. Wyman, B. T., D. J. Harvey, K. Crawford, and et al., "Standardization of analysis sets for reporting results from adni mri data," *Alzheimer's & Dementia*, Vol. 9, pp. 332–337, 5 2013.

34. Avants, B., C. Epstein, M. Grossman, and J. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, Vol. 12, pp. 26–41, 2 2008.

35. Henschel, L., S. Conjeti, S. Estrada, and et al., "Fastsurfer - a fast and accurate deep learning based neuroimaging pipeline," *NeuroImage*, Vol. 219, p. 117012, 10 2020.

36. Fischl, B., "Freesurfer," *NeuroImage*, Vol. 62, pp. 774–781, 8 2012.

37. Kahhale, I., N. J. Buser, C. R. Madan, and et al., "Quantifying numerical and spatial reliability of hippocampal and amygdala subdivisions in freesurfer," *Brain Informatics*, Vol. 10, p. 9, 12 2023.

38. SÃ¤mann, P. G., J. E. Iglesias, B. Gutman, and et al., "freesurfer based segmentation of hippocampal subfields: A review of methods and applications, with a novel quality control procedure for enigma studies and other collaborative efforts," *Human Brain Mapping*, Vol. 43, pp. 207–233, 1 2022.

39. Huang, C., and O. Mitchell, "A euclidean distance transform using grayscale morphology decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, pp. 443–448, 4 1994.

40. Kimmel, R., N. Kiryati, and A. M. Bruckstein, "Sub-pixel distance maps and weighted distance transforms," *Journal of Mathematical Imaging and Vision*, Vol. 6, pp. 223–233, 6 1996.

41. Hatt, M., M. Vallieres, D. Visvikis, and A. Zwanenburg, "Ibsi: an international community radiomics standardization initiative," 2018.

42. van Griethuysen, J. J., A. Fedorov, C. Parmar, and et al., "Computational radiomics system to decode the radiographic phenotype," *Cancer Research*, Vol. 77, pp. e104–e107, 11 2017.

43. Probst, P., A.-L. Boulesteix, and B. Bischl, "Tunability: Importance of hyperparameters of machine learning algorithms," *The Journal of Machine Learning Research*, Vol. 20, pp. 1934–1965, 2019.

44. Elgeldawi, E., A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis," *Informatics*, Vol. 8, p. 79, 11 2021.

45. Ying, X., "An overview of overfitting and its solutions," *Journal of Physics: Conference Series*, Vol. 1168, p. 022022, 2 2019.

46. Ray, P., S. S. Reddy, and T. Banerjee, "Various dimension reduction techniques for high dimensional data analysis: a review," *Artificial Intelligence Review*, Vol. 54, pp. 3473–3515, 6 2021.

47. Pini, L., M. Pievani, M. Bocchetta, and et al., "Brain atrophy in alzheimer's disease and aging," *Ageing Research Reviews*, Vol. 30, pp. 25–48, 9 2016.

48. Brabec, J., A. Rulseh, B. Hoyt, and et al., "Volumetry of the human amygdala an anatomical study," *Psychiatry Research: Neuroimaging*, Vol. 182, pp. 67–72, 4 2010.

49. Huang, L., K. Chen, X. Hu, and Q. Guo, "Differential atrophy in the hippocampal subfield volumes in four types of mild dementia," *Frontiers in Neuroscience*, Vol. 14, 7 2020.

50. Padulo, C., C. Sestieri, M. Punzi, and et al., "Atrophy of specific amygdala subfields in subjects converting to mild cognitive impairment," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, Vol. 9, 10 2023.

51. Li, J., E. Antonecchia, M. Camerlenghi, and et al., "Correlation of [18f]florbetaben textural features and age of onset of alzheimer's disease: a principal components analysis approach," *EJNMMI Research*, Vol. 11, p. 40, 12 2021.

52. Feng, Q., Y. Chen, Z. Liao, and et al., "Corpus callosum radiomics-based classification model in alzheimer's disease: A case-control study," *Frontiers in Neurology*, Vol. 9, 7 2018.

53. Feng, F., P. Wang, K. Zhao, and et al., "Radiomic features of hippocampal subregions in alzheimer's disease and amnestic mild cognitive impairment," *Frontiers in Aging Neuroscience*, Vol. 10, 9 2018.

54. Wang, Z., M. Xia, Z. Dai, and et al., "Differentially disrupted functional connectivity of the subregions of the inferior parietal lobule in alzheimer's disease," *Brain Structure and Function*, Vol. 220, pp. 745–762, 3 2015.

55. Baran, T. M., and F. V. Lin, "Amyloid and fdg pet of successful cognitive aging: Global and cingulate-specific differences," *Journal of Alzheimer's Disease*, Vol. 66, pp. 307–318, 10 2018.

56. Du, Y., S. Zhang, Y. Fang, and et al., "Radiomic features of the hippocampus for diagnosing early-onset and late-onset alzheimer's disease," *Frontiers in Aging Neuroscience*, Vol. 13, 1 2022.

57. Ge, T., M. Reuter, A. M. Winkler, and et al., "Multidimensional heritability analysis of neuroanatomical shape," *Nature Communications*, Vol. 7, p. 13291, 11 2016.

58. Shi, J., N. Lepore, B. A. Gutman, and et al., "Genetic influence of apolipoprotein e4 genotype on hippocampal morphometry: An surface based alzheimer's disease neuroimaging initiative study," *Human Brain Mapping*, Vol. 35, pp. 3903–3918, 8 2014.

59. Guiot, J., A. Vaidyanathan, L. Deprez, and et al., "A review in radiomics: Making personalized medicine a reality via routine imaging," *Medicinal Research Reviews*, Vol. 42, pp. 426–440, 1 2022.