SIMILARITY-BASED ANALYSIS OF FDG-PET IMAGES OF ALZHEIMER'S DISEASE PATIENTS: A METHOD FOR AUTOMATED DIAGNOSIS AND SEVERITY PREDICTION WITH THE AIM OF THERAPY RESPONSE MONITORING

by

Ceren Yüksel

B.S., in Medical Engineering, Acibadem Mehmet Ali Aydinlar University, 2019

Submitted to the Institute of Biomedical Engineering
in partial fulfillment of the requirements
for the degree of
Master of Science
in
Biomedical Engineering

Boğaziçi University
2022

ACKNOWLEDGMENTS

I would like to express my deepest thanks to my thesis advisor, Prof. Dr. Albert Guvenis for his guidance, mentorship, and advice, which still shapes my career. I would like to thank Assoc. Prof. Daniela Schulz and Assoc. Prof. Mehmet Kocaturk for their contributions to my career and for taking the time to be on my defense committee.

I would like to thank my mom Huriye, dad Ferit, sister Aslihan, brother Oguzhan and nephew Mete for their endless support, curiosity, and understanding.

I would to thank college friends for their contributions and time in answering my questions, and for their support and friendship.

This study has been accepted by the Turkish Physical Society-38th International Physics Congress (TPS-38) 2022 (**http://tfd38.org**/).

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, Ceren. Yuksel, 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

SIMILARITY-BASED ANALYSIS OF FDG-PET IMAGES OF ALZHEIMER'S DISEASE PATIENTS: A METHOD FOR AUTOMATED DIAGNOSIS AND SEVERITY PREDICTION WITH THE AIM OF THERAPY RESPONSE MONITORING

This study aimed to evaluate 18-Fluorodeoxyglucose positron emission tomography (18F-FDG-PET) images of the brain for the computer-aided characterization detection of Alzheimer's disease (AD) intuitive image similarity-measure-based approach. The first objective was to diagnose AD automatically. The second objective was to determine the association between the similarity measure and neuropsychological assessments. Therefore, we aimed to develop a new AD evaluation algorithm that can give an early diagnosis of the disease and define an objective severity index that correlates with well-known neuropsychological tests. 125 patients with AD, 132 Cognitively Normal (CN), and a total of 257, FDG-PET data were obtained from ADNI. We then found a distance value indicative of the similarity between any 3D image to available CN and AD patient images in the database using the mutual information method. The diagnosis was based on a threshold value for the distance value. Then, the Mini Mental State Examination (MMSE) and Clinical Dementia Rating (CDR) results of all patients and the distance values obtained from FDG-PET were analyzed using an analysis of variance. The algorithm achieved an AUC ROC of 0,969 using a leave-one-out method for the original dataset (n=197) and 0,873 using the independent testing dataset (n=60). The correlation was 0,642 between MMSE scores and imaging scores, and for CDR global test correlation between imaging and testing was 0,677. A simple and intuitive similarity-based algorithm can be used for the early detection of AD using molecular imaging as well as determining an objective severity index. No ROI and feature computations should be performed.

Keywords: Alzheimer's disease, 18F-FDG-PET, Similarity Index, Neuropsychological Assessments, Severity Index.

ÖZET

ALZHEİMER HASTALIĞININDA FDG-PET GÖRÜNTÜLERİNİN BENZERLİGE DAYALI ANALİZİ: TEDAVİ YANITININ İZLENMESİ AMACIYLA OTOMATİK TANI VE ŞİDDET TAHMİNİ İÇİN YÖNTEM

Bu çalışma, Alzheimer hastalığının bilgisayar destekli karakterizasyon tespiti ile beynin Florodeoksiglukoz pozitron emisyon tomografisi (18F-FDG PET) görüntülerini benzerlik ölçüsüne dayalı yaklaşım kullanarak değerlendirmeyi amaçlamıştır. İlk amaç, hastalığı otomatik olarak teşhis etmekti. İkinci amaç benzerlik ölçüsü ile nöropsikolojik değerlendirmeler arasındaki ilişkiyi belirlemekti. Böylelikle, hastalığın erken teşhisini sağlayabilecek ve iyi bilinen nöropsikolojik testlerle ilişkili objektif bir hastalık şiddet indeksi tanımlayabilen yeni bir alzheimer hastalığı değerlendirme algoritması geliştirmeyi amaçladık. ADNI'den 125 hasta, 132 sağlıklı ve toplam 257, FDG-PET verisi olan yüz yirmi beş hasta görüntüsü alındı. Daha sonra, karşılıklı bilgi yöntemini kullanarak veritabanındaki mevcut sağlıklı ve hasta görüntülerine herhangi bir 3B görüntü arasındaki benzerliği gösteren bir mesafe değeri bulduk. Tanı, mesafe değeri için bir eşik değerine dayanıyordu. Daha sonra tüm hastaların mini mental durum ve klinik demans derecelendirme testi sonuçları ve FDG-PET'den elde edilen mesafe değerleri varyans analizi kullanılarak analiz edildi. Algoritma, orijinal veri seti (197 hasta) için birini dışarıda bırak yöntemi kullanılarak 0,969 ve bağımsız testing veri seti (60 hasta) kullanılarak 0,873 AUC elde etti. MMSE ve görüntüleme skorları arasında korelasyon 0,642, görüntüleme ve CDR global test arasındaki korelasyonu ise 0,677 idi. Moleküler görüntüleme ile AD'nin erken saptanması ve objektif önem indeksinin belirlenmesi için basit ve görüntü benzerlik temelli bir algoritma kullanılabilir. İlgi alanı ve özellik hesaplamaları yapılmasına gerek yoktur.

Anahtar Kelimeler: Alzheimer hastalığı, 18F-FDG-PET, benzerlik indeksi, nöropsikolojik değerlendirmeler, hastalık şiddet indeksi

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

M_j	AD patient
N_j	CN patient
k	Selected number of AD and CN patients

LIST OF ABBREVIATIONS

AD	Alzheimer Disease
ADNI	Alzheimer Disease Neuroimaging Initiative
CAD	Computer Aided Diagnosis
CT	Computed Tomography
FDG	Fluoro-Deoxy Glucose
MI	Mutual Information
MMSE	Mini Mental State Exam
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
CN	Cognitively Normal
ROI	Region of Interest
AI	Artificial Intelligence
ANN	Artificial Neural Network
CDR	Clinical Dementia Rating Scale
MMSE	Mini-Mental State Examination
SPSS	Statistical Package for the Social Sciences
ANOVA	Analysis of Variance
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
PCCs	Pearson Correlation Coefficients
CIs	Confidence Intervals
ML	Machine Learning
MCI	Mild Cognitive Impairment

1. INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease that progresses over time and eventually leads to death by destroying memory and other cognitive abilities [1]. Even though the cause of the disease has not been understood clearly yet, it has been viewed that it is partially characterized by the extracellular deposition of misfolded amyloid beta plaques and the intracellular formation of neurofibrillary tangles in the brain [2]. Over 135 million people are expected to have Alzheimer's disease by the year 2050 [3]. Moreover, the annual cost of AD care is estimated to increase from $335 billion to $1 trillion over the next 15 years [4]. Given these facts, finding a way to halt the course of Alzheimer's disease is critical.

Some exams to diagnose Alzheimer's disease are performed by physicians using tools such as neuropsychological tests, and neurological and brain imaging tests [5]. When the patient is examined by a doctor, physicians review the patient's family and medical history. Then, they perform physical and neurological exams. The doctor examines the patient's balance, and reflexes during these examinations. In addition, clinicians use neuropsychological tests to check memory, reasoning, and rudimentary problem-solving skills. **http://www.alz.org/alzheimers-dementia/diagnosis/**.

In addition, physicians carry out brain scans using MRI (Magnetic Resonance Imaging), CT (Computerized Tomography), and PET. The loss of volume of the brain can be seen using MRI [6]. 18-Fluorodeoxyglucose positron emission tomography (FDG-PET) is another imaging technique. FDG-PET shows the density of glucose uptake by tissues, it is preferred in Alzheimer's disease, because it is very sensitive and gives quantitative results. FDG-PET images have shown promise in revealing Alzheimer's disease metabolic patterns. In fact, they show metabolic alterations to the brain. Moreover, it can be predicted the progression from cognitively normal (CN) to Alzheimer's disease by using FDG-PET [7].

1.1 Computer-Aided Diagnosis of AD Using FDG-PET and Artificial Intelligence

Today, computer-aided diagnostic (CAD) systems are implemented frequently. Numerous researchers have tried to create an automated CAD system to track the functional changes caused by AD. These systems aim to assist doctors in making a diagnosis. The probability of an early diagnosis of the disease increases by using these systems [8]. Moreover, computer-aided diagnosis is used to manage a patient with a genetic risk [9]. CAD can be considered a second opinion for physicians.

Artificial intelligence (AI) is frequently used to detect Alzheimer's disease [10]. In the literature, to detect Alzheimer's disease from imaging tests, support vector machine (SVM), artificial neural network (ANN), and deep learning are used in Alzheimer's studies [11].

Making deep learning models more interpretable is becoming increasingly important as they gain cutting-edge performance in a variety of fields. Two primary factors affect interpretability. First, a model with breakthrough performance might have found data patterns that experts in the field would like to know more about. However, if the model is a black box, this would not be possible. Second, trust depends on interpretability. It's crucial to check that a model is making conclusions based on solid evidence and isn't concentrating on a data artifact while making a medical diagnosis. Physicians want to know and interpret the decision structure of artificial intelligence, but the decision mechanism is not fully understood. In addition, artificial intelligence is prone to error. Algorithms often result in about 5% error [12]. This score carries risks in the clinic, so the doctor cannot fully trust artificial intelligence [13].

The severity score of Alzheimer's disease is very important. This way, we can have a classification of the disease. Tracking the development of the illness and the effectiveness of treatment can both be done with the help of an objective severity index

with an objective figure obtained from FDG-PET images [7].

Besides, artificial intelligence divides the disease into classes but does not give an objective degree of severity [14]. In addition, since early diagnosis of the disease is very important, we give the disease severity index by looking at the metabolic activity of the brain with FDG-PET in this study [15].

1.2 Correlation Between Imaging Method And Neuropsychological Assessments

The regression of neuropsychological assessment and imaging tests is important in terms of understanding the reliability of the tests. Revealing the relationship between the FDG-PET image and the assessment results will convey the sustainability of the test and increase the reliability of diagnosing results.

In this study, we aimed to assess FDG-PET images of the brain and we tested whether there is a regression between brain images and neuropsychological assessments. Additionally, we tested how strong a regression is. We first found an Alzheimer's Disease similarity index for diagnosing patients with mutual information (MI) method using patients' FDG-PET images. Then, we monitored an objective severity index by finding the association between the similarity index and neuropsychological assessment.

This study is very important for the clinic. By using mutual information, we increase the interpretability of disease and we give severity score for disease with distance value. Moreover, we correlate severity scores and neuropsychological test scores to increase reliability. There is not enough effort to both increase reliability and understanding of the disease area.

2. LITERATURE REVIEW

Machine learning algorithms that do diagnoses using information from PET scans have been published in the literature. Moreover, in contrast to traditional machine learning techniques, deep learning algorithms are particularly effective at identifying a variety of Alzheimer's disorders. Medical images are entered into CNN models, which then perform the classification process to produce the results, without the need for feature engineering. For diagnosing Alzheimer's and other dementia diseases, deep learning models are divided into 2D and 3D image processing methods. MRI and PET scans are frequently used in the generated models. According to certain research, [16], in comparison to the 3D CNN that receives images of the complete brain, the 2D CNN model that receives images of the brain's surface perfusion performs a classification task better. On the other hand, three-dimensional methods have also been widely employed [17],[18]. These studies use volumetric brain scans as the input and voxels as the processing unit. Additionally, was said that the performance was enhanced by the use of numerous imaging modalities [19].

In another recent study, it was demonstrated that our model is responsive to the underlying metabolic processes using explainable deep neural networks. They employed 520 AC-PC aligned FDG-PET images from AD and healthy control (CN) groups from the ADNI cohort. The deep neural network (ResNet, 3.4 million parameters) used to classify each scan is then given these data. The average accuracy of the algorithm was 89.2 percent [20].

In a comprehensive study, using brain FDG-PET data from the ADNI, researchers attempted to develop a deep learning (DL) model that could distinguish between Alzheimer's disease (AD) and healthy controls (CN). In order to distinguish between individuals with dementia and those without dementia, the model was immediately applied to a cohort of patients who had been prospectively enrolled. AUC-ROC results were 0.75 [21].

Additionally, researchers developed a special method that takes into account the diagnosis interpretability for the early detection of AD using multi-scale discriminative regions in FDG-PET imaging. A multi-scale region localization (MSRL) module is specifically presented in order to automatically and unsupervisedly find disease-related discriminative regions in full-volume FDG-PET images. This information is used to generate a confidence score that evaluates how regions should be prioritized based on the density distribution of abnormalities. Using 146AD + 184CN FDG-PET images, accuracy was 97.8 [22].

In a recent study that was published in 2020, the association between FDG absorption and cognitive evaluation in Alzheimer's disease was examined. In conclusion, brain FDG uptake in the AD population is only weakly correlated with the neuropsychological evaluation, indicating a negligible influence on the statistical analysis of brain glucose metabolism [23].

In a study published in 2021, according to research that examined the relationship between brain 18F-AV45 and 18F-FDG PET distribution characteristics and performed a positive correlation between neuropsychological tests, the range and degree of reduced FDG metabolism in different regions were proposed to be positively correlated with the overall score of the MMSE or MOCA [24].

3. PROBLEM STATEMENT

THIS THESIS CONSISTS of 2 steps. The first one is automatically detecting AD and CN, the second one is finding an association between disease scores and neuropsychological tests. For the first one, past studies were based mostly on machine learning and deep learning based. However, support vector machines, artificial neural networks, convolutional neural networks, and random forests are the classification methods most frequently utilized in AD. Black box algorithms are a term that describes machine learning (ML) techniques. In ML learning, the computer just predicts that the animal is a unicorn without offering any explanation [25]. Moreover, finding a correlation between FDG-PET and test studies shows that there is a positive correlation. However, couldn't find a good correlation between them in previous studies.

In this study, we used FDG-PET because the error rate of FDG-PET is very low because of objectiveness. FDG-PET can be used to determine whether a therapeutic approach, such as medication therapy, is effective or to determine whether the disease is advancing or not. Also, we can diagnose AD automatically, and determine the association between the similarity measure and neuropsychological assessments. Therefore, aimed to develop a new AD evaluation algorithm that can give an early diagnosis of the disease as well as can define an objective severity index that correlates with well-known neuropsychological tests.

4. MATERIALS AND METHODS

4.1 Participants

The Alzheimer's Disease Neuroimaging Initiative (ADNI) database was utilized to gather the information for this article. A total of 257 patients (cognitively normal = 132; Alzheimer's patient = 125) with a mean age of 71.7 ±6.7 years, were prospectively selected based on FDG-PET images and neuropsychological assessments. ADNI is a multicenter longitudinal study aimed at developing clinical, imaging, genetic, and biochemical indicators for Alzheimer's disease early diagnosis and tracking [26]. The ADNI 1, ADNI-GO, ADNI 2, and ADNI 3 studies are all aimed at discovering the relationships between clinical, cognitive, imaging, genetic, and biochemical biomarkers across the Alzheimer's disease spectrum. In this article, we used 257 patient data from ADNI 1.

4.2 FDG-PET Data Acquisition

FDG-PET, which monitors glucose metabolism in the brain, and pilot research employing amyloid PET, which uses radioactive compound pictures, were obtained from specific locations and gathered at the University of Michigan's center. All images that are used in this article are 3D images and in DICOM format. The images are with a maximum grey level of 32700, and have 96 slices for each patient. 185 MBq (5 mCi), dynamic 3D scan of six 5-min frames 30-60 min post-injection. The University of Michigan is in charge of data pre-processing. Data is co-registered, averaged, standardized image and voxel size, and consistent resolution dataset after pre-processing. Following preprocessing, post-processing is done in Utah by the Foster Group. Using the NeuroStat package, all images were under the process named standardized to baseline FDG. With the same package, all images are warped into Talairach space. Finally, the image intensities are normalized once more. The ADNI website has comprehensive

information on FDG-PET capture and pre-processing.

4.3 Calculation Of 3D-mutual Information

This study used the degree of similarity between a specific image and the other images in the database to evaluate patient similarity. We calculated the mutual information for all 96 slices of FDG-PET images using kernel density estimation with a Gaussian kernel [27]. The mutual information concept is a method in computer-aided diagnosis, and it is a measurement of two variables' mutual dependency [28]. The similarity index is calculated based on mutual information to detect similarities between two images. The formula for calculating the mutual information is as follows:

$$I(X;Y) = H(X) - H(Y) = H(X) + H(Y) - H(X,Y) \tag{4.1}$$

$$I(X;Y) = \Sigma_x \Sigma_y p(x,y) log \frac{p(x,y)}{P(x)p(y)} \tag{4.2}$$

The two random variables are X and Y, and their respective probability density functions, $P(X)$ and $P(Y)$, are determined by the pixel values of the two images, X and Y. Their distributions were discovered using the X and Y pictures' intensity histograms. The two images' combined probability density function, $P_{XY}(X,Y)$, was calculated using the pixel values of the images.

To find a similarity index, we calculated mutual information of all same slices from each pair of patients. The result, calculations from 257 patients and 96 slices, we got 38612×38612 computing. After that, we found a MI score from the average of these slice-based calculations, so we obtained a 3D-Mutual Information score.

4.4 The Distance Value

The decision index was calculated using mutual information scores, for the decision algorithm of computer-aided diagnosis. The formula for calculating the distance value is as follows:

$$D(Qi) = \frac{1}{k}\Sigma_{j=1}^{k} MI(Qi, Mi) - \frac{1}{k}\Sigma_{y=1}^{k} MI(Qi, Ni) \tag{4.3}$$

In the formula, M_j is AD patient, and N_j is CN patient, k is the selected number of AD and CN patients. After calculating MI for all slices, we computed distance values for the selected Q image compared with others. We repeated this process for every image.

4.5 Neuropsychological Assessments

Neuropsychological assessments are tests by which memory and thinking skills are measured. The severity and presence of cognitive dysfunction can be measured, with these tests [29]. In this study, we used Clinical Dementia Rating Scale (CDR) and Mini-Mental State Examination (MMSE) neuropsychological assessments. A widely used test of cognitive performance for the elderly is the mini-mental state examination, which includes assessments of orientation, attention, memory, language, and visual-spatial abilities. A score of 20 to 24 indicates mild dementia, a score of 13 to 20 indicates moderate dementia, and a score of less than 12 indicates severe dementia [30].

The Clinical Dementia Rating is based on evaluating six distinct cognitive and behavioral domains, including memory, orientation, problem-solving and judgment

skills, performance in communal, domestic, recreational activities, and personal care. The CDR was initially created as a staging tool to divide dementia severity into normal, dubious, mild, moderate, and severe categories. Clinicians conduct a semi-structured interview with the subject as well as a trustworthy informant or collateral source before rating the severity of the symptoms across six domains. There are three functional domains and three domains of cognition (community affairs, personal care, and home & hobbies) (orientation, memory, judgment, and problem-solving). Five levels of impairment are described by the response alternatives for each area: 0 = None, 0.5 = Questionable (not present in the personal care domain), 1 = Mild, 2 = Moderate, and 3 = Severe. The dementia stage is identified by the CDR Global score, which ranges from 0 to 3. Scores for the Sum of Boxes, which range from 0 to 18, are a continuous indicator of dementia severity [31].

4.6 Statistical Analysis

All statistical analyses were performed with IBM SPSS (Statistical Package for the Social Sciences, version 28.0.1) and Excel (version 2022) which is used to determine the numerical distribution of data. Mean values, standard deviation, coefficient of variation, and minimum, and maximum values were calculated for demographic and test variables. An association of similarity index with neuropsychological assessment results is investigated. Associations were investigated by correlation and analysis of variance (ANOVA) in SPSS and Excel.

First, all statistical tests were based on a significance level of $p < 0.05$. Demographic and clinical characteristics were analyzed using the correlation test for differences between CN and AD patients.

Second, the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) are analyses for discrimination between CN and AD. The ability of a model to discriminate between classes can be determined and informed by using the AUC-ROC statistic.

The ROC curve, also known as a probability curve, and AUC, which stand for the level of measurement of separability, respectively, show how well the model can vary across classes [32]. In a ROC curve, when two curves completely avoid each other, the model's separability is at its maximum. It can tell the difference between a positive class and a negative class perfectly. The ROC curve shows the progression of the true positive rate against the false positive rate at various threshold values. On the ROC curve, TPR is plotted against FPR with FPR on the x-axis and TPR on the y-axis.

A table called a confusion matrix is used to describe how well the ROC measurement method, calculates the False Negative (FN) and False Positive (FP) ratios on the image of the segmentation by comparing the results of the test image segmentation on the original image, to ensure that a segmentation application has a sufficient level of accuracy. TN is a true negative (truth-value between pictures of detected segmentation as non-background) with a background in the reference image), and TP is a true positive (truth-value between image result of segmentation identified as a foreground with reference image foreground). False Positive (FP) stands for false positive, and False Negative (FN) for false negative [33]. The formula for calculating the true positive rate, false negative rate, true negative rate, and false negative rate is given as follows:

$$TPR = \frac{TP}{ActualPosition} = \frac{TP}{TP + FN} \tag{4.4}$$

$$FNR = \frac{FN}{ActualPosition} = \frac{FN}{TP + FN} \tag{4.5}$$

$$TNR = \frac{TN}{ActualPosition} = \frac{TN}{TN + FN} \tag{4.6}$$

$$FPR = \frac{FP}{ActualPosition} = \frac{FP}{TN + FP} \tag{4.7}$$

4.7 Computation Time

The complexity of computing mutual information and its high computational time for large datasets led us to implement a GPU-based method using PyTorch (an open-source machine learning framework that accelerates the path from research prototyping to production deployment). To compute histograms, our implementation uses kernel density estimation with a Gaussian kernel. This implementation uses kernel density estimation with a Gaussian kernel to compute histograms. It decomposes the multivariate kernel into the product of each univariate kernel using the diagonal bandwidth matrix [34],[35],[36]. Based on the dataset size we need to compute 100x100x96 mutual information for training the model, so we were able to reduce the computation time of each mutual information from 63 seconds to 0.018 seconds on a computer with a Corei7 Gen10th processor. RTX206 VGA card (1920 CUDA cores with 240 tensors) and 16 GB DDR5 RAM.

5. RESULTS

5.1 Demographic

Table 5.1 provides the mini-mental state examination scores and demographic information for the datasets utilized in this investigation, as well as the training and testing set distributions. The dataset included 257 cases including 125 AD and 132 CN from ADNI. The average age of the patients was 75,368 ±7,224, and total average MMSE score is 26,17 ±3,629.

Table 5.1
Demographics and mini-mental state examination scores of the datasets.

Clinical diagnosis	No. cases	Male/Female	Age (mean ±SD)	MMSE score
Total Data Set:				
AD	125	74/51	75,244 ±7,012	24,36 ±3,56
CN	132	58/74	75,492 ±6,874	28,95 ±1,124
Total	257	132/125	75,368 ±7,224	26,17 ±3,629
training Set:				
AD	95	57/38	76,795 ±7,693	23,431 ±2,142
CN	102	62/40	77,328 ±4,849	28,941 ±1,115
Total	197	119/78	77,061 ±6,974	26,28 ±3,23
testing Set:				
AD	30	17/13	73,693 ±8,053	22,66 ±1,174
CN	30	18/12	73,656 ±7,203	29 ±4,808
Total	60	35/25	73,675 ±7,174	25,83 ±4,71

5.2 Histogram

The histogram graph of the disease severity index (distance value) of both AD and CN patients calculated in the study is given in Figure 5.1 as can be seen from the histogram graph, the disease severity index of 98.7% of healthy patients has below zero disease severity index. 88,5% of AD patients have distance value scores above 0. The most optimal threshold value separating CN and AD images turned out to be -0.000402052.

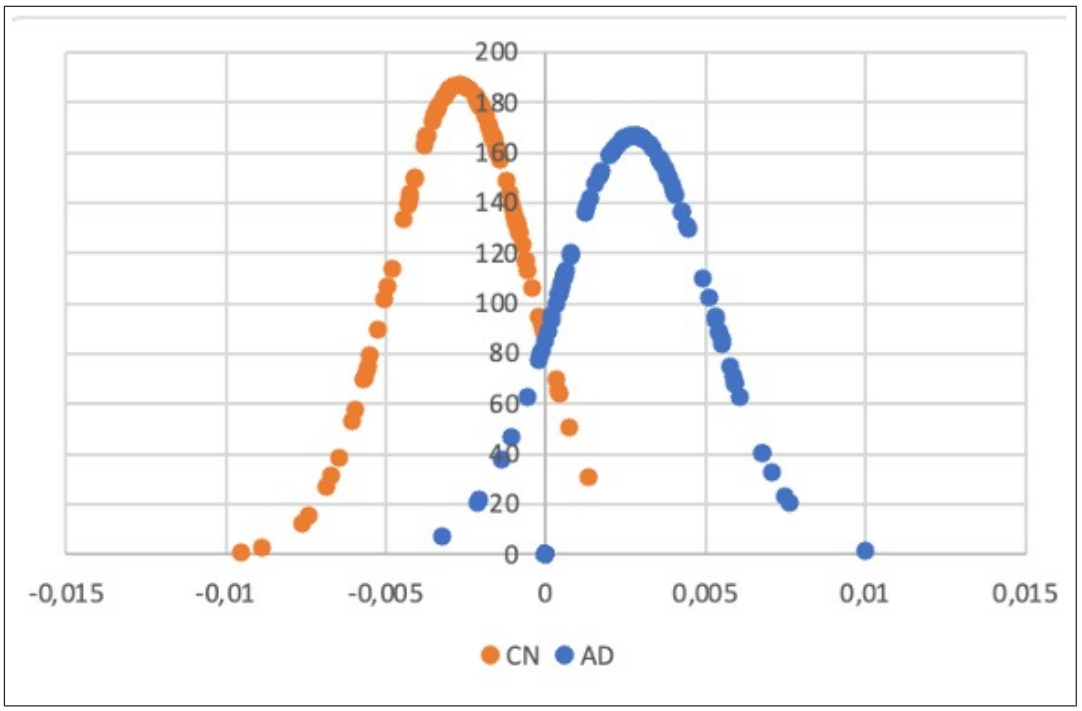

Figure 5.1 Histogram of disease severity index of AD and CN patients.

5.3 Evaluations

The ROC curve is a two-dimensional representation of the performance of a classifier. A common method used to calculate a classifier's performance values is the area under the ROC curve, called AUC [37].

The ROC curves of distance values of AD and CN are shown in Figure 5.2 the highest accuracy of 96,9% was achieved with 197 samples, and the accuracy was computed by not including the remaining 60 samples of the training set. In the testing results, the ROC curve turned out to be similar to the training data shown in Figure 5.2 The accuracy for CN and AD testing data was 87,3% output.

An excellent model has an AUC close to 1, indicating that it has a high level of separability [38]. In this study, our AUC for the prediction of AD, and CN was 0,969 and 0.873 for the testing and training data. This means that it has a good measure of separability.

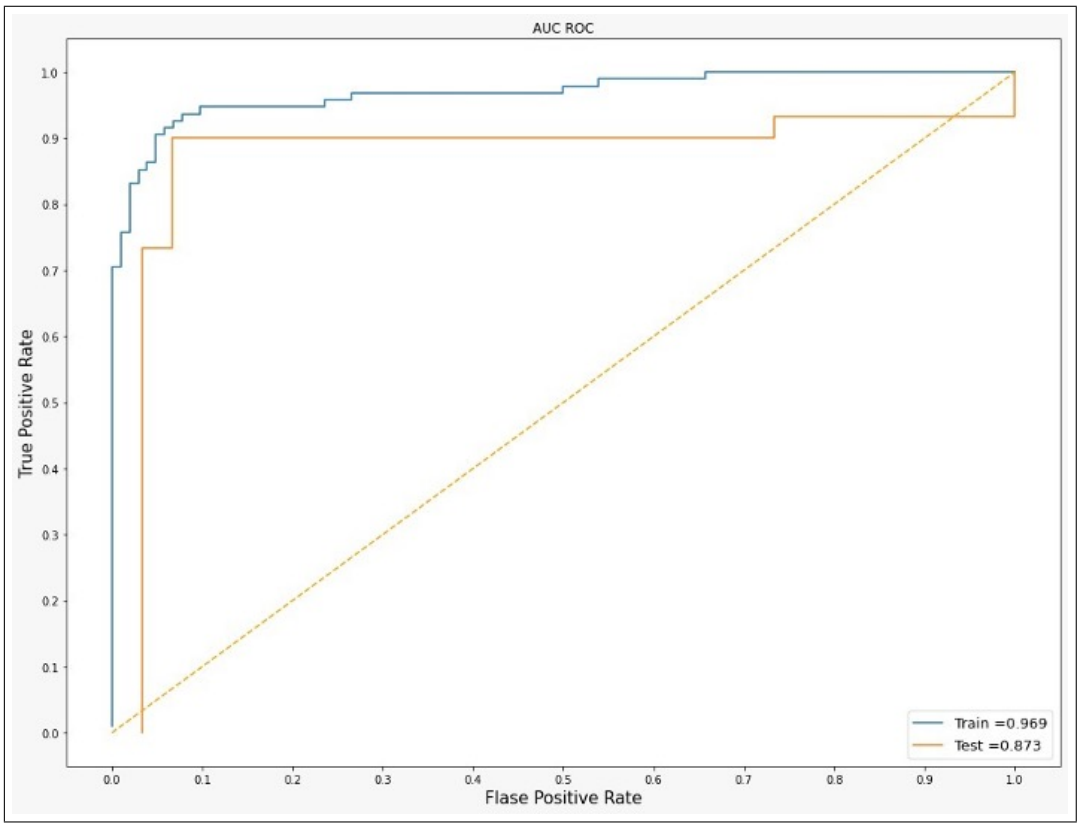

Figure 5.2 ROC Curve of training and testing data.

5.4 MMSE Assessment

One method used in cognitive neurology for Alzheimer's Disease screening is the Mini-Mental State Examination. The MMSE is designed to assess overall cognitive functioning and is frequently used in clinical settings, particularly for Alzheimer's disease screening [39]. Alzheimer's identification with the MMSE rating between 0 and 30. A normal cognitive result is obtained when the MMSE score is 25 and above and all CN patients in this study turned out to be 25 and over as expected in the training data, in AD patients, 62.5% had an MMSE score between 0 and 24, and the remaining by 37.5% had an MMSE score, 25 and over. Moreover, when we looked at the results in the testing data, 66.7% of AD patients had an MMSE score between 0 and 24, while all CN patients had a score of 25 and above.

To investigate the relationship between the MMSE scores and the disease severity index (distance value), Pearson correlation coefficients (PCCs) and ANOVA were

computed. Data has a normal distribution. Estimated regression coefficients, their corresponding 95% Confidence Intervals (CIs), standard error, and p-values were summarized in Table 5.2.

Table 5.2
Regression coefficients, standard error, corresponding 95% CIs and p-values of MMSE scores and distance value.

	Coefficients	Standard Error	P-value	Lower 95 (%)	Upper 95 (%)
Intercept	262,090,122	0,178244849	4.26E-197	258,574,543	265,605,702
X Variable 1	-5,857,185	5,034,581,484	4.76E-20	-68,501,714	-48,641,985

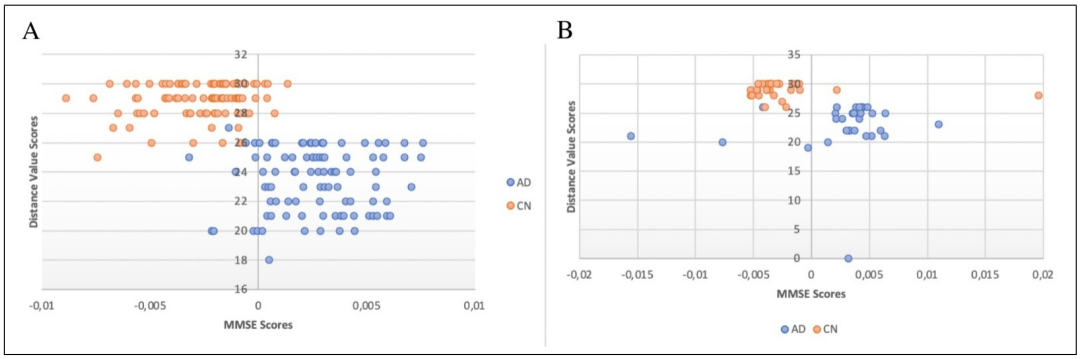

Figure 5.3 MMSE Scores of training and testing data. (A) MMSE Scores of the training. (B) MMSE Scores of the testing.

Table 5.3
ANOVA Analysis of MMSE score and distance value.

	df	SS	MS	F	Significance F
Regression	1	8,380,940,717	8,380,940,717	135,347,778	4.76E-20
Residual	196	1,195,085,415	6,192,152,412		
Total	197	2,033,179,487			

The statistical difference PCCs for each severity index and the MMSE domain scores for testing and training data were also investigated by employing the identical statistical tests used in the correlation analyses described above (Figure 5.3A, 5.3B) Also, the regression coefficient (r) is 0,642 for training data and 0,38 for testing data as shown in Figure 5.4. Moreover, the statistical model used in this work is based on an analysis of variance (ANOVA) to assess the association between each disease severity index and the MMSE scores. The relative comparisons among degrees of freedom (df), sum-of-squares (SS), mean squares (MS), F value, and significance F were also included in ANOVA analyses (Table 5.3).

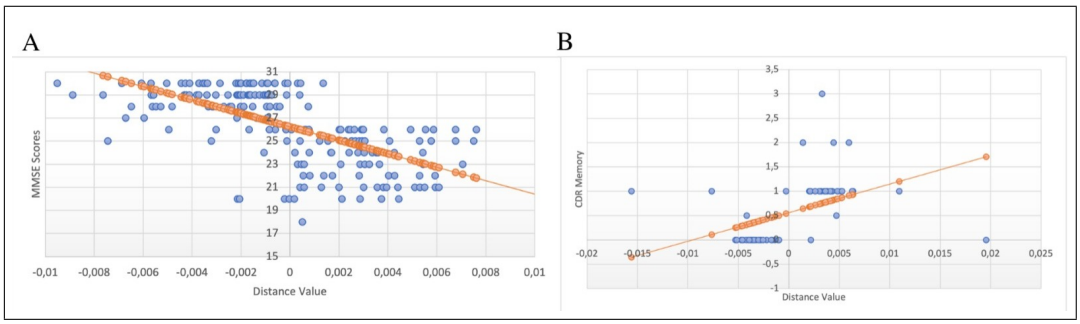

Figure 5.4 Regression plots representing the relationship between MMSE and Distance Value. (A) Relationship between MMSE and Distance Value of training data (r = 0.642, p < 0.05). (B) Relationship between MMSE and Distance Value of testing data (r = 0.38, p < 0.05).

5.5 CDR Assessment

The CDR domains in this study are CDR Global, CDR-SOB, CDR-Orient, and CDR-Memory tests. The CDR rates only impairments due to cognitive deficits rather than physical disabilities. CDR Global score is about the severity of the illness scale, so it is an important parameter in the detection of early diagnosis [40]. It is based on assessments of how well the patient functions in six areas that are frequently impacted by Alzheimer's disease. CDR Global Test gave good results are shown in Figure 5.5A, all CN patient has a '0' score (None) as expected, and Alzheimer's patients have shown varying scores between 0 and 1 [41].

Moreover, Pearson correlation coefficients were obtained to examine the relationship between the disease severity index (distance value) and the CDR Global scores. Estimated regression coefficients, their corresponding 95% CIs, standard error, and p-values were summarized in Table 5.4.

Table 5.4
Regression coefficients, standard error, corresponding 95% CIs and p-values of CDR Global and distance value.

	Coefficients	Standard Error	P-value	Lower 95 (%)	Upper 95 (%)
Intercept	0,39415085	0,022998265	9.13E-37	0,34879358	0,43950813
X Variable 1	839,410,996	6,525,949,415	8.11E-24	710,705,957	968,116,034

Disease severity indexes were highly correlated with CDR scores ranging from

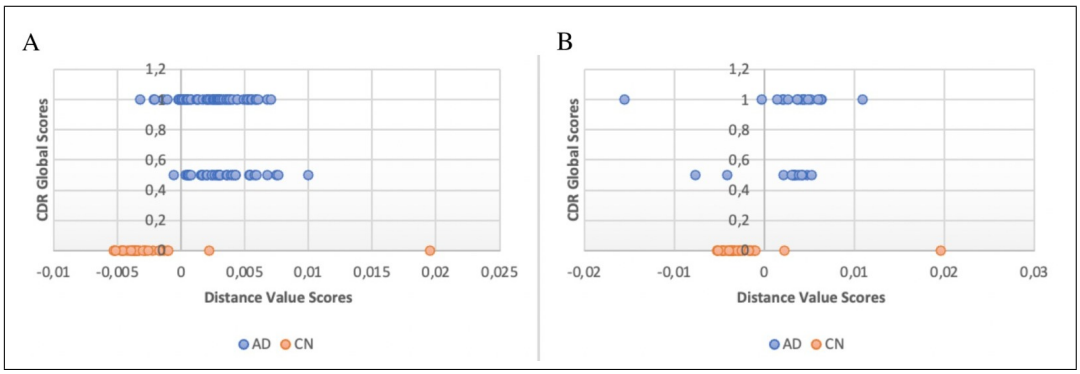

Figure 5.5 CDR Global scores of training and testing data. (A) CDR Global Scores of the training. (B) CDR Global Scores of the testing.

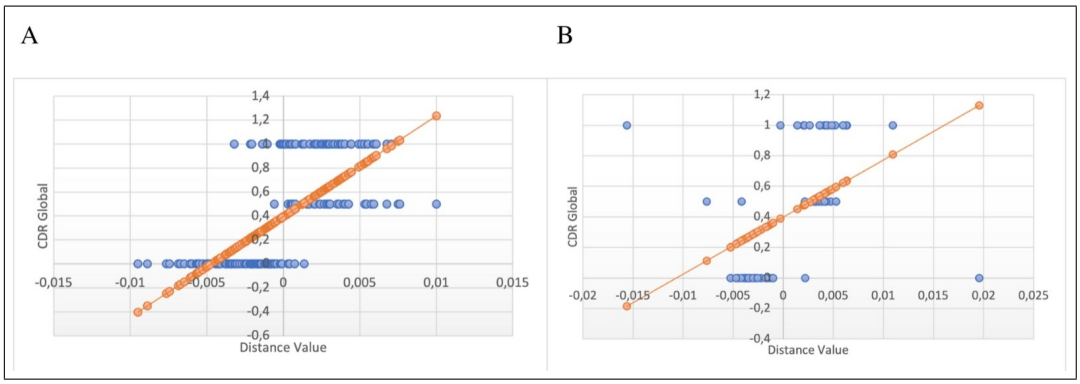

Figure 5.6 Regression plots representing the relationship between CDR Global Scores and Distance Value. (A) Relationship between CDR Global Scores and Distance Value of training data ($r = 0.677$, $p < 0.05$). (B) Relationship between CDR Global Scores and Distance Value of testing data ($r = 0.51$, $p < 0.05$).

0 to 1. The regression coefficient (r) is 0,677 for training and 0,51 for testing data. As shown in Figure 5.6 and Table 5.5, has a normal distribution and there was a strong positive correlation between disease severity index, assessed with the CDR Global score. Moreover, we performed an analysis of variance analyses to evaluate the association between each disease severity index and the CDR Global Score. ANOVA results are given in Table 5.5.

The Sum of Boxes score is a continuous measure of dementia severity and ranges from 0-18 [40]. Pearson correlation coefficient is 0,672. As a result, the CDR-SOB test is a total of tests that gave a good correlation with the disease severity index. Estimated regression coefficients, their corresponding 95% CIs, standard error, and p-values of CDR-SOB were summarized in Table 5.6.

Table 5.5
ANOVA of CDR Global.

	df	SS	MS	F	Significance F
Regression	1	1,723,255,937	172,325,594	165,448,145	8.11E-24
Residual	196	2,031,058,784	0,10415686		
Total	197	3,754,314,721			

Table 5.6
Regression coefficients, standard error, corresponding 95% CIs and p-values of CDR-SOB and distance value.

	Coefficients	Standard Error	P-value	Lower 95 (%)	Upper 95 (%)
Intercept	217,644,416	0,127276059	1.19E-36	19,254,298	242,745,852
X Variable 1	45,780,336	3,611,564,319	2.99E-23	386,575,942	529,030,777

Moreover, we performed ANOVA analyses to assess the association between each disease severity index and the CDR-SOB Score. The relative comparisons among degrees of freedom, sum-of-squares, mean squares, F value, and significance F were also included in ANOVA analyses (Table 5.7).

Table 5.7
ANOVA of CDR-SOB.

	df	SS	MS	F	Significance F
Regression	1	512,576,187	512,576,187	160,682,005	2.99E-23
Residual	196	622,050,716	319,000,367		
Total	197	11,346,269			

Moreover, the CDR Memory score has a high correlation solely (rtraining= 0,70217454 and rtesting= 0,512) as seen in Figure 5.8. The correlation studies were then carried out with a population of participants with CDR Memory scores ranging from 0 to 1 (mild dementia) in order to acquire insight into the initial stages of cognitive decline [42]. Disease severity index scores in these two subject groups were still significantly correlated with cognitive status, with a confidence interval of 0.446-0.559 (Table 5.8 and Figure 5.7).

Furthermore, the CDR Orient score has also a good correlation solely (rtrain-

ing= 0,643 and rtesting= 0.501) as shown in Figure 5.10.

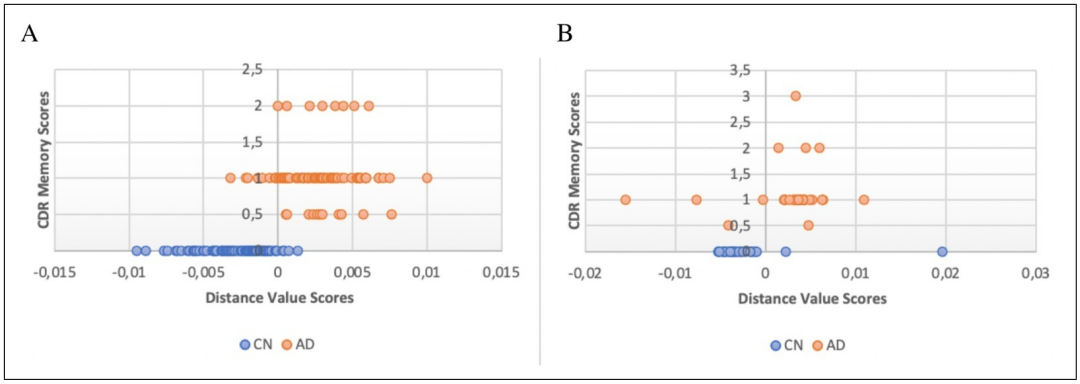

Figure 5.7 CDR-Memory Scores of training and testing data. (A) CDR Memory Scores of the training. (B) CDR Memory Scores of the testing.

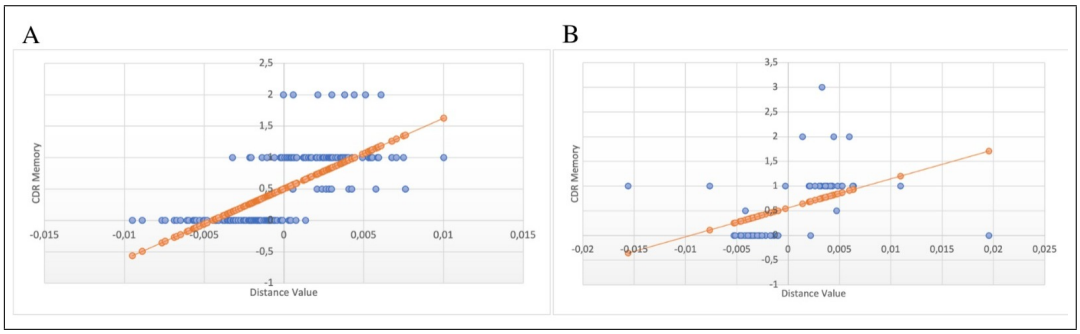

Figure 5.8 Regression plots representing the relationship between CDR Memory Scores and Distance Value. (A) Relationship between CDR Memory Scores and Distance Value of training data ($r = 0.702$, $p < 0.05$). (B) Relationship between CDR Memory Scores and Distance Value of testing data ($r = 0.512$, $p < 0.05$).

Table 5.8
ANOVA of CDR-Memory.

	df	SS	MS	F	Significance F
Regression	1	3,081,306,481	308,130,648	189,652,622	1.39E-25
Residual	196	3,168,185,905	0,16247107		
Total	197	6,249,492,386			

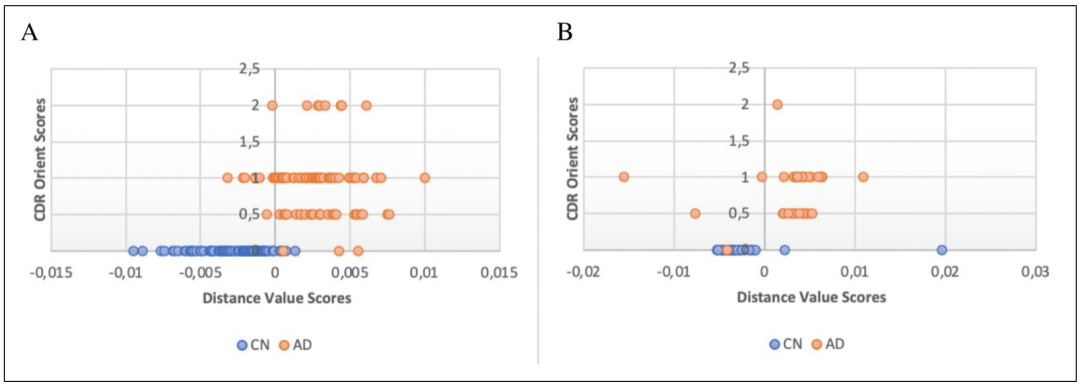

Figure 5.9 CDR-Orient Scores of training and testing data. (A) CDR Orient Scores of the training. (B) CDR Orient Scores of the testing.

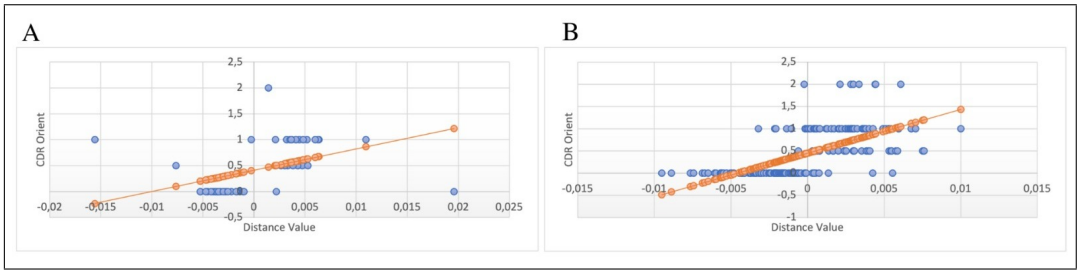

Figure 5.10 Regression plots representing the relationship between CDR Orient Score and Distance Value. (A) Relationship between CDR Orient Score and Distance Value of training data ($r = 0.643$, $p < 0.05$). (B) Relationship between CDR Orient Score and Distance Value of testing data ($r = 0.501$, $p < 0.05$).

6. DISCUSSION

Alzheimer's disease is one of our most unknown diseases, and many studies are still in the experimental stage. Currently, available drugs specifically, that target AD, elicit low response rates [43]. The mechanisms through which AD escapes AD-targeted approaches remain unclear. However, the majority of patient trials result in death [44]. Improving the interpretability and monitoring response for AD treatments will facilitate the development of novel therapies.

Our first major finding is to improve the interpretability and monitoring and response of AD. In recent studies, artificial intelligence is frequently used in the detection of Alzheimer's disease, however, it is a black box. Besides, promising results of medical AI systems in diagnosing and predicting are not completely fully trusted by physicians. Medical experts want to know and interpret the decision structure of the mechanism. Moreover, the severity score of the disease is very important. It can be understood whether the disease is mild or severe, with the severity score. Artificial intelligence divides the disease into classes but does not give an objective degree of severity [45]. With this study, by computing mutual information between AD and CN patients we find a similarity index, whereupon we calculated the distance value, which gives severity index through a distance value of FDG-PET data and enhances interpretability and monitoring of AD with two simple formulas.

Our second major finding is increasing the reliability of diagnostic results of CDR and MMSE tests. We found a correlation between the similarity index and the neuropsychological assessment scores. It shows that FDG-PET can be a well-performed biomarker to validate CDR and MMSE tests. In addition, we increased the reliability of diagnosing by test results. In [46], a similar result has been reported. So, given the correlated results of FDG-PET with CDR and MMSE tests, this method can be used to increase the reliability of the diagnosis and the severity as assessed by neuropsychological tests by making use of objective imaging results.

Previous studies demonstrated that rates of decline in CDR were also significantly correlated with whole brain atrophy and/or ventricular enlargement by using MRI [47]. Another study demonstrates that mice exhibit serious neuronal loss, cognitive deficits, and an age-dependent decline in glucose metabolism. Young mice have already shown an early decline in FDG uptake, similar to AD patients [48]. When we analyze the previous studies in the literature, our method has several advantages: We offer some interpretability of results by relying on image similarity. At the same time, we confirm the CDR and MMSE test results by using the similarity index. This index can be potentially used to monitor the progress of the patient in time while allowing a more objective and earlier diagnosis of AD [49].

This study we have done can be important for clinical applications. We can use FDG-PET to investigate whether a treatment method, such as drug treatment, is working or see if the disease is progressing. It is probably possible to see the progression of the disease within weeks with FDG-PET because it shows a metabolic reduction in the brain. Future work should tackle the diagnostic performance of the method for other stages of AD such as MCI and the classification accuracy for distinguishing AD from MCI. There is also a likelihood that this method can better distinguish severity between healthy and MCI patients with respect to MMSE tests. The sensitivity of the similarity index to changes in a patient's status may also be investigated.

For limitations, although significant advances in prediction, diagnosis, and prognosis have been made in recent years using various approaches, such as similarity index, MI, and Deep learning, there are challenges and limitations in terms of standards in image acquisition and reproducibility [50]. In addition, one of the serious problems is that environmental conditions are important when taking a PET image, factors that can cause artifacts or noise in the environment can cause major changes. At the same time, device quality is one of the important parameters during shooting [51].

For future works, The relationship of CDR and MMSE test questions and the different parts of the brain will be explored. For example, the MMSE was related to the cerebral cortex, accumbens, cerebral white matter, inferior lateral ventricles, and

hippocampus [52]. Since we used the whole brain image in the study, the correlation in the MMSE, CDR tests, and distance value in AD patients was relatively lower and there were some overlaps in different values. With this method, accuracy will increase, the noise will be reduced, and we will possibly get higher reliability results. Moreover, we will easily detect earlier (MCI) with FDG-PET, since it is diagnosed from brain metabolic activity, and will be able to assess disease severity that should correspond to neuropsychological assessment.

In conclusion, given the importance of diagnosing AD and monitoring its severity, we were able to improve the interpretability and monitor the response to therapy by using the severity index from an FDG-PET study as a complement of CDR and MMSE. Extracting the region of interests (ROIs) and investigating the relationship between them and neuropsychological evaluations using the proposed method can be conducted as a future study.

REFERENCES

1. Association, A. s. , "2018 Alzheimer's disease facts and figures," *Alzheimer's & Dementia,* Vol. 14, pp. 367-429, 2018.
2. Tiwari, S., V. Atluri, A. Kaushik, A. Yndart, "Alzheimer's disease: pathogenesis, diagnostics, and therapeutics," *International Journal of Nanomedicine*, Vol.14, pp. 5541-5554, 2019.
3. Pedroza, P., K. Molly, et al., "Global and regional spending on dementia care from 2000-2019 and expected future health spending scenarios from 2020-2050: An economic modelling exercise," *eClinical Medicine*, Vol. 45, pp. 101-337, 2022.
4. Vasileios, M., A. Alexiou, "Biomarkers in Alzheimer's Disease," *Alzheimer Res.*, Vol. 14, pp. 1149-1154, 2016.
5. Brookmeyer, R., N. Abdalla, CH. Kawas, MM. Corrada, "Forecasting the prevalence of preclinical and clinical Alzheimer's disease in the United States," *Alzheimers Dement*, Vol. 14, pp. 121-129, 2018.
6. Keijzer, H. M., et al. "Brain imaging in comatose survivors of cardiac arrest: Pathophysiological correlates and prognostic properties," *Resuscitation Vol.*, Vol. 133, pp. 124-136, 2018.
7. Wabik, A. et al., "Comparison of dynamic susceptibility contrast enhanced MR and FDG-PET brain studies in patients with Alzheimer's disease and amnestic mild cognitive impairment," *Journal of Translational Medicine*, Vol. 7, pp. 20- 259, 2022.
8. Yanase, J., E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Systems with Applications*, pp. 138, 2019.
9. Zhang, C., "Genetic Basis of Alzheimerâs Disease and Its Possible Treatments Based on Big Data" , *International Conference on Big Data and Social Sciences (ICBDSS)*, Vol. 2, pp. 10-14, 2020.
10. Vinny, P. W., V. Y. Vishnu, "Artificial Intelligence shaping the future of neurology practice," *Medical Journal Armed Forces India*, Vol. 77, pp. 276-282, 2021.
11. Ching, T., D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, "Opportunities and Obstacles for Deep Learning," *Biology and Medicine J. R. Soc. Interface*, pp. 15- 141, 2018.
12. Tanveer, M. et al., "Machine learning techniques for the diagnosis of Alzheimer's disease: A review," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, Vol. 16, pp. 1-35, 2020.
13. Kumar, S. S., K. R. Swamy, "Role of Computational Intelligence Techniques in Diagnosing Alzheimerâs disease at Early Stages: A Systematic Literature Review," *International Conference on Electronics and Renewable Systems (ICEARS)*, Vol. 2, pp.1233-1240, 2022.
14. Fabrizio, C., A. Termine, C. Caltagirone, "Artificial Intelligence for Alzheimer's Disease: Promise or Challenge?," *Diagnostics*, Vol. 14, pp. 11-1473, 2021.
15. Smailagic, N., et al., "F-FDG PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)," *The Cochrane Database of Systematic Reviews*, Vol. 1, 2015.
16. Lizuka, T., M. Fukasawa, M. Kameyama, "Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies," *Scientific Reports*, Vol. 9, pp. 1-8944, 2019.
17. Cheng, D., M. Liu, "Classification of Alzheimer's disease by cascaded convolutional neural networks using PET images," *International Workshop on Machine Learning in Medical Imaging Springer, Cham*, Vol. 4, pp.106-113, 2017.
18. Choi, H., K.H. Jin, "Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging," *Behav. Brain Res.*, Vol. 344, pp. 103-109, 2018.

19. Huang, Y., J. Xu, Y. Zhou, T. Tong, X. Zhuang, "Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network," *Front. Neurosci.*, Vol. 4, pp. 13-509, 2019.

20. Weatheritt, J., A. Palombit, R. Manber, R. Wolz, "Alzheimer's disease detection using explainable AI on PET images," *Alzheimer's Association*, Vol. 17, 2022.

21. Lee, R., H. Choi, "Prediction of post-stroke cognitive impairment using brain FDG PET: deep learning-based approach," *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 49, pp. 1254-1262, 2022.

22. Zang, B., Y. Lin, Z. Liu, X. Gao, "Deep learning method for single-trial EEG classification in RSVP task based onspatiotemporal features of ERPs," *Journal of Neural Engineering*, Vol 18, pp. 4 2021.

23. Chiaravalloti, A., M. Ricci, D. Di Biagio, "The Brain Metabolic Correlates of the Main Indices of Neuropsychological Assessment in Alzheimer's Disease," *Journal of Personalized Medicine*, Vol. 10, pp. 2:25, 2020.

24. Jiaojiao, J., et al., "Correlation Between Brain 18F-AV45 and 18F-FDG PET Distribution Characteristics and Cognitive Function in Patients with Mild and Moderate Alzheimer's Disease," *Journal of Alzheimer's Disease*, Vol. 79, pp. 1317 - 1325, 2021.

25. Sudar, K. M., P. Nagaraj, S. Nithisaa, R. Aishwarya, M. Aakash and S. I. Lakshmi, "Alzheimer's Disease Analysis using Explainable Artificial Intelligence (XAI)," *International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, Vol. 5, pp. 419-423, 2022.

26. Weiner, M. W., et al., "The Alzheimer's Disease Neuroimaging Initiative: a review," *Alzheimer's Association*, Vol. 8, pp. 1-68, 2015.

27. Botev, Z.I., J. F., Grotowski, "Kernel density estimation via diffusion," *The Annals of Statistics*, Vol. 38, pp. 2916-2957, 2010.

28. Isaac, A., H. Khanna Nehemiah, "Computer-Aided Diagnosis System for Diagnosis of Cavitary and Miliary Tuberculosis Using Improved Artificial Bee Colony Optimization," *IETE Journal of Research*, Vol. 1, pp. 1-20, 2021.

29. Rosen, H. J., et al., "Neuropsychological and functional measures of severity in Alzheimer's disease, frontotemporal dementia, and semantic dementia," *Alzheimer's Disease and Associated Disorders*, Vol. 18, pp. 202-207, 2004.

30. Tobis, S., K. Wieczorowska-Tobis, D. Talarska, M. Pawlaczyk, A. Suwalska, "Needs of older adults living in long-term care institutions: an observational study using Camberwell Assessment of Need for the Elderly, " *Clin Interv Aging*, Vol. 13, pp. 2389-2395, 2018.

31. Mendez, M. F., A. K. Chen, "Acquired Sociopathy and Frontotemporal Dementia," *Dement Geriatr Cogn Disord*, Vol. 20, pp. 99-104, 2005.

32. Hajian-Tilaki, K. "Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation," *Caspian Journal of Internal Medicine*, Vol. 4, pp. 627-635, 2013.

33. Li, Z., X. Jiang, Y. Wang, Y. Singh Kim, "Applied Machine Learning in Alzheimer's Disease Research: Omics, Imaging, and Clinical Data," *Emerg Top Life Sci.*, Vol. 21, pp. 765-777, 2021.

34. Rosenblatt, M., "Remarks on Some Nonparametric Estimates of a Density Function," *The Annals of Mathematical Statistics*, Vol. 27, pp. 832-837, 1956.

35. Emanuel, P., "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, Vol. 3, pp. 1065-1076, 1962.

36. Botev, Z. I., J. F. Grotowski, "Kernel density estimation via diffusion," *The Annals of Statistics*, Vol. 38, pp. 2916-2957, 2010.

37. Hu, Q., et al., "Noved Classification of Benign and Malignant Breast Lesions Using Deep Feature Maximum Intensity Projection MRI in Breast Cancer Diagnosis Using Dynamic Contrast-enhanced MRI," *Artificial Intelligence*, Vol. 3, 2021.

38. Thabtah, F., R. Spencer, "The correlation of everyday cognition test scores and the progression of Alzheimer's disease: a data analytics study," *Health Information Science and Systems*, Vol. 8, pp. 1-24, 2020.

39. Arevalo-Rodriguez, I., N. Smailagic, "Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI) ," *Cochrane Database Syst*, Vol. 7, 2015.

40. O'Bryant, S. E., et al., "Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's research consortium study," *Staging dementia using Clinical Archives of Neurology*, Vol. 65, pp. 1091-1095, 2008.

41. Tan, J. E., E. Strauss, "Encyclopedia of Clinical Neuropsychology," *Springer New York*, Vol. 5, pp. 587-590, 2011.

42. Duara, R., et al., "Diagnosis and staging of mild cognitive impairment, using a modification of the clinical dementia rating scale: the mCDR," *International Journal of Geriatric Psychiatry*, Vol. 25, pp. 282-289, 2010.

43. Jeremic, D., L. Jimenez-Diaz, "Past, present and future of therapeutic strategies against amyloid beta peptides in Alzheimer's disease: a systematic review," *Ageing Research Reviews*, Vol. 72, 2021.

44. Qian, C., C. Yuan, C. Li, "Multifunctional nano-enabled delivery systems in Alzheimer's disease management," *Biomaterials Science*, Vol. 8, pp. 5538-5554, 2020.

45. Ceccarini, J., et al., "Direct prospective comparison of 18F-FDG PET and arterial spin labelling MR using simultaneous PET/MR in patients referred for diagnosis of dementia" , *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 47, pp. 2142-2154, 2020.

46. Nagaraj, S., S. & Duong, "Deep Learning and Risk Score Classification of Mild Cognitive Impairment and Alzheimer's Disease," *Journal of Alzheimer's Disease*, Vol. 80, pp. 1079-1090, 2021.

47. Chiaravalloti, A., et al., "The Brain Metabolic Correlates of the Main Indices of Neuropsychological Assessment in Alzheimer's Disease," *Journal of Personalized Medicine*, Vol. 10, pp. 1-25, 2020.

48. Bouter, C., et al., "F-FDG-PET Detects Drastic Changes in Brain Metabolism in the Tg4-42 Model of Alzheimer's Disease," *C. Frontiers in Aging Neuroscience*, Vol. 10, 2018.

49. Martin, S., et al., "F-FDG PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)," *Cochrane Database of Systematic Reviews*, Vol. 1, 2015.

50. Timmeren, J.E., D. Cester, S. Tanadini-Lang, H. Alkadhi, B. Baessler, "Radiomics in medical imaging- "how-to" guide and critical reflection," *Insights Imaging*, Vol. 1, 2020.

51. Muhlberg, A., A. Katzmann, V. Heinemann, R. Kargel, "The Technome - A Predictive Internal Calibration Approach for Quantitative Imaging Biomarker Research," *Sci Rep*, Vol. 10, pp. 1103, 2020.

52. Fjell, A. M., I. K. Amlien, L. T. Westlye, "Mini-mental state examination is sensitive to brain atrophy in Alzheimer's disease," *Dementia and Geriatric Cognitive Disorders*, Vol. 28, pp. 252-258, 2009.