

Original Paper

Longitudinal Magnetic Resonance Imaging as a Potential Correlate in the Diagnosis of Alzheimer Disease: Exploratory Data Analysis

Afreen Khan*, MCA; Swaleha Zubair*, PhD

Aligarh Muslim University, Aligarh, India

*all authors contributed equally

Corresponding Author:

Swaleha Zubair, PhD

Aligarh Muslim University

Department of Computer Science

Adjacent Computer Centre

Anoopshahr Road

Aligarh, 202002

India

Phone: 91 9410059635

Email: swalehaowais123@gmail.com

Abstract

Background: Alzheimer disease (AD) is a degenerative progressive brain disorder where symptoms of dementia and cognitive impairment intensify over time. Numerous factors exist that may or may not be related to the lifestyle of a patient that result in a higher risk for AD. Diagnosing the disorder in its beginning period is important, and several techniques are used to diagnose AD. A number of studies have been conducted on the detection and diagnosis of AD. This paper reports the empirical study performed on the longitudinal-based magnetic resonance imaging (MRI) Open Access Series of Brain Imaging dataset. Furthermore, the study highlights several factors that influence the prediction of AD.

Objective: This study aimed to correlate the effect of various factors such as age, gender, education, and socioeconomic background of patients with the development of AD. The effect of patient-related factors on the severity of AD was assessed on the basis of MRI features, Mini-Mental State Examination (MMSE), Clinical Dementia Rating (CDR), estimated total intracranial volume (eTIV), normalized whole brain volume (nWBV), and Atlas Scaling Factor (ASF).

Methods: In this study, we attempted to establish the role of longitudinal MRI in an exploratory data analysis (EDA) of AD patients. EDA was performed on the dataset of 150 patients for 343 MRI sessions (mean age 77.01 [SD 7.64] years). The T1-weighted MRI of each subject on a 1.5-Tesla Vision (Siemens) scanner was used for image acquisition. Scores of three features, MMSE, CDR, and ASF, were used to characterize the AD patients included in this study. We assessed the role of various features (ie, age, gender, education, socioeconomic status, MMSE, CDR, eTIV, nWBV, and ASF) on the prognosis of AD.

Results: The analysis further establishes the role of gender in the prevalence and development of AD in older people. Moreover, a considerable relationship has been observed between education and socioeconomic position on the progression of AD. Also, outliers and linearity of each feature were determined to rule out the extreme values in measuring the skewness. The differences in nWBV between CDR=0 (nondemented), CDR=0.5 (very mild dementia), and CDR=1 (mild dementia) are significant (ie, $P < .01$).

Conclusions: A substantial correlation has been observed between the pattern and other related features of longitudinal MRI data that can significantly assist in the diagnosis and determination of AD in older patients.

(*JMIR Biomed Eng* 2020;5(1):e14389) doi: [10.2196/14389](https://doi.org/10.2196/14389)

KEYWORDS

Alzheimer disease; dementia; longitudinal; magnetic resonance imaging; exploratory data analysis; Mini-Mental State Examination; Clinical Dementia Rating; Atlas Scaling Factor

Introduction

Alzheimer disease (AD) is a degenerative brain ailment characterized by the development of dementia and other related cognitive impairments [1-3]. It is a heterogeneous, irreversible neurodegenerative disorder that may find an association with genetic complexity in the individual. The Alzheimer's Association describes dementia as a syndrome comprising a cluster of symptoms that encompass several features including age, gender, education, and the Mini-Mental State Examination (MMSE) of the inflicted patients [4].

There has been a significant increase in the number of AD cases in recent years. It has been reported that it is the sixth most diagnosed disease in the United States. As of 2018, 5.7 million Americans of all ages have been diagnosed with AD [4]. Approximately 44 million people worldwide are living with AD or an associated kind of dementia [5].

With the advancement of technology pertaining to treatment methodologies and development of novel diagnostic tools, many of the modern age diseases are being diagnosed earlier and treated successfully. In contrast, AD still remains a poorly diagnosed ailment with little success in treatment.

In the information technology era, machine and deep learning tools have found a wide scope in medical diagnosis [6]. Although medical expert opinion, disease symptom, and other related data from the patient remain the prime parameters that help in the diagnosis of a particular disease, machine learning predictions, data analytics visualizations, and other artificial intelligence techniques have emerged as alternate ways to predict diseases and help the current state of the medical world in a great way [7,8].

The occurrence of cognitive disorders is a common feature observed in elderly people, and this can be considered a primary indication of a growing dementing syndrome like AD [9]. Individuals with cognitive disorders experience mild cognitive impairment (MCI) [10-12]. Various biomarkers or related parameters may evolve that can help in the diagnosis of AD in patients. Similarly, techniques like magnetic resonance imaging (MRI) studies, positron emission tomography scans, and neurochemical testing of the cerebrospinal fluid can also help in the diagnosis of AD [13,14].

In this study, we systematically examined the distinct and interactive impact of age, gender, education, socioeconomic status (SES), Mini Mental State Examination (MMSE), Clinical Dementia Rating (CDR), estimated total intracranial volume (eTIV), normalized whole brain volume (nWBV), and Atlas Scaling Factor (ASF) on the basis of several longitudinal MRI sessions of various patients. The information was retrieved from the Open Access Series of Brain Imaging (OASIS-2) dataset. We performed exploratory data analysis (EDA) to understand the correlation between various feature sets. Consistent with the literature, we predicted that men were more likely to be diagnosed with AD compared with women. The gender bias can be correlated to the dataset dependency. The $\epsilon 4$ allele of the apolipoprotein E gene (APOE- $\epsilon 4$) has also been reported to play a major role in the occurrence of AD. We did not include

APOE- $\epsilon 4$ data in the study in order to avoid the complexity. A significant relationship has been observed among educational background and SES of the patients and emergence of dementia. Anomalies and linearity of each of the features were resolved to remove extreme values in determining the skewness.

Methods

Subjects

The dataset used in this study consists of a longitudinal collection of MRI data in demented and nondemented older adults. A total of 150 subjects aged 60 to 96 years participated in 373 MRI sessions. The data included in this study were based on the subjects reported to a longitudinal collection of MRI scans at the Washington University Alzheimer Disease Research Center [15].

The T1-weighted MRI acquisition of each subject was performed on a 1.5 Tesla Vision scanner (Siemens). Related technical details are as follows: sequence of magnetization prepared rapid acquisition gradient echo, repetition time=9.7 msec, echo time=4.0 msec, flip angle=10°, inversion time=20 msec, delay time=200 msec, orientation is sagittal, thickness=1.25 mm, gap=0 mm, slice number=128, and resolution=256×256 (1×1 mm) [15].

Methodology

In this analysis, we cataloged previous EDA. The general objective of the study was to report the relative association between the target group (demented or nondemented) and other features that play a major role in the diagnosis of AD. Furthermore, we examined the risk of AD induction in inflicted patients. We analyzed longitudinal MRI data of both healthy patients and patients with AD [15].

Scoring Rules

In this study, we used the following instruments to determine the state of the healthy versus inflicted brain.

- SES: according to the Hollingshead Index of Social Position, the SES is classified into groups of highest status (1) and lowest status (0) [16]
- MMSE: values range from 0 to 30; 0 to 9 indicates extreme impairment, 10 to 18 demonstrates moderate dementia, 19 to 23 mild dementia, and 24 to 30 is considered normal [17]
- CDR: scored after a semistructured discussion with the patient, with scores ranging from 0 to 3 (ie, 0=none, 0.5=very mild, 1=mild, 2=moderate, 3=extreme dementia) [18]

Experiment Environment

Empirical analysis of the dataset described in this paper was performed using Python libraries conducted on the Jupyter platform of Anaconda Navigator. The Jupyter platform presents a well-defined skeleton for developers to process, develop, and assess their models. Python is an interpreted and high-level programming language comprising dynamic semantics. It includes Seaborn, a visualization library through which statistical graphs can be plotted with the aim of performing univariate and multivariate analyses.

Exploratory Data Analysis

EDA is a data analysis methodology using techniques that are usually graphical. It maximizes understanding of the dataset, reveals underlying structure, detects anomalies and outliers, extracts imperative features, and ascertains ideal factor settings [19]. EDA is not similar to statistical graphics despite the fact that the two terms are used interchangeably. It is a more direct approach that allows the data to reveal the underlying model and its structure [19].

In this study, we focused on establishing a correlation between attributes of MRI tests and patient classification groups. The primary objective of performing this exploratory analysis was to determine the association of data among the features before performing the data analysis or data extraction process. It was supposed to assist in understanding the data subclassification and facilitate choosing the proper analysis technique for the model later.

Dataset Description

The dataset comprised 373 observations and 15 attributes, out of which group was the target variable while the rest were the independent variables in this empirical study. Table 1 provides a description of the dataset attributes.

Figure 1 outlines the dataset attributes in terms of the total count of each attribute for 15 columns on the basis of null/nonnull and data type of respective attributes. It can be seen from the figure that SES and MMSE consist of values less than the total 373 MRI sessions, marked by the red right bracket in the figure. This is what missing values relates to. The rest of the features, marked by the blue right brackets, do not contain any missing values (ie, for the total 373 sessions, all recorded MRI features emerged as nonnull and without any missing value).

The *P* values used for comparison in the study are shown in Table 2.

Table 1. Detail of dataset attributes.

Number	Attribute name	Attribute description
1	Subject ID ^a	Patient's identification number
2	MRI ^b ID	Patient's imaging identification number
3	Group	Demented, nondemented, or converted
4	Visit	Number of visits of each patient
5	MR ^c Delay	Magnetic resonance delay is the delay time given before the image procurement in real time
6	M/F ^d	Patient's gender
7	Hand	Right-handed or left-handed
8	Age	Patient's age at the scanning
9	EDUC ^e	Educational level of the patient
10	SES ^f	Socioeconomic status of the patient
11	MMSE ^g	Mini-Mental State Examination score
12	CDR ^h	Clinical Dementia Rate score
13	eTIV ⁱ	estimated total intracranial volume result
14	nWBV ^j	normalized whole brain volume result
15	ASF ^k	Atlas Scaling Factor

^aID: identification.

^bMRI: magnetic resonance imaging.

^cMR: magnetic resonance.

^dM/F: male or female.

^eEDUC: educational level of the patient.

^fSES: socioeconomic status.

^gMMSE: Mini-Mental State Examination.

^hCDR: Clinical Dementia Rating score.

ⁱeTIV: estimated total intracranial volume.

^jnWBV: normalized whole brain volume.

^kASF: Atlas Scaling Factor.

Figure 1. Dataset Information.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 373 entries, 0 to 372
Data columns (total 15 columns):
Subject ID    373 non-null object
MRI ID       373 non-null object
Group        373 non-null object
Visit        373 non-null int64
MR Delay     373 non-null int64
M/F          373 non-null object
Hand         373 non-null object
Age          373 non-null int64
EDUC         373 non-null int64
SES          354 non-null float64
MMSE         371 non-null float64
CDR          373 non-null float64
eTIV         373 non-null int64
nWBV         373 non-null float64
ASF          373 non-null float64
dtypes: float64(5), int64(5), object(5)

```

Table 2. *P* value for the corresponding attribute.

Attribute name	<i>P</i> value
EDUC: educational level of a patient	<.001
MMSE: Mini Mental State Examination	<.001
CDR: Clinical Dementia Rating	<.01
nWBV: normalized whole brain volume	<.01

Summary Statistics

Statistical information includes count, mean, standard deviation, first quartile, second quartile (median), third quartile, and minimum and maximum values of each attribute as shown in [Table 3](#).

From the data depicted in [Table 3](#), we can infer that the mean value is less than the median on some features and greater than

the median value on certain other sets of features. The median value is represented by 50% (50th percentile) in the index column. The median value of each feature aids in the data preprocessing when dealing with the imputation step. There is a large difference in the 75th percentile and maximum values of predictors in MR delay, CDR, and eTIV. The observation suggests the occurrence of extreme values (ie, outliers) in the dataset.

Table 3. Summary statistics of each attribute.

Attribute	Count	Mean (SD)	Min-max ^a	Quartiles		
				25%	50%	75%
Visit	373.00	1.88 (0.92)	1.00-5.00	1.00	2.00	2.00
MR ^b delay	373.00	595.10 (635.49)	0-2639.00	0	552.00	873.00
Age	373.00	77.01 (7.64)	60.00-98.00	71.00	77.00	82.00
EDUC ^c	373.00	14.60 (2.88)	6.00-23.00	12.00	15.00	16.00
SES ^d	354.00	2.46 (1.13)	1.00-5.00	2.00	2.00	3.00
MMSE ^e	371.00	27.34 (3.68)	4.00-30.00	27.00	29.00	30.00
CDR ^f	373.00	0.29 (0.37)	0-2.00	0	0	0.50
eTIV ^g	373.00	1488.13 (176.14)	1106.00-2004.00	1357.00	1470.00	1597.00
nWBV ^h	373.00	0.73 (0.04)	0.64-0.84	0.70	0.73	0.76
ASF ⁱ	373.00	1.20 (0.14)	0.88-1.59	1.10	1.19	1.29

^aMin-max: minimum and maximum values.

^bMR: magnetic resonance.

^cEDUC: educational level of the patient.

^dSES: socioeconomic status.

^eMMSE: Mini-Mental State Examination.

^fCDR: Clinical Dementia Rating.

^geTIV: estimated total intracranial volume.

^hnWBV: normalized whole brain volume.

ⁱASF: Atlas Scaling Factor.

Data Exploration

Initially, the dataset consisted of 373 MRI sessions out of which there were nondemented (n=72), demented (n=64), and converted patients (n=14). On the first visit, patients were grouped as nondemented and were categorized as demented at a later visit. The 14 converted patients are those patients which were found to be nondemented in the first visit, but in their second and third visits, they were diagnosed with dementia. Therefore, only the subjects of the first visit are being considered throughout the study, and total of 150 subjects have been explored under this analysis.

The dataset consists of many missing values (ie, some of the rows of certain attributes consist of no value, which is determined during the EDA step). To locate exactly which column comprises missing values, a heat map is plotted for all 373 MRI sessions initially, consisting of all the patient visits (Figure 2A). The SES and MMSE columns contain missing values (represented by yellow lines on a purple background). Figure 2B delineates the count of missing values in numeric form for all attributes. Figures 3A and 3B highlight the heat map and count of missing values for the 150 subjects for visit 1. SES is the only feature that consists of 8 missing values, while the rest of the features have all values filled.

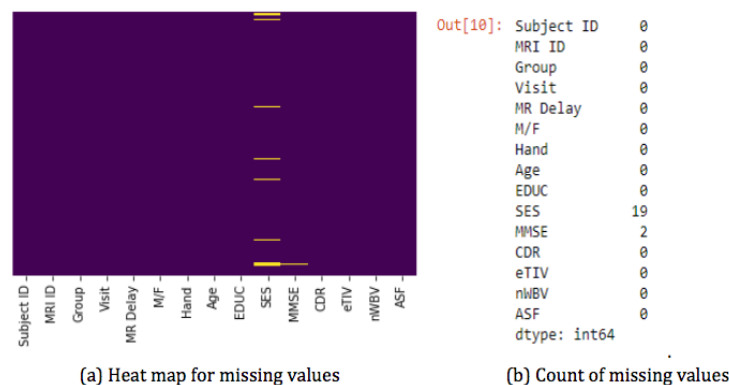
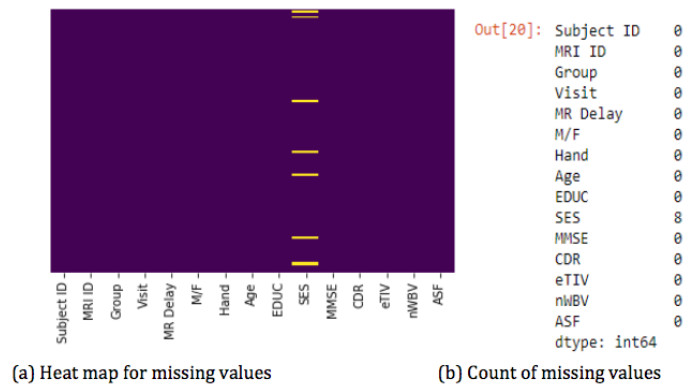
Figure 2. Illustration of missing values for 373 magnetic resonance imaging sessions for all patient visits.

Figure 3. Outline of missing values for 150 patients for the first visit.



Results

In this section, the results of the EDA are reported. Subsequent to applying the preprocessing and data preparation strategies, we attempted to break down the data outwardly and make sense of the dispersion of features as far as adequacy and effectiveness are concerned. By breaking down data, we have tried to make

it more simple and meaningful. This helped in increasing the efficiency of the analysis.

Patient Demographic Profiles

The study comprised 62 males and 88 females within the age range of 60 to 96 years. Table 4 illustrates the demographic summary of patients who were examined for AD.

Table 4. Demographic profile of the study population (n=150).

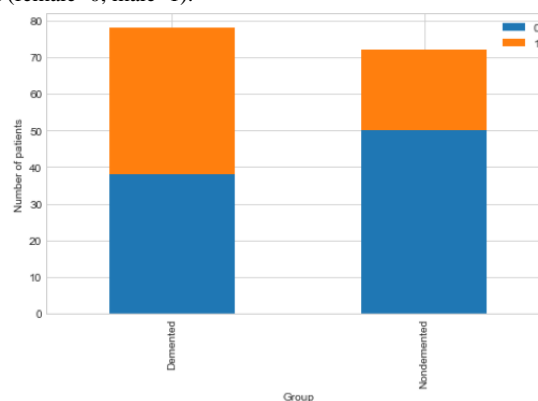
Characteristic	Values
Gender, n (%)	
Male	62 (41.3)
Female	88 (58.7)
Age in years, mean (SD)	77.01 (7.64)
Age in years, median	77

Gender and Demented Proportion

The bar chart as demonstrated in Figure 4 confirms that men are more prone to dementia than women. The blue color, coded

as 0, represents female, while the orange color, coded as 1, represents male. Of the 150 patients, 78 are in the demented category. Of the 78 demented patients, 40 are male.

Figure 4. Gender and demented proportion (female=0, male=1).

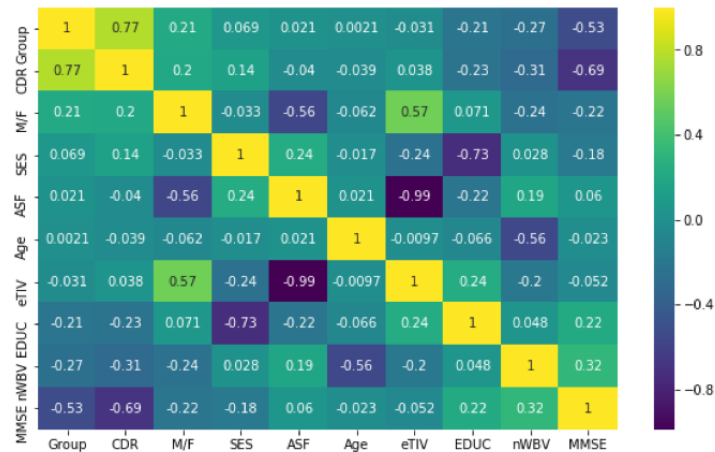


Correlation Matrix With a Heat Map

In order to build the model, an essential condition is to eliminate the correlated variables. Correlations were obtained by applying the Python Pandas corr() function, which aided us in visualizing the correlation grid built using a heat map.

The correlation matrix with heat map is illustrated in Figure 5. The dark shades represent positive correlation while lighter shades represent negative correlation. We exclude the target variable (ie, group) and then checked for the correlated independent variables. Thus we can infer that eTIV has a strong positive correlation with male/female (M/F) whereas it has a strong negative correlation with ASF among all.

Figure 5. Heat map illustrating the correlations among the dataset features.

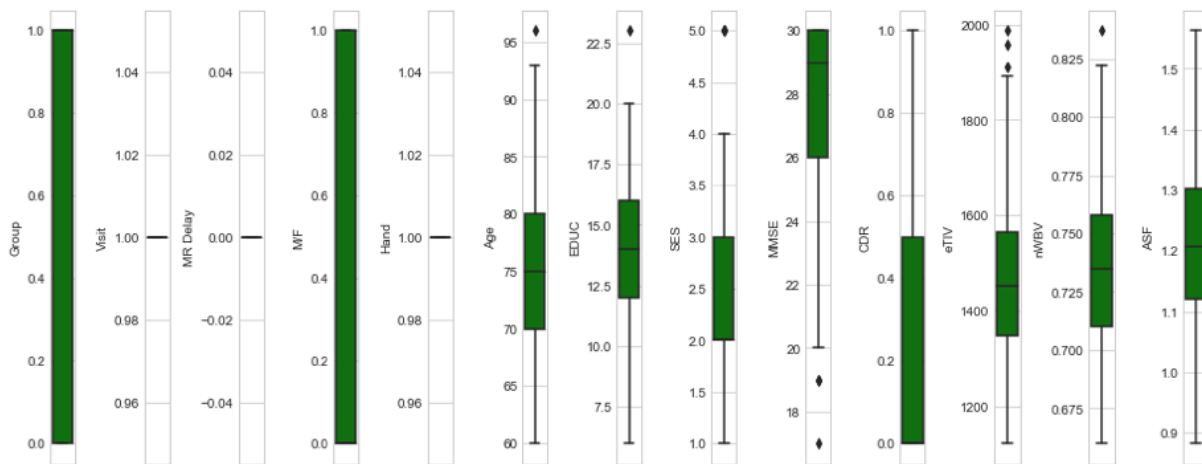


Outliers Check With Box-Whisker Plot

A box-whisker plot displays the spread of quantitative data in a manner that facilitates comparisons between attributes. In Figure 6, the box illustrates the dataset’s quartiles whereas the whiskers stretch out to demonstrate whatever remains of the dispersion. The box-whisker schema is a standardized method for displaying the data distribution, which is dependent on 5 major aspects: minimum value, first quartile, median value (second quartile), third quartile, and maximum value. The

middle rectangle traverses the first quartile to the third quartile, known as interquartile range (IQR). A fragment inside the rectangle demonstrates the median value. Whiskers above and beneath the rectangle demonstrate the areas of the minimum and greatest value. Outliers are either 3×IQR or progressively over the third quartile or 3×IQR or more beneath the first quartile. Thus, we can infer from Figure 6 that age, patient education level (EDUC), SES, MMSE, eTIV, and nWBV feature columns show outliers.

Figure 6. Box-whisker plot demonstrating outliers.

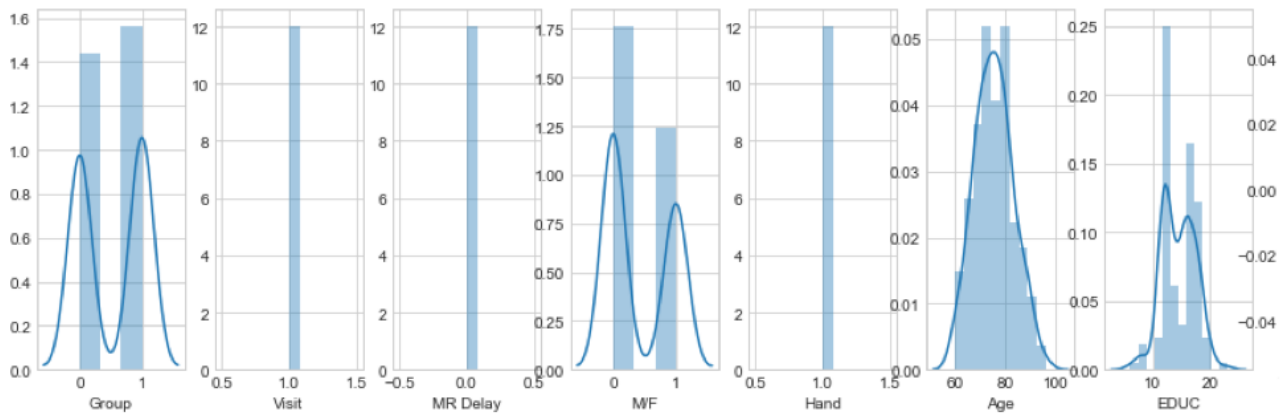


Skewness and Distribution Plot

The linearity of the attributes was determined by plotting a distribution graph. The graph was used to study the skewness of both the target variable and the independent variables. From

Figure 7, it can be concluded that group, visit, MR delay, M/F, hand, and age feature columns appear to be normally distributed while all the remaining independent variables are discovered to be experiencing skewness.

Figure 7. Distribution plot of the dataset features.



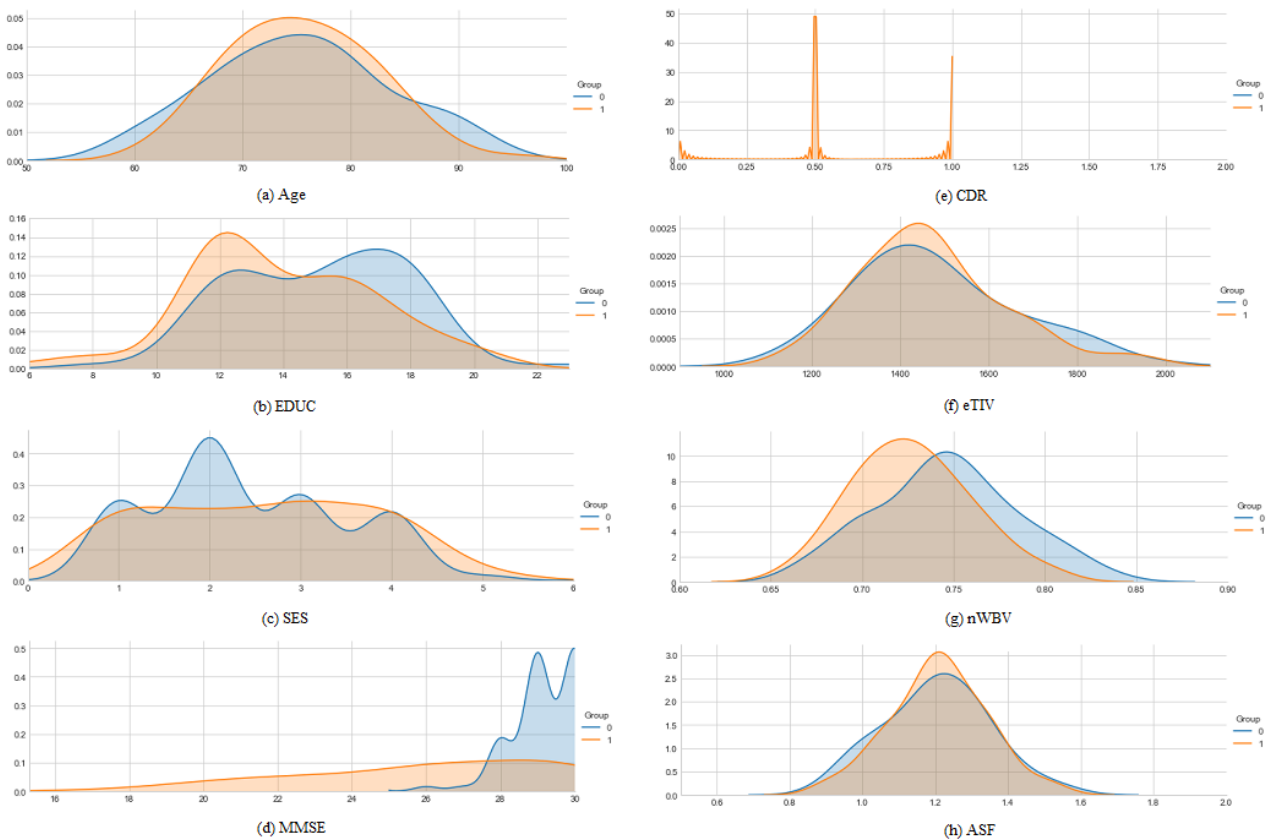
Effect of Independent Variable on Dependent Variable

A graph was plotted between the target variable (ie, group: demented/nondemented) and independent variables. We plotted 8 such graphs, for age, EDUC, MMSE, ASF, eTIV, nWBV, SES, and CDR, shown in Figure 8.

The following features were inferred: (1) age: between 60 and 90 years; (2) EDUC: demented patients were less educated; (3) SES: considerable increment in the prevalence of dementia as we move from highest status to lowest status; (4) MMSE:

nondemented group got much higher MMSE scores than the demented group; (5) CDR: more individuals with a score of 0.5 (ie, very mild dementia), fewer individuals with a score of 1 (ie, mild dementia), and very few with a score of 0 (ie, no dementia); (6) eTIV: higher for demented patients; (7) nWBV: nondemented group has higher brain volume ratio than demented group; and (8) ASF: demented patients have higher score than nondemented ones. The differences in nWBV between CDR=0 (nondemented), CDR=0.5 (very mild dementia), and CDR=1 (mild dementia) comes out to be significant (ie, $P < .01$).

Figure 8. A plot between the target variable and each independent variable (nondemented=0, demented=1).



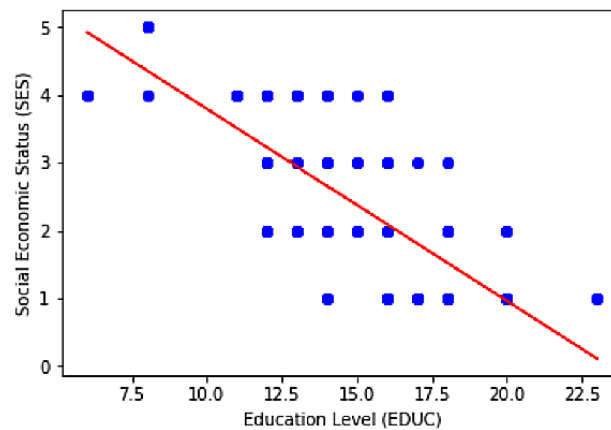
Impact of Socioeconomic Status and Education Level in the Demented Group

The relationship between SES and EDUC on dementia can be inferred from Figure 9, which shows that individuals with the

highest status (1) exhibit higher education levels while individuals with the lowest status (5) exhibit lower education level. Thus, years of education have an immense effect on dementia. The scatter plot with linear regression lines for SES

and EDUC display a positive correlation among EDUC and SES.

Figure 9. Scatter plot for socioeconomic status and level of education.

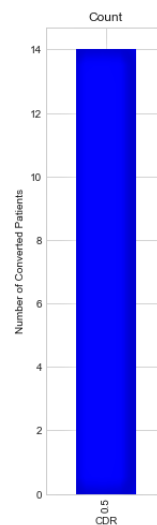


Correlation Between Converted Patients and Clinical Dementia Rating

The data shown in Table 3 suggest that 14 patients converted. These patients were earlier classified as nondemented and in a later visit found to have dementia. We tended to draw a relationship among these 14 converted patients with their

respective CDR values on subsequent (second and third) visits. For developing a correlation between dementia and other related factors, we focused on changes incurred in CDR values. For earlier visits, it was 0.0, signifying that the patient was nondemented, while at a later visit, it changed to 0.5, indicating the patient had very mild dementia. Figure 10 shows a correlation between converted patients with their CDR values.

Figure 10. Distribution plot for converted patients and their Clinical Dementia Rating value.



Discussion

Principal Findings

This study provides an understanding of attributes related to AD in older adults. We observed that men are more likely to have AD compared with women. There are several major differences that frequently appear between men and women in the occurrence, presentation, and development of psychiatric disorders [20]. Earlier studies suggested that women are more prone to develop AD since they are at greater risk of depression compared with men [21]. The genetic factor APOE- ϵ 4 has also been reported to affect men and women differently [21]. Riedel et al [22] stated that age, APOE- ϵ 4, and sex are the most serious risk factors in the development of AD. Further, the rate of AD

is practically identical in women and men until late age when the frequency becomes more prominent in women [22].

We performed an empirical analysis on the dataset comprising longitudinally obtained T1-weighted MRI data of 150 patients aged between 60 to 96 years. Among the 15 studied features, we found that only gender, age, educational years, SES, MMSE, CDR, eTIV, and nWBV were significantly associated with making an impact on the occurrence of AD in both demented and nondemented subjects. Our analysis demonstrated that patients aged between 70 and 90 years exhibit a higher clustering of dementia than nondemented patients. Since AD has a lower survival rate, it is the reason why data available in the aged patient is scarce. All patients examined were right-handed, thus handedness doesn't have an effect in this analysis. Of the 150

patients, demented patients were found to be less educated compared with nondemented patients (Figure 8B).

We found an independent link between various features in both demented and nondemented groups and found that there were numerous correlated indicators of AD. Unfortunately, this study lacks an adequate feature set that could have helped in uncovering related associations efficiently.

We observed that over the change from higher (score 1) to lower (score 5) SES, there was a considerable decrease in the prevalence of dementia. In general, education has been found to be directly associated with SES. In fact, there seems to be a high to moderate level of association between education and occupation-based SES [23]. Social epidemiology relates education with SES by defining education as “the transition from a socioeconomic position largely received from parents to an achieved socioeconomic position as an adult” [16]. Various components of SES, viz education, income and occupational status, can influence AD development in the aged patients [24].

The MMSE, a complete measure of cognitive impairment, has been widely used in the detection of AD. Arevalo-Rodriguez et al [25] performed an analysis to determine the MMSE accuracy for the detection of AD in people with mild MCI. In fact, the MMSE score cannot aid in categorizing people as demented or nondemented [25]. In contrast to this, we identified that the nondemented study group got a much higher MMSE score than the demented group.

The scoring of the CDR have been widely used in clinical trials and longitudinal studies to determine the state of dementia [18]. We found the CDR peaks at 0.5 (very mild dementia), followed by 1 (mild dementia) and 0 (no dementia). Unsurprisingly, our results are in agreement with those illustrated by Marcus et al [15], which states that patients who were categorized to be nondemented in the first visit were found to be demented in later visits with a CDR of greater than 0.

The plot for eTIV summarizing various data shows that demented patients have more eTIV compared with nondemented patients (Figure 8F). The intracranial volume, describing brain size, is found to be less in AD patients. Earlier, Tate et al [26] reported that there were certain patients for which the total intracranial volume emerged to have an impact on dementia prediction when the data were examined in a nonparametric manner.

Acknowledgments

The data used in the preparation of this article were obtained from the OASIS database [33], made available by the Washington University Alzheimer Disease Research Center. Longitudinal MRI data were retrieved from the following published NIH grants: P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, and R01 MH56584.

Conflicts of Interest

None declared.

References

1. Wilson RS, Segawa E, Boyle PA, Anagnos SE, Hizez LP, Bennett DA. The natural history of cognitive decline in Alzheimer's disease. *Psychol Aging* 2012 Dec;27(4):1008-1017 [FREE Full text] [doi: [10.1037/a0029857](https://doi.org/10.1037/a0029857)] [Medline: [22946521](https://pubmed.ncbi.nlm.nih.gov/22946521/)]

In line with an earlier study using a subset of the data [27], we found that the nondemented group had a higher nWBV than the demented group. This could be attributed to the fact that AD may lead to shrinkage of neuronal tissues of the brain. Marcus et al [15] exploited nWBV as an approach to evaluate the anatomical features of the brain to determine the level of dementia. Several other studies suggested that nWBV declines upon advancement of AD stage and growing age of the patients [27-31].

Our findings suggest that demented patients have a higher ASF when compared with nondemented ones. The scaling factor changes the skull and native-space brain to the atlas target, which is determined by calculating the determinant of the transform matrix [32].

On the basis of data analysis, we infer that there was no correlation between the repeated measures. In longitudinal data analysis, it seems to be an easy and straightforward approach but an unrealistic alternative. To this end, we can justify it as a fair approach to assess the relationship among covariates irrespective of the visits. This structure was chosen at the commencement of the analysis, and we suggested that it bears a resemblance to the experimental correlations for improved estimate of standard errors.

Limitation

More feature set brain mapping is required to strengthen the robustness of the results and discover the causal methods underlying the relation between distinct features of both longitudinal and cross-sectional MRI data and the consequence on the late-life health.

Conclusion and Future Work

This study highlights the relationship between the target and the independent features in MRI sessions of AD patients. It can be argued that whatever effect the independent features have on the prediction of the target variable (demented/nondemented), it is unlikely to be dependent on the sample size relationship. We infer that men are more likely to suffer from AD than women. The study also finds that attributes such as eTIV, nWBV, and ASF have a greater correlation in the prevalence of AD in women compared with men. Finally, we conclude that imaging biomarkers play a major role in the diagnosis of AD.

2. Barker WW, Luis CA, Kashuba A, Luis M, Harwood DG, Loewenstein D, et al. Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. *Alzheimer Dis Assoc Disord* 2002;16(4):203-212. [doi: [10.1097/00002093-200210000-00001](https://doi.org/10.1097/00002093-200210000-00001)] [Medline: [12468894](https://pubmed.ncbi.nlm.nih.gov/12468894/)]
3. Kim H. Understanding internet use among dementia caregivers: results of secondary data analysis using the us caregiver survey data. *Interact J Med Res* 2015;4(1):e1 [FREE Full text] [doi: [10.2196/ijmr.3127](https://doi.org/10.2196/ijmr.3127)] [Medline: [25707033](https://pubmed.ncbi.nlm.nih.gov/25707033/)]
4. 2018 Alzheimer's disease facts and figures. URL: <https://www.alz.org/media/homeoffice/facts%20and%20figures/facts-and-figures.pdf> [accessed 2019-02-17]
5. Alzheimer's disease statistics. 2018. URL: <https://alzheimersnewstoday.com/alzheimers-disease-statistics/> [accessed 2019-02-19]
6. Cleret de Langavant L, Bayen E, Yaffe K. Unsupervised machine learning to identify high likelihood of dementia in population-based surveys: development and validation study. *J Med Internet Res* 2018 Dec 09;20(7):e10493 [FREE Full text] [doi: [10.2196/10493](https://doi.org/10.2196/10493)] [Medline: [29986849](https://pubmed.ncbi.nlm.nih.gov/29986849/)]
7. Farhan W, Wang Z, Huang Y, Wang S, Wang F, Jiang X. A predictive model for medical events based on contextual embedding of temporal sequences. *JMIR Med Inform* 2016 Nov 25;4(4):e39 [FREE Full text] [doi: [10.2196/medinform.5977](https://doi.org/10.2196/medinform.5977)] [Medline: [27888170](https://pubmed.ncbi.nlm.nih.gov/27888170/)]
8. Celi LA, Davidzon G, Johnson AE, Komorowski M, Marshall DC, Nair SS, et al. Bridging the health data divide. *J Med Internet Res* 2016 Dec 20;18(12):e325 [FREE Full text] [doi: [10.2196/jmir.6400](https://doi.org/10.2196/jmir.6400)] [Medline: [27998877](https://pubmed.ncbi.nlm.nih.gov/27998877/)]
9. Khan A, Zubair S. Machine learning tools and toolkits in the exploration of big data. *ijcse* 2018 Dec 31;6(12):570-575. [doi: [10.26438/ijcse/v6i12.570575](https://doi.org/10.26438/ijcse/v6i12.570575)]
10. Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2001 May 08;56(9):1133-1142. [doi: [10.1212/wnl.56.9.1133](https://doi.org/10.1212/wnl.56.9.1133)] [Medline: [11342677](https://pubmed.ncbi.nlm.nih.gov/11342677/)]
11. Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, de Mendonça A. Data mining methods in the prediction of dementia: a real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Res Notes* 2011 Aug 17;4:299 [FREE Full text] [doi: [10.1186/1756-0500-4-299](https://doi.org/10.1186/1756-0500-4-299)] [Medline: [21849043](https://pubmed.ncbi.nlm.nih.gov/21849043/)]
12. Bott N, Kumar S, Krebs C, Glenn JM, Madero EN, Juusola JL. A remote intervention to prevent or delay cognitive impairment in older adults: design, recruitment, and baseline characteristics of the Virtual Cognitive Health (VC Health) study. *JMIR Res Protoc* 2018 Aug 13;7(8):e11368 [FREE Full text] [doi: [10.2196/11368](https://doi.org/10.2196/11368)] [Medline: [30104186](https://pubmed.ncbi.nlm.nih.gov/30104186/)]
13. Portet F, Ousset PJ, Visser PJ, Frisoni GB, Nobili F, Scheltens P, MCI Working Group of the European Consortium on Alzheimer's Disease (EADC). Mild cognitive impairment (MCI) in medical practice: a critical review of the concept and new diagnostic procedure. Report of the MCI Working Group of the European Consortium on Alzheimer's Disease. *J Neurol Neurosurg Psychiatry* 2006 Jun;77(6):714-718 [FREE Full text] [doi: [10.1136/jnnp.2005.085332](https://doi.org/10.1136/jnnp.2005.085332)] [Medline: [16549412](https://pubmed.ncbi.nlm.nih.gov/16549412/)]
14. Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol* 2007 Aug;6(8):734-746. [doi: [10.1016/S1474-4422\(07\)70178-3](https://doi.org/10.1016/S1474-4422(07)70178-3)] [Medline: [17616482](https://pubmed.ncbi.nlm.nih.gov/17616482/)]
15. Marcus DS, Fotenos AF, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. *J Cogn Neurosci* 2010 Dec;22(12):2677-2684 [FREE Full text] [doi: [10.1162/jocn.2009.21407](https://doi.org/10.1162/jocn.2009.21407)] [Medline: [19929323](https://pubmed.ncbi.nlm.nih.gov/19929323/)]
16. Lynch J, Kaplan G. Socioeconomic position. In: Berkman L, Kawachi I, editors. *Social Epidemiology*. New York: Oxford University Press; 2000:13-35.
17. Magni E, Binetti G, Padovani A, Cappa SF, Bianchetti A, Trabucchi M. The Mini-Mental State Examination in Alzheimer's disease and multi-infarct dementia. *Int Psychogeriatr* 1996;8(1):127-134. [doi: [10.1017/s1041610296002529](https://doi.org/10.1017/s1041610296002529)] [Medline: [8805093](https://pubmed.ncbi.nlm.nih.gov/8805093/)]
18. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993 Nov;43(11):2412-2414. [doi: [10.1212/wnl.43.11.2412-a](https://doi.org/10.1212/wnl.43.11.2412-a)] [Medline: [8232972](https://pubmed.ncbi.nlm.nih.gov/8232972/)]
19. NIST/SEMATECH e-handbook of statistical methods. 2012 Apr. URL: <http://www.itl.nist.gov/div898/handbook/> [accessed 2019-02-10]
20. Sukel K. BrainFacts/SfN. 2018 Nov 15. Figuring out why Alzheimer's disease strikes more women than men URL: <https://www.brainfacts.org/diseases-and-disorders/topic-center-alzheimers-and-dementia/figuring-out-why-alzheimers-disease-strikes-more-women-than-men-1115183> [accessed 2020-03-06]
21. Hara Y. How does Alzheimer's affect women and men differently?. 2018 Jul 02. URL: <https://www.alzdiscovery.org/cognitive-vitality/blog/how-does-alzheimers-affect-women-and-men-differently> [accessed 2020-03-06]
22. Riedel BC, Thompson PM, Brinton RD. Age, APOE and sex: triad of risk of Alzheimer's disease. *J Steroid Biochem Mol Biol* 2016 Jun;160:134-147 [FREE Full text] [doi: [10.1016/j.jsbmb.2016.03.012](https://doi.org/10.1016/j.jsbmb.2016.03.012)] [Medline: [26969397](https://pubmed.ncbi.nlm.nih.gov/26969397/)]
23. Karp A, Kåreholt I, Qiu C, Bellander T, Winblad B, Fratiglioni L. Relation of education and occupation-based socioeconomic status to incident Alzheimer's disease. *Am J Epidemiol* 2004 Jan 15;159(2):175-183. [doi: [10.1093/aje/kwh018](https://doi.org/10.1093/aje/kwh018)] [Medline: [14718220](https://pubmed.ncbi.nlm.nih.gov/14718220/)]

24. Evans DA, Hebert LE, Beckett LA, Scherr PA, Albert MS, Chown MJ, et al. Education and other measures of socioeconomic status and risk of incident Alzheimer disease in a defined population of older persons. *Arch Neurol* 1997 Nov;54(11):1399-1405. [doi: [10.1001/archneur.1997.00550230066019](https://doi.org/10.1001/archneur.1997.00550230066019)] [Medline: [9362989](https://pubmed.ncbi.nlm.nih.gov/9362989/)]
25. Arevalo-Rodriguez I, Smailagic N, Roqué IFM, Ciapponi A, Sanchez-Perez E, Giannakou A, et al. 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 Rev* 2015 Mar 05(3):CD010783. [doi: [10.1002/14651858.CD010783.pub2](https://doi.org/10.1002/14651858.CD010783.pub2)] [Medline: [25740785](https://pubmed.ncbi.nlm.nih.gov/25740785/)]
26. Tate DF, Neeley ES, Norton MC, Tschanz JT, Miller MJ, Wolfson L, et al. Intracranial volume and dementia: some evidence in support of the cerebral reserve hypothesis. *Brain Res* 2011 Apr 18;1385:151-162 [FREE Full text] [doi: [10.1016/j.brainres.2010.12.038](https://doi.org/10.1016/j.brainres.2010.12.038)] [Medline: [21172323](https://pubmed.ncbi.nlm.nih.gov/21172323/)]
27. Fotenos AF, Snyder AZ, Girton LE, Morris JC, Buckner RL. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology* 2005 Mar 22;64(6):1032-1039. [doi: [10.1212/01.WNL.0000154530.72969.11](https://doi.org/10.1212/01.WNL.0000154530.72969.11)] [Medline: [15781822](https://pubmed.ncbi.nlm.nih.gov/15781822/)]
28. Storandt M, Grant EA, Miller JP, Morris JC. Longitudinal course and neuropathologic outcomes in original vs revised MCI and in pre-MCI. *Neurology* 2006 Aug 08;67(3):467-473. [doi: [10.1212/01.wnl.0000228231.26111.6e](https://doi.org/10.1212/01.wnl.0000228231.26111.6e)] [Medline: [16894109](https://pubmed.ncbi.nlm.nih.gov/16894109/)]
29. Killiany RJ, Gomez-Isla T, Moss M, Kikinis R, Sandor T, Jolesz F, et al. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. *Ann Neurol* 2000 Apr;47(4):430-439. [Medline: [10762153](https://pubmed.ncbi.nlm.nih.gov/10762153/)]
30. Killiany RJ, Hyman BT, Gomez-Isla T, Moss MB, Kikinis R, Jolesz F, et al. MRI measures of entorhinal cortex vs hippocampus in preclinical AD. *Neurology* 2002 Apr 23;58(8):1188-1196. [doi: [10.1212/wnl.58.8.1188](https://doi.org/10.1212/wnl.58.8.1188)] [Medline: [11971085](https://pubmed.ncbi.nlm.nih.gov/11971085/)]
31. Fox NC, Freeborough PA. Brain atrophy progression measured from registered serial MRI: validation and application to Alzheimer's disease. *J Magn Reson Imaging* 1997;7(6):1069-1075. [doi: [10.1002/jmri.1880070620](https://doi.org/10.1002/jmri.1880070620)] [Medline: [9400851](https://pubmed.ncbi.nlm.nih.gov/9400851/)]
32. Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage* 2004 Oct;23(2):724-738. [doi: [10.1016/j.neuroimage.2004.06.018](https://doi.org/10.1016/j.neuroimage.2004.06.018)] [Medline: [15488422](https://pubmed.ncbi.nlm.nih.gov/15488422/)]
33. Open Access Series of Brain Imaging (OASIS) database. URL: <https://www.oasis-brains.org> [accessed 2020-03-06]

Abbreviations

- AD:** Alzheimer disease
- APOE-ε4:** ε4 allele of the apolipoprotein E gene
- ASF:** Atlas Scaling Factor
- CDR:** Clinical Dementia Rating
- EDA:** exploratory data analysis
- EDUC:** patient education level
- eTIV:** estimated total intracranial volume
- IQR:** interquartile range
- MCI:** mild cognitive impairment
- M/F:** male/female
- MMSE:** Mini-Mental State Examination
- MRI:** magnetic resonance image
- nWBV:** normalized whole brain volume
- OASIS:** Open Access Series of Brain Imaging
- SES:** socioeconomic status

Edited by G Eysenbach; submitted 15.04.19; peer-reviewed by Y Hu, F Lanfranchi; comments to author 16.07.19; revised version received 08.09.19; accepted 09.02.20; published 14.04.20

Please cite as:

Khan A, Zubair S

Longitudinal Magnetic Resonance Imaging as a Potential Correlate in the Diagnosis of Alzheimer Disease: Exploratory Data Analysis
JMIR Biomed Eng 2020;5(1):e14389

URL: <http://biomedeng.jmir.org/2020/1/e14389/>

doi: [10.2196/14389](https://doi.org/10.2196/14389)

PMID:

©Afreen Khan, Swaleha Zubair. Originally published in JMIR Biomedical Engineering (<http://biomedeng.jmir.org>), 14.04.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Biomedical Engineering, is properly cited. The complete bibliographic information, a link to the original publication on <http://biomedeng.jmir.org/>, as well as this copyright and license information must be included.