

# Alzheimer's Disease Early Detection Using Machine Learning Techniques

Roobaea Alroobaea (✉ [roobaea.alroobaea@gmail.com](mailto:roobaea.alroobaea@gmail.com))

Taif University

Seifeddine Mechti

University of Sfax: Universite de Sfax

Mariam Haoues

University of Sfax: Universite de Sfax

Saeed Rubaiee

University of Jeddah

Anas Ahmed

University of Jeddah

Murad Andejany

University of Jeddah

Nicola Luigi Bragazzi

York University

Dilip Kumar Sharma

Jaypee Institute of Information Technology

Bhanu Prakash Kolla

Koneru Lakshmaiah Education Foundation

Sudhakar Sengan

PSN College of Engineering and Technology

---

## Research Article

**Keywords:** Alzheimer's Disease, Machine Learning, Dementia, Disease Prediction, Classification

**Posted Date:** June 17th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-624520/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Alzheimer's is the main reason for dementia, that affects frequently older adults. This disease is costly especially, in terms of treatment. In addition, Alzheimer's is one of the deaths causes in the old-age citizens. Early Alzheimer's detection helps medical staffs in this disease diagnosis, which will certainly decrease the risk of death. This made the early Alzheimer's disease detection a crucial problem in the healthcare industry. The objective of this research study is to introduce a computer-aided diagnosis system for Alzheimer's disease detection using machine learning techniques. We employed data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Open Access Series of Imaging Studies (OASIS) brain datasets. Common supervised machine learning techniques have been applied for automatic Alzheimer's disease detection such as: logistic regression, support vector machine, random forest, linear discriminant analysis, etc. The best accuracy values provided by the machine learning classifiers are 99.43% and 99.10% given by respectively, logistic regression and support vector machine using ADNI dataset, whereas for the OASIS dataset, we obtained 84.33% and 83.92% given by respectively logistic regression and random forest.

## 1. Introduction

Dementia is *"the loss of memory, language, problem solving and other thinking abilities that are severe enough to interfere with daily life tasks"* (Alzheimer's Association, 2019). It is not considered as a specific disease, but as a set of symptoms with an increasing decline in memory or other thinking and reasoning skills (Creative Caregivers LLC, 2020). In an aging society, dementia is a priority in health and social care. In fact, dementia usually affects the elderly, whereas 2% of people with dementia do not exceed 65 years (Alickovic & Subasi, 2020). Worldwide, around 50M people are with dementia and approximately 10M new cases every year (World Health Organization, 2020b). It is predicted that by 2030, the number of people having dementia will be around 75M, which will cost the society nearly US\$ 2 trillion (Prince et al., 2015). In today's world, dementia is more significant in terms of healthcare compared to other diseases (World Health Organization, 2020a). In addition, there is an important difficulty in diagnosing dementia due to the absence of a standardized test for its detection (Stamate et al., 2020). Therefore, there is no treatment until now to cure dementia, although, some treatments are available to support and improve the life of those patients as well as their caregivers.

Alzheimer's is a frequent type of dementia, which is considered as a main threat for the healthcare industry in today's world (Jo et al., 2019). It accounts for 60–80% of the population with dementia a (Creative Caregivers LLC, 2020). In 2018, around 50M people have been affected by Alzheimer's (Patterson, 2018). Moreover, Alzheimer's is the sixth leading cause of death in the United States (World Health Organization, 2020a). Alzheimer's disease can be clinically diagnosed by physical and neurological examination, which can be costly and time consuming. Alzheimer's symptoms generally develop slowly and get bad over time, which can become more severe and affect the daily activities (Alzheimer's Association, 2019). However, the early detection of this disease, before most of its symptoms are observable, is difficult. The prediction of Alzheimer's at pre-symptomatic stages is

recommended to slow down the disease progression. Currently, Alzheimer's disease is diagnosed by calculating the Multi Slice Multi Echo (MSME) score and by the manual study of the Magnetic Resonance Imaging (MRI) scan (Janghel & Rathore, 2020). This may require analyzing thousands of slides of brain tissue, which is lengthy and costly. Learning-based techniques can be used to speed up the diagnosis process and reduce its cost. In the recent years, intelligence scientists investigated the use of advanced technologies to improve Alzheimer's detection quality and precision. Thus, several machine learning models have been applied successfully for early disease detection (Kumar, 2019).

In this paper, we aim to propose a computer-aided detection system for Alzheimer's early detection using machine learning. The remainder of this research paper is as follows. In Sect. 2, we provide background information about Alzheimer's disease and review previous works on the use of learning-based methods in the Alzheimer's early detection. Section 3 describes the use of machine learning in the early detection of Alzheimer's disease. Section 4 presents and discusses the experimental results. We conclude this work in Sect. 5.

## 2. Alzheimer'S Disease And Related Work

In this section, we provide background information about Alzheimer's disease and survey previous research studies that proposed learning-based approaches for the early detection of this disease.

### 2.1. Alzheimer's Disease

Alzheimer's disease, mentioned directly as Alzheimer's, is a *"progressive neurological brain disease, which is caused due to the damage of nerve cells in parts of the brain"* (Alzheimer's Association, 2019).

Alzheimer's has mostly severe physical and psychological effects on the person with Alzheimer's and his family. Alzheimer's normally starts with a slow progression and get worse increasingly as time progresses. At the beginning of this disease, the first symptom that appears is the memory loss. In the advanced stages, Alzheimer's symptoms become more serious. Hence, a person with Alzheimer's may suffer from emotional changes (*e.g.*, depression, apathy, etc.), changes in behavior and even the decrease in physical abilities (*e.g.*, coordination, managing self-care, etc.). Until-to-days, there is no cure for Alzheimer's despite the worldwide effort to find better ways for treating this disease. Nevertheless, treatments for Alzheimer's symptoms are available. Those treatments are not able to prevent Alzheimer's progression, but they are used to temporarily reduce the worseness of its symptoms. Hence, the earlier a person is diagnosed with Alzheimer's, the sooner help he can receive. Recently, Alzheimer's disease received a remarkable focus in recent scientific research studies since it eventually leads to the people death. However, Alzheimer's diagnosis needs a good clinical assessment based on patient's medical history, several neuropsychological tests, and other pathological evaluations (Kundaram & Pathak, 2021). Those examinations can be costly and time-consuming.

### 2.2. Related Work

In the recent years, learning-based techniques such as supervised machine learning have been progressively being applied in healthcare. In particular, computer-aided detection systems that use learning-based techniques have been successfully applied in multiple diseases' detection (*e.g.*, heart disease (Barik et al., 2020), breast cancer (Asri et al., 2016), etc.). This information, if earlier correctly detected, can be beneficial for clinicians and patients. Table 1 presents a summary of the previous research studies that used machine learning techniques for Alzheimer's disease early detection [Alickovic & Subasi, 2020; Shahbaz et al., 2019]. Some of those studies are outlined in brief below.

Regarding the use of machine learning for the Alzheimer's detection, different models have been commonly used. For instance, Alickovic & Subasi conducted a comparative study to evaluate how well supervised machine learning models can be used in the Alzheimer's disease prediction (Alickovic & Subasi, 2020). This study focused mainly on: support vector machine, naïve bayes, k-nearest neighbours, random forest, artificial neural network and logistic regression. They conducted their experiments using the ADNI data repository (ADNI, 2017). The highest performances have been given by the random forest classifier with an accuracy of 85.77%, and the k-nearest neighbours classifier with an accuracy of 84.27%. Shahbaz et al., used five machine learning models (k-nearest neighbours, decision tree, rule induction, naïve bayes, and generalized linear model), and deep learning model to classify the five stages of Alzheimer's disease (Shahbaz et al., 2019). The authors performed their experiments using the ADNI data repository (ADNI, 2017). The highest performances have been given by the generalized linear model classifier with accuracy of 88.24%, and deep learning with an accuracy of 78.32%.

Table 1  
Learning-based techniques for the Alzheimer's disease early detection

Research study	Dataset	Models	Accuracy
Albright et al., 2019	ADNI	Neural network model	86.60%
Alickovic & Subasi, 2020	ADNI	Random forest	85.77%
		K- Nearest Neighbours	84.27%
		Naïve Bayes	75.16%
		Artificial Neural Network	76.03%
		Logistic Regression	75.28%
		Support Vector Machine	83.15%
Shahbaz et al., 2019	ADNI	K-Nearest Neighbours	43.26 %
		Decision Tree	74.22%
		Rule Induction	69.69%
		Naïve Bayes	74.65%
		Generalized Linear Model	88.24%
		Deep Learning	78.32%
Islam & Zhang, 2018	OASIS	Deep Convolutional Neural Network	93.18%

On the other hand, deep learning models have been successfully used in the Alzheimer's early detection. For instance, Albright et al., investigated the application of learning-based techniques in using clinical data to predict the progression of Alzheimer's in the future years (Albright et al., 2019). They proposed a multi-layered neural network model that includes one input layer, one hidden layer and one output layer. The authors performed their experiments using the ADNI data repository (ADNI, 2017). The accuracy value of the proposed model in this paper is equal to 86.60%. Finally, Islam & Zhang proposed to use a deep convolutional neural network in Alzheimer's diagnosis using brain Magnetic Resonance Imaging (MRI) data analysis (Islam & Zhang, 2018). Their experiments conducted using the Open Access Series of Imaging Studies (OASIS) dataset (Marcus et al., 2007) have shown that the deep convolutional neural network model proposed in this paper identifies the four stages of Alzheimer's (non-demented, very mild, mild and moderate dementia) with an accuracy of 93.18%.

As it is illustrated in Table 1, the frequently used dataset is the ADNI data repository (ADNI, 2017). Moreover, some studies conducted their experiments using numerical datasets that include pre-extracted features from MRI, while other studies conducted their experiments using MRI without handcrafted features. Regarding the machine learning models, the best accuracy values were given by random forest (85.77% in (Alickovic & Subasi, 2020)), and generalized linear model (88.24% in (Shahbaz et al., 2019)),

for the ADNI dataset. For the OASIS dataset, the best value was given by the deep convolutional neural network provided by Islam & Zhang with 93.18% (Islam & Zhang, 2018).

### 3. Machine Learning For Alzheimer'S Detection

#### 3.1. The Alzheimer's Disease Neuroimaging Initiative Dataset

The data used in this research study have been obtained from the widely used data repository, ADNI (Alzheimer's disease Neuroimaging Initiative) (ADNI, 2017). As it is illustrated in Table 1, this dataset has been widely used mainly to detect Alzheimer's disease at the earliest stages (ADNI, 2017). A detailed description of the ADNI dataset is given in Table 2.

Table 2  
A detailed description of the ADNI dataset

Examination Result	Number of Records
Cognitively Normal	2665
Mild Cognitive Impairment	3924
Alzheimer's Disease	1731
<b>Total</b>	<b>8320</b>

ADNI dataset includes data recorded from the North American male and female individuals that are "Cognitively Normal", with "Early Mild Cognitive Impairment", with "Late Mild Cognitive Impairment", or with "Alzheimer's Disease". The dataset used in this paper contains 502 attributes for 1737 participants. This dataset is longitudinal since it contains data from multiple visits per patient. In fact, ADNI contains records of individuals' examination, at different monthly intervals (*i.e.*, from 0 to 120 months), from July 2005 to May 2017. Consequently, this dataset contains a total of 8320 examinations (see Table 2). 2665 examinations are cognitively normal, 3924 examinations are with early or late mild cognitive impairment, and 1731 examinations are with Alzheimer's disease. The ADNI dataset has various parameters (*i.e.*, features) that can be used in the Alzheimer's disease detection such as: the Mini-Mental State Examination (MMSE) score, person age, person gender, number of visits, etc. As we mentioned previously, the MMSE score is the main parameter used for the Alzheimer's disease detection. In fact, if a person is affected by Alzheimer's, the MMSE score is reduced periodically.

Figure 1 presents the visualization of the ADNI dataset according to the DXCHANGE, PTGENDER, PTEDUCAT, and AGE parameters. DXCHANGE presents the examination result at a specific visit. 1.0 indicates that the patient is cognitively normal, whereas, 3.0 indicate that the patient is diagnosed with Alzheimer's. As it is illustrated in this Figure, patient with Alzheimer's are mostly male between 70 and 80

years old. Moreover, people with high education level are mostly diagnosed cognitively normal compared to people with lower education level.

## 3.2. ADNI Dataset Preprocessing

Data pre-processing is an important task in this research work. It refers to all the necessary transformations done on the data to be used.

### 3.2.1. ADNI dataset cleaning

As we mentioned in Sect. 3.1, the ADNI dataset contains a total of 502 parameters (*i.e.*, attributes). However, the 22 most relevant attributes that are mainly used in the Alzheimer's disease detection are represented in Table 3. They include String (*e.g.*, PTGENDER and PTMARRY), Number (*e.g.*, AGE, PTEDUCAT, etc.) and Boolean (*e.g.*, APOE4) data type. The class of people diagnosed with Alzheimer's disease is assigned as DXCHANGE = 3.0. The class of people with Mild cognitive impairment is assigned as DXCHANGE = 2.0. Finally, the class of people cognitively normal is assigned as DXCHANGE = 1.0. However, the ADNI dataset contains missing values. Hence, before being used by the different learning-based models, this dataset needs to be cleaned.

Table 3  
ADNI dataset attributes and their definitions

Attributes	Type	Definition	Missing values
PTGENDER	String	The patient's gender ( <i>e.g.</i> , Female, Male)	0
PTMARRY	String	The patient's marital status ( <i>e.g.</i> , Married, Divorced, etc.)	0
AGE	Number	The patient's age ( <i>e.g.</i> , 66, 75, etc.)	0
PTEDUCAT	Number	The patient's education ( <i>e.g.</i> , 10, 19, 20, etc.)	0
APOE4	Boolean	The APOE4 gene presence ( <i>e.g.</i> , 1.0, 0.0, etc.)	3
CDRSB	Number	The score of the Clinical dementia rating sum of boxes ( <i>e.g.</i> , 0.1, 0.5, etc.)	0
ADAS11	Number	The score of the 11 item-AD Cognitive Scale ( <i>e.g.</i> , 4, 19, etc.)	2
ADAS13	Number	The score of the 13 item-AD Cognitive Scale ( <i>e.g.</i> , 6, 30, etc.)	8
MMSE	Number	The score of the Mini-Mental State Examination	0
RAVLT_Immediate	Number	The score of the Rey's Auditory Verbal Learning Test for immediate response ( <i>e.g.</i> , 74.0, 13.0, etc.)	5
RAVLT_Learning	Number	The score of the Rey's Auditory Verbal Learning Test for learning ( <i>e.g.</i> , 1, 4, etc.)	5
RAVLT_Forgetting	Number	The score of the Rey's Auditory Verbal Learning Test for forgetting ( <i>e.g.</i> , 2, 5, etc.)	5
RAVLT_Perc_Forgetting	Number	The percentage of forgetting ( <i>e.g.</i> , 80, 40, etc.)	9
FAQ	Number	The functional activities questionnaire ( <i>e.g.</i> , 0, 2, etc.)	1
Ventricles	Number	The Volume of ventricles ( <i>e.g.</i> , 118233.0, 84599.0, etc.)	48
Hippocampus	Number	The volume of hippocampus ( <i>e.g.</i> , 8336.0, 5319.0, etc.)	112
WholeBrain	Number	The volume of Brain ( <i>e.g.</i> , 1229740.0, 1129830.0, etc.)	28
Entorhinal	Number	The volume of the entorhinal cortex ( <i>e.g.</i> , 4177.0, 1791.0, etc.)	133
Fusiform	Number	The volume of the fusiform gyrus ( <i>e.g.</i> , 16559.0, 15506.0, etc.)	133

Attributes	Type	Definition	Missing values
MidTemp	Number	The volume of the middle temporal gyrus ( <i>e.g.</i> , 27936.0, 18422.0, etc.)	133
ICV	Number	The Intra Cranial Volume ( <i>e.g.</i> , 1984660.0, 1920690.0, etc.)	7
DXCHANGE	Number	The diagnosis of the patient at the corresponding visits ( <i>e.g.</i> , 1 for cognitively normal, 2 for mild cognitive impairment and 3 for Alzheimer's disease.)	0

Since our objective is the Alzheimer's early detection, the first step for the data cleaning, we keep only the cognitively normal people (DXCHANGE = 1.0) and people with Alzheimer's (DXCHANGE = 3.0). Moreover, we removed redundant attributes that do not provide any relevant information in the disease detection. For instance, ADNI dataset includes six attributes to represent the examination time and date (*e.g.*, EXAMDATE, EXAMDATE\_bl, etc.). All those information have been removed, we kept only the VISCODE to represent the visit number. On the other hand, we checked the number of null values for each attribute (*i.e.*, feature) as provided in Table 3. We removed rows that include missing values. However, we kept those that include data in the main features mentioned in Table 3 in order to avoid data lost.

Finally, the ADNI dataset includes String data type (*e.g.*, PTGENDER, PTMARRY, etc). However, the learning-based models require that the inputs must be numeric; we used the one hot encoder to convert the non-numeric values into numeric values. After applying all the data cleaning tasks mentioned above, the data that will be used by our learning-based models includes 1000 instances with 22 features, where 521 patients are cognitively normal and 479 patients are with Alzheimer's.

### 3.2.2. Correlation Matrix

To understand the relationships between the different features in our dataset, we used the correlation matrix. This matrix highlights the most correlated attributes in the ADNI dataset.

Figure 2 shows the correlation matrix between the 22 selected features from the ADNI dataset. As shown in this Figure, there are few features correlated to each other (> 0.85). For instance, FAQ is correlated to CDRSB with 0.9; ADAS13 is correlated to ADAS11 with 0.98, etc. However, we decided to keep those features since the total number of features is restricted to 22.

## 4. Experiments And Discussion

This section presents and discusses the experimental results performed using the ADNI and OASIS datasets. We compare the results provided by the common machine learning models.

### 4.1 Experimental Results

We have performed several experiments with different parameters. In fact, after the pre-processing phase, the conversion of all the variables into numerical features and after keeping the pertinent features to be used by the machine learning models, we can now split the data into training and test sets. For this purpose, we used 5-fold cross-validation model. Finally, we evaluate the ability of using machine learning models in the Alzheimer’s disease detection using the accuracy, precision, recall, and F-measure metrics given by, respectively Eq. 1, Eq. 2, Eq. 3, and Eq. 4. The references and introduced labels in the evaluation are given in Table 4.

$$Accuracy (\%) = \frac{|TP+TN|}{|TP+TN+FP+FN|} \quad (1)$$

$$Precision(\%) = \frac{TP}{TP+FP} \quad (2)$$

$$Recall (\%) = \frac{TP}{TP+FN} \quad (3)$$

$$F1-score(\%) = \frac{2 * Recall * Precision}{Precision + Recall} \quad (4)$$

**Table 4. Reference and introduced labels**

		Algorithm Label	
		Cognitively Normal	Alzheimer’s Disease
Referenced label	Cognitively Normal	Cognitively Normal - Cognitively Normal (TP)	Cognitively Normal - Alzheimer’s Disease (FN)
	Alzheimer’s Disease	Alzheimer’s Disease - Cognitively Normal (FP)	Alzheimer’s Disease - Alzheimer’s Disease (TN)

In Table 5, we compare the results given by the machine learning models using the ADNI dataset. As it is provided in this Table, the best values given by the machine learning models have been provided by the logistic regression and the support vector machine models with 99.43% and 99.10%, respectively. The lowest accuracy value has been given by the naïve bayes classifier with 87.07%.

The same learning-based models have been also used with the OASIS dataset. In Table 6, we compare the results provided by the selected machine learning models using OASIS dataset. As it is illustrated in this table, the best accuracy values have been provided by the logistic regression classifier and random forest with respectively, 84.33% and 83.92%. Whereas, the lowest accuracy value has been given by the naïve bayes classifier with 71.91%.

**Table 5. Comparison of the performances of the machine learning models using ADNI**

Algorithms	Dataset	Performances (%)			
		F- measure	Accuracy	Precision	Recall
Logistic Regression	ADNI	99.46%	99.43%	99.30%	99.70%
Decision Tree	ADNI	98.10%	97.53%	96.62%	99.70%
Support Vector Machine	ADNI	99.19%	99.10%	99.30%	99.29%
K-nearest Neighbours	ADNI	98.09%	97.55%	96.96%	99.29%
Random Forest	ADNI	99.08%	98.89%	98.89%	99.28%
Naïve Bayes	ADNI	87.13%	87.07%	86.94%	87.34%
Linear Discriminant Analysis	ADNI	98.88%	98.62%	98.48%	99.19%

**Table 6. Comparison of the performances of the machine learning models using OASIS**

Algorithms	Dataset	Performances (%)			
		F- measure	Accuracy	Precision	Recall
Logistic Regression	OASIS	84.27%	84.33%	84.54%	84.14%
Decision Tree	OASIS	79.00%	79.46%	80.00%	79.00%
Support Vector Machine	OASIS	78.00%	77.67%	80.00%	78.00%
K-Nearest Neighbours	OASIS	82.39%	82.93%	84.13%	81.80%
Random Forest	OASIS	84.00%	83.92%	84.00%	84.00%
Naïve Bayes	OASIS	71.91%	71.97%	72.18%	71.78%
Linear Discriminant Analysis	OASIS	83.46%	83.72%	84.13%	83.32%

#### 4.2 Comparative Evaluation

In this section, we compare the results given in this paper to those provided in the related work section. We selected the research studies that used the same ADNI database (*cf.*, (Albright et al., 2019), (Alickovic & Subasi, 2020), and (Shahbaz et al., 2019)) or the OASIS dataset (*cf.*, (Islam & Zhang, 2018)).

**Table 7. Comparison of accuracy of the selected machine learning and state-of-the-art models over ADNI dataset**

Research Studies	Models	Database	Accuracy
Albright et al., 2019	Neural Network Model	ADNI	86.60%
Alickovic & Subasi, 2020	Random Forest	ADNI	85.77%
	K-Nearest Neighbours	ADNI	84.27%
	Support Vector Machine	ADNI	83.15%
Shahbaz et al., 2019	K-Nearest Neighbours	ADNI	43.26%
	Decision Tree	ADNI	74.22%
	Random Forest	ADNI	69.69%
Our Study	Random Forest	ADNI	98.89%
	<b>Support Vector Machine</b>	<b>ADNI</b>	<b>99.10%</b>
	K-Nearest Neighbours	ADNI	97.55%
	Decision Tree	ADNI	97.53%
	<b>Logistic Regression</b>	<b>ADNI</b>	<b>99.43%</b>
	Random Forest	OASIS	83.92%
	Logistic Regression	OASIS	84.33%
Islam & Zhang, 2018	Deep Convolutional Neural Network	OASIS	93.18%

Table 7 provides the results of the comparison between our obtained results with the state-of-the-art models. As it is illustrated in this table, we obtained better results than the state-of-the-art models. For instance, compared to the results obtained by (Alickovic & Subasi, 2020), that used random forest, k-nearest neighbours, and support vector machine, we still achieve better performances for the k-nearest neighbours with 97.55% compared to 84.27%, for the support vector machine with 99.10% compared to 83.15%, and for the random forest classifier with 99.43% compared to 85.77%. Finally, we compared our obtained results with those given by (Shahbaz et al., 2019), that used random forest, k-nearest neighbours, and decision tree. We achieved better performance for the k-nearest neighbours with 97.55% compared to 43.26%, for the random forest with 98.89% compared to 69.69%, and for the decision tree with 97.53% compared to 74.22%.

## 5. Conclusion

Dementia, in particular Alzheimer's disease has an important impact on the society healthcare. However, the early detection of this disease is recommended to slow down the symptoms progression and avoid brain damage. Hence, such information, if earlier detected, can help people with Alzheimer's having a healthy life as well as their families' members. In the herein presented work, we proposed to use machine

learning models for Alzheimer's disease detection. The evaluation of the classification models is performed using the ADNI and OASIS datasets. The experimental results shown that the best accuracy values provided by the machine learning models are 99.43% and 99.10% given by respectively, logistic regression and support vector machine using ADNI dataset, whereas for the OASIS dataset, we obtained 84.33% and 83.92% given by respectively logistic regression and random forest.

Further research work could be conducted to focus on other diseases. It will be also interesting to be restricted on MRI without any handcrafted features.

## Declarations

**Funding-** Not applicable

**Conflicts of Interest/Competing Interests-**Not applicable

**Availability of Data and Material-**Not applicable

**Code Availability-**Not Applicable

## References

1. ADNI. Alzheimer's disease neuroimaging initiative. (2017).
2. Albright, J., Initiative, A. D. N., et al. (2019). Forecasting the progression of alzheimer's disease using neural networks and a novel preprocessing algorithm. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 5:483–491.
3. Alickovic, E. and Subasi, A. (2020). Automatic detection of alzheimer disease based on histogram and random forest. In Badnjevic, A., Škrbic, R., and Gurbeta Pokvić, L., editors, *CMBEBIH*. Springer International Publishing.
4. Alzheimer's Association (2019). Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 15(3):321–387.
5. Asri, H., Mousannif, H., Al Moatassime, H., and Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, 83:1064–1069.
6. Barik, S., Mohanty, S., Rout, D., Mohanty, S., Patra, A. K., and Mishra, A. K. (2020). Heart disease prediction using machine learning techniques. In *Advances in Electrical Control and Signal Systems*, pages 879–888. Springer.
7. Creative Caregivers LLC (2020). What is dementia?
8. Islam, J. and Zhang, Y. (2018). Brain mri analysis for alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain informatics*, 5(2):2.
9. Janghel, R. and Rathore, Y. (2020). Deep convolution neural network-based system for early diagnosis of alzheimer's disease. *IRBM*.

10. Jo, T., Nho, K., and Saykin, A. J. (2019). Deep learning in alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. *Frontiers in aging neuroscience*, 11:220.
11. Kumar, U. (2019). Applications of machine learning in disease pre-screening. In *Pre-Screening Systems for Early Disease Prediction, Detection, and Prevention*, pages 278–320. IGI Global.
12. Kundaram, S. S. and Pathak, K. C. (2021). Deep learning-based alzheimer disease detection. In *Springer Science and Business Media LLC*, pages 587–597. Springer.
13. Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2007). Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9):1498– 1507.
14. Patterson, C. (2018). The state of the art of dementia research: New frontiers. *World Alzheimer Report*.
15. Prince, M., Wimo, A., Guerchet, M., Ali, G., Wu, Y., Prina, M., et al. (2015). The global impact of dementia. *World Alzheimer Report*, pages 1–82.
16. Shahbaz, M., Ali, S., Guergachi, A., Niazi, A., and Umer, A. (2019). Classification of alzheimer's disease using machine learning techniques. In *DATA*, pages 296–303.
17. Stamate, D., Smith, R., Tsygancov, R., Vorobev, R., Langham, J., Stahl, D., and Reeves, D. (2020). Applying deep learning to predicting dementia and mild cognitive impairment. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 308–319. Springer.
18. World Health Organization (2020a). Alzheimer's disease fact sheet.
19. World Health Organization (2020b). Dementia.

## Figures

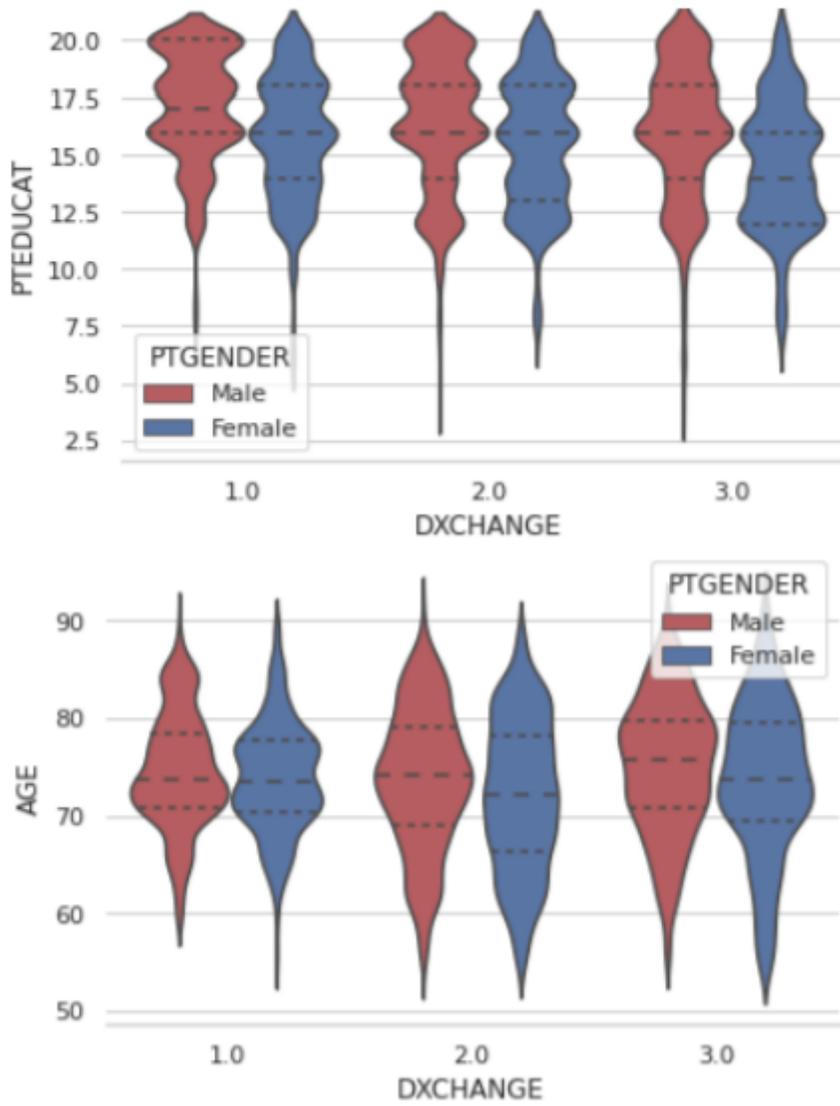


Figure 1

Visualization of the ADNI dataset according to DXCHANGE, PTGENDER, PTEDUCAT, and AGE attributes

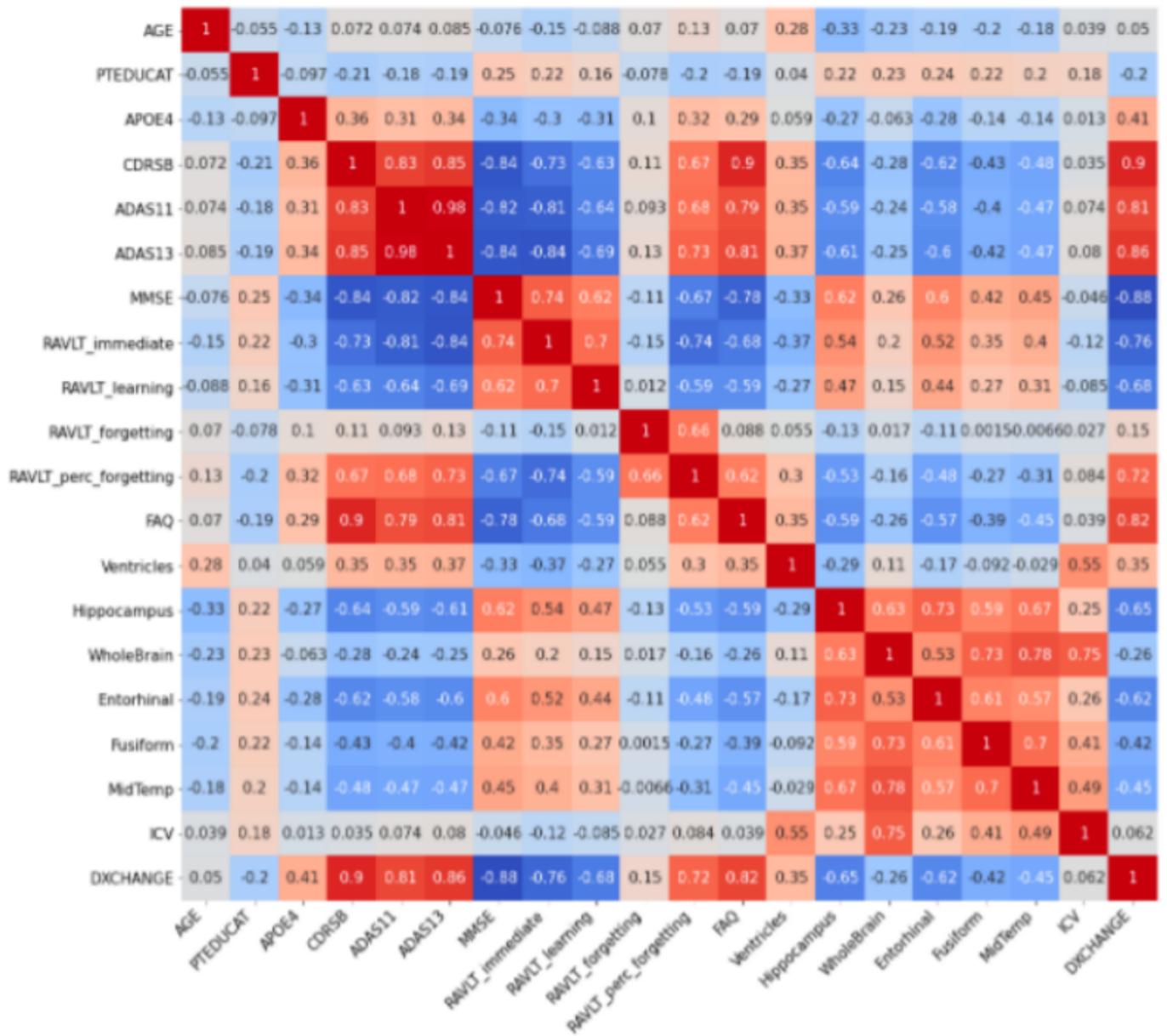


Figure 2

Correlation Matrix for the ADNI dataset