



Multiple Machine Learning Models for Detection of Alzheimer's Disease Using OASIS Dataset

Preety Baglat, Ahmad Waleed Salehi, Ankit Gupta, Gaurav Gupta

► To cite this version:

Preety Baglat, Ahmad Waleed Salehi, Ankit Gupta, Gaurav Gupta. Multiple Machine Learning Models for Detection of Alzheimer's Disease Using OASIS Dataset. International Working Conference on Transfer and Diffusion of IT (TDIT), Dec 2020, Tiruchirappalli, India. pp.614-622, 10.1007/978-3-030-64849-7_54 . hal-03701760

HAL Id: hal-03701760

<https://inria.hal.science/hal-03701760>

Submitted on 22 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Multiple Machine Learning Models for Detection of Alzheimer's Disease using OASIS Dataset

Preety Baglat¹, Ahmad Waleed Salehi², Ankit Gupta³, Gaurav Gupta^{*}

^{1,2,*} Yogananda School of AI, Computers and Data Science, Shoolini University, Himachal Pradesh 173229, India

¹Universita della Svizzera Italiana, Switzerland

³Interactive Technologies Institute (IT/LARSyS) and M-ITI, 9020-105 Funchal, Portugal

Abstract. Alzheimer's Disease (AD) is the most common form of dementia that can lead to a neurological brain disorder that causes progressive memory loss as a result of damaging the brain cells and the ability to perform daily activities. This disease is one of kind and fatal. Early detection of AD because of its progressive threat and patients all around the world. The early detection is promising as it can help to predetermine the condition of lot of patients they might face in the future. So, by examining the consequences of the disease, using MRI images we can get the help of Artificial intelligence (AI) technology to classify the AD patients if they have or may not have the deadly disease in future. In recent years, AI-based Machine Learning (ML) techniques are very useful for the diagnosis of AD. In this paper, we have applied different machine learning techniques such as Logistic Regression, Decision Tree, Random forest classifier, Support Vector Machine and AdaBoost for the earlier diagnosis and classification of Alzheimer 's disease using Open Access Series of Imaging Studies (OASIS) dataset, in which a significant performance and result gained on classification with Random Forest classifier.

Keywords: Alzheimer's Disease, Machine Learning, SVM, Logistic Regression, Random Forest, Decision Tree, AdaBoost, OASIS.

1 Introduction

Alzheimer Disease (AD) is a type and the most common form of dementia that can leads to neurological or brain disorder that causes cognitive deterioration and progressive memory loss as a result of brain cells death. Normally the symptoms and signs in AD cases develop very slowly and get terrible enough that involve affecting in their daily routine life [1]. Even though the main cause of AD is not only problem with old age but as well as in its early stages, the loss of the memory is mild, and their skills ability are dramatically changed [2]. The possible effective treatment of this disease is early diagnosis. Especially at an initial stage of AD diagnosis is a challenging task. It has been estimated that 1 out of 85 people around the world can affect with AD disease by the year 2050 [3]. Those are in the final stage of this disease are bed-bound which also require care around-the-clock.

Previous studies showed that in the most patients of AD the language function is lost. So, for early diagnosis of Alzheimer usually a neuropsychological examination is used. And the researchers believe and stated that the key for preventing from the disease is the early detection of AD [4]. The Symptoms of Alzheimer's disease vary among individuals.

1.1 Common and initial symptom which are gradually worsening are [5]

- Memory loss that affect daily life
- Difficulty in completing familiar tasks
- Confusion with place or time
- Trouble in understanding of visual image and spatial relationships
- New problems and troubles with words in writing or speaking
- Misplacing in things and losing the capability to retrace steps

Diagnosing of AD requires careful and a complete medical evaluation. the psychological cognitive test accuracy is completely depends on the capability and experience of the clinician. Make use of this test with huge number of AD patients will make the use of more effort, money, and time. So, it is important to build or develop Artificial Intelligence-based automatic detection model for classification of the disease [6]. Medical specialists are responsible in analyzing the interpreted of medical data, this technique is quite hard and limited for a medical specialist to interpret the images due to image subjectivity and complexity. So, in other areas of the real-world application the usage of Artificial Intelligence (AI), Machine Learning (ML) algorithms has shown promising and accurate in providing output for medical data [7]. With the speedy growth of AI-based approaches machine learning algorithms are capable of classification, feature extraction and will also assist in the accurate diagnosing of AD patients with less amount of time using images such as MRI, PET[8].

The aimed of this paper is to train an automated model based on the combination of ML algorithms includes Support Vector Machine (SVM), Random Forest, Logistic Regression for AD subjects. The classification is based on Open Access Series of Imaging Studies (OASIS) library of MRI patient scans. In this paper, the first section includes the experimental data, method and the algorithms used for classification of AD in the second section, then followed with next section results and discussion, and the last section is the conclusion part.

2 Experimental Data

The data is from OASIS dataset MRI scans obtained from <https://www.oasis-brains.org>. OASIS is a project which is aimed at making and providing neuroimaging datasets of the brain available freely to the scientific researchers [9]. By collecting and distributing neuroimaging datasets freely, we hope to facilitate and help future discoveries and clinical neuroscience like other initiatives for example Alzheimer's Disease Neuroimaging Initiative (ADNI). In this study, we have describe T1-

weighted MRI data from OASIS with and without AD subjects from older adults. In this MRI scans images dataset, we have taken a longitudinal collection of 150 individuals aged from 60 to 96 years.

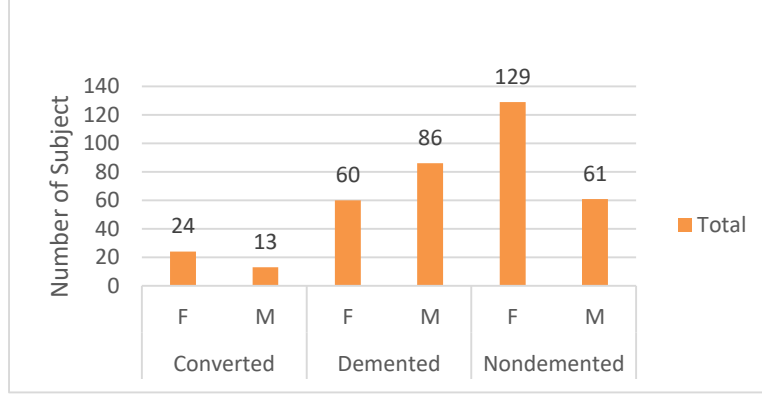


Fig 1. Count of subject ID by group and gender (F=Female, M=Male)

3 Method and MRI Classifiers

In this work, we have implemented different machine learning models such as Logistic regression, Decision tree, Random forest classifier, Support vector machine and AdaBoost for the classification of Alzheimer's disease using Anaconda for Python and TensorFlow with 8 GB RAM and Graphics Intel HD 6000 1536 MB. 5-fold cross-validation is used due to the small dataset.

3.1 Support Vector Machine Classifier

A support vector machine is a type of supervised learning method that is used for the classification and the regression problem. Prediction of Disease the most popular binary classification algorithm SVM is used for the structural MRI scan images. SVM is machine learning derived classifiers which is used for the mapping a vector of predictors into a higher dimensional plane by either using two different linear and non-linear kernel functions. The binary classification problem these two-dimensional objects which belong to two classes (+1,-1) that are simple to find out the line that can separate them perfectly in a higher dimension hyperplane according to structural risk minimization principle [10]. The objective of SVM is to build a discriminating function that use training data which will classify new examples (m, n). When there is no training data possible for linear separation then support vector machines can work efficiently in combination with kernel models using the kernel trick. So, that hyperplane can defining the SVMs communicate to a nonlinear decision boundary. Only this way the decision function can communicate in terms of the support vectors:

$$f(m) = \text{sign} \left(\sum a_i y_i K(s_i, x) + w_0 \right)$$

Where K is a kernel function and s_i is the support vectors and α_i is a weight constant that is derived from the support vector machine [11].

3.2 Decision Tree Classifier

Decision tree classifier is a type of supervised machine learning that splits the dataset based on certain parameter that maximizes the separation of data and gives the result in the form of a tree-like structure [12]. Decision tree algorithm is used for classification and regression problem and the algorithm creates a binary tree and each node has two edges for finding the best categorical and numeric feature to split by using suitable impurity criterion. For decision tree classification, use Gini and Entropy impurity. In Gini impurity.

$$\sum_{i=1}^c f_i(1 - f_i)$$

Where n is the number of unique labels and f_i is the frequency of label x at a node. In Entropy, f_i is the frequency of label x at a node and n is the number of unique labels.

$$\sum_{i=1}^c -f_i \log(f_i)$$

3.3 Logistic Regression Classifier

Logistic regression is a type of machine learning which comes under the supervised learning technique. This technique is a predictive analysis and used in a way to describe data and explain the relationship between a dependent binary variable and one or more ordinal, or nominal variables [13]. The outcome is measured with a dichotomous variable [14].

$$p = b_0 + b_1X_1 + b_2X_2 + \text{odds} = \frac{p}{1+p}$$

$$\log \text{it}(p) = \ln\left(\frac{p}{1-p}\right)$$

3.4 Random Forest Classifier

Random forest is an ensemble learning method used for the classification. which creates a set of decision trees together. A decision tree is the most important element in a random forest classifier which is based on bagging which is a sampling method and the same samples are using again and it will be put back into the dataset [15]. Suppose there is a dataset with n samples and if m decision tree is needed for that then m data sets for bagging will be built for training. For the decision tree, each node will be represented by the features used for the classification.

3.5 Adaptive Boosting Classifier

The Adaptive Boosting (AdaBoost) is a classifier that is based on the enhancement methods and boosting methods increase the accuracy of weak classifiers by taking the repeated iteration. Aims of every iteration are to be decreasing the classified data and increasing the weight of missorted data [16].

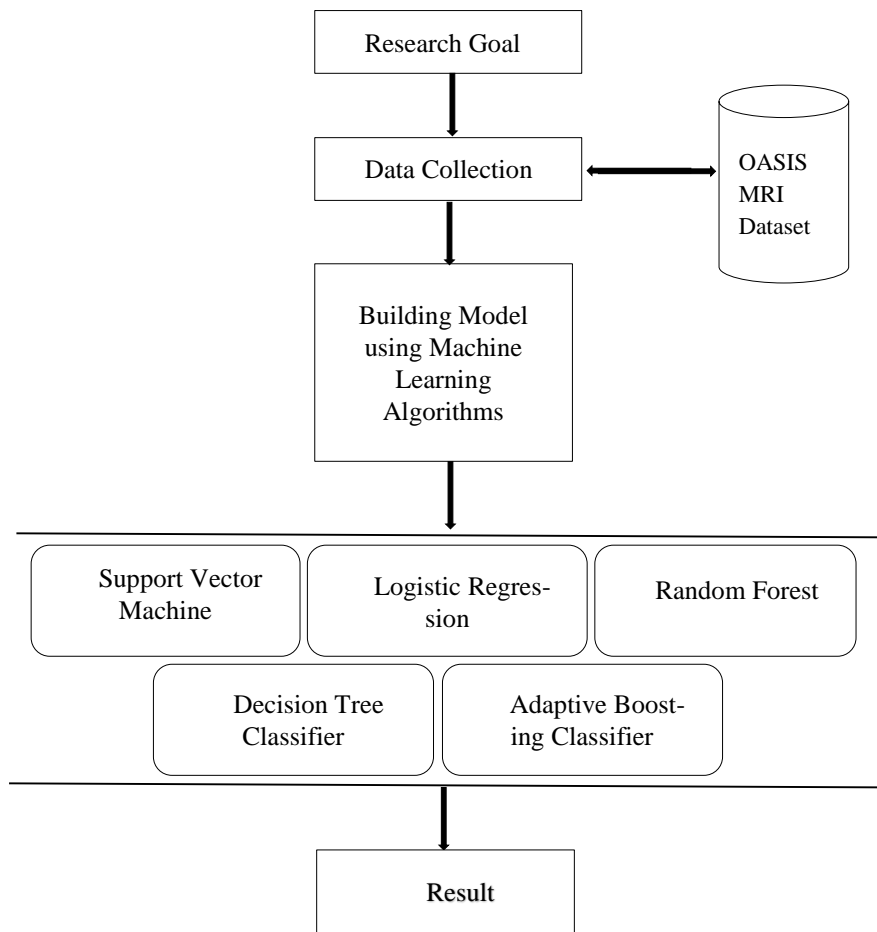


Fig 2. Workflow diagram

4 Results

For classification of Alzheimer's disease using machine learning models on OSIAS dataset the current performance of our models using Logistic Regression with the accuracy rate of 78%, 81% form SVM, Decision tree and 86% from Random Forest and Adaptive Boosting classifier.

Table 1. Performance metrics for each model.

SN	Model	Accuracy	Recall	AUC
1	Logistic Regression(w/imputation)	0.789474	0.75	0.791667
2	Logistic Regression(w/dropping)	0.750000	0.70	0.700000
3	SVM	0.815789	0.70	0.822222
4	Decision Tree	0.815789	0.65	0.825000
5	Random Forest	0.868421	0.80	0.872222

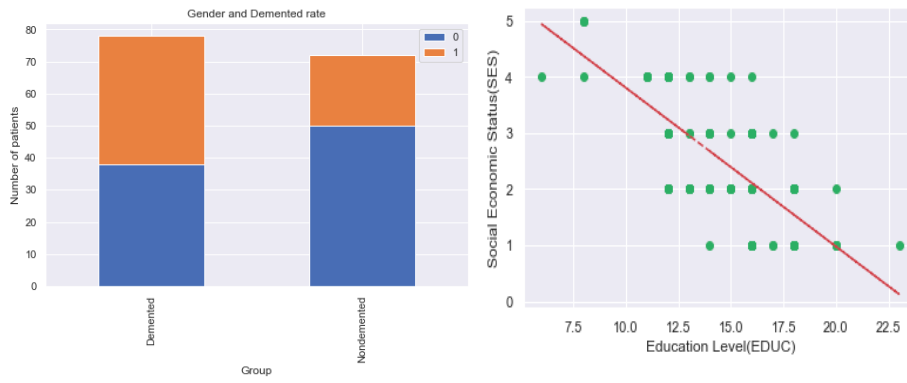


Fig. 3. a) Gender and Demented rate (Male=0, Female=1), b) SEES and EDUC

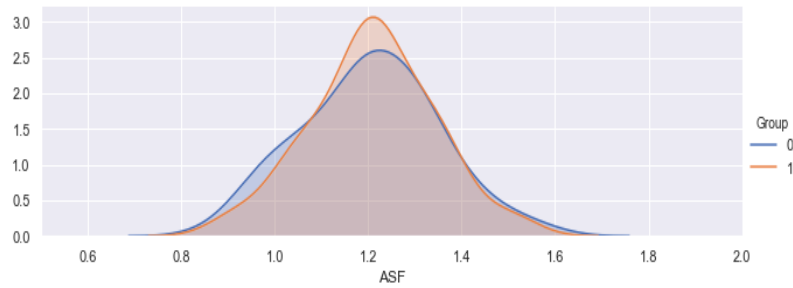


Fig 4. Bar chart of Atlas Scaling Factor (ASF) (Nondemented=0, demented= 1)

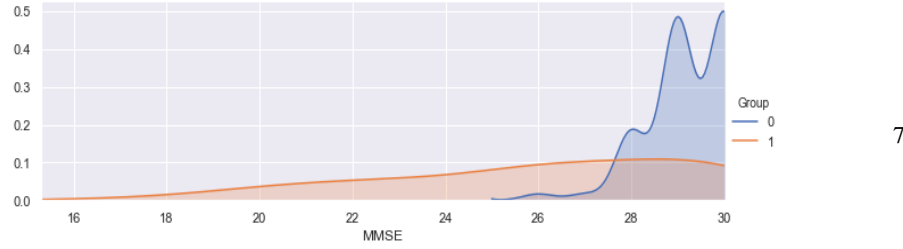


Fig 5. Mini-Mental State Examination (MMSE)

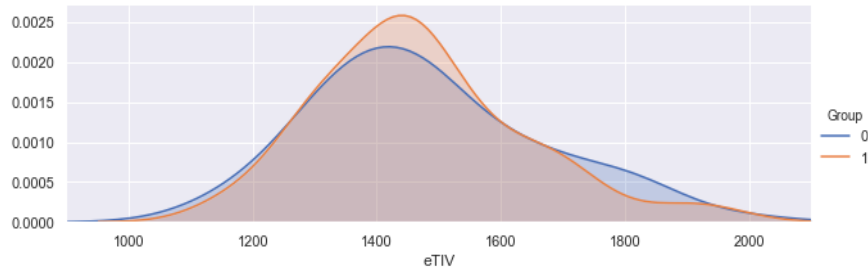


Fig 6. Estimated Total Intracranial Volume (eTIV)

5 Discussion and Conclusions

The recent developments in the course of biomedical engineering the analysis of medical images become one of the major research areas. Application of ML in analyzing of medical images is one of the reasons for this development [17]. In the last years classification using ML has been highly used and the most important AI-based development for automatic diagnosis of disease in its early stage to satisfy clinicians primary goals and needs [18]. An automated detection and classification framework for Alzheimer's disease using medical images is very crucial for the early detection of the AD patients.

In this study, we have proposed different ML algorithms for classification of AD patients using T1-weighted MRI data from OASIS dataset using different models such as Logistic Regression, Decision Tree, Random Forest classifier, SVM, and AdaBoost. Out of all these models, a significant result achieved when dealing with Random Forest and Adaptive boosting classifier with the accuracy rate of 86%.

We are aiming at encouraging of future work and looking forward. Therefore, in future studies these results could be further improved by carrying out other advanced AI-based techniques such as deep learning which is a subset of ML and has recently shown its potentiality in neuroimaging investigations. Thus, the combination of multi-approaches like Convolutional Deep Neural Network and SVM using multi-sites MRI data like ADNI would significantly enhance the ability of the model in early detection of disease.

References

1. A. W. Salehi, P. Baglat, and G. Gupta, "Alzheimer's Disease Diagnosis using Deep Learning Techniques," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 3, pp. 874–880, 2020, doi: 10.35940/ijeat.c5345.029320.
2. F. Bature, B. A. Guinn, D. Pang, and Y. Pappas, "Signs and symptoms preceding the diagnosis of Alzheimer's disease: A systematic scoping review of literature from 1937 to 2016," *BMJ Open*, vol. 7, no. 8, 2017, doi: 10.1136/bmjopen-2016-015746.
3. S. Rathore, M. Habes, M. A. Ifthikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *Neuroimage*, vol. 155, no. April, pp. 530–548, 2017, doi: 10.1016/j.neuroimage.2017.03.057.
4. J. Gaugler, B. James, T. Johnson, K. Scholz, and J. Weuve, "2016 Alzheimer's disease facts and figures," *Alzheimer's Dement.*, vol. 12, no. 4, pp. 459–509, 2016, doi: 10.1016/j.jalz.2016.03.001.
5. Alzheimer's Association, "Memory Loss & 10 Early Signs of Alzheimer's," *Alzheimer*. 2009, [Online]. Available: http://www.alz.org/alzheimers_disease_10_signs_of_alzheimers.asp.
6. G. Lyu, "A Review of Alzheimer's Disease Classification Using Neuropsychological Data and Machine Learning," *Proc. - 2018 11th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2018*, no. 2017, pp. 1–5, 2019, doi: 10.1109/CISP-BMEI.2018.8633126.
7. G. Gupta, A. Gupta, V. Jaiswal, and M. D. Ansari, "A review and analysis of mobile health applications for alzheimer patients and caregivers," *PDGC 2018 - 2018 5th Int. Conf. Parallel, Distrib. Grid Comput.*, pp. 171–175, 2018, doi: 10.1109/PDGC.2018.8745995.
8. S. Sarraf, D. D. DeSouza, J. Anderson, G. Tofghi, and for the A. D. N. Initiative, "DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI," *bioRxiv*, no. August 2016, p. 070441, 2017, doi: 10.1101/070441.
9. and T. A. R. Hoda Badr, Cindy L. Carmack, Deborah A. Kashy, Massimo Cristofanilli, "基因的改变NIH Public Access," *Bone*, vol. 23, no. 1, pp. 1–7, 2011, doi: 10.1161/CIRCULATIONAHA.110.956839.
10. J. Maroco, D. Silva, ... A. R.-B., and U. 2011, "Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic," *Bmcresnotes.Biomedcentral.Com*, 2011.
11. A. Savio, "Supervised classification using deformation-based features for Alzheimer's disease detection on the OASIS cross-sectional database," *Front. Artif. Intell. Appl.*, vol. 243, pp. 2191–2200, 2012, doi: 10.3233/978-1-61499-105-2-2191.
12. S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: A methodology review," *J. Biomed. Inform.*, vol. 35, no. 5–6, pp. 352–359, 2002, doi: 10.1016/S1532-0464(03)00034-0.
13. Solutions. Statistics, "What is Logistic Regression? - Statistics Solutions," *Solutions Statistics*. p. 1, 2016, doi: 10.1016/j.radonc.2012.09.019.
14. S. Bhol, "Comparative Analysis for the Detection of Alzheimer's using Multiple Machine Learning Models," pp. 4–6.
15. L. Xu, G. Liang, C. Liao, G. Den Chen, and C. C. Chang, "K-SkIP-N-Gram-RF: A random forest based method for Alzheimer's disease protein identification," *Front. Genet.*, vol. 10, no. FEB, pp. 1–7, 2019, doi: 10.3389/fgene.2019.00033.

16. R. B. Patil, R. Piyush, and S. Ramakrishnan, "Identification of brain white matter regions for diagnosis of Alzheimer using Diffusion Tensor Imaging," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 6535–6538, 2013, doi: 10.1109/EMBC.2013.6611052.
17. S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical Image Analysis using Convolutional Neural Networks: A Review," *J. Med. Syst.*, vol. 42, no. 11, pp. 1–13, 2018, doi: 10.1007/s10916-018-1088-1.
18. L. Sharma, G. Gupta, and V. Jaiswal, "Classification and development of tool for heart diseases (MRI images) using machine learning," *2016 4th Int. Conf. Parallel, Distrib. Grid Comput. PDGC 2016*, no. January, pp. 219–224, 2016, doi: 10.1109/PDGC.2016.7913149.