Zeelkumar Gajera, 2023, 11:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Early Detection of Alzheimer's Disease: Through Neuroimaging Using Machine Learning Techniques

Zeelkumar Gajera,
Machine Learning & Artificial Intelligence ,
PP Savani University

YugantKumar Gajera,
MSc in Computer Science,
Pace University

Anshu Patel
Computer Engineering
PP Savani University

zeelgajera1010@gmail.com,yugantgajera@gmail.com,anshupatel3009@gmail.com

Abstract- Alzheimer's disease (AD) is a progressive neurodegenerative disorder that affects millions of people worldwide and has no cure. Early diagnosis of AD can help patients and their families to plan for the future, access appropriate care and support, and participate in clinical trials that may slow down the disease progression. In this paper, we review the current state of the art in using machine learning techniques for early detection of AD, and identify the challenges and opportunities for future research. We first provide an overview of AD and its diagnosis criteria.We then describe the main types of data and features that can be used for early detection of AD, such as structural MRI images, cognitive tests, or biomarkers, and the machine learning techniques that can be applied to them, such as classification, regression, clustering, or dimensionality reduction. We also compare and evaluate the performance and limitations of different machine learning techniques for early detection of AD using various evaluation metrics and criteria. Finally, we discuss the implications and applications of using machine learning techniques for early detection of AD, and suggest some directions for future research. We hope that this paper will inspire and inform researchers, practitioners, and policymakers who are interested in using machine learning techniques for early detection of AD, and contribute to the advancement of this important and promising field.

Keywords **- Alzheimer's, Neuroimaging, Machine Learning.**

I. INTRODUCTION

Imagine if you could detect Alzheimer's disease before it causes irreversible damage to your brain and memory. Imagine if you could use a simple test or scan to diagnose your risk of developing this devastating condition. Imagine if you could prevent or delay the onset of cognitive decline and dementia with early intervention and treatment. This is not a fantasy, but a realistic possibility with the help of machine learning techniques. Machine learning is a branch of artificial intelligence that enables computers to learn from data and make predictions or decisions without explicit programming. Machine learning techniques can be applied to various types of data, such as images, texts, sounds, or biological signals, to extract meaningful patterns and insights. Machine learning techniques can also be used to model complex phenomena, such as human behaviour, cognition, or disease progression, and to provide personalized recommendations or solutions. In this paper, we explore how machine learning techniques can be used for early detection of Alzheimer's disease, a progressive neurodegenerative disorder that affects millions of people worldwide and has no cure. Alzheimer's disease is characterized by the accumulation of amyloid plaques and neurofibrillary tangles in the brain, which impair the function and communication of neurons.

Alzheimer's disease leads to gradual loss of memory, language, reasoning, and other cognitive abilities, as well as changes in personality and behaviour. Alzheimer's disease is usually diagnosed at a late stage, when the symptoms are already severe and the brain damage is irreversible.

However, recent studies have shown that Alzheimer's disease can be detected at an early stage, before the onset of clinical symptoms, using various types of data and machine learning techniques. For example, structural magnetic resonance imaging (MRI) images can reveal the shrinkage or atrophy of brain regions affected by Alzheimer's disease. Cognitive tests can measure the performance or decline of mental functions related to Alzheimer's disease. Biomarkers can indicate the presence or level of amyloid plaques or neurofibrillary tangles in the brain or cerebrospinal fluid. Machine learning techniques can analyze these data and features to classify or predict the diagnosis or prognosis of Alzheimer's disease with high accuracy and reliability.

The aim of this paper is to review the current state of the art in using machine learning techniques for early detection of Alzheimer's disease, and to identify the challenges and opportunities for future research. We first provide an overview of Alzheimer's disease and its diagnosis criteria. We then describe the main types of data and features that can be used for early detection of Alzheimer's disease, and the machine learning techniques that can be applied to them. We also compare and evaluate the performance and limitations of different machine learning techniques for early detection of Alzheimer's disease. Finally, we discuss the implications and applications of using machine learning techniques for early detection of Alzheimer's disease, and suggest some directions for future research.

II. DATASETS

Data used in the preparation of this article were obtained from the publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) database and the Open Access Series of Imaging Studies (OASIS) project database. The most recent visit in which a diagnosis was made was considered the best available "ground-truth" to train the classifiers.

Furthermore, the most recent diagnosis visit must have been at least 1 year after the selected scan for classifier training. The maximum follow-up time was of 3 years. Furthermore, diagnosis transitions must have occurred at least 6 months after the MRI scan. For each subject, we selected their earliest available structural MRI scan that fulfilled our study's requirements. Differences between diagnoses in age and sex (which are known to impact brain structure) and "time from MRI to most recent diagnosis" were estimated with 1-way ANOVAs (or, if residuals had nonparametric distributions, Mann-Whitney tests) and chisquare.

III. ADNI

The ADNI was launched in 2003 as a public-private partnership, led by principal investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. From ADNI, 570 subjects were included (211 HC, 188 MCI, 171 AD). To ensure diagnostic criteria equivalence across the different ADNI samples (ADNI, ADNI2, ADNIGO, and ADNI3), besides the Diogo et al. Alzheimer's Research & Therapy (2022) diagnosis attributed by the clinician based on a clinical interview and exam results, subjects had to fulfill additional criteria based on the ADNI2 procedures manual. Specifically, HC must have a Mini-Mental State Exam (MMSE) score of at least 24 and a Clinical Dementia Rating (CDR) of 0; MCI patients must have an MMSE score of at least 24 and a CDR of 0.5 with a Memory Box score of at least 0.5; and AD patients must have an MMSE score below 27 and a CDR of at least 0.5.

IV. OASIS

From the OASIS-3 dataset of the OASIS project, 531 subjects were included (463 HD, 70 AD). OASIS subjects had to fulfill the same MMSE and CDR requirements used for ADNI subjects and must have been diagnosed by a clinician based on a clinical interview and exam results.

V. MACHINE LEARNING MODEL

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 ai,2019	ADNI	Neural network model	86.60%
Alickovic& Subasi,2020	ADNI	Support Vector Machine	83.15%
		Random forest	85.77%
		K-Nearest Neighbours	84.27%
		Naïve Bayes	75.16%
		Artificial Neural Network	76.03%
Shahbaz et ai,2019	ADNI	Logistic Regression	75.28%
		K-Nearest Neighbours	43.26%
		Decision Tree	74.22%
		Rule Induction	69.69%
		Naïve Bayes	74.65%
		Generalized Linear Model	88.24%
Islam & Zhang,2018	OASIS	Deep Learning	78.32%
		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%. 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)), Page 6/16 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).

VI. ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE DATASET

Table 2 Detailed description of the ADNI dataset.

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

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.

VI. DATA CLEANING

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.

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, wekept those that include data in the main features mentioned in Table 3 in order to avoid data lost.

VII. CORRELATION MATRIX

To understand the relationships between the different features in our dataset, we used the correlation matrix. 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.

VIII. EXPRIMENTAL 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 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%. Thesame learning-based models have been also used with the OASIS dataset. 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%.

AD On Oasis Dataset

Algorithms	Dataset	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%

AD On Adni Dataset

Algorithms	Dataset	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%

IX. CONCLUSION

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

rnal

Algorithms	Dataset	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%

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.

DECLARATION

Funding: Not Applicable
Consent for publication: Not applicable
Availability of data and materials: All ADNI data presented is available through the ADNI website: http://adni.loni.usc.edu/datasamples/access-data/. All OASIS data is available through the OASIS website at :https://www.oasisbrains.org/
HYPERLINK "https://www.oasis-brains.org/" \h

REFERENCE

[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 GurbetaPokvi ´ c, 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.

[5] Barik, S., Mohanty, S., Rout, D., Mohanty, S., Patra, A. K., and Mishra, A. K. (2020). Heart disease prediction using machine Creative Caregivers LLC (2020). What is dementia?

[6] 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.

[7] Janghel, R. and Rathore, Y. (2020). Deep convolution neural network-based system for early diagnosis of alzheimer's disease. IRBM.

[8] 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.

[9] 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.

[10] Kundaram, S. S. and Pathak, K. C. (2021). Deep learning-based alzheimer disease detection. In Springer Science and Business Media LLC, pages 587–597.

[11] Springer.

[12] 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.

[13] Patterson, C. (2018). The state of the art of dementia research: New frontiers. World Alzheimer Report.

[14] 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.

[15] 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.

[16] 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.

[17] World Health Organization (2020a). Alzheimer's disease fact sheet. World Health Organization (2020b). Dementia.