OPEN ACCESS

Manuscript ID:
ASH-2022-10025087

Volume: 10

Issue: 2

Month: October

Year: 2022

P-ISSN: 2321-788X

E-ISSN: 2582-0397

Received: 19.08.2022

Accepted: 28.09.2022

Published: 01.10.2022

Citation:
Suja, GP. "Machine Learning Approach for Early Detection of Alzheimer's Disease." *Shanlax International Journal of Arts, Science and Humanities*, vol. 10, no. 2, 2022, pp. 38–43.

DOI:
https://doi.org/10.34293/sijash.v10i2.5087

Machine Learning Approach for Early Detection of Alzheimer's Disease

G P Suja
Department of Computer Science
Muslim Arts College, Thiruvithancode, Kanyakumari, Tamil Nadu, India

Abstract
Alzheimer's Disease (AD) is thought to be the most generally known cause of dementia, and it is estimated that only one in four people with Alzheimer's are properly identified. While there is no definitive cure, the negative effects may be ignored while the weakening is mild. Treatment is most effective when it begins before major downstream injury occurs, i.e., at the stage of moderate cognitive impairment (MCI) or much earlier. Physical and neurological examinations and neuropsychological and cognitive testing are used to diagnose Alzheimer's disease. There is a need for improved diagnostic tools, which is what this postulation addresses. Kaggle is an online open-access dataset for improving the Alzheimer's disease diagnosis technique. The information acquired during the conference is documented. One goal of this theory's research is to look at machine learning methodologies to create a classifier that can help screen new persons for different stages of Alzheimer's disease. In comparison to previous work methods, our methodology is suitable for breaking down diverse classes in a single setting and needs less distinct training samples and inconsequential prior knowledge. In our tests, we saw a significant improvement in categorizing all diagnostic categories. Initially, the model is trained on 64 evaluated examples from the Kaggle database. We next test our created model on the whole set of entries supplied by the Kaggle dataset to confirm our framework's finding. Our results show 96.37% accuracy in AD detection and categorization.
Keywords: AD, ML, Classification, Dementia, Clinical Dementia Rating, Early Detection, MRI

Introduction

Alzheimer's disease (AD) is the most well-known cause of dementia. It is a degenerative brain disease that affects humans. Dementia, on the other hand, maybe caused by a variety of illnesses and disorders. A reduction in memory depicts it, as does the ability to figure out and use language, critical thinking, and other intellectual abilities, all of which impact an individual's ability to conduct regular activities. This loss in human capabilities occurs because nerve cells (neurons) in the parts of the brain involved with intellectual ability have been damaged and will never perform normally again. In Alzheimer's disease, neuronal damage affects areas of the brain that enable a person to accomplish basic things like walking and swallowing in the long term. Alzheimer's disease is a lethal disease with no cure known yet. Dementia is a catch-all phrase for a wide range of adverse effects. Several types of dementia, including Alzheimer's disease, vascular dementia, dementia with Lewy bodies, and others. However, dementia of Alzheimer's (AD) is the most well-known cause of dementia.

Most forms of mental illnesses have been resolved based on clinical perception. These include the differentiating evidence of manifestations, which will typically classify the rate and bias of the symptoms to resolve, relapse, or become recurring. There is no cure for Alzheimer's disease, and we lack strong early demonstrative tools. AD is clinically detected by doing physical and neurological exams and examining several indicators of academic incapacity using conventional neuropsychological and psychological testing. Regardless of the aforesaid clinical parameters, the overall process is based on analysis by disposal, for example, administering everything else out until AD is the final option.

Figure 1 Alzheimer's Symptoms

Measures from magnetic resonance imaging (MRI), positron emission tomography (PET), cerebrospinal fluid (CSF) protein profiles, and a study of familial risk profiles are included. However, they are expensive and difficult to scale to massive quantities. Unmistakably, there is a need to develop improved diagnostic tools for Alzheimer's disease diagnosis, maybe using data mining and data analysis processes, which we examine in this paper. If new drugs or aversion approaches are developed and shown to be effective, an early analysis may enable mediation at an earlier stage, which would be of demonstrable benefit. However, it is arduous when the clinical decision is reached based only on the disease's symptoms and side effects. No one test can determine whether or not a person has Alzheimer's disease. While physicians can typically determine whether or not a person has dementia, determining the exact cause might be difficult.

The paper is organized according to Section II literature review and critical evaluation. Section III describes the proposed machine learning techniques. Section IV describes the results and discussion. Finally, conclusions are drawnin Section V.

Literature Survey

Much study has recently been done to precisely identify psychiatric disorders, such as Alzheimer's, and several methodologies have been presented for this purpose. In general, data extracted from structural and functional brain imaging data or cerebrospinal fluid is employed for a more accurate evaluation. Furthermore, the group has made many efforts, and several AD stages have lately been projected. The following are the absolute most concentrated works that have recently been done near there:

Farouk et al. [2]: The authors demonstrated an image analysis technique for AD forecasting. The system uses support vector machine classifier techniques to categorize Alzheimer's patients based on texture characteristics extracted from gray-level co-occurrence matrices and voxel-based morphometry neuroimaging data. The dataset utilized by the authors in this study is ADNI [3].

Khajehnejad et al. [4] provide a technique for evaluating the OASIS [5] dataset. The approach is based on semi-supervised learning, which takes just a small portion of the dataset as training data to accurately predict the labels for the remainder of the test material.

Hosseini-Asl et al. [6] offer an approach based on a 3D convolutional auto-encoder. This model employs a deep 3D convolutional neural network to extract and benefit from AD-related variables. Finally, the classification assignment is completed for different binary combinations of three topic groups (AD, MCI, and NC) and a ternary arrangement among them. The authors utilize the ADNI dataset in this case.

Moradi et al. [7]: A semi-supervised learning strategy is used in this approach to transform another biomarker of MCI to AD. The aging effects are removed from the MRI images using regularised logistic regression during feature selection. Finally, the developed biomarker is bonded with age and cognitive data concerning MCI participants using a supervised learning approach for the final classification done using a random forest classifier. The ADNI dataset was used to acquire data.

Suk et al. [8]: In this research, the authors suggest a deep learning technique for major level inactive and shared feature portrayal using neuro imaging modalities. They used a Deep Boltzmann Machine (DBM), a deep network with a limited Boltzmann machine as a structure obstruct, to find a latent hierarchical feature portrayal from a 3D fix, and then devised a precise technique for a joint feature portrayal from the combined patches of MRI and PET with a multimodal DBM. To validate the feasibility of the suggested method, they conducted experiments on the ADNI dataset compared to cutting-edge techniques.

Savio et al. [9]: The authors used the OASIS dataset and conducted deformation-based features derived from non-linear registration deformation vectors. The connection between the scalar values processed from the deformation maps and the control variable determines feature selection.

Problem Formation

These prior research are only a few instances of how machine learning experiments should be carried out. Other excellent and similarly outstanding studies with good findings exist. This research demonstrated how outcomes should be evaluated and documented, particularly in the prognosis and prediction of Alzheimer's disease. Identifying possible flaws in the input data, experimental design, validation, or implementation, on the other hand, is crucial, particularly for those evaluating various research and those wanting to employ machine learning.

Data quality and essential attribute selection are also critical for successful machine learning outcomes creation. Unfortunately, the authors' procedures for ensuring data integrity and quality were seldom mentioned. The importance of feature selection is equal to that of data quality. However, certain clinical data elements, such as histological evaluations, may become obsolete with time. As a result, a classifier must keep feature sets up to date in terms of time.

Similarly, the training and testing data should be specified. Most algorithms prioritize categorizing large classes over misclassifying or disregarding minorities; this class imbalance leads to selecting the dominant class with poor class prediction, lowering classification quality.

Proposed Model

The first step in effectively classifying AD data is pretreatment. The pathologically demonstrated data set is treated to prevent class imbalance before being transformed to a readable data format. Machine learning algorithms perform effectively when the number of instances of one class is almost equal to that of other classes. To prevent class imbalance, data is oversampled using machine learning approaches such as the synthetic minority oversampling methodology, and data is oversampled to avoid class imbalance (SMOTE). The input data type is changed from numeric to nominal/numeric to nominal values for the algorithms that employ said data type to be implemented.

Dataset Description

We will be using the longitudinal MRI data.

- The dataset consists of longitudinal MRI data of 150 subjects aged 60 to 96.
- Each subject was scanned at least once.
- Everyone is right-handed.
- 72 of the subjects were grouped as 'Non demented' throughout the study.
- 64 of the subjects were grouped as 'Demented' at their initial visits and remained so throughout the study.
- 14 subjects were grouped as 'Non-demented' at the time of their initial visit and were subsequently characterized as 'Demented' at a later visit. These fall under the 'Converted' category.

Attribute Selection

Attribute selection combes all potential attribute combinations in the data to determine which subset of attributes works best for prediction and categorization. It is useful for reducing dimensionality and removing unnecessary characteristics. It may lead to improved classification accuracy or lower computing expenses. The third phase is based on categorization with little help and confidence utilizing Attribute mining.

Machine Learning Tools and Scikit-Learn

We looked at programming options such as Python and the popular SciKit-Learn module. Again, this is a free source and often ranks high in polls, such as those on the KDnuggets site, detecting the use of machine learning languages. R, a statistical programming language, was investigated as an alternative programming language. However, Python was our preferred choice, owing to the SciKit-Learn library enhancements and documentation available with it. Scikit-learn has been one of the most popular open-source machine learning packages since its release in 2007. Scikit-Learn (also known as sklearn) offers machine learning algorithms for classification, regression, dimensionality reduction, and clustering applications. It also includes tools for feature extraction, data management, and model evaluation. It provides pre-written code to a huge variety of algorithms. Scikit-Learn documentation is vast, well-known, and well-maintained. Sklearn is built on Python libraries such as NumPy, SciPy, and Matplotlib. It has a working improvement network with frequent library update arrivals.

Feature Extraction

We have followed the following steps for suitable extraction:

- At first, we have downloaded the data from the KAGGLE website. Then we have observed the dataset and identified the unnecessary entries. After that, we have cleaned those unnecessary entries.
- Thus, the standardization of the dataset is performed by obtaining a standard scale.
- Finally, the dependent and non-dependent variables of the dataset are figured out.

Mini-Mental State Examination (MMSE)

The Mini-Mental State Examination (MMSE) or Folstein test is a 30-item questionnaire to assess cognitive impairment in clinical and research contexts. It is often used to test for dementia in medicine and allied health. It is also used to determine the degree and development of cognitive impairment and track the trajectory of cognitive changes in a person over time, making it a valuable tool for documenting an individual's response to therapy. The MMSE's objective has not been to offer a diagnostic for any specific nosological entity on its own.

Clinical Dementia Rating (CDR)

The CDR is a 5-point scale used to evaluate six cognitive and functional performance categories relevant to Alzheimer's disease and associated dementias: Memory, Orientation, Judgment & Problem Solving, Community Affairs, Home & Hobbies, and Personal Care. A semi-structured interview with the patient and a trusted informant or collateral source is used to collect the information needed to make each evaluation (e.g., family member).

Estimated Total Intracranial Volume (eTIV)

Total intracranial volume (TIV/ICV) is a critical covariate in volumetric assessments of the brain and brain areas, particularly in studying neuro degenerative illnesses, where it may serve as a proxy for maximal pre-morbid brain volume.

Classification

The use of 10-fold cross-validation accomplishes classification. That is, data is separated into ten sections. One component serves as a test, while the remaining nine serve as training data, and the procedure is repeated ten times to confirm the findings. To determine the precise parameters, the training set is utilized for classification. The association rules produce distinct connections among the qualities utilized in the next phase. In the last stage, a particular threshold is applied to the resulting rules to categorize the instances into two groups: Control and AD.

Results and Discussion

The proposed method has been implemented by using a python programming language. The datasets are collected from the kaggle.com website.

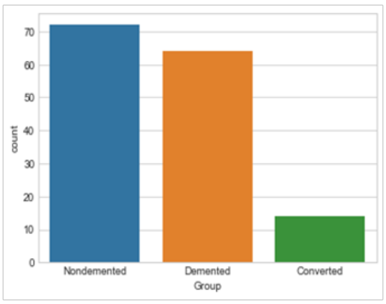

Figure 2 Counting Demented and Non-Demented Data's

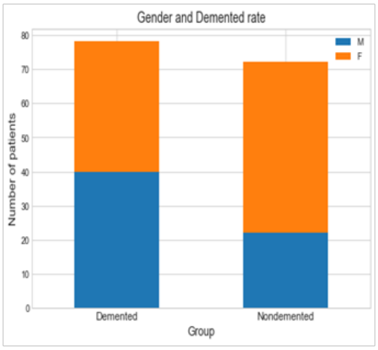

Figure 3 Number of Patients in Gender Wise

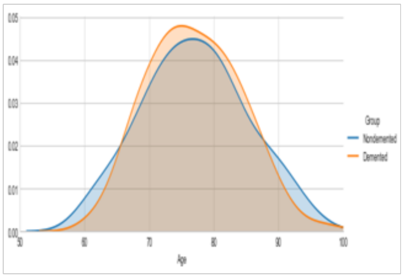

Figure 4 By Age Group-Wise Classification

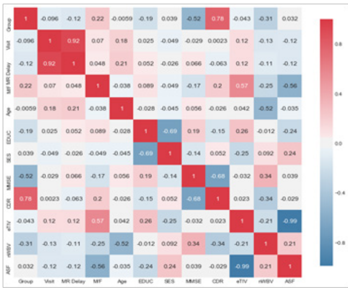

Figure 5 Confusion Matrix

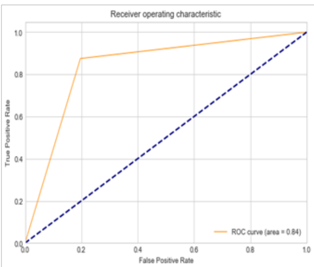

Figure 6 ROC Curve value

Figures 2 to 6 represent the proposed data classification and confusion matrix with Receiver operating characteristics.

Conclusion

Early detection of Alzheimer's disease may prevent or slow the course of the disease. Existing research on the early diagnosis of Alzheimer's disease has proven beneficial in recognizing the disease. However, no one has attempted to classify Alzheimer's disease into different degrees. We created an AD diagnostic system using a mathematical method linked to machine learning. Even though our algorithm provides superior information for diagnosing Alzheimer's at a young age, it is limited to the Kaggle dataset. A larger dataset might help us provide increasingly rigid results. Our point-by-point testing on the Kaggle dataset yields a better recognition rate of 96.37%. Because our framework employs SVM (Support Vector Machine), we are convinced that if more training data (than the 64 individuals) is used to create the machine learning model, the classification accuracy will be much greater.

Future Work

In the current study, we used the clinical data, including neuropsychological assessments, demographic information, physical and neurological examination, cognitive assessments, patient medical history, and baseline diagnosis and symptoms.

- In the future, we will try to develop an embedded system with real-time data feeds. The device will automatically collect necessary information from the subjects through wearable sensors. Then it will use the data against the machine learning model for disease diagnosis.
- Also, we can use more advanced brain images in conjunction with the clinical data to improve the diagnosis and prediction processes.

References

Dubois, Bruno, et al. "Advancing Research Diagnostic Criteria for Alzheimer's Disease: The IWG-2 Criteria." *The Lancet Neurology*, vol. 13, no. 6, 2014, pp. 614-29.

Farouk, Yasmeen, et al. "Statistical Features and Voxel-Based Morphometry for Alzheimer's Disease Classification." *International Conference on Information and Communication Systems,* 2018, pp. 133-38.

Hollingshead, A. *Four factor Index of Social Status*. Yale University, 1975.

Hosseini-Asl, Ehsan, et al. "Alzheimer's Disease Diagnostics by Adaptation of 3D Convolutional Network." *IEEE International Conference on Image Processing,* 2016.

Khajehnejad, M., et al. "Alzheimer's Disease Early Diagnosis using Manifold-Based Semi-Supervised Learning." *Brain Sciences*, vol. 7, 2017.

Moradi, Elaheh, et al. "Machine Learning Framework for Early MRI-based Alzheimer's Conversion Prediction in MCI Subjects." *Neuroimage*, vol. 104, 2015, pp. 398-412.

Savio, Alexandre, et al. "Deformation based Features for Alzheimer's Disease Detection with Linear SVM." *International Conference on Hybrid Artificial Intelligence Systems,* 2011, pp. 336-343.

Suk, Heung, et al. "Hierarchical Feature Representation and Multimodal Fusion with Deep Learning for AD/MCI Diagnosis." *NeuroImage*, vol. 101, 2014, pp. 569-82.

Westman, Eric, et al. "Combining MRI and CSF Measures for Classification of Alzheimer's Disease and Prediction of Mild Cognitive Impairment Conversion." *Neuroimage*, vol. 62, no. 1, 2012, pp. 229-38.

Westman, Eric, et al. "Multivariate Analysis of MRI data for Alzheimer's Disease, Mild Cognitive Impairment and Healthy Controls." *Neuroimage*, vol. 54, no. 2, 2011.

Author Details

G P Suja, *Department of Computer Science, Muslim Arts College, Thiruvithancode, Kanyakumari, Tamil Nadu, India,* ***Email ID****: sujamaran89@gmail.com*