ISSN: 2469-5866

Battineni et al. Int J Brain Disord Treat 2020, 6:033
DOI: 10.23937/2469-5866/1410033
Volume 6 | Issue 1
Open Access

International Journal of
Brain Disorders and Treatment

ORIGINAL RESEARCH

Late-Life Alzheimer's Disease (AD) Detection Using Pruned Decision Trees

Gopi Battineni*, Nalini Chintalapudi and Francesco Amenta

E-Health and Telemedicine Center, University of Camerino, Italy

***Corresponding author:** *Gopi Battineni, E-health and Telemedicine Center, University of Camerino, Camerino, 62032, Italy, Tel: +39-3331728206*

Abstract

Machine Learning (ML) is a contemporary technique of artificial intelligence. These methods are exponentially rising in the medical field, especially in diagnosis and disease predictions. The present study was aimed to develop a decision tree model to predict late-life Alzheimer's disease (AD). A dataset of 150 subjects along with 373 MRI sessions demographic values were considered in this paper. Pruned decision trees (J48) were employed to do predictive analysis on AD subjects. Model validation was conducted with cross fold (k = 10) methods. Performance measures were evaluated by accuracy, precision, and receiver operating characteristic (ROC) curve. Results were provided an accuracy of 88.7%, precision of 86.7%, and ROC of 91.8% was recorded.

Keywords

Machine learning, AD, Cross-validation, Decision tree, ROC

Introduction

There are some established plans and proposals for a medical practice on some external examinations and hard-coded into their software. However, these programs are restrained the data precision because they are generated from different people and conditions. Dementia is one of the global medical issues that was high in demand. Most of the studies are related to dementia causes explaining the risk reduction, early medication, and immediate disease finding in older adults. Therefore, it is mandatory to conduct some advanced studies dealing with these diseases.

In general, subjects with Mild Cognitive Impairment (MCI) are relevant groups for the cure as they are at the prodromal stages and a higher risk of Alzheimer's disease (AD). AD and different kinds of dementia were becoming a global challenge and tending to the death in one of three elder peoples in the USA. While the reasons for these diseases have not yet completely understood, they can effectively affect discourse, memory, and other essential psychological abilities.

Machine learning (ML) is a category of an algorithm that allows software applications to become more accurate in outcome prediction without being explicitly programmed [1]. The basic premise of these methods is to build algorithms that can receive input data and use statistical analysis to predict an output. Nowadays, it is hard to exclude these techniques because most of them used in real-time purposes, and many researchers are thinking that it is an ideal approach to gain grounds toward human-level artificial intelligence (AI) [2,3]. Furthermore, ML methods are similar to data mining, and prediction algorithms as of both require data exploration to search for examples and to change program activities in the same manner [2,4]. Recently, these techniques are gradually increasing in the medical filed for prediction or visualization of patient data [5], development of medical diagnostics case studies [6,7]. Present concentrate on late-life AD detection with the help of MRI demographic data and AD prediction were evaluated with feature characteristics. Pruned decision trees (J48) model was employed to conduct this analysis, and model performance was assessed by accuracy, precision, and receiver operating characteristic curve (ROC).

Methods

Patients

Citation: Battineni G, Chintalapudi N, Amenta F (2020) Late-Life Alzheimer's Disease (AD) Detection Using Pruned Decision Trees. Int J Brain Disord Treat 6:033. doi.org/10.23937/2469-5866/1410033

Accepted: February 06, 2020: **Published:** February 08, 2020

A dataset of consists of 150 patients (i.e., subjects) of demographic MRI data with age ranging from 60 to 96 were considered. All Subjects are giving in informed consent and extracted from the open Access Series of Imaging Studies (OASIS) [8]. Each subject was exposed at least two scanning sessions and a total of 373 MR session information was available. All subjects associated with right hand irrespective gender. Present AD status was decided by the clinical dementia ratio (CDR) and each session was categorized into 3 groups of 146 AD (demented), 190 AD_{non} (Non-demented), and 37 AD_{con} (converted).

Decision trees

Decision trees are the conventional model of machine learning techniques and produce results with higher accuracy when compared to others. An algorithmic methodology developed these that data splitting was done by distinct conditions [9]. Many studies were considered decision trees as a great approach to conducting a predictive analysis. In AD prediction, we begin from the tree root feature and compare this feature with other tree node features. Based on the correlation, we pursue the branch relating to that value and jump to the next node [10]. It is important to keep different AD groups and other tree internal nodes until we achieve a leaf node with a predicated class.

Results

Based on the AD group, all the features were exposed to the J48 decision trees model. Cross (k-fold) validation (CV) techniques were employed to validate the model. The CV is a resampling technique with a unique parameter 'k,' which used in model evaluation on a limited data sample. Based 'k' value data can split into test and train groups. The cross-validation was conducted with k = 5 to avoid fitting issues, which means of five data folds (or subgroups) for testing and k-5 folds for training purposes had used. For generating pruning decision tree, we considered limited features of CDR, MMSE, n-WBV, gender, and MR delay since these are highly correlated with the group category.

Model performance was evaluated by accuracy, precision, and area under receiver operating characteristic curve (AU-ROC). Data preprocessing was conducted by selection of highly correlated features coupled with AD group. Model training was held between the target AD group and rest of the features, model drive the operation of dementia forecasting along performance measures and confusion matrix (Figure 1).

From Figure 1, it is evident 331 were correctly classified among 373 MRI sessions with an accuracy of 88.7%. Weighted average of true positive prediction (i.e., precision) of 86.7% was recorded. Precision (or sensitivity) was calculated by the ratio of true positives and a total number of positive predictions. For example, the precision of true AD subjects is evaluated as 91.3% (Equation 1).

True AD predictive AD

$$\frac{TrueADsubjects}{True(AD+ADnon+ADcon)subjects}=\frac{188}{188+18}*100=91.3\% \quad (1)$$

Correctly Classified Instances	331	88.7399%
Incorrectly Classified Instances	42	11.2601%
Kappa statistic	0.7992	
Mean absolute error	0.1085	
Root mean squared error	0.2609	
Relative absolute error	28.1314%	
Root relative squared error	59.4574%	
Total Number of Instances	373	

--------- Detailed Accuracy By Class ----------

	Precision	ROC Area	Class
	0.913	0.937	Nondemented
	0.924	0.962	Demented
	0.409	0.650	Concerted
Weighted Avg.	0.867	0.918	

---------- Confusion Matrix ----------

a	b	c	< -- classified as
188	1	1|	a = Nondemented
0	134	12|	b = Demented
18	10	9|	c = Converted

Figure 1: Model outcome.

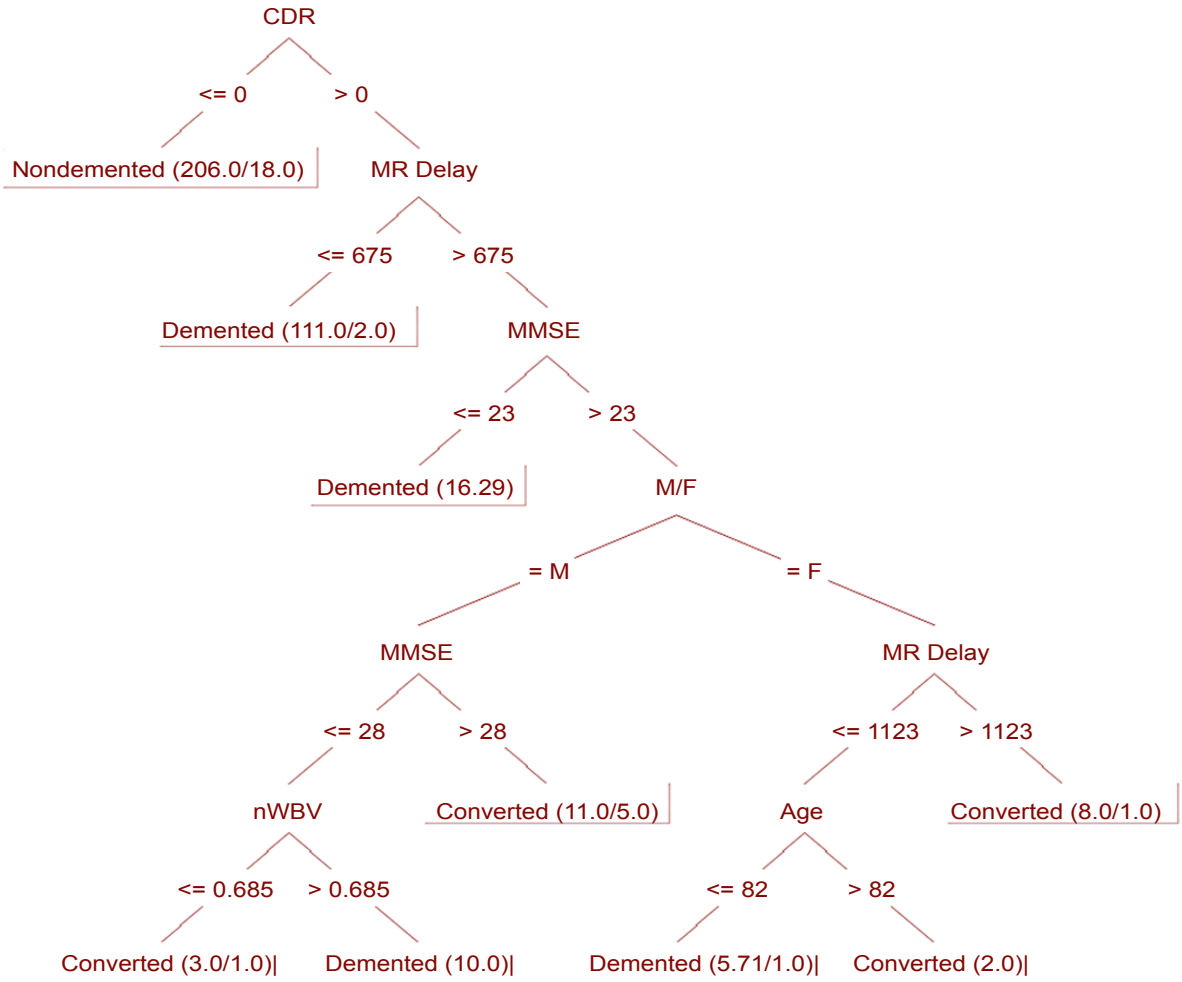

Figure 2: Pruned decision tree outcome (CDR: Clinical Dementia Ratio; MMSE: Mini Mental State Examination; n-WBV: Total Brain Volume; MRI: Magnetic Resonance Imaging).

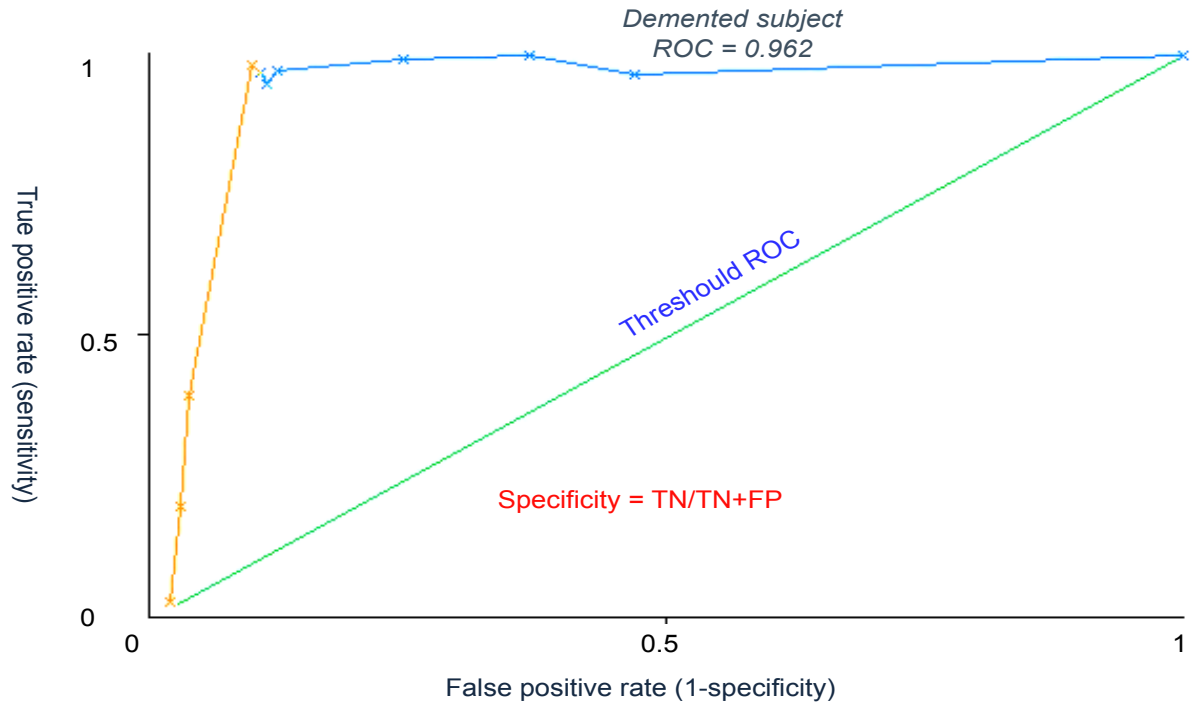

Figure 3: ROC curve of AD subjects.

The J48 pruned decision tree with a central node of CDR can be observed (Figure 2). If the branch CDR ≤ 0, MRI session classification was done as AD_{non} with an accuracy of 92%. The second branch CDR > 0, splitting

into two branches of MR delay as the central node. It generated AD subject accuracy of 98.2%, along with another branch with an MMSE central node. This tree follows the bottom node with a group category. As mentioned, there are some specific cases AD_{con} (i.e., characterized as non-demented at first visit and subsequently described into demented at a later visit and vice versa), which are having a more significant effect on other dementia factors. Generated decision tree predictions have correctly mapped and analyzed with confidence values of dementia status. Eventually, the highest confidence value of dementia will predict the future dementia status of the particular patient, and the mentioned model explains and predict the patient's condition by utilizing specific benefits to help patients by assisting them in advance.

Discussion

In AD diagnosis of most MCI studies, MRI demographic information along with other features highly important in AD forecasting [11]. In this study, we have developed a machine-learning model with a feature reduction (pruning) technique to enhance classification accuracy. Distinct medical diagnostics have developed with the connection of ML implementation. But, few studies were only associated with AD classification. AD is one of the complex data analysis because it requires test information, physical test, cognitive testing, research facility studies, and MR images [11-13]. As of this, we consider specific features such as CDR, MR delay, MMSE, and n-WBV.

At first, the AD group was mapped with the rest of the features, which were highest correlated with present AD status. The CDR value evaluated late-life AD prediction. Despite age, if CDR ≤ 0, then subjects were classified as $AD_{non,}$ and CDR > 0 highest percent of subjects were classified as AD, and rest were as AD_{con}. The outcome tree was generated with different sub-branches and left a decision at the end, considered as a leaf of the corresponding branch. In the end, outcomes suggesting that pruned decision tree models are one of the best approaches with an accuracy of 88.7%.

ROC curve value was evaluated as fundamental analysis in medical diagnosis [14], and it's a plot of true positive rate on y-axis and false positive rate on x-axis (Figure 3). According to [15], in diagnosis classification an excellent model possess ROC near to one that means it has effective measure of separability. If it near to zero said to have worst measure of separability. In this experiment, we got ROC of AD classification is 0.962, which means that comprehensive classification of AD patients was done.

Conclusions

Highest percent of mortality rates were happened due to the lack of early disease diagnosis, and AD is one among them. Especially, old patients were facing the dementia problem; such patients can overcome this issue by some extent through early doctor approach. At the same time, reduction of MR delay could also a comprehensive precaution to overcome probability of AD happening. Therefore, there is more chance to save AD patients in future before they turn into helpless situations.

Conflicts of Interest

No author possesses any conflict of information.

References

1. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI (2015) Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal 13: 8-17.
2. Domingos P (2012) A few useful things to know about machine learning. Communications of the ACM 55: 10.
3. Learning M, Zheng A (2015) Evaluating Machine Learning Models: A Beginner's Guide to Key Concepts and Pitfalls. O'Reilly.
4. Bhatia P (2019) Introduction to Data Mining. Data Mining and Data Warehousing.
5. Darcy AM, Louie AK, Roberts LW (2016) Machine learning and the profession of medicine. JAMA 315: 551-552.
6. Battineni G, Chintalapudi N, Amenta F (2019) Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). Informatics in Medicine Unlocked 16: 100200.
7. Giger ML (2018) Machine Learning in Medical Imaging. Journal of the American College of Radiology 15: 512-520.
8. Smith SS (2009) Predicting Alzheimer's dementia mortality using Medicare Outcome Assessment and Information Set (OASIS).
9. Podgorelec V, Kokol P, Stiglic B, Rozman I (2002) Decision trees: An overview and their use in medicine. Journal of Medical Systems 26: 445-463.
10. Ritchie LJ, Tuokko H (2011) Clinical decision trees for predicting conversion from cognitive impairment no dementia (CIND) to dementia in a longitudinal population-based study. Archives of Clinical Neuropsychology 26: 16-25.
11. Eckerström C, Olsson E, Borga M, Ekholm S, Ribbelin S, et al. (2008) Small baseline volume of left hippocampus is associated with subsequent conversion of MCI into dementia: The Göteborg MCI study. J Neurol Sci 272: 48-59.
12. Facal D, Valladares-Rodriguez S, Lojo-Seoane C, Pereiro AX, Anido-Rifon L, et al. (2019) Machine learning approaches to studying the role of cognitive reserve in conversion from mild cognitive impairment to dementia. International Journal of Geriatric Psychiatry 34: 941-949.
13. Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, et al. (2011) 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 Research Notes 4: 299.
14. Hajian-Tilaki K (2013) Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 4: 627-635.
15. Kumar R, Indrayan A (2011) Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatrics 48: 277-287.