International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 7
July 2017

Feature Selection Techniques for Alzheimer's Disease: A Review

Mrs. S. Hannah Immanuel
SSN College of Engineering
Dr. Shomona Gracia Jacob
SSN College of Engineering

ABSTRACT

Dementia is a broad category of brain disorders. Alzheimer's is the most common form of dementia, a general term for memory loss and other cognitive abilities, serious enough to interfere with daily life. Alzheimer's disease accounts for 60 to 80 percent of dementia cases. Alzheimer's disease is a progressive, degenerative disorder that attacks the brain's nerve cells, or neurons. It is a permanent and progressive brain disease which slowly decrease memory, thinking, remembering, and reasoning skills. In the current era, neuroimaging based Alzheimer's disease (AD) or mild cognitive impairment (MCI) diagnosis has attracted researchers in the field, due to the increasing prevalence of the diseases. We will have to implement dimensionality reduction techniques to avoid the curse of dimensionality, as the number of features appear to be quite high in genetic information when compared to other clinical datasets. In this paper we have surveyed the various feature selection techniques that have been used to predict the onset of Alzheimer's disease in an accurate manner.

KEYWORDS: Alzheimer's disease; dementia; mild cognitive impairment; machine learning; dimensionality reduction; feature selection;

INTRODUCTION

Alzheimer's disease is a memory impairment disease which mostly affects elderly people. Currently, about 27 million patients are affected by this disease and the prevalence is expected to grow fourfold by 2050 [1]. The occurrence of Alzheimer's disease is expected to quadruple by the year 2020. Alzheimer's disease cannot be cured nor its progression stopped. As it eventually leads to death every clinical treatment aims to only delay its progression. There are seven stages of the Alzheimer's disease and usually the initial stages are misinterpreted as a normal aging process or as a result of high stress [3]. We start to notice changes only in a stage of mild decline of cognitive functions [2]. There is no cure for Alzheimer's disease, but medications, sensory therapy and more can help its symptoms. And to get the full benefit of the treatments, early diagnosis is important. Early diagnosis of the disease helps the patients, the caregivers and health institutions to save time, cost and minimize patient's suffering. Alzheimer's disease (AD) is being diagnosed using different techniques such as examination of the medical history, physical examination, laboratory test, neuropsychological or cognitive function testing and brain imaging scan.

The clinical datasets that are used in various studies were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database and Gene Expression data from the NCBI. This data is high dimensional in nature and the number of features are usually greater than the number of samples present in them. Hence we will have to perform dimensionality reduction in order to use the data in an effective manner and not over-fit them. Dimensionality reduction is the process of reducing the number of features under consideration by combining, transforming or selecting the various features and can be used for exploring and understanding the data that helps us to build a simple model [4]. Dimensionality reduction can be broadly classified into

- Feature reduction where all the original features are used so that we obtain transformed features by combining them in a linear fashion.
- Feature Selection where only a subset of the features are used.

FEATURE SELECTION

Feature selection is one of the most important preprocessing steps in data mining and machine learning. Feature selection techniques are often used in domains where there are many features and comparatively few samples of data [5]. A typical case for the application of feature selection is the analysis of DNA microarray data or the clinical data for the Alzheimer's disease, where there are many thousands of features, and a few tens to hundreds of samples. The objective of variable selection is three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data. There are three general classes of feature selection algorithms: filter methods, wrapper methods and embedded methods.

Filter methods are generally used as a preprocessing step. The selection of features is independent of any machine learning algorithm. Instead, the features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable [5]. We rank each feature according to some univariate metric and select the highest ranking features. Filter methods use a proxy measure instead of the error rate to score a feature subset. This measure is chosen to be fast to compute, while still capturing the usefulness of the feature set.

- Pearson's Correlation: It is used as a measure for quantifying linear dependence between two continuous variables X and Y. Its value varies from -1 to +1 [6].
- LDA: Linear discriminant analysis is used to find a linear combination of features that characterizes or separates two or more classes (or levels) of a categorical variable [6].
- ANOVA: ANOVA stands for Analysis of variance. It is similar to LDA except for the fact that it is operated using one or more categorically independent features and one continuous dependent feature. It provides a statistical test of whether the means of several groups are equal or not [6].
- Chi-Square: Chi-Square is a statistical test applied to the groups of categorical features to evaluate the likelihood of correlation or association between them using their frequency distribution [14].
- Information gain: The Information gain of an attribute tells us how much information with respect to the classification target the attribute gives you. That is, it measures the difference in information between the cases where we know the value of the attribute and where we do not know the value [2].
- Gain Ratio: Gain Ratio is a score that is similar to Information Gain and measures the quantity of information content of a set of tuples D to alleviate the overestimation for multi-valued features that biased the Information Gain score. As a matter of fact, the Information Gain score is prone to select attributes that have a large number of values [8]. However such partitioning is clearly biased. In order to fix this bias, Gain Ratio score applies a normalization to the Information Gain score by computing a "split information" value, which represent the amount of information obtained by using the attribute A as split. Gain Ratio score is then computed as ratio between Information Gain and the entropy of the feature's value distribution.
- Kolmogorov-Smirnov (K-S) test: The Kolmogorov- Smirnov test is used to decide if a sample comes from a population with a specific distribution [13]. The Kolmogorov Smirnov (K-S) test is based on the empirical distribution function. In statistics, the Kolmogorov Smirnov test is a nonparametric test of the equality of continuous, one dimensional probability distributions that can be used to compare a sample with a reference probability distribution, or to compare two samples. An attractive feature of this test is that the distribution of the K-S test statistic itself does not depend on the underlying cumulative distribution function being tested. Another advantage is that it is an exact test.

Wrapper: Wrapper methods use a predictive model to score feature subsets. Each new subset is used to train a model, which is tested on a hold-out set [6]. Counting the number of mistakes made on that hold-out set (the error rate of the model) gives the score for that subset. As wrapper methods train a new model for each subset, they are very computationally intensive, but usually provide the best performing feature set for that particular type of model. We search for the best subset of features by assessing the quality of a set of features using a specific classification algorithm by internal cross validation. The common examples of wrapper methods are -

- **Forward Selection:** Forward selection is an iterative method in which we start with having no feature in the model [6]. In each iteration, we keep adding the feature which best improves our model, till an addition of a new variable does not improve the performance of the model.
- **Backward Elimination:** In backward elimination, we start with all the features and remove the least significant feature at each iteration which improves the performance of the model. We repeat this until no improvement is observed on removal of features.
- **Recursive Feature elimination:** It is a greedy optimization algorithm which aims to find the best performing feature subset. It repeatedly creates models and keeps aside the best or the worst performing feature at each iteration. It constructs the next model with the left features until all the features are exhausted. It then ranks the features based on the order of their elimination [6].

Embedded methods combine the qualities of filter and wrapper methods. It is implemented by algorithms that have their own built-in feature selection methods. The most common type of embedded feature selection methods are regularization methods. Regularization methods are also called penalization methods that introduce additional constraints into the optimization of a predictive algorithm that bias the model toward lower complexity. Some of the most popular examples of these methods are LASSO and RIDGE regression which have inbuilt penalization functions to reduce over fitting.

REVIEW OF LITERATURE

The various feature selection techniques that have been used to predict the onset of the Alzheimer's disease have been reviewed and presented in this section.

Mohamed M. Dessouky et al, in [3] have used the CAD tools for the automated diagnosis of the Alzheimer's disease to extract the most significant features.They select the features that have different intensity level at all images and discard the features that have the same intensity level to reach the fewer subset of features that are distinctive and have a high impact. A binary classification on the Open AccessSeries of Imaging Studies (OASIS) database was performed by a Linear Support Vector Machine (SVM) classifier.The proposed feature selection approach here was the rejection of the similar voxels that had a same intensity level in all images. A random 5 fold cross validation was performed on the dataset after which the predicted labels were used as the input for the SVM classifier. The sensitivity (SEN), specificity (SPE), accuracy (ACC) and Matthew's correlation coefficient (MCC) were calculated for the proposed approach along with the traditional Principal Component Analysis (PCA) and Linear Discrimination Method (LDA). The proposed methodology had outperformed the results of the PCA and the LDA.

A.V. Lebedev et al, in their study [7] have investigated theefficacy of Random Forest classifiers trained using differentstructural MRI measures, with and without neuroanatomicalconstraints in the detection and prediction of AD in termsof accuracy and between-cohort robustness. The dataset hasbeen obtained from the Alzheimer's Disease NeuroimagingInitiative (ADNI) database. The inclusion criteria forthe clinical procedures was that the Mini-Mental StateExamination (MMSE) score should be between 24 and 30,while the Clinical Dementia Rating (CDR) should have ascore of 0.5. A random subsample of the data was selectedto form the training and the testing dataset. The recursivefeature elimination (RFE) based on the Gini-criterion with5-fold cross-validation was performed. The data with zero ornear zero variance were removed after which the stepwiseRFE was performed. RFE was performed based on featureimportance vector derived from the first forest, by removinglowest-ranked 5 percent of the features at each step. Afterselection of the optimal feature subset, parameter adjustmentwas performed using 1000 trees. The ensembles classifierdemonstrated a more accurate AD detection than the LinearSVM model.

JussiTohka et al, in their study [8] have a frameworkto identify the high risk of the Mild Cognitive Impairment(MCI) subjects into Alzheimer's disease. They have used asemi supervised learning to construct the MRI biomarkers.A feature selection on the MRI data from the ADNI datasetwas performed by the regularized logistic regression. Themost informative voxels (features) were selected while thenon-informative ones were discarded. The aging effects fromthe MRI data were removed while prediction was performedby the random forest classifier. A 10 fold cross validationwas performed to obtain the area under the receiver operatingcharacteristic curve (AUC) score of 0.9020 in discriminatingthe progressive from the stable MCI patients. The resultsfor Accuracy,Sensitivity and Specificity were 82%, 87% and74%.

Abhinav Grover in his comprehensive study [9] hasidentified the putative Alzheimer-associated genes usingmachine learning approaches. He has identified the candidateAD associated genes by integrating topological propertiesof the genes from the protein-protein interaction networks,sequence features and functional annotations.The data fromthe Entrez Gene, an online database that incorporatesextensive gene-specific information has been used for theinvestigations. Seven feature selection techniques were usedthat include a gain-ratio based attribute evaluation, OneRalgorithm, chi square based selection, correlation based selection, information gain-basedattribute evaluation andrelief-based selection, to select the important attributes. Gainratiobased attribute selection approach measures the gainratio regarding the prediction class while info-gain attributeevaluation uses Info Gain Attribute Evaluator and measuresthe information gain with respect to the prediction class.Chi-squared Attribute Evaluator calculates the chi-squarestatistic with respect to the class. OneR algorithm usesOneR classifier for attribute selection and generates one rulefor each attribute followed by selecting the attribute withsmallest-error to be used for classification. Correlation-basedselection employs CfsSubsetEval and measures the worthof a subset of attributes by evaluating each predictor. Thealgorithm finally selects the subset in which the predictors arehighly correlated with the prediction class while are poorlycorrelated to other predictors. A 10 fold cross validation wasused to train the eleven machine learning classifier modelspredict candidate Alzheimer's genes.

Ali Khazaee and AtaollahEbrahimzadeh in their study[10] of brain network on the basis of resting-state functionalmagnetic resonance imaging (fMRI) have combined graphtheoretical approaches with advanced machine learningmethods to predict the functional brain network alteration inpatients with Alzheimer's disease (AD). The data of twentypatients with AD and twenty age-matched healthy controlsfrom the Alzheimer's Disease Neuroimaging Initiative(ADNI) database whose Mini-Mental State Examination(MMSE) score was between 14 and 26, and the ClinicalDementia Rating (CDR) score of 0.5 or 1 were included forthe study. The automated anatomical labeling (AAL) atlasparcellates the brain into 90 distinct regions that are used inthe construction of the brain network. The computation ofthe graphs was performed using the measures of clusteringcoefficient, local efficiency and normalized local efficiency.As fMRI is a high dimension data, the filter method of featureselection was used to make the computations simple andfast. Seven different filter methods like Fisher score, minimalredundancy maximal relevance (MRMR), t-test, Chi-square,information gain, Gini and Kruskal-Wallis are incorporated. Asoft margin SVM classifier with linear, polynomial, sigmoidand Gaussian radial basis function kernels were used alongwith the leave out one cross validation algorithm to obtain aprediction accuracy of 97.5%.

Javier Escudero [11] has used machinelearning conceptsto classify the MRI Features of Alzheimer's Disease and MildCognitive Impairment in patients so that the sample size ofthe clinical trials can be greatly reduced. The data set hasbeen sourced from ADNI whose primary goal has been totest whether serial MRI, PET, other biological markers, andclinical and neuropsychological assessment can be combinedto measure the progression of MCI and early AD. FreeSurferv.4.3 was used to compute features from the MRI scansobtained. Then forward selection, a feature selection methodthat is independent of the classifier was applied. As the selection of the best classifier for diagnosis is an open problem,four classifiers namely the Logistic Regression (LR), SVM,Radial Basis Function (RBF), and C4.5 tree learner wereused. The average accuracy performance

(Mean + Standarddeviation) of the of the classifiers for the experiment normalcognition versus Alzheimer's disease was calculated as

- Logistic Regression (LR) = 85.63%
- Support Vector Machine (SVM) = 89.17%
- Radial Basis Function (RBF) = 87.94%
- C4.5 tree learner = 83.93%

Massimo Brescia et al, in their study of feature selectionin MRIs for Hippocampal Segmentation [13] have shownhow the MRI scans can act as a supporting feature inpredicting Alzheimer's disease. The hippocampus, a wellknown biomarker has been scanned in a robust mannerto have accurate predictions. The main goal of the studywas to exemplify and demonstrate the benefits of applyingfeature selection algorithms in hippocampus segmentationfield. Hence a profound study was made on four different featureselection methods namely (i) univariate filter method basedon the Kolmogorov Smirnov test, (ii) sequential forwardselection, a deterministic wrapper method, (iii)sequentialbackward elimination, a deterministic wrapper method and(iv) embedded method based on the RandomForestClassifier.Filter method is a technique based on the measurement ofthe importance of each single feature of the given parameterspace. They are independent from the classification algorithmand therefore their results can be used for all types ofclassifier. Wrapper methods basically integrate the twoaspects of the workflow, that is, the model hypothesis andfeature search. This procedure involves the generation andevaluation of various subsets of features. Every generatedfeature subset is associated to a classification criterion. Thedata set that is used is a 10 T1-weighted brain MRIs on anindependent set of 25 subjects from the Open Access Seriesof Imaging Studies (OASIS). A 5-fold cross validation wasperformed on 10 of 35 images in the database.The goodnessof the selected group was tested on the remaining 25 images.The K-S test allowed us to select only the features whichhave a correlation between the two hippocampus and not hippocampus classes less than 5%, resulting in a total of57 features. When using the Sequential Forward Selectionand Backward Elimination the feature achieving the bestperformance is chosen, when used in combination with theselected features in the previous step. The comparison of theproposed methodology of feature selection methods with thewidely used PCA demonstrates the very low performance ofthe PCA technique.

Sang-Woong Lee et al., in their diagnosis of Alzheimer'sDisease Based on Structural MRI Images Using a RegularizedExtreme Learning Machine and PCA Features [14] havecompared AD diagnosis approaches using structural magneticresonance (sMRI) images to discriminate AD, mild cognitiveimpairment (MCI), and healthy control (HC) subjects using a support vector machine (SVM), an import vector machine(IVM), and a regularized extreme learning machine (RELM).The experiments were conducted on the sMRI Datasetfrom ADNI. The focus here is mainly on the sMRI based(structural) feature extraction as the intensity and stage of theneurodegeneration can be identified by the help of atrophymeasured by sMRI. Extreme learning machine (ELM) is alearning algorithm implemented without iteratively tuning theartificial hidden nodes, thus decreasing the computation time,an effective solution for Single hidden-layer feed forwardneural networks. The main impetus of the study was tocompare representative classifiers, SVM, IVM, and RELMfor binary and multi class classification tasks. Trivially, theaccuracy of the binary classification cases was higher thanthe corresponding multi class classification cases. Also, theexperimental results on large dataset of 214 subjects verifiedthat RELM based AD diagnosis framework outperform theothers with higher accuracy.

CONCLUSIONS

The high dimensionality versus the small sample size of the datasets have been a major hurdle in predicting the onset of the Alzheimer's disease. The various feature selection techniques have been exploited by all researchers in order to efficiently extract the most informative features.

Machine learning approaches are a better option for discriminating the patient as an Alzheimer's disease or a Mild Cognitive Impairment or a Healthy person. We have also analyzed the state of art classification techniques that have been used by the various researchers. In this review, we have made a comparison and evaluation of recent work done in the prognosis and prediction of Alzheimer's disease. With this bird's eye view, in future we propose to implement novel feature selection techniques to effectively predict Alzheimer's disease.

ACKNOWLEDGEMENT

This research work is a part of the Science and Engineering Research Board (SERB), Department of Science and Technology (DST) funded project under Young Scientist Scheme – Early Start-up Research Grant- titled "Investigation on the effect of Gene and Protein Mutants in the onset of Neuro-Degenerative Brain Disorders (Alzheimer's and Parkinson's disease): A Computational Study" with Reference No- SERB – YSS/2015/000737..

REFERENCES

[1] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, 2007. Forecasting the global burden of Alzheimers disease, Alzheimers Dement., vol. 3, no. 3, pp. 186191.
[2] A. Khan and M. Usman, 2015. Early diagnosis of Alzheimer's disease using machine learning techniques: A review paper, 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), Lisbon, pp. 380-387.
[3] Mohamed M. Dessouky,Mohamed A. Elrashidy, Taha E. Taha and Hatem M. Abdelkader, 2013. Selecting and Extracting Effective Features for Automated Diagnosis of Alzheimer's Disease, International Journal of Computer Applications (0975 8887).
[4] SaptarsiGoswami and AmlanChakrabarti, 2014. Feature Selection: A Practitioner View, I.J. Information Technology and Computer Science, pp. 66-77.
[5] Machine Learning. URL: www.http://machinelearningmastery.com/anintroduction-to-feature-selection
[6] Feature Selection Techniques. URL: www.http://analyticsvidhya.com
[7] A.V. Lebedev et Al, 2014. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between cohort robustness, Elsevier NeuroImage: Clinical.
[8] ElahehMoradi, Antonietta Pepe, Christian Gaser, HeikkiHuttunen and JussiTohka, 2014. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects Elsevier, (NeuroImage),
[9] Salma Jamal, Sukriti Goyal, AsheeshShanker and Abhinav Grover, 2016. Integrating network, sequence and functional features using machine learning approaches towards identification of novel Alzheimer genes, BMC Genomics.
[10] Ali Khazaee, AtaollahEbrahimzadeh and Abbas Babajani-Feremi, 2014. Automatic classification of Alzheimers disease with resting-state fMRI and graph theory, 21st Iranian Conference on Biomedical Engineering.
[11] Javier Escudero, John P. Zajicek and Emmanuel Ifeachor, 2011. Machine Learning Classification of MRI Features of Alzheimers Disease and Mild Cognitive Impairment Subjects to Reduce the Sample Size in Clinical Trials, 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA.
[12] Edward Challis, Peter Hurley, Laura Serra, Marco Bozzali, Seb Oliver, Mara Cercignani, 2015. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI, Elsevier NeuroImage.
[13] Sabina Tangaro, Nicola Amoroso, Massimo Brescia, StefanoCavuoti, AndreaChincarini, RosangelaErrico, PaoloInglese, GiuseppeLongo, RosaliaMaglietta, Andrea Tateo, Giuseppe Riccio and Roberto Bellotti, 2015.Feature Selection Based on Machine Learning in MRIsfor Hippocampal Segmentation,Hindawi Publishing Corporation,Computational and Mathematical Methods in Medicine.
[14] Ramesh Kumar Lama, JeonghwanGwak,Jeong-Seon Park and Sang-Woong Lee, 2017. Diagnosis of Alzheimer's Disease Based on Structural MRIImages Using a Regularized Extreme Learning Machine and PCA Features,Hindawi, Journal of Healthcare Engineering.
[15] S. R. Bhagya Shree, H. S. Sheshadri and Muralikrishna, 2016. Diagnosis ofAlzheimer's Disease using Rule based Approach, Indian Journal of Scienceand Technology.