Prediction For Incidence of Alzheimer's Disease Using Neural Networks

Abstract

Alzheimer's Disease (AD) is the 6th leading cause of death in US, according to the Alzheimer's Association. Prediction of incidence of AD can lead to early diagnosis and delaying the onset in patients. In this project, a machine learning technique called neural networks is used to predict the CDR (Clinical Dementia Rating) value from the patient's MRI data. The CDR in one aspect is a 5-point scale used to characterize six domains of cognitive and functional performance applicable to AD and Dementia. The CDR score is useful for tracking a patient's level of impairment or dementia. The dataset is used from the OASIS (Open Access Series of Imaging Studies). It included a longitudinal MRI dataset of 416 subjects and cross-sectional MRI dataset of 150 from multiple visits. This paper explores a naive way of how a neural network model was configured for the CDR prediction, its evaluation, and the ways to improve the performance of the model. In the end, the paper compares this approach with other models from the literature to find which model best fits the given data.

Introduction

Alzheimer's Disease is the type of dementia that affects memory, thinking and behavior. Symptoms eventually grow severe enough to interfere with daily tasks. The greatest known risk factor is ageing. AD is a progressive disease, where the dementia symptoms worsens over a number of years. AD has no known cure until today, but the symptoms can be treated, and researchers are continuously studying the cure for the disease.

For the AD prediction, the prior models are mostly tree-based and different models focus on different features. Some very early prediction models explore a set of features from a particular domain such as sociodemographic (age, sex, education), lifestyle (physical activity) and cognitive profile [1]. Depending on these small set of features for the prediction would compromise the results and hence there was a need to explore a larger feature space while considering this problem. Recently, a relationship between a combination of features and the accuracy of prediction model by y Schouten et al. [2]. Based on this, many machine learning techniques have been built that can classify the AD that have provided encouraging results.

Machine Learning enthusiasts have built everything from decision trees, SVM models, random forests, and Gradient Boosting Trees for AD prediction, each with different results and features. Currently, the major problem in this field is not, finding the right model for prediction but the luxury of finding proper data for analysis. Biomedical and health data rarely comes in a uniform shape and size. It has too many varied forms with different standards and formats depending on the region, doctors, patients, healthcare industries and other clinical surveys. This kind of data usually is not readily available and even if it is available, it needs a lot of data pre-processing. When it comes to pre-processing a large amount of data, a whole level of approximations come

into the picture, which is enough to trade-off with the compute intensity and the performance of the model in general.

This paper compares the best derived results based on the available data and the performance of the neural network model to predict the AD as compared to other model implementations.

Dataset and Features

The dataset was used from OASIS (Open Access Series of Imaging Studies). OASIS released 2 datasets OASIS-1 and OASIS-2. The OASIS-1 dataset has the cross-sectional MRI data in young, middle-aged, non-demented and Demented older adults that contains 416 subjects collected around 434 MR sessions. The cross-sectional data has 12 features that constitutes the cross-sectional MR image of the patients aged 18-96. The subjects are all right-handed and include both men and women. 100 of the included subjects over the age of 60 have been clinically diagnosed with very mild to moderate AD. The other dataset, OASIS-2 has longitudinal MRI data in non-demented and demented older adults. It consists of 50 subjects from 373 MR sessions. For each subject, 3 or 4 individual T1-weighted MRI scans obtained in single scan sessions are included. The subjects are all right-handed and include both men and women. 72 of the subjects were characterized as nondemented throughout the study. 64 of the included subjects were characterized as demented at the time of their initial visits and remained so for subsequent scans, including 51 individuals with mild to moderate Alzheimer's disease. Another 14 subjects were characterized as nondemented at the time of their initial visit and were subsequently characterized as demented at a later visit. The longitudinal dataset had 15 features.

The cross-sectional features include identification (ID), Gender(M/F), Hand (Dominant hand), Age (in years), Educ (Education level of patient), SES (Socio Economic Status), MMSE (Mini Mental State Examination), eTIV (Estimated Total Inter-Cranial Volume), nWBV (Normalize Whole Brain Volume), CDR (Clinical Dementia Rating), ASF (Atlas Scaling Factor), MR Delay (MR Delay time). The longitudinal features include all the features from cross-sectional and some additional features, that are, MRI ID (MRI Exam Identification), Group (Class: Demented or non-demented), Visit (No. of visits).

The eTIV is the Estimated total Inter- Cranial Volume in mm^3. The nWBV is the Normalized Whole Brain Volume, expressed as a percent of all voxels in the atlas-masked image that are labeled as gray or white matter by the automated tissue segmentation process. The ASF is Atlas Scaling Factor (unitless), computed scaling factor that transforms native-space brain and skull to the atlas target (i.e., the determinant of the transform matrix). The Mini–Mental State Examination (MMSE) or Folstein test is a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment. It is commonly used in medicine and allied health to screen for dementia. It is also used to estimate the severity and progression of cognitive impairment and to follow the course of cognitive changes in an individual over time; thus, making it an effective way to document an individual's response to treatment. The MMSE's purpose has been not, on its own, to provide a diagnosis for any particular oncological entity. Any score greater than or equal to 24 points (out of 30) indicates a normal cognition. Below this, scores can indicate severe (≤9 points), moderate (10–18 points) or mild (19–23 points) cognitive impairment. That is,

a maximal score of 30 points can never rule out dementia. Low to very low scores correlate closely with the presence of dementia, although other mental disorders can also lead to abnormal findings on MMSE testing. The presence of purely physical problems can also interfere with interpretation if not properly noted; for example, a patient may be physically unable to hear or read instructions properly or may have a motor deficit that affects writing and drawing skills [3]. The CDR™ in one aspect is a 5-point scale used to characterize six domains of cognitive and functional performance applicable to Alzheimer disease and related dementias: Memory, Orientation, Judgment & Problem Solving, Community Affairs, Home & Hobbies, and Personal Care. The necessary information to make each rating is obtained through a semi-structured interview of the patient and a reliable informant or collateral source (e.g., family member) referred to as the CDR™ Assessment Protocol. The CDR™ Scoring Table provides descriptive anchors that guide the clinician in making appropriate ratings based on interview data and clinical judgment. In addition to ratings for each domain, an overall CDR™ score may be calculated through the use of an CDR™ Scoring Algorithm. This score is useful for characterizing and tracking a patient's level of impairment/dementia is shown below in Table 1.1.

CDR Value	Clinical Judgement
0	Normal
0.5	Very mild Dementia
1	Mild Dementia
2	Moderate Dementia
3	Severe Dementia

Table 1.1 CDR scores and their clinical meanings

The eTIV is the Estimated total intracranial volume. The ICV measure, sometimes referred to as total intracranial volume (TIV), refers to the estimated volume of the cranial cavity as outlined by the supratentorial dura matter or cerebral contour when dura is not clearly detectable. ICV is often used in studies involved with analysis of the cerebral structure under different imaging modalities. ICV, along with age and gender are reported as covariates to adjust for regression analyses in investigating progressive neurodegenerative brain disorders, such as Alzheimer's disease, aging and cognitive impairment. ICV has also been utilized as an independent voxel based morphometric feature to evaluate age-related changes in the structure of premorbid brain, determine characterizing atrophy patterns in subjects with mild cognitive impairment (MCI) and Alzheimer's

disease (AD), delineate structural abnormalities in the white matter (WM) in schizophrenia, epilepsy, and gauge cognitive efficacy [4].

This paper focuses on dataset that combines both the cross-sectional and longitudinal considering the common columns in both the datasets.

Methods

Data Pre-Processing: The first step in data pre-processing was removing the columns that did not contribute to the prediction process such as Subject ID, Handedness (since everyone was right-handed in the provided dataset), Delay and the Visits to the doctor. So, the total no. of features that are present in the final dataset is 8 which includes M/F, Age, Educ, SES, MMSE, eTIV, nWBV and ASF. The feature Group that had information about demented, non-demented and converted categories of the patient was an important feature in the prediction process but it was not present in both the datasets. It was just present in longitudinal dataset.

The next step in pre-processing was imputing the missing values. After combining both the datasets, the total number of data samples were 809 and total no. of features were 8, until now. The major problem in the combined dataset was that there were a lot of missing vales in certain columns. For example, in the column Educ, there were 203 missing values out of 809 and in the column SES, there were about 239 missing values out of 809. Around 25% of the data in both the columns was missing and dropping these columns would lead to a huge loss proportion, thus reducing the performance of the model. So, instead of dropping the columns, they were replaced by the most frequently occurring values (mode) of the columns, thus approximating, or generalizing some part of the total dataset. There was one more column, MMSE, that had around 99 missing values out of 809 samples. Now, since this value was less as compared to the other missing value columns (SES and Educ), the missing values here were replaced with median of the column.

The next step in data pre-processing was Label Encoding. The final dataset just had one column in the string format. Since, the project was implemented in GNU Octave, the string to integer transformation was necessary. The only column that had to be transformed into string to integer was Gender, M/F. All the column entries with value ‘M’ were converted as 1 and all the ‘F’ entries were converted into 0.

Since most of the columns had wide scale of distribution, feature scaling seemed necessary on the dataset. The common method of Mean Normalization was used for feature scaling. The formula used for feature scaling is provided below in fig.1.1. ‘z’ represents the new value after normalization, ‘x_i’ is the original value of the data sample and μ is the mean value of the column and σ is the standard deviation of the column.

$$z = \frac{x_i - \mu}{\sigma}$$

Fig. 1.1 Mean Normalization for Feature Scaling

The neural network architecture developed initially was very simple and consisted of 8 input nodes (1 for each feature), 1 hidden layer with 14 nodes and 4 out nodes (1 for each CDR value). The CDR prediction had values 0,0.5,1 and 2, but since the implementation was done in Octave, and Octave is not compatible with decimals and zeros, the values 0 and 0.5 were replaced by 3 and 4. So, now the prediction values in Octave were 1,2,3 and 4. The sigmoid function was used as the activation function of the nodes. The model was general, using both forward propagation and backward propagation, that was used to calculate the cost function. Later, this cost function was minimized by using an advanced optimization algorithm ***fmincg()***. After implementation of the model following the above architecture, the results received were quite disappointing. The accuracies that I received even with experimenting with best of the parameters (*lambda* being 0.5 and *Max_Iters* = 500) were very low, being, training accuracy = 63.23%, the cross-validation accuracy = 55.12% and the Test accuracy = 52%.

After evaluating the model, a lot of drawbacks were found in both, the dataset, and the model. Analyzing them needed diving more into the literature survey and more experimentations. Taking a quick look at the results, the model was suffering from what is usually known as 'High Bias Problem'. This problem occurs when the model does not generalize well with the training data and hence fails learning or generalizing the cross-validation and test set. It can cause the model to miss the relevant relations between features and the target output (underfitting). It usually occurs when the model is too simple for the dataset provided and needs more complex implementation. Analyzing the dataset and the neural network model implemented, there existed a lot of drawbacks that needs to be focused on. First, the dataset had a lot a smaller number of data samples and features, which is disadvantageous when using a neural network model. Neural network is a complex model and needs considerably larger amount of data to properly fit the data set. Second, the dataset consisted of a lot of missing values, which in turn led to more approximations, there by underfitting the original dataset. Third, the network architecture was too simple for the given dataset, it just had input layer size as 8, one hidden layer of size 14 and one output layer of size 4. Also, the model used sigmoid function, which is a simple activation function. A more complex activation function like tanh or ReLu.

The model was further improvised by tuning the parameters and increasing the complexity for better results. The first step towards solving the problem of High Bias was to reduce the regularization factor *lambda.* The new *lambda* value was set to 0.2. The next step was introducing polynomial features in the dataset to increase the learning complexity of the model. When a dataset has multiple features, it is common to have different weights for each feature when predicting a particular value. These weights can be calculated by the correlation plots or values between the features and the value to be predicted. This shows how each column affects the output value and how the output value can be predicted more accurately based on assignment of different weights to different features. The other method that can be used to get accurate predictions is the naïve way of experimenting with different weights assigned to features. Introducing polynomial degrees to the features can increase their weights in the final prediction. I experimented with polynomial degrees of 2 and 3 with the features that gave high correlation values with CDR column as compared to others. Following in Table1.2, are the correlation values I received with all the initial

features with respect to the CDR values. The closer the correlation value is to 1 or -1, greater is the correlation between the 2 variables. According to the table, values M/F, MMSE, SSE and Educ are closer to 1 or -1, hence signifying greater correlation between the variables. So, these values can be manipulated to create polynomial degree features in the dataset. The polynomial features that can be introduced can be anywhere from features raised to degree of 2 to features raised to degree of 10, depending upon how complex model is needed. In the project, only the options of polynomial degrees of 2 and 3 were explored, but more polynomial degree features can be used for prediction.

Feature (column in dataset)	Correlation value w.r.to CDR column
Gender (M/F)	6.5100e-02
Age	5.6289e-02
Educ	4.3836e-02
SES	3.2928e-02
MMSE	1.4174e-01
eTIV	-8.4124e-03
nWBV	2.6559e-02
ASF	1.7078e-02

Table 1.2 Correlation values of all features w.r.to CDR

The final columns that were taken into consideration were the initial 8 columns with more 7 columns namely, Age_squared, MMSE_squared, SSE_squared, eTIV * ASF, Age * SES, Age * MMSE, Age * EDUC. So, in total, now the features were expanded from 8 to 15 in number. Automatically, the input layer size increased from 8 to 15. The network architecture was improvised from one hidden layer of size 14 to 2 hidden layers of sizes 14 and 10 respectively. The output layer size was same due to the 4 output values of CDR.

Results

The results after improvising the model were as follows, with parameters, *lambda* = 0.2, *Max_Iters* = 500, degree of polynomial = 2, the training accuracy was 97.18%, the cross-validation accuracy was 83.11% and the test accuracy was 80% which had significantly increased from the previous results. All the previous research conducted for prediction of AD considered just one of the datasets, either OASIS longitudinal or OASIS cross-sectional. So, comparison with any of the previously built models would not be valid. Based on the previously built models, the prediction for AD using tree-based approach [5], compares various tree-based models like decision trees, random forests, and Gradient Boosting Machine (GBM). The training and the cross-validation accuracies received by the author were 78% and 71%, and the AUC being 0.829 and 0.832 respectively for decision trees. The AUC for same dataset using random forest model was 0.88 and using GBM was 0.91 and the accuracy being about 70%. Since only a part of dataset was used for these tree-based models, the accuracies received by the author were comparatively better or mostly equivalent to the ones received in this neural network approach.

Conclusion and Future Work

In conclusion, we can say that neural network works pretty good with the given dataset but needs a lot of pre-processing, tuning of parameters, experimenting around the feature selection and may be compute intensive if over tuned. Taking a closer look at the dataset, a large amount of data is missing, and that may affect the performance of the model. On the other hand, if same dataset was evaluated using a random forest or a GBM, probably, the results would have been better, since most of the missing data would be irrelevant in smaller datasets bagged during the process.

Some of the future work that needs focus on would be, a more efficient feature selection process, collecting and working with more samples (OASIS recently released the 3rd set of MRI data named OASIS-3), synthesis of data from the MR images directly, instead of working on the processed data from the MR images, using a complex activation and implementing the algorithm for the dataset using random forest and GBM and therefore comparing the results from all these different models in order to get the best working model.

References

[1] Predicting Alzheimer's disease progression using deep recurrent neural networks

MinhNguyen, TongHe, LijunAn, Daniel C.Alexander, JiashiFeng, B.T, ThomasYeo

[2] T. M. Schouten et al., 'Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease', NeuroImage Clin., 2016.

[3] https://www.kaggle.com/ruslankl/dementia-prediction-w-tree-based-models

[4] A practical guideline for intracranial volume estimation in patients with Alzheimer's disease Saman Sargolzaei, Arman Sargolzaei,[1] Mercedes Cabrerizo,[1] Gang Chen,[2] Mohammed Goryawala,[3] Shirin Noei,[4] Qi Zhou,[1] Ranjan Duara,[5] Warren Barker,[5] and Malek Adjouadi[1]

[5] https://www.kaggle.com/ruslankl/dementia-prediction-w-tree-based-models

[6] Prediction of Alzheimer's disease (AD) Using Machine Learning Techniques with Boruta Algorithm as Feature Selection Method

[7] https://www.kaggle.com/vishal2develop/dementia-prediction-python

[8] https://www.alz.org/alzheimers-dementia/what-is-alzheimers

[9] https://www.oasis-brains.org/

[10] Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data Ji Hwan Park, Han Eol Cho, Jong Hun Kim, Melanie M. Wall, Yaakov Stern, Hyunsun Lim, Shinjae Yoo, Hyoung Seop Kim & Jiook Cha