AlzheTect: Bio-Marker Analysis for Early Alzheimer's Prediction

Raidel Hernandez
Computer Science
Florida State University
Tallahassee, Florida
rh13k@my.fsu.edu

***Abstract* -- One of the most obscure yet significantly debilitating diseases currently facing humanity is alzheimer's disease (AD). In 2010, there were 4.7 million individuals with alzheimer, it is predicted that by 2050 the figure will be 13.8 million [8]. To be able to effectively treat AD, it is important that we are able to identify it early in individuals. This study will focus on identifying biomarkers (features) that are strong indicators of progressive AD disease through supervised machine learning models. Then, the selected features will be used along with deep learning models to properly classify mild cognitively impaired (MCI) patients. In this research we aim to predict if a patient, believed to have MCI, is still cognitively normal (CN) or may already be showing potential signs of AD. As well as a confidence value that the individual will start showing symptoms of AD within a relatively short period of time (1-5 years).**

Keywords -- mild cognitive impairment (MCI), Alzheimer's Disease (AD), SVM, Deep Learning (DNN), MRI, ADNI, classification.

Ongoing Project: https://github.com/raidel123/AlzheTect

I. INTRODUCTION

Alzheimer's disease is the leading cause of dementia. Dementia is a term used to describe the set of conditions affecting an individual's memory, thinking, and social abilities. Statistics show dementia affect 7% of individuals over the age of 65, and 20% of individuals above 80%. Alzheimer's is the most common type of dementia seen and contributes 60% to 70% of all cases [1]. Alzheimer's is a progressive neurodegenerative disease which worsens over time. However, alzheimer is not only a disease that affects the older population, it also affects many individuals under the age of 65, approximately 200,000 cases in the U.S alone, these individuals are said to have younger-onset (or early-onset) alzheimer's. It is also a disease for which there is no cure or proper treatment. However, early prediction of alzheimer's allow patients to reap the benefits of resources available to them, and reduce symptoms that affect their everyday life.

Due to this increasingly prevalent disease, being able to identify different biomarkers for early onset diagnosis is a crucial area of research. Many previous works have already shown the effectiveness of using neuroimaging techniques, especially those associated with outlining the structural features of the brain to identify differences between patients with alzheimer's and cognitively normal patients. Other strategies have recently been proposed to help accurately measure certain interest regions of the brain, which typically atrophy as the condition progresses, such as the hippocampus. Providing techniques to measure certain atrophied regions has enabled researchers with the ability to properly classify individuals by severity of dementia, or cognitive normality.

In this research, we aim to study individuals with mild cognitive impairment (MCI), specifically the early (EMCI) to late (LMCI) stages of the disease. The motivation for the research came from a challenge posted on *Kaggle* [2], by a UCL group focusing on Progression Of Neurodegenerative Disease (POND). The challenge is called The *A*lzheimer's *D*isease *P*rediction *O*f *L*ongitudinal *E*volution (TADPOLE)**, which aims at the goal of predicting which subject currently believed to have mild cognitive impairment will likely turn into an alzheimer's patient in a relatively short period of time (1-5 years). The data provided for the challenge contains labels and detailed information for 3 main categories of patients: cognitively normal (CN), several stages of mild cognitive impairment (MCI), and alzheimer's disease (AD).

From the set of data points for each patient, per disease category, a feature selection models such as Support Vector Machine (SVM) or cross validation techniques, will be used to extract a subset of important data points for each category, CN and AD. Once the subset of features is selected, a model will then be trained to classify patients into classes. Particularly, we will be training a deep neural network (DNN), on the

subset of features extracted, on patients that are cognitively normal or have alzheimer's disease. Once a model has been trained we will test the models classification accuracy with unlabeled mild cognitive impaired patient information.

Training the DNN model on CN and AD patients was performed due to the fact that MCI patients may belong to other classifications of dementia that may not necessarily be alzheimer. Currently, diagnosing alzheimer's is typically done by ruling out other diseases or types of dementia first, there is no definite method to diagnose the disease until after death when the brain could be further analyzed. Due to the nature of how alzheimer's disease is diagnosed, classifying MCI patients as AD subject is not trivial. Therefore, by training a model with CN and AD patients, the deep neural network will be able to have a better classification accuracy on the conversion of patients from MCI to AD.

The remainder of the paper will focus on the following. In the next section, we provide a breakdown of the data provided for the TADPOLE challenge, and the motivation for the research. In section 3, we will lay out a technical plan describing the methodology used to obtain our results, and the implementation of the DNN, including data transformations performed. Section 4 will cover the results obtained with our trained neural network, including training/testing performance and some probability distributions. We provide a look at the motivating study behind this research and other related works, further advancing brain imaging techniques as well as potential advantages to the ongoing research efforts and conclude our work, in section 5.

II. BACKGROUND

Mild cognitive impairment (MCI) is the transition stage from cognitive decline due to age, and alzheimer's disease (AD). MCI is also the earliest detectable stage of progression todards AD [3]. Annually, around 10% to 15% of MCI patients, referred from a memory clinic, will eventually become AD patients [3]. In this research, data will be obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI)* database. The database contains many sets of data values regarding each patient analyzed, including information such as: patients age, demographics, MRI scans, PET scans, neuropsychological test results, cerebrospinal fluid, cognitive information, and many other data values.

A. Data

The ADNI study provides a data set (CSV) files for the TADPOLE challenge. The data set contains a list of individuals who have been clinically studied, and have agreed to follow-up for the next round of clinical study. Therefore, the data sets contain ADNI rollover individuals, those who have provided data to ADNI in at least two separate visits, for whom they have obtained longitudinal data from in the ADNI databases. Three sets of data are provided, which are obtained from 3 different studies and thus different databases (ADNI-1, ADNI-GO, ADNI-2). Data-set 1 (D1) is a collection of measurements and the outcomes for each patient, whether they are CN, or had MCI or AD. Data-set 2 (D2) is a collection of only baseline measurements, without outcomes for patients, for which we will forecast an outcome. Lastly, data-set 3 (D3) is the outcome for each patient against which we will evaluate the forecasted results. Therefore, D1 will be used as the training data set, D2 will be the prediction data set, and D3 will be the test data set we will use to compare results and obtain the classification accuracy.

B. Data Breakdown

ADNI provides a massive list of relevant biomarkers, for each individual, believed to have some potential influence on AD classification. Two main categories of biomarkers include: measures of amyloid beta protein, measures of damage to nerve cells. Amyloid beta proteins group together in clumps, forming plaque and blocking cell-to-cell signaling at synapses when they mis-fold. Beta-amyloid and tau levels can be measured using cerebrospinal fluid (CSF) or positron emission tomography (PET) scan, both of which we have values for in the data-sets, for each patient. Damaged nerve cells can be measured by quantifying brain metabolism using fluorodeoxyglucose (FDG) PET scan, or brain atrophy using magnetic resonance imaging (MRI), both of which we have values for in the data-sets. All the biomarkers/features of interest retrieved for analysis are detailed in *Table 1* along with the key of the field belonging to that feature.

MRI results are obtained after they are preprocessed by a software called FreeSurfer [4], which is used to measure brain structural integrity and calculate measures of volumes, cortical thicknesses, and surface areas [6]. Another type of MRI scan performed was

Features	Field Name
Cognitive tests:	MMSE, MMSE_bl (Baseline), ADAS11, ADAS13, CDRSB, RAVLT_immediate
MRI measures	Hippocampus, WholeBrain, Entorhinal, MidTemp
PET measures	FDG, AV45
Cerebral-spinal fluid measures	ABETA_UPENNBIOMK9_04_19_17 (amyloid-beta) TAU_UPENNBIOMK9_04_19_17 (tau level) PTAU_UPENNBIOMK9_04_19_17 (phosphorylated tau level)
Risk factors	APOE4, AGE, PTGENDER
Entorhinal Thickness (R)	ST83TA_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Average) ST83TS_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Standard Deviation)
Entorhinal Volume (R)	ST83CV_UCSFFSX_11_02_15_UCSFFSX51_08_01_16
Entorhinal Thickness (L)	ST24TA_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Average) ST24TS_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Standard Deviation)
Entorhinal Volume (L)	ST24CV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16
Rostral anterior cingulate volume	ST113CV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 ST54CV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16
Medial orbito-frontal thickness (L)	ST39TA_UCSFFSX_11_02_15_UCSFFSX51_08_01_16 (Average) ST39TS_UCSFFSX_11_02_15_UCSFFSX51_08_01_16 (Standard Deviation)
Medial orbito-frontal area (L)	ST39SA_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16
Hippocampus (R)	ST88SV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Volume) RIGHT_HIPPOCAMPUS_UCBERKELEYAV45_10_17_16 RIGHT_HIPPOCAMPUS_SIZE_UCBERKELEYAV45_10_17_16 RIGHT_HIPPOCAMPUS_UCBERKELEYAV1451_10_17_16
Hippocampus (L)	ST29SV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Volume) LEFT_HIPPOCAMPUS_UCBERKELEYAV45_10_17_16 LEFT_HIPPOCAMPUS_SIZE_UCBERKELEYAV45_10_17_16
Inferior Lateral Ventricle (R)	ST89SV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Volume)
Inferior Lateral Ventricle (L)	ST30SV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Volume)
Insula thickness (L)	ST130TA_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Average) ST130TS_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Standard Deviation)
Cuneus Thickness (R)	ST82TA_UCSFFSX_11_02_15_UCSFFSX51_08_01_16 (Average) ST82TS_UCSFFSX_11_02_15_UCSFFSX51_08_01_16 (Standard Deviation)
Amygdala (L)	ST12SV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Volume) LEFT_AMYGDALA_UCBERKELEYAV45_10_17_16 LEFT_AMYGDALA_SIZE_UCBERKELEYAV45_10_17_16
Middle temporal thickness (L)	ST40TA_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Average) ST40TS_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Standard Deviation)
Superior temporal area (R)	ST117SA_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16
Posterior cingulate thickness	ST109TA_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (R, Average) ST109TS_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (R, Standard Deviation) ST109TA_UCSFFSX_11_02_15_UCSFFSX51_08_01_16 (L, Average) ST109TS_UCSFFSX_11_02_15_UCSFFSX51_08_01_16 (L, Standard Deviation)
Pre-central thickness (R)	ST110TA_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Average) ST110TS_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Standard Deviation)
Total Features: 23	**Total Fields: 59**

Table 1: Extracted features used, and the keys for the feature in the data set.

diffusion tensor imaging (DTI), which help obtain values regarding the brains white matter tracts, related to cells and axons. DTI is used to estimate the brains mean diffusivity, axial diffusivity, and radial diffusivity. FDG-PET scans are used to obtain average measurements of cell metabolism, AD affected individuals tend to show reduced metabolism.

Also, a number of neuropsychological tests were conducted, by clinical experts, to assess a variety of skills such as general cognition, language, memory, vision, and many others. The cognition tests are used as an indicator of cognitive decline in AD individuals. If there is no decline in mental health, the next time taking the test, individuals should already remember how to approach it and thus perform better, however the opposite is seen in AD patients as one would expect with dementia. Other data provided include: genetic data, demographics, and medical diagnosis among many others. The results of all tests are included in the data for

each patient entry, in the data-sets provided, unless otherwise specified (D2 set doesn't include diagnosis).

III. TECHNICAL PLAN

A. Extracting Data

The data provided by ADNI for the challenge is one large .csv formatted file, containing a massive set of measurements obtained from the different biomarkers analyzed. The file contains data for 12,741 longitudinal patient measurements, between multiple studies each containing close to 2,000 different measurements. Among those, training and testing data were split by ADNI studies, there were 12,671 measurements obtained during D1 (training set), and 7,660 measurements during D2 (testing set). The number of patients who were found to be cognitively normal (CN) in D1 was 3,810. While the number of patients who were found to be cognitively normal (AD) in D1 was 1,544. For testing (D2), the number of CN measurements was 2,797 and the number of AD measurements was 362. The remaining classification data not used were found to potentially have cMCI, or MCI, who were not necessary for the scope of this research.

The current implementation of *Alzhetect* the features were manually obtained from the data sets provided. The features used for training and predicting of the DNN model (more in sub-section B) were obtained from multiple sources [2][3][6] and provided in *Table 1*. The table is comprised of 23 main features, each may contain multiple different measurements, for a total of 59 different features extracted. Parsing was performed using *Pandas*, an open source, high-performance, data structure and data analysis tool. Pandas was very helpful in transforming the original data (in the .csv files) into the proper subset for each class (AD, CN), then creating .csv files from each of those for future ease of use.

There were many necessary transformations needed to be done to the data in order to properly train a working DNN model. The first of which was described in the previous paragraph. Further transformation required involve converting patient gender into binary representations (0-Female, 1-Male) using a LabelEncoder provided by scikit-learn. Further normalization of the data was performed using an Imputer, also provided with scikit-learn, to complete (pad) null values in the tables.

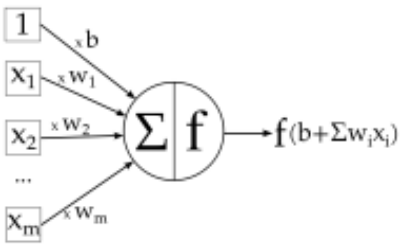

Fig. 1. ANN Neuron

Future work on this research will focus on using several feature extraction models to test and see which one produces better results. Some of the models I intend to use are: SVM, cross-validation, random-forest [2], logistic regression [3]. They will be used to reduce the number of data values provided, and produce a smaller set of correlated values for each patient data-set provided. Feature extraction may be performed to every value in the training data set as well as any data we will be predicting on using this model. The results obtained (relevant fields) will then be used as inputs to a deep neural network, which will be used to accurately classify individuals into one of 2 (potentially more) classes: CN, and AD.

B. Training a Deep Neural Network

An artificial neural network is a mathematical model used to solve problem related to optimization. An artificial neural network is an interconnected network of neurons, which take an input, computes on that input, and produce a value. More specifically, a neuron sums the values of previous neurons with the weight of each connection. Initially a neurons must be initiated with help of an activation function. A very commonly used activation function uses a sigmoid neuron which output a value between [0, 1]. After a neuron calculates all possible values it must add a bias at every layer of a neural network. Fig. 1 is an example of a neuron receiving values from previous layers and applying an activation function to produce a single output for the next layer.

The model will be trained using the D1 data-set provided by ADNI for the challenge. Once the features have been extracted from the original file, using a supervised learning classifier, they will be used for training using a Deep neural network. According to a

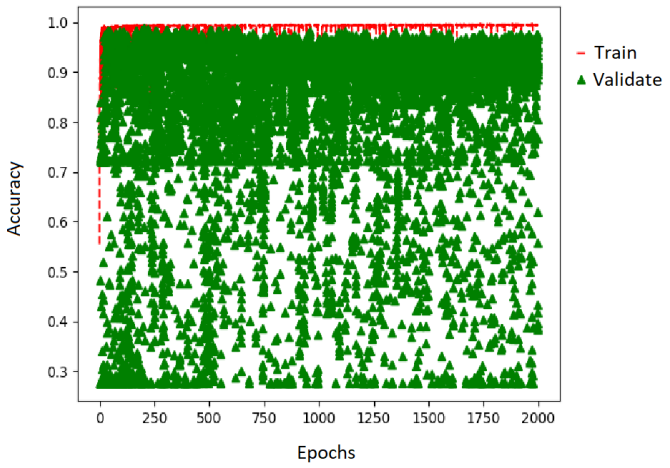

Fig. 2: Training graph produced by DNN, showing the training and validation accuracy increasing as the epochs increase.

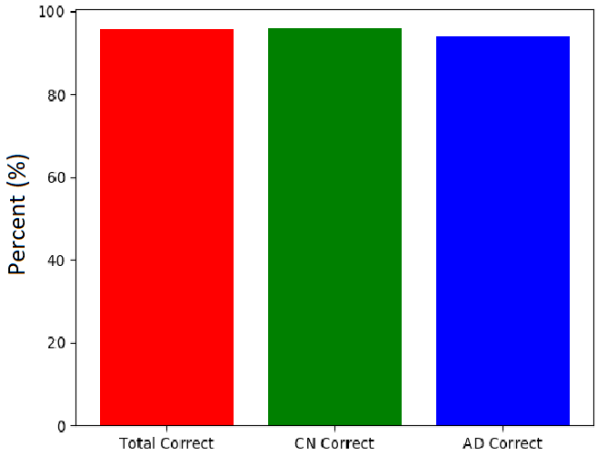

Fig. 3: Testing accuracy for all classifications together, as well as CN and AD individually.

previous study, using the data provided by ADNI, a DNN with around 11 layers, including hidden and visible layers, and reduced number of neurons will perform best [2]. In their research any added layers after 11, or more neurons did not yield much improvement. Amoroso et al. [2] used an input layer consisting of 2056 units, and a 4 unit output, since they were trying to classify 4 different classes: CN, cMCI, MCI, AD. However, for our research and methods the neural network consists of only 10 hidden units, as well as an input layer and an output layer consisting of 2 possible outputs and a confidence for the result that was predicted, the possible prediction classes are: CN, or AD. Also, The DNN also produce a weighted graph that

RID	Baseline DX	Prediction
4024	AD	AD
4039	AD	AD
...	...	...
125	CN	CN
315	CN	AD

Table 2: Contains an example output of the DNN. Containing patient diagnosis, model's prediction of patients diagnosis.

may be kept and used for further predictions without the need to re-train, we may even further train the model if we wish as well through the graph produced. The graph produced by the DNN is provided in *Fig. 2*. To optimize the parameters of the DNN we used AdamOptimizer during training. Testing accuracy on model prediction is also provided in *Fig. 3* for multiple cases, the accuracy of predicting correctly the total data, and the accuracy of predicting correctly for CN and AD. Future works will focus on providing prediction of more types of dementia classification.

IV. RESULTS

We analyzed large quantities of data trying to find the features that would improve the performance of a trained deep neural network. The accuracy graph in *Fig. 2* shows the results of the training, the model has an accuracy of 95.63% according to the test results. Also, as we can see the model predict accurately on the test data as well, specifically in *Fig. 3*, we can see the individual prediction results for AD and CN individuals was actually 93.92% and 95.85% respectively. These results are very good at first glance. However, the model appeared to be performing too well under the number of parameters it must calculate weights for. After investigating this issue, we realized that the data was longitudinal and therefore contains data from the same individuals on multiple occasions, which introduces the potential of overfitting our model. Having similar data for the same patients means that the testing data also contained individuals which the model trained with. Therefore, this account for the high prediction accuracy.

The trained model was also developed to output a results file associated with any input data entered (as long as it is contains the proper measurements). A condensed version of the output file can be seen in *Table 2*, where the model provides the roster id of each patient (unique), along with the diagnosis already given to the individual, and prediction value from the DNN. This table was used to obtain model training accuracy. For future work we will use a confusion matrix to obtain more precise model accuracy.

V. CONCLUSION

The work presented a python tool, Alzhetect, which demonstrated the potential of deep learning models for early detection of alzheimer's disease. We were also able to obtain a comprehensive list of biomarkers based on several different studies presented in this paper. While the model might be overfitting on the data, due to the time constraints involved in completing this project, it provides a proof of concept idea. Providing a base for further improvements and optimizations to aid in early detection. This is an ongoing area of research where very little is still known about the disease. Future works will primarily focus on removing the longitudinal constraints of the project, for the sake of training our model and obtaining higher accuracy without the cost of overfitting. Also, normalizing the data must also be done properly as the values in some fields can range greatly from one another.

REFERENCES

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. *http://adni.loni.usc.edu*

** TADPOLE Contest details: *https://tadpole.grand-challenge.org*

[1] "Dementia Fact sheet N°362". World Health Organization. March 2015. Archived from the original on 18 March 2015. Retrieved 13 January 2016.

[2] Nicola Amoroso, Domenico Diacono, Annarita Fanizzi, Marianna La Rocca, Alfonso Monaco, Angela Lombardi, Cataldo Guaragnella, Roberto Bellotti, Sabina Tangaro, Deep learning reveals Alzheimer's disease onset in MCI subjects: Results from an international challenge, Journal of Neuroscience Methods, 2017

[3] Moradi, Elaheh & Pepe, Antonietta & Gaser, Christian & Huttunen, Heikki & Tohka, Jussi. (2014). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage. 104. 10.1016/j.neuroimage.2014.10.002.

[4] Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B. 2012. Within-Subject Template Estimation for Unbiased Longitudinal Image Analysis. Neuroimage 61 (4), 1402-1418.

[5] Harper, Lorna et al. "MRI Visual Rating Scales in the Diagnosis of Dementia: Evaluation in 184 Post-Mortem Confirmed Cases." Brain 139.4 (2016): 1211–1225. PMC. Web. 23 Mar. 2018.

[6] Rachael, Scahill, et al. Mapping the evolution of regional atrophy in Alzheimer's disease: Unbiased analysis of fluid-registered serial MRI, Proceedings of the National Academy of Sciences Apr 2002, 99 (7) 4703-4707; DOI:10.1073/pnas.052587399

[7] Yang, E, et al. "Quantifying the Pathophysiological Timeline of Alzheimer's Disease." Journal of Alzheimer's Disease : JAD., U.S. National Library of Medicine,2011,www.ncbi.nlm.nih.gov/pubmed/21694449.

[8] Hebert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. Neurology. 2013;80(19):1778-1783. doi:10.1212/WNL.0b013e31828726f5.