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

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

***Abstract* -- Alzheimer's disease (AD) is one of the most obscure yet significantly debilitating diseases currently facing humanity. The ability to effectively treat AD is dependent on identifying the disease early on. This study uses data from an international challenge (TADPOLE), which aims to predict future AD progression. The research presented in this paper will focus on identifying and using biomarkers (features), that are strong indicators of Alzheimer's disease. Through the use of multiple supervised machine learning models, the attempt to identify features of importance will be made. Those same features will then be input into another model in order to properly classify individuals into different stages of the disease. The stages used in this research consist of individuals who are cognitively normal (CN), mild cognitively impaired (MCI), AD, and converting from MCI-to-AD. A random forest regressor was adapted to identify features of importance. The top 20 features pinpointed were used for classification. Multiple models were adopted to classify individuals, including: KNN, SVC, and a DNN. The results of the models were then compared using testing data, demonstrating that an SVC reached greater classification accuracy than the other aforementioned models. In this research we also aim to predict when an individual, diagnosed as MCI, may convert to AD. A secondary classification model, using a DNN, was trained to further classify individuals. This model will only classify patients predicted to be MCI-to-AD. The smaller pool of patients were classified by the predicted amount of months remaining until they convert to AD. Our secondary model reached 36.7% accuracy when determining the amount of time until transition to AD. The accuracy of the top 10 teams who submitted a model to the challenge, range from 38% to 42%.**

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

Project Wiki: https://github.com/raidel123/AlzheTect/wiki

Keywords -- K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Deep Neural Network (DNN) mild cognitive impairment (MCI), AD, ADNI, classification.

I. INTRODUCTION

During 2010, there were 4.7 million individuals affected by Alzheimer's Disease (AD). By 2050, it is predicted the figure will be 13.8 million. Alzheimer's is the leading cause of dementia[8]. Dementia is a term used to describe the set of conditions affecting an individual's memory, thinking, and social abilities. Statistics show dementia affects 7% of individuals over the age of 65, and 20% of individuals above 80. Alzheimer's is known as the most common type of dementia observed, and contributes 60% to 70% of all cases [1]. Alzheimer's exists as a neurodegenerative, progressive disease, meaning it worsens over time. Contrary to popular belief, Alzheimer's does not only affect the older population, many individuals under the age of 65 are affected, approximately 200,000 cases in the U.S alone. These individuals are said to have younger-onset (or early-onset) Alzheimer's, 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 a possibility to reduce symptoms affecting their everyday life.

Due to this increasingly prevalent disease, the ability to identify different biomarkers for early onset diagnosis is a crucial area of research. Many previous efforts 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 been recently proposed to help accurately measure certain interest regions of the brain. Interest regions of the brain, typically atrophy as the condition progresses, such as the hippocampus. Providing techniques to measure certain atrophied regions has enabled researchers to properly classify individuals by severity of dementia, or cognitive normality.

Motivation for this research stemmed from a challenge posted on *Kaggle* [2], by a UCL group focusing on Progression Of Neurodegenerative Disease

(POND). The challenge, named The *A*lzheimer's *D*isease *P*rediction *O*f *L*ongitudinal *E*volution (TADPOLE)**, aims at predicting which MCI subject 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).

The data set provided by the challenge contained 1,907 features for each patient visit. A reduced list of relevant fields known to indicate AD, suggested by the challenge coordinators, as well as other research performed on the topic [2], was compiled. A random forest regressor was then utilized on the reduced list of features to calculate feature importance. The top 20 resulting features from the random forest regressor were input into a variety of classifiers. The classifiers were used to predict the current stage of a patient in AD progression. If they are MCI-to-AD, it will predict a time range (in months), in which the patient is expected to convert from MCI to AD. The classifiers trained include a K-Nearest Neighbors (KNN) classifier, Support Vector Classifier (SVC), and a Deep Neural Network (DNN).

The remainder of the paper will focus on the following: a breakdown of the data provided by the TADPOLE challenge and the motivation for the research. In Section 3, we describe the data transformations performed along with a plan on how features were extracted and split for training and testing. A layout of the technical plan describing the methodology used to obtain our results is provided, along with the implementation of the KNN, SVM, and DNN in section 4. Section 5 will cover the results obtained with our trained classifiers, including training/testing performance and detailed tables and graphs. The efforts made through this research are concluded in section 6.

II. Data

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, as well as other data values.

A. Obtained Data Files

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 clinicals. 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 and stored 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, MCI, or AD. Data-set 2 (D2) is a collection of only 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. Biomarker Description

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 (field name) of the field belonging to that feature.

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.

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 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 and diagnosis of all tests are included for each patient entry, in the data-sets provided, unless otherwise specified.

III. Transforming the Data

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 csv file contains 12,741 longitudinal measurements, between multiple studies. Each entry contains 1,907 different biomarkers or features, including a diagnosis for each individual. Therefore, it is imperative that the data provided is separated into classes, based on the diagnosis. A total of 1,577 unique individuals were used in this study, they are distributed based on diagnosis (DX) as follows: CN (423 individuals), MCI (483 individuals), AD (332 individuals), MCI-to-AD (339 individuals). The groups diagnosed as CN, MCI, and AD were given their respective baseline diagnosis and never deviated throughout the study. On the other hand, individuals diagnosed as MCI-to-AD were at some point diagnosed with MCI, and then converted to AD.

For individuals who are MCI-to-AD, a second set of training and testing datasets were obtained from the MCI-to-AD dataset. The 339 individuals known to have converted from MCI to AD were also separated into classes. The individuals were classified based on the time elapsed from initial diagnosis of MCI until they converted to AD, in months. Four classifications were created, based on the statistical quartiles of the data. The classes were labeled with integer values (0, 1, 2, 3) and correspond to the ranges in table 2. The 339 MCI-to-AD individuals were split into each class and they are distributed as follows: 0 - (76 individuals), 1 - (110 individuals), 2 - (64 individuals), 3 - (89 individuals).

In the current implementation of *Alzhetect*, a novel Alzheimer's detection tool, a large quantity of features were manually obtained from the data sets provided. These manual features were recommended as important for Alzheimer's classification from multiple sources [2][3][6], and provided in *Table 1*. The table is made up of 59 features that were extracted and used for

Class	Months to AD Conversion
0	less than 12 months
1	$12 < X \leq 24$ months
2	$24 < X \leq 36$ months
3	More than 36 months

Table 2 : Classifications of early MCI-to-AD prediction, by months.

further processing. Parsing was performed using *Pandas*, an open source, high-performance, data structure and data analysis tool. Pandas was very useful when transforming the original data (in the .csv files) into the proper subset for each class (CN, MCI, AD, MCI-to-AD). The pandas dataframe was then queried for the subset of fields manually obtained through research. Once the csv file was properly analyzed and filtered, it was used as input to a random forest regressor explained in the next subsection.

Other necessary transformations need to be performed on the data in order to properly train working classifiers. 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. The last preprocessing operations performed before fitting the model were mean removal and variance scaling using scikit-learn preprocessing scaler.

B. Feature Selection

To reduce the number of features required to train the model, a Random Forest Regressor was used to evaluate features in terms of importance. Selecting only highly relevant/correlated features helps improve the prediction accuracy of the classifiers. A Random Forest Regressor is an ensemble of decision tree classifiers that uses averaging to improve accuracy of predictions and helps avoid overfitting of a model. Overfitting occurs when a model learns the training data too well but performs poorly on testing data. The regressor used in

Measurement_Type	Field_Name
Cognitive Test	MMSE_bl
Cognitive Test	CDRSB
Cognitive Test	ADAS13
Cognitive Test	ADAS11
Cognitive Test	MMSE
Cognitive Test	RAVLT_immediate
Risk factor	APOE4
PET measures	AV45
Amygdala (L)	LEFT_AMYGDALA_UCBERKELEYAV45_10_17_16
MRI measures	Hippocampus
MRI measures	WholeBrain
Entorhinal Volume (R)	ST83CV_UCSFFSX_11_02_15_UCSFFSX51_08_01_16
Inferior Lateral Ventricle (L)	ST30SV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Volume)
Hippocampus (R)	ST88SV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Volume)
Cuneus Thickness (R)	ST82TA_UCSFFSX_11_02_15_UCSFFSX51_08_01_16 (Average)
Cuneus Thickness (R)	ST82TS_UCSFFSX_11_02_15_UCSFFSX51_08_01_16 (Standard Deviation)
Hippocampus (L)	ST29SV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16 (Volume)
Hippocampus (L)	LEFT_HIPPOCAMPUS_UCBERKELEYAV45_10_17_16
Posterior cingulate thickness (L)	ST109TA_UCSFFSX_11_02_15_UCSFFSX51_08_01_16 (Average)
Medial orbito-frontal thickness (L)	ST39TA_UCSFFSX_11_02_15_UCSFFSX51_08_01_16 (Average)

Total Fields: 20

Table 3: Selected Features using the random forest classifier, along with the keys for the featuresi n the dataset

this study evaluates the features of importance in terms of mean squared error. The adopted configuration of the regressor is very similar to that of [2], a 500 tree ensemble and 20 features randomly considered when selecting the best split. Using this Random Forest Regressor configuration the 20 most important features were selected and provided in table 3.

C. *Training/Testing Datasets*

Once data extraction and feature selection occurred, the data was shuffled and split into train and test datasets. The data was split as follows: 80% for training and 20% for testing. Therefore, out of the total 1,577 individuals: 1,261 (80%) were used for training and 316 (20%) were used for testing and validating the accuracy of the classifier. The *training dataset* consisted of the following number of individuals from each stage of progressive Alzheimer's disease: 327 (CN), 395 (MCI), 270 (AD), 269 (MCI-to-AD). Likewise, the *testing dataset* consisted of the following number of individuals from each stage of the disease: 96 (CN), 88 (MCI), 62 (AD), 70 (MCI-to-AD).

Since we also created a second classifier model to determine the number of months until an individual converts MCI-to-AD, we must also shuffle and split the datasets. The data was also split 80:20 for training and testing. That means that out of the 339 MCI-to-AD individuals: 271 (80%) were used for training the models, and 68 (20%) were used for testing. The *training dataset* consisted of the following number of individuals for each class (in table 2): 59 (0), 92 (1), 50 (2), 70 (3). The *testing dataset* consisted of the following number of individuals for each class (in table): 17 (0), 18 (1), 14 (2), 19 (3).

IV. Methods of Classification

A. *K-Nearest Neighbors (KNN)*

KNN is a supervised, lazy machine learning classifier. This classifier is based on feature similarity: how closely related a data point is to K of its neighbors (where K is a natural number). It is lazy due to its nature of having an extremely minimal training stage, or not having an explicit training stage at all. Therefore, it does not use any of the training data to do any generalization. This study adopted a general KNeighborsClassifier, provided by scikit-learn. A default value of 5 was used for K. Euclidean distance to compare the distance between neighbors was also implemented. During training, the adopted model achieved a 63.6% accuracy using a KNN classifier. The model was saved into a pickle to be loaded and used at anytime without needing to retrain the model.

B. *Support Vector Classifier (SVC)*

A Support Vector Machine (SVM) is a classifier formally defined by a separating line or hyperplane, depending on the number of dimensions (classes). An SVM effectively separates all the classes in a multi-dimensional space. This study implements a Support Vector Classifier (SVC) provided by scikit-learn

library. The SVC applied uses a radial basis function (rbf) kernel, allowing this model be effective in high dimensional spaces as the one that will be classified. The model was also trained to provide probability estimates for each classification attempted, which entailed the use of five-folds cross-validation. The overall SVC model achieved a 73.9% accuracy in training. The trained classifier was also saved into a pickle to be loaded at a later time for future predictions.

D. Deep Neural Network (DNN)

An artificial neural network is a mathematical model used to solve problems 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 neuron must be initiated with help of an activation function. A very commonly used activation function uses a sigmoid neuron which outputs a value between [0, 1]. Once 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 based on a random forest regressor, they will be used for training using a Deep neural network. According to a 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. Each output corresponds to one of the 4 different diagnosis class/stage of AD progression.

The adopted DNN model for this research is very similar to that of [2]. The model contains 11 layers: 1 input (relu activator) layer + 9 hidden layers (mix of relu and tanh activators) + 1 output layer (softmax). The

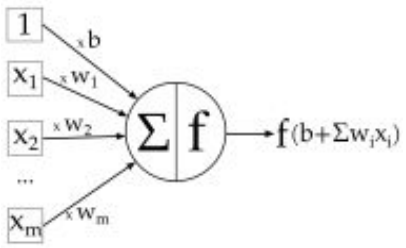

Fig. 1. ANN Neuron

first 3 layers consist of 128 neurons, layers 4-6 consist of 64 neurons, layers 7-9 consist of 32 neurons, layer 10 consist of 16 neurons, and the output layer consist of 4 neurons. The output of the softmax function applied to the last layer provides us with a probability prediction of which class in the model the input belongs to. A weighted graph is then created by the DNN, which is kept and used for further predictions without the need to re-train. The model may be further trained as well, through the graph produced. To optimize the parameters of the DNN, an AdamOptimizer was used during training, which is known to be a good deep learning optimizer to iteratively update network weights. The overall trained network achieved a 62% diagnosis classification accuracy in training. The model and the respective trained weights were also saved to be loaded for future predictions without the need to retrain the network.

V. RESULTS

In our study we achieved results close to that of the top 10 competitors in the TADPOLE challenge. The overall final classification models, used to determine the number of months left to AD, were very similar to those achieved by [2] which had a 38% prediction accuracy. However, our tool (Alzhetect) applied a two level system for prediction during classification. The first level consists of predicting the patient's diagnosis and determining whether they are CN, MCI, AD, or MCI-to-AD. The second level of our model will take the output of the first level and use it as it's input. The second level, or MCI-to-AD classifier, was used to classify MCI-to-AD conversion time, in months.

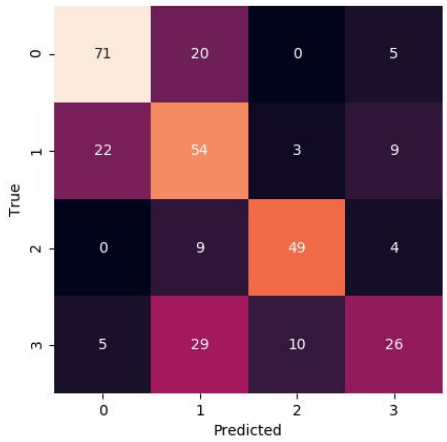

Fig. 2: Confusion matrix for KNN diagnosis of AD progression, the labels refer to the stage of the disease. (0: CN, 1: MCI, 2: AD, 3: MCI-to-AD)

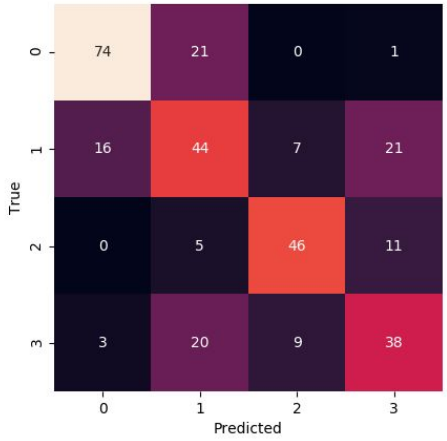

Fig. 3: Confusion matrix for DNN diagnosis of AD progression, the labels refer to the stage of the disease. (0: CN, 1: MCI, 2: AD, 3: MCI-to-AD)

Diagnosis classification was performed using several models: KNN, SVC, and DNN to compare results. Conversion time classification was performed using a DNN as detailed by [2]. The models were tested using 'real-world data' provided by ADNI in D2 and D3 studies.

Firstly, let's look at the results of the KNN classifier. Testing the model with the 'real-world' test data, provided by ADNI in D2, yields a 62% accuracy. More specifically, according to the confusion matrix in Fig. 2 the following are the correct prediction percentages for each stage of progressive Alzheimer's disease: CN 74%, MCI 61%, AD 79%, and MCI-to-AD 37%. The percent correct for each prediction is also provided in Fig. 5. The KNN classifier performed well for CN, MCI, and AD classes. However, it performed

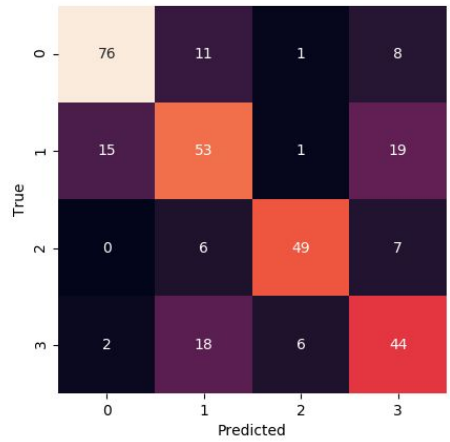

Fig. 4: Confusion matrix for SVC diagnosis of AD progression, the labels refer to the stage of the disease.
(0: CN, 1: MCI, 2: AD, 3: MCI-to-AD)

poorly on MCI-to-AD classifications as it would incorrectly classify 41% of individuals as MCI.

The support vector classifier (SVC) and deep neural network (DNN) were then tested with the same D2 dataset as the KNN classifier. The results were similar for both classifiers, the DNN achieved 67% accuracy, while the SVC achieved 70%. More specifically, according to the confusion matrix in Fig. 3 for the DNN classifier the following are the correct prediction percentages for each stage of progressive Alzheimer's disease: CN 77%, MCI 50%, AD 68%, and MCI-to-AD 54%. Likewise, for the SVC classifier the confusion matrix in Fig. 4 yielded the following correct prediction percentages for each stage: CN 79%, MCI 60%, AD 79%, and MCI-to-AD 63%. The percent correct for each stage prediction is also provided in Fig. 5 for each classifier. Although the increase in accuracy for CN, MCI, and AD individual seems minimal, it is still significant. The largest prediction accuracy increases were seen for MCI-to-AD individuals where the number of incorrect MCI classifications, seen in the KNN results, decreased.

Classifying MCI-to-AD subjects by months-left to AD conversion, resulted in a DNN model that was 36.7% accuracy. The results for each classification of months-left to conversion can be found in Fig. 6. The low accuracy can be attributed to the size of the dataset available for training and testing. The accuracy of this classifier is similar to that of the top 10 submissions of the TADPOLE challenge, which vary from 38% to 42%.

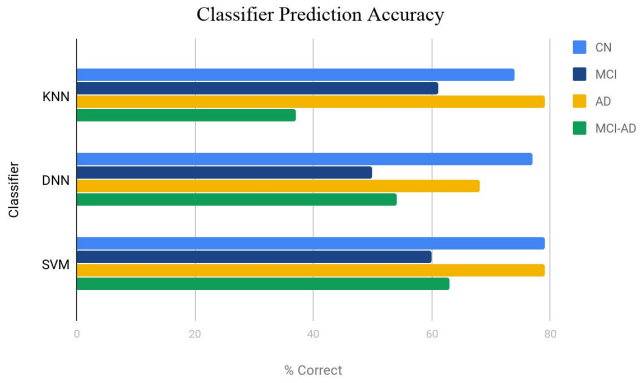

Fig. 5: Percent of correctly classified individuals for each stage of progressive AD classification, using the classifiers detailed in this paper.

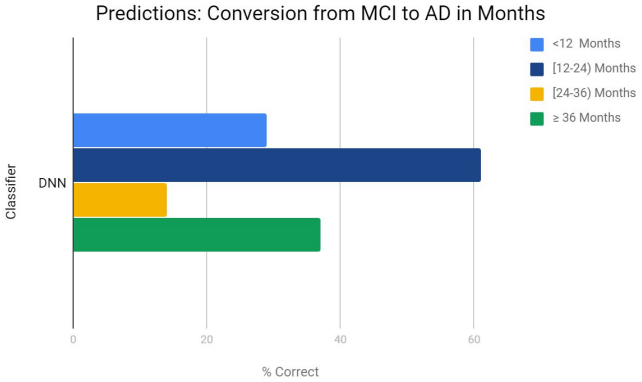

Fig. 6: Percent of correctly classified individuals for each time period classification, using the DNN detailed in this paper.

RID	DX	Actual MCI-to-AD Conversion	Predicted DX	Predicted MCI-to-AD Conversion
182	MCI-AD	6 months	MCI-AD	0-12 months
324	MCI	N/A	AD	N/A
108	MCI-AD	18 months	MCI-AD	12-24 months
57	MCI-AD	19 Months	MCI-AD	24-36 months
66	CN	N/A	CN	N/A
889	AD	N/A	AD	N/A
2079	MCI	N/A	MCI-AD	0-12 months

Table 4: Contains acondensed example output of the DNN multi-level classification. Contains patient's actual diagnosis and conversion times, as well the models predictions of patients diagnosis and conversion time to AD.

The trained models were also developed to output a results file associated with any input data entered (as long as it is contains the proper features , in table 3, required for classification). A condensed version of the output file can be seen in *Table 4*, where the model provides the roster id of each patient (unique), along with the diagnosis already given to the individual, and prediction values from the classifiers.

V. CONCLUSION

Through this research an original python tool, Alzhetect, is introduced. Alzhetect demonstrates the potential of deep learning models for early detection of Alzheimer's disease. The ability to obtain a comprehensive list of biomarkers based on several studies allows for concise prediction, contrary to including ambiguous data that may skew results.

This research reveals that using an SVC classifier provided slightly better accuracy (3% better) than when using a deep neural network for diagnostic classification. Additionally, it contributes enhanced accuracy compared to a KNN classifier. SVC performed superior to the rest, particularly when classifying MCI-to-AD. The DNN and KNN classifiers on the other hand, incorrectly classified a large amount of individuals as MCI when in actuality they were diagnosed as MCI-to-AD, as seen in Figures 2, 3, and 4.

The tool developed in this study, Alzhetect, provides a base for further improvements and optimizations to aid in early Alzheimer's disease detection. This continues to be an ongoing area of research, where minimal is known about the disease.
Future efforts will primarily focus on removing a larger portion of the longitudinal constraints of the project. ADNI is currently performing another round of studies on a group of individuals, which will be implemented on further iterations of this research effort..

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.