Investigating the Conversion between Alzheimer's Disease and Mild Cognitive Impairment

Meltem Arslan, 2016400117
Yunus Kardaş, 2014400120

I. Introduction

More than 30 million people have a clinical diagnosis of Alzheimer's disease (AD) worldwide, and this number is expected to triple by 2050.[1] Alzheimer's disease (AD) has been known as one of the progressive neurodegenerative diseases characterized by progressive memory loss. It has currently been ranked as one of the most common causes of death in developed countries. The progression and manifestation of AD cause a long and undetected preclinical stage, followed by the problem of dementia and the Mild Cognitive Impairment (MCI) stage, which eventually results in death. Because the progression of neuropathology in AD begins many years before clinical symptoms of the disease become evident and progressive neurodegeneration irreversibly damages the brain, emerging treatments will likely have the greatest impact when provided in the earliest stages of the disease. Therefore, rapid identification of people at high risk of developing AD is of great importance.

The ability to identify declining individuals at the prodromal AD stage provides a critical time window for early clinical management, treatment & care planning, and design of clinical drug trials.[2] Identification of individuals at risk and initiating early treatment in these individuals will ensure more efficient clinical results and reduce health costs and costs in clinical practice. However, simulations also suggest that the health care system is not prepared to handle the potentially high volume of patients who would be eligible for treatment.[3]

Mild Cognitive Impairment (MCI) is an ill-defined problem that causes memory problems that are more difficult to detect than AD. Around 10%–15% of MCI patients per year convert to AD over a relatively short time. However, the annual conversion rate tends to diminish progressively and the mean conversion rate from MCI to AD is approximately 4% per year.[4] MCI patients who do not revert to AD tend to remain stable at the disease level or to evolve into other stages of dementia or even return to their healthy positions.

In the modern world, many machine learning and artificial intelligence algorithms have been developed to objectively predict the conversion of MCI patients to AD using today's technologies and big data. Although algorithms for the health sector are still viewed with suspicion in the medical world, the fact that objective results can be obtained makes this situation an opportunity and development.

So far, many studies have focused on predicting the conversion of AD in MCI patients using different combinations of data, including brain imaging, CSF

biomarkers, genotyping, demographic and clinical information, and cognitive performance, achieving varying levels of accuracy.[5] As the reasons for the different results, it can be shown that applications such as brain imaging are very diverse and expensive and that the optimum state of all of them cannot be achieved due to cost.

This study also looks at the previous studies and considers the right and wrong or the deficiencies and redundancies in those studies. In this way, it aims to improve the results and accuracy rates obtained in the past, as well as to measure the connection more accurately between MCI and AD.

II. Related Work

The application of artificial intelligence and machine learning algorithms in the field of medicine with big data is possible thanks to today's technologies. For this reason, the use of machine learning and artificial intelligence in medicine dates back to the last 15-20 years. In the previous 5-10 years, these studies have accelerated.

In the first studies, standard machine learning algorithms such as SVM, kNN, Naive Bayes were generally used. However, efficient and accurate results could not be obtained with these applications. However, Net algorithms, which are more advanced versions of these algorithms, have started to be used and implemented recently.

It is seen that 3D-DenseNets was applied in the study conducted in 2018. With this application, it has achieved an accuracy rate of almost 80 percent on certain data.[6] In a study conducted in 2019, Net algorithms such as AlexNet and Xception were applied in one.[7] This Net algorithm is also supported by mathematical and statistical equations. In addition, in another study conducted in the same year, ElasticNet was applied and besides this, Naive Bayes, kNN, SVM, etc.[8] This machine learning technique is supported by algorithms. A study conducted in 2020 used MudNet and CNN to use brain images and the outputs of image processing technologies in general.[9] On the other hand, another study conducted in 2020 conducted a study using questionnaires made with people in a certain age range of AD and the characteristics of these people. It is seen that ANN and DT models are used in this study and have an accuracy of about 80-90 percent.[10]

As a result, each study aimed to measure the MCI - AD connection and to estimate the accuracy of this conversion at a higher rate, despite the use of different types of data and different techniques and aimed to increase the accuracy of the previous studies and make the prediction close to perfect. This study aims to measure the MCI-AD connection more accurately and predict the conversion of MCI to AD more accurately by taking the previous studies as a reference.

III. Dataset

The dataset used in this project was obtained from ADNI (Alzheimer's Disease Neuroimaging Initiative), a center established in 2004 to collect data about Alzheimer's Disease and research on it. The data contains 12741 patient records with labels:

"NL", "MCI", "Dementia", "NL to MCI", "NL to Dementia", "MCI to Dementia", "Dementia to MCI", "MCI to NL". The meanings of the labels: "NL": a person at a healthy state, "Dementia": a patient diagnosed with dementia, "MCI": a patient diagnosed with Mild Cognitive Impairment (MCI), which is an early stage of dementia.

Number of samples in the data set and for each set is shown below:

Class name	# of Samples
Normal	2745
MCI	4052
Dementia	2107

IV. Methodology

IV.I Preprocessing & Understanding the data

In the dataset, there were hundreds of features that we could not use in the first step, such as entry or registration dates of the patients, and the values of these features were not filled completely; some were even mostly empty. So, first, the features that can be useful for the study had to be selected. A separate dataset was created with the features that were found to be important in previous related studies on the subject. Selected features for this study include some demographical information, neuropsychological assessment results, measurement values of some proteins, data from MRI, and their baseline values. Selected feature names are: "AGE", "PTGENDER", "PTEDUCAT", "APOE4", "FDG", "CDRSB", "ADAS11", "ADAS13", "MMSE", "RAVLT_immediate", "RAVLT_learning", "RAVLT_forgetting", "RAVLT_perc_forgetting", "FAQ", "Ventricles", "Hippocampus", "WholeBrain", "Entorhinal", "Fusiform", "Midtemp", "ICV", "CDRSB_bl", "ADAS11_bl", "ADAS13_bl", "MMSE_bl", "FDG_bl", "RAVLT_immediate_bl", "RAVLT_learning_bl", "RAVLT_forgetting_bl", "RAVLT_perc_forgetting_bl", "FAQ_bl", "Ventricles_bl", "ICV_bl", "Hippocampus_bl", "WholeBrain_bl", "Entorhinal_bl", "Fusiform_bl", "Midtemp_bl".

Since our study will be carried out using only NL, MCI and Dementia distinction, the labels containing state changes are arranged to include only the last states:

NL to MCI	> MCI
NL to Dementia	> Dementia
MCI to Dementia	> Dementia
Dementia to MCI	> MCI
MCI to NL	> NL

In order to better understand the features, different pairs of features were selected, scatter plots were drawn and examined according to their labels. While doing this, only patients having values for both features were selected. Therefore, the number of patients were observed was roughly half of the entire dataset, although it varied for selected features.

Clustering is applied with c=3 to the features selected in this way. Clustering results for "FDG" and "CDRSB" features and ground truth values can be seen from the figures:

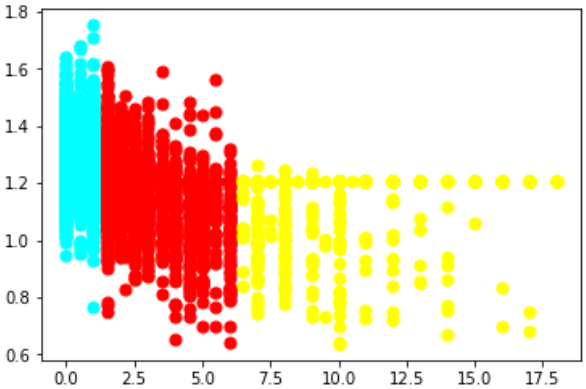


Figure1.1: Clustering result

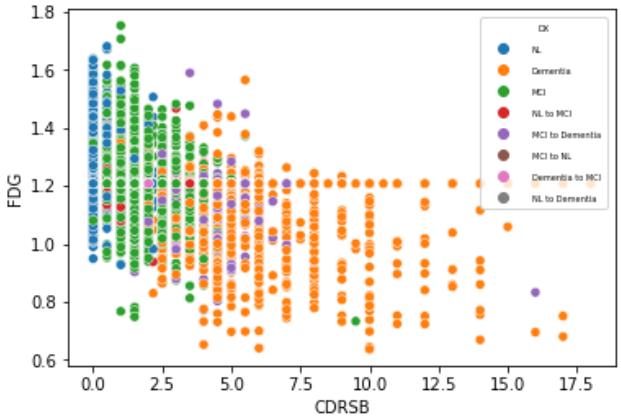


Figure1.2: Ground truth values (with seven labels)

Clustering results for “ADAS11” and “APOE4”:

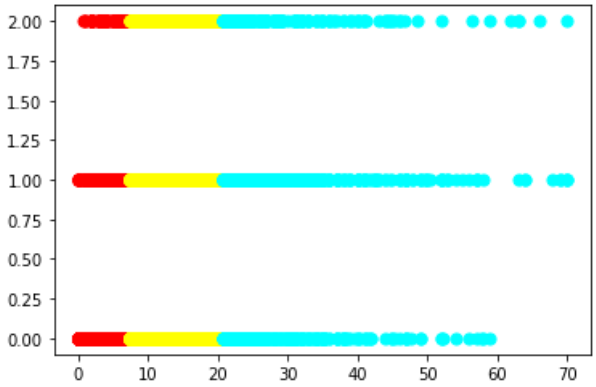


Figure2.1: Clustering result

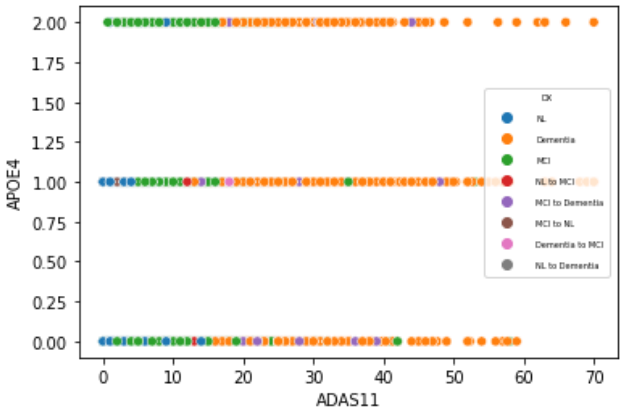


Figure2.2: Ground truth values (with seven labels)

As we can see in the figures below, due to the outliers in the “ADAS13” and “RAVLT_immediate” k-means offered a better group separation with the number of clusters equal to 4 (Figure3.2) instead of 3 (Figure3.1). So, these outlier values must be removed at some stage.

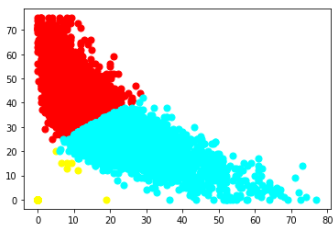


Figure3.1

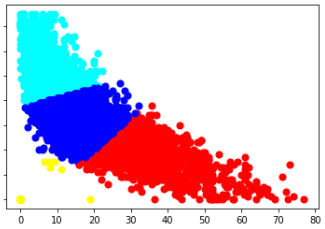


Figure3.2

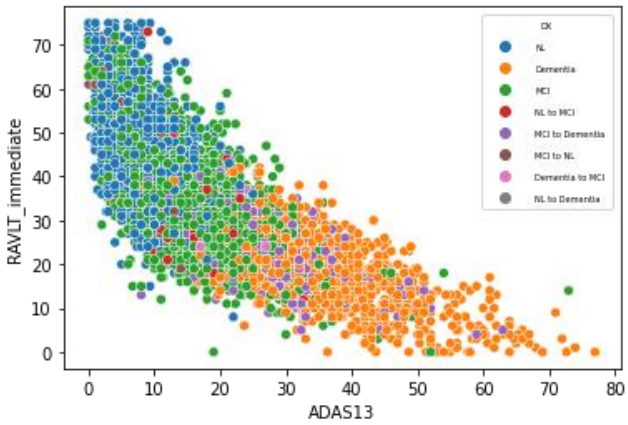


Figure3.3: Ground truth labels (with seven labels)

When clustering was wanted to be applied to all features, empty values had to be filled (data imputation) since the intersection set of the full features corresponds to a tiny portion of the data we have. The empty values were filled by the average of all values found in that feature (mean imputation). In this way, clustering would be able to run using all features. It was possible to examine the result of just two features quickly with a scatter plot, now this must be done for all features. In order to do this, T-SNE was used, which enables the visualization of multidimensional data. The k-means (k=3) results can be seen from the figures below:

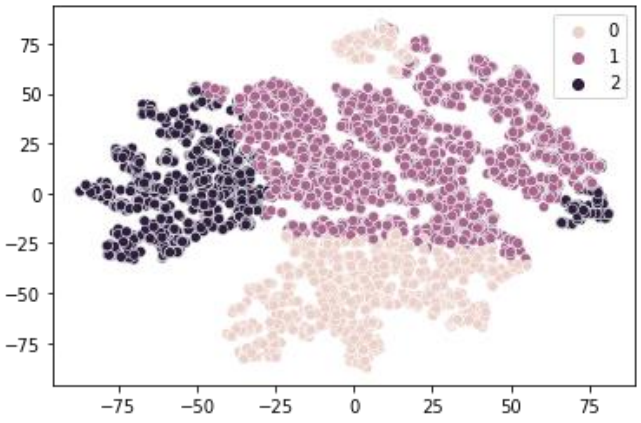


Figure4.1: Clustering result

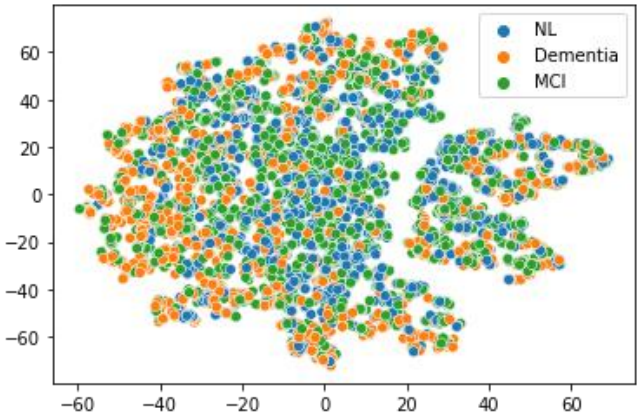


Figure4.2: Ground truth labels (with three labels)

However, as can be seen from the figures, the labels were intertwined, and no meaningful result could be observed. Two significant mistakes were causing this:

1- There was an imbalance between different feature values because the data were not normalized.

2- The empty fields in the data were filled without considering the labels, which had the opposite impact on clustering.

After the mistakes were noticed, the empty spaces were filled by taking the average values of features within each label (NL, MCI, Dementia) and used by normalizing. Then just 17 features were selected since they were the most essential ones. The selected features are: 'RAVLT_immediate', 'ADAS11', 'ADAS13', 'MMSE', 'Hippocampus', 'WholeBrain', 'FDG', 'MidTemp', 'Entorhinal', 'Fusiform', 'ICV', 'APOE4', 'Ventricles', ‘FAQ’, ‘CDRSB’, ‘AGE’, ‘PTGENDER’.

The average values of attributes for each class is shown in the following table (the averages obtained from normalized values):

Feature Name	***Normal***	***MCI***	***Dementia***
PTGENDER	0.494	0.343	0.430
AGE	0.543	0.507	0.542
APOE4	0.143	0.302	0.445
FDG	0.594	0.505	0.361
CDRSB	0.009	0.101	0.358
ADAS11	0.078	0.156	0.346
ADAS13	0.101	0.206	0.416
MMSE	0.968	0.904	0.692
RAVLT_immediate	0.613	0.422	0.256
FAQ	0.007	0.128	0.639
Ventricles	0.191	0.246	0.323
Hippocampus	0.564	0.485	0.358
WholeBrain	0.438	0.451	0.363
Entorhinal	0.489	0.424	0.278
Fusiform	0.446	0.435	0.313
MidTemp	0.491	0.472	0.346
ICV	0.573	0.605	0.603

Table4.3: Feature averages for each class

The values in the green rows are constantly increasing from normal to dementia and the values in the orange rows are constantly decreasing. On average, the difference between Dementia and MCI is lower than the difference between MCI and Normal which can make it harder to distinguish a normal person from an MCI patient.

The visualization of the k-means clustering result(with c=3) and ground truth values with T-SNE is as follows:

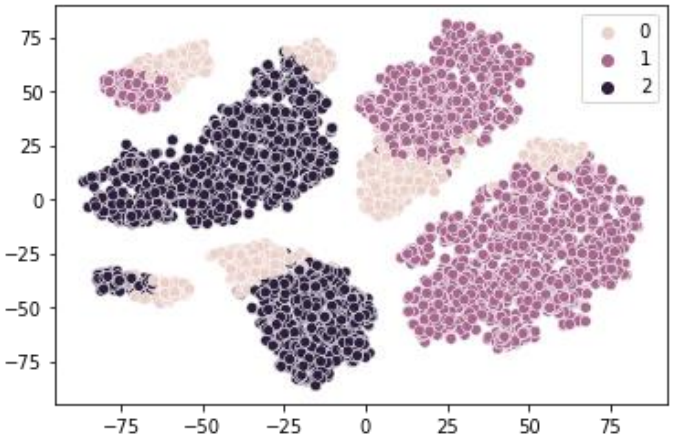


Figure5.1: Clustering result

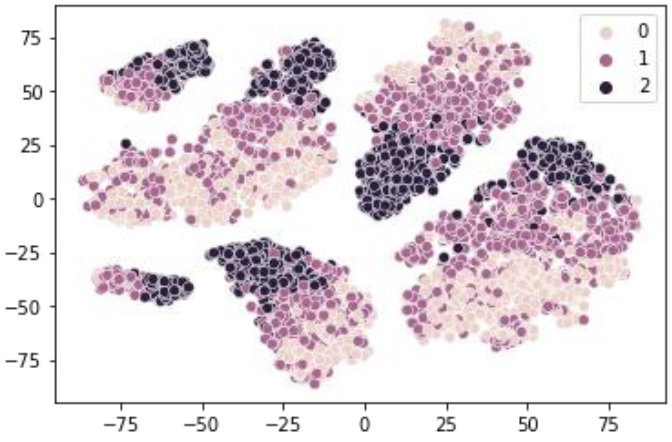


Figure5.2: Ground truth labels

K-means still did not classify the classes well. Having blobs at six different places as in Figures 5.1 and 5.2 is not an expected or desired result. To get a meaningful result from classifying, the reason must be found before continuing to develop the model. One reason for this could be the "PTGENDER" feature because it has just two discrete values as 0 and 1. When the dataset is visualized for just females or males, we got the following figures:

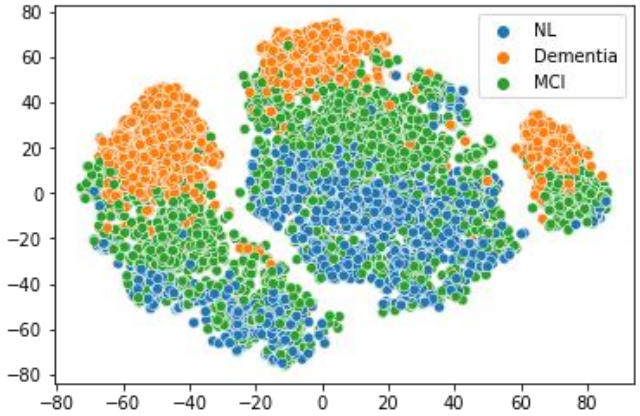


Figure6.1: Data Distribution – Female

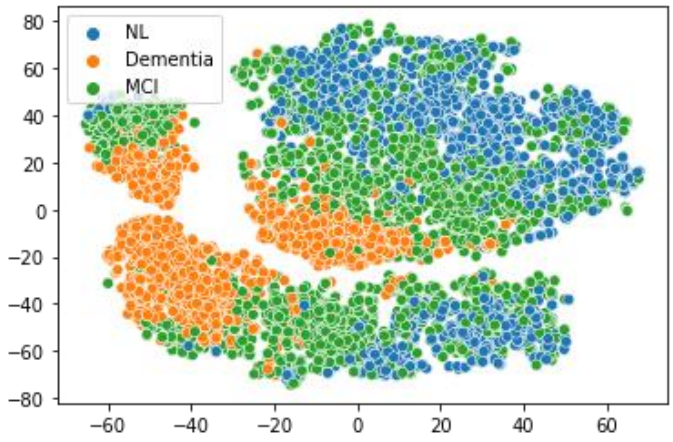


Figure6.2: Data Distribution - Male

When plots are drawn for two genders separately, it seems that one of the reasons for clustering is the

values of the “PTGENDER” feature. Since there are still three different clusters in each plot, it can be considered that there is at least one more feature that causes this. As a result of examining the dataset for the features we use, it is seen that the “APOE4” protein only takes 0, 1, and 2 values. Therefore, “APOE4” is removed from the feature set as well as “PTGENDER” and the development of the model is started.

IV.II PCA + K-MEANS

After getting clearer data representation, PCA was applied. To determine the number of components for PCA “n_components“ was tried with different values beginning from 1.

n_components	explained_variance_ratio_
1	0.44
2	0.66
3	0.75
4	0.83
5	0.87

When applying PCA, a threshold value of 0.85 was determined so that information loss would not be high, and the number of components was chosen to be at least 5. After PCA, k-means was applied with n_clusters = 3, max_iter = 2000 and n_init = 50. The model accuracies with different number of components are shown below:

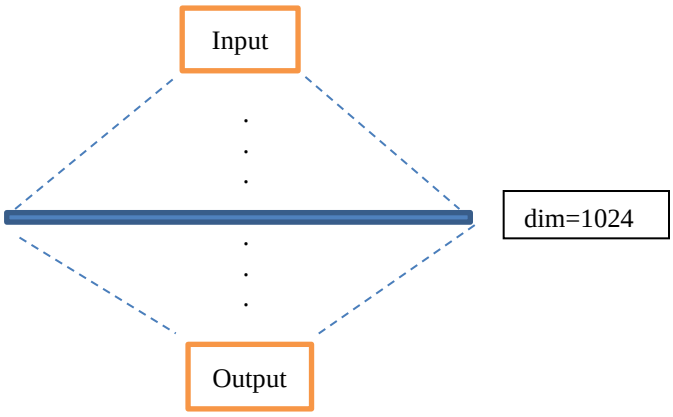
Model	Accuracy
PCA(c=5) + k-means	0.6834
PCA(c=6) + k-means	**0.6842**
PCA(c=7) + k-means	0.6834

IV.III AUTOENCODER + K-MEANS

As another method, autoencoder was used, not to reduce the dimension, but to map the features to a larger space and separate them from each other. At first, the dataset was split into two as training set and test set. The class distributions according to the sets are as follows:

Set name	**Class name**	**# of Samples**
Training Set	Normal	2168
Training Set	MCI	3175
Training Set	Dementia	1657
Test Set	Normal	577
Test Set	MCI	877
Test Set	Dementia	450

"nn” module of the PyTorch was used to create our autoencoder. For that, Leaky Relu was chosen as an activation function and linear transformation was applied. The autoencoder was trained with learning_rate = 1e-3, batch size = 64, number of iterations = 16000. Mean square error was used as loss function. The training loss was equal to 8.54e-05. The autoencoder structure is as follows:



After training, encoded data was given to k-means with number of clusters = 3, random_state = 1, algorithm = ‘full’, init = ‘random’. K-means result with the test set can be shown in the figure:

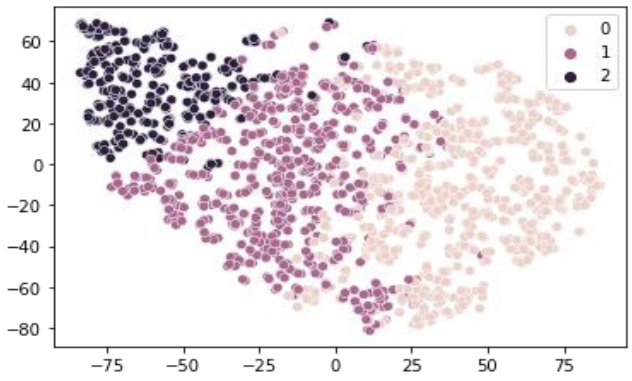


Figure7.1: K-means result

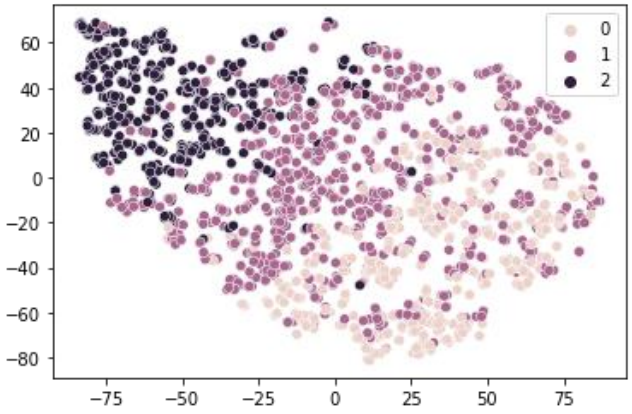


Figure7.2: Ground truth labels

The model accuracy was 0.6749 for predicting 3 classes.

If only the patient had been investigated for dementia, the model with PCA would have given 0.93, and the model with autoencoder would have given 0.91 accuracies.

IV.IV SVM

Until now, 2 different models were created with unsupervised methods. If support vector machines with ‘rbf’ kernel are used as a supervised method (test set/train set = 1/4), the accuracy increases to 0.83 and the classification results can be shown şn the figures below:

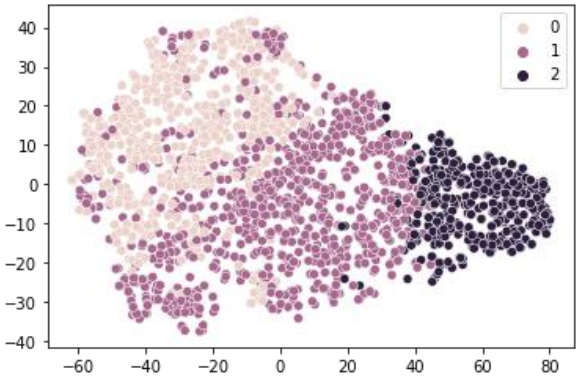


Figure8.1: SVM result

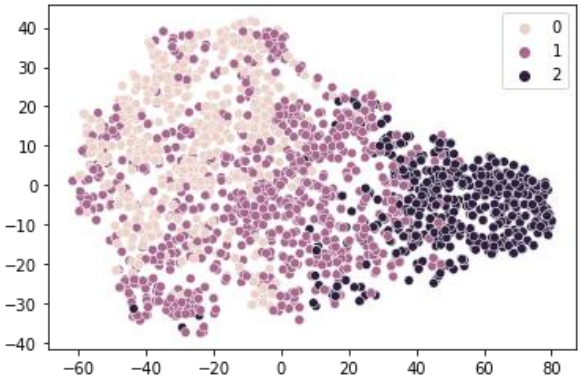


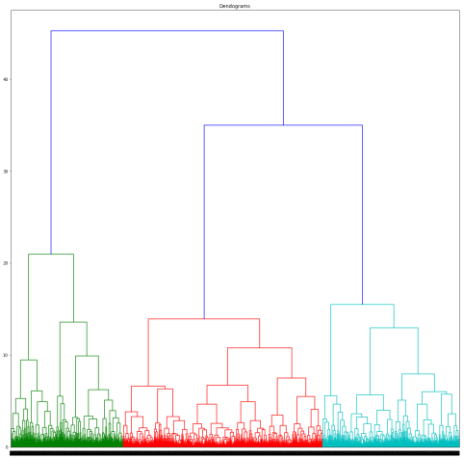
Figure8.2: Ground truth labels

IV.V HIERARCHICAL CLUSTERING

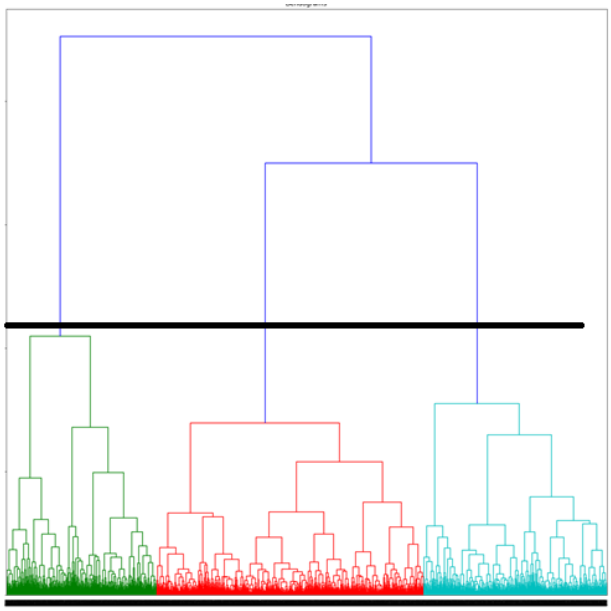
As another method, Hierarchical Clustering and its Ward method were used. In hierarchical clustering, this step also includes finding the optimal number of clusters. Dendrogram was used for this. The Ward method is actually a method that tries to minimize the variance within each cluster. In K-means we try to minimize wcss to plot our elbow method chart, the only difference is almost the same here, we are

minimizing within-cluster variables instead of minimizing wcss.

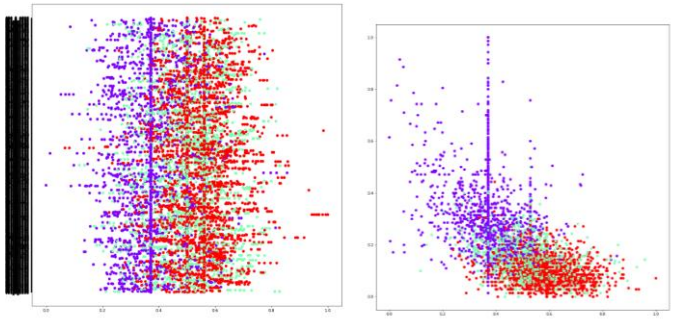
This is the variance within each set:



According to this dendrogram, the y-axis consists of the Euclidean distance between the clusters. If we look at the longest vertical distance in the dendrogram, the vertical line found will be our reference. When we cut the point where this reference intersects with the first horizontal line with a line, we can reach the number of intersections of this line with the vertical lines. This result gives us the optimal number of clusters.



Looking at the final state of the dendrogram, the number of clusters for this dataset is 3. When we used the number of clusters on the algorithm and also used the ward method, we obtained some figures.



Our expectation in these results is the formation of clusters. But when we look at these figures, it is seen that the clusters are intertwined, and the separations do not occur. For this reason, we can reach meaningless results when we use hierarchical clustering. Therefore, hierarchical clustering cannot be used to solve this problem.

V. Obstacles

Although we have been able to obtain some essential feature names through other sources at the moment, it is difficult to examine the features and make the right choice according to the occupancy rate as there are too many features. Also, although we have a dataset consisting of 12741 patient records, a substantial part of the fields are empty, even for a small feature set we use.

Although it is not clearly seen in the final version of our study that we used all the features, outlier values adversely affect the clustering result, as we observed in our comparisons with two features.

While using the PTGENDER feature in our model, “female” values were converted into one, and “male” values were converted into zero since k-means can’t use string values as input. APOE4 values were 0, 1, and 2. The fact that the PTGENDER and APOE4 protein values were not continuous caused the data to be fragmented. Therefore, k-means could not give meaningful results, as shown in figures *5.1, 5.2, 6.1, 6.2*. By removing these features from the feature set, we were able to achieve meaningful results, but this time we could not benefit from two critical features for the diagnosis of Dementia.

VI. Conclusion

As we can see in *Figure5.1* and *Figure5.2* after normalizing the dataset and imputing data according to the three labels, the result of k-means clustering can distinguish dementia patients from others, but we could not say this for the other two conditions. After removing PTGENDER and APOE4 from the feature set, we could get 0.675 accuracy with autoencoder + k-means and 0.684 accuracy with PCA + k-means. Although the removal of APOE4 protein and PTGENDER features enables some MCI patients to be predicted, the model is still not successful enough to distinguish between MCI and Normal.

Since the model cannot accurately distinguish between MCI and Normal, it cannot give a high accuracy at that point. However, it reaches 0.93 accuracy by using PCA + K-means in detecting patients with Dementia.

The MCI phase can be intertwined with the normal state at some points. It is a critical phase as it is difficult to detect. Therefore, good follow-up is required for the early diagnosis of dementia. This clinical situation can also be observed in our experiments.

It is difficult to classify patients because there is a correlation between attributes. Although it can be separated using supervised methods, it seems impossible to achieve similar accuracy levels with linear and unsupervised methods.

VII. Future Work

According to many studies, the presence of Apoe4 protein in the patient is an important factor in the diagnosis of dementia, so the first step in developing the model should be to use gender and apoe4 features in the model. The importance of APOE4 can also be observable from table 4.3. The values of the protein are constantly increasing from normal dementia classes:

0.143 -> 0.302 -> 0.445.

The fact that the difference between Normal and MCI is higher than the difference between MCI and Dementia is also promising for the model to be able to distinguish between MCI and Dementia where it fails.
A similar situation can be observable for the PTGENDER feature as well. The averages from Normal to Dementia:

0.543 -> 0.507 -> 0.542

Although it is not constantly increasing, the fact that the MCI average is separate from the others can make it easier to distinguish from the other two classes.

VIII. REFERENCES

[1] Simeon Spasov, Luca Passamonti, Andrea Duggento, Pietro Liò, and Nicola Toschi. "A Parameter-Efficient Deep Learning Approach to Predict Conversion from Mild Cognitive Impairment to Alzheimer's Disease." *NeuroImage* 189 (2019): 277.

[2] Massimiliano Grassi, Nadine Rouleaux, Daniela Caldirola, David Loewenstein, Koen Schruers, Giampaolo Perna, and Michel Dumontier. "A Novel Ensemble-Based Machine Learning Algorithm to Predict the Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Socio-Demographic Characteristics, Clinical Information, and Neuropsychological Measures." *Frontiers in Neurology* 10 (2019): 2.

[3] Massimiliano Grassi, "A Novel Ensemble-Based Machine Learning Algorithm to Predict the Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Socio-Demographic Characteristics, Clinical Information, and Neuropsychological Measures." 2.

[4] Simeon Spasov, "A Parameter-Efficient Deep Learning Approach to Predict Conversion from Mild Cognitive Impairment to Alzheimer's Disease." 277.

[5] Massimiliano Grassi, "A Novel Ensemble-Based Machine Learning Algorithm to Predict the Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Socio-Demographic Characteristics, Clinical Information, and Neuropsychological Measures." 2.

[6] Wang, Hongfei, Yanyan Shen, Shuqiang Wang, Tengfei Xiao, Liming Deng, Xiangyu Wang, and Xinyan Zhao. "Ensemble of 3D Densely Connected Convolutional Network for Diagnosis of Mild Cognitive Impairment and Alzheimer's Disease." *Neurocomputing* 333 (2019): 147.

[7] Simeon Spasov, Luca Passamonti, Andrea Duggento, Pietro Liò, and Nicola Toschi. "A Parameter-Efficient Deep Learning Approach to Predict Conversion from Mild Cognitive Impairment to Alzheimer's Disease." *NeuroImage* 189 (2019): 277.

[8] Massimiliano Grassi, Nadine Rouleaux, Daniela Caldirola, David Loewenstein, Koen Schruers, Giampaolo Perna, and Michel Dumontier. "A Novel Ensemble-Based Machine Learning Algorithm to Predict the Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Socio-Demographic Characteristics, Clinical Information, and Neuropsychological Measures." *Frontiers in Neurology* 10 (2019): 5.

[9] Sijan S Rana, Xinhui Ma, Wei Pang, and Emma Wolverson. "A Multi-Modal Deep Learning Approach to the Early Prediction of Mild Cognitive Impairment Conversion to Alzheimer's Disease." *2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)*, 2020: 9.

[10] Kuang, Jie, Pin Zhang, TianPan Cai, ZiXuan Zou, Li Li, Nan Wang, and Lei Wu. "Prediction of Transition from Mild Cognitive Impairment to Alzheimer's Disease Based on a Logistic Regression–Artificial Neural Network–Decision Tree Model." *Geriatrics & Gerontology International* 21, no. 1 (2020): 43.

[11] https://www.datacamp.com/community/tutorials/autoencoder-classifier-python