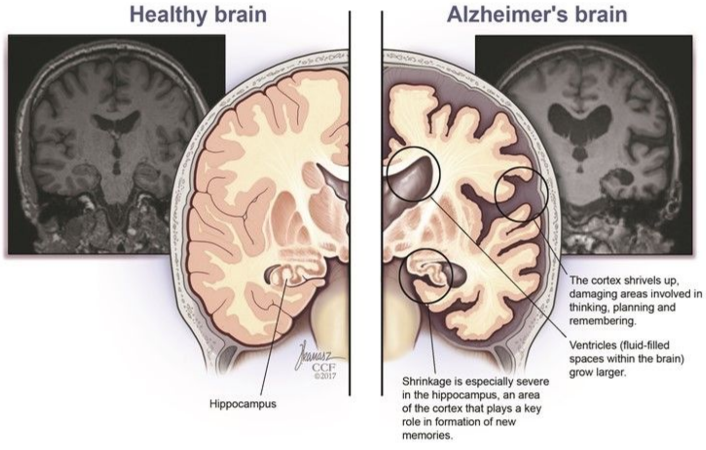
Alzheimer Detection

By- Nimesh Tripathi

Objective

- To detect with good precision whether the patient has Alzheimer’s Disease or not.
- In the MRI scans, the tissue shrinkage is visible between a healthy patient and a patient with Alzheimer's’.



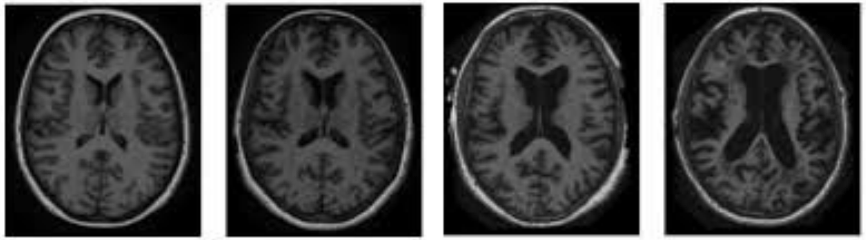
MRI scans (gray) and illustrations (color) show the differences between a brain affected by Alzheimer’s disease and a normal brain.

Objective

- Early onset of Alzheimer’s can be predicted through machine learning algorithms, but are not to the state of art yet.
- Accurately predicting the different stages of dementia is still being worked and improved upon.
- The machine learning and deep learning algorithms, have proved certain success in this field, but its not scalable enough currently.
- Medical practitioners are still searching for this to be downscaled enough (in terms of processing required) so that it is available easily.
- The main goal of this project is to predict the early onset of Alzheimer’s.

Dataset

- The Datasets used in the project are from OASIS.
- The data is longitudinal MRI scans (150 Scans), across different age groups.
- The patients in the dataset, were all right handed and have been scanned at least once, per patient.
- The classification categories in the Data are- 1. Demented 2. Non Demented 3. Converted.
- The converted patients, were detected to be non-demented at the first visit, but were later found to have dementia, on a later visit.



An example of the MRI scans in the OASIS Database

OASIS
OPEN ACCESS SERIES OF IMAGING STUDIES

Dataset: Column Descriptors

The column descriptors in the dataset are:

1. Clinical Dementia Rating
2. Socioeconomic Status
3. Years of Education
4. Mini-mental State Examination
5. Atlas Scaling Factor
6. Estimated Total Intracranial volume
7. Normalize whole brain volume

Dataset: Clinical Dementia Rating

The CDR or Clinical Dementia Rating is used to categorize people into several categories of dementia:

- CDR=0 is equivalent to no dementia
- CDR=0.5 is equivalent to questionable dementia
- CDR=1 is equivalent to moderate cognitive impairment
- CDR=3 is equivalent to severe cognitive impairment

Exploratory Data Analysis

- While trying to work with the data, I referenced Kaggle, for getting better correlation in the data.
- Kaggle had a detailed explanation on Exploratory Data Analysis, which I referenced.(https://www.kaggle.com/ekami66/detailed-exploratory-data-analysis-with-python)
- With EDA, I found the minimum, maximum and average values of each of the column descriptors.
- The effect of this was observed with different models, with the dependencies on different variables.

Methodology: Preprocessing

Missing Values

- The dataset has 8 missing values, which can be solved by two ways.
- The first one is dropping the rows, containing the missing values, however since there are only 150 rows, dropping 8 rows, would be losing 5% of the dataset.
- Hence the missing values were replaced. While replacing the values, certain columns had discrete values and needed to be replaced using the median values.
- They were then grouped and replaced.

Algorithms used

I used different models for testing out which model helped detect the most accurate model. The models I worked with which the accuracy was good-

1.Logistic Regression

2.SVM

Other Methods Used-

Decision Tree

Random Forest

Logistic Regression

- Several Logistic Regression models were tested to get the optimal hyperparameter.
- The hyperparameter tuning was done for obtaining the optimal value for parameter C(in the hyperparameter space), as the regularization parameter.
- Overfitting and underfitting was a major problem, since with large values of the hyperparameter,the model was overfitting.

<u>Logistic Regression</u>

Cross Validation set Accuracy: 80%

Test Accuracy with best parameters: 82%

```
accuracy_list = []
opt_score = 0
opt_var=2.0
folds_no = 10
Regression_1 = LogisticRegression(C=2)
#accuracy
metrics = cross_val_metrics_val(Regression_1, train_x, train_y, var1=folds_no,
                         flag='accuracy')
#model
R = LogisticRegression(C=opt_var).fit(train_x_scaled, Y_trainval)

metric_x = LR.metrics_val(X_test_scaled, Y_test)
y_final = LR.predict(X_test_scaled)
#recall
recall_data_test = recall_metrics_val(Y_test, y_final, label_pos=1)
```

SVM

- For support vector machine, I used 2 kernels, which were - Radial Basis Function(rbf) and the Linear Kernel(LK).
- The different models had different kernels(rbf, LK),degrees, shrinking coefficients etc.
- The different combinations were done to the minimize the classifications.

<u>SVM</u>

Cross Validation set Accuracy: 81%

Test Accuracy with best parameters: 85%

```python
# Replacing Missing Values
data = data.replace(np.NaN, 0)
var = data[1:]
data = data['G']
var_2 = data[2:]

train, test, label_tr, test_labels = train_test_split(var, var_2, test_size=0.4, random_state=3)


kernel_v = SVC(kernel='rbf', C=1.5)
#kernel can be changed to poly, sigmoid, linear, by changing rbf to the other ones in the line above.
```

Decision Tree

- Decision Trees cause the maximum separation,due to division of data into a tree-like structure.
- With its help, classification was done using Gini-impurity.

Gini Impurity = $$\sum_{i=1}^{C} fi(1 - fi)$$

f represents the frequency of a particular label at a particular node x.

<u>Decision Tree</u>

Cross Validation

set Accuracy: 72%

```
# Replacing Missing Values
data = data.replace(np.NaN, 0)
var = data[1:]
data = data['G']
var_2 = data[2:]

train, test, label_tr, test_labels = train_test_split(var, var_2, test_size=0.4, random_state=3)


kernel_v = DecisionTreeClassifier()
```

Performance Comparison

Model	Accuracy(Overall)(%)	Accuracy(with best parameters)(%)
Logistic Regression	82	83
SVM	81	85
Decision Tree	72	72
Random Forest	69	72

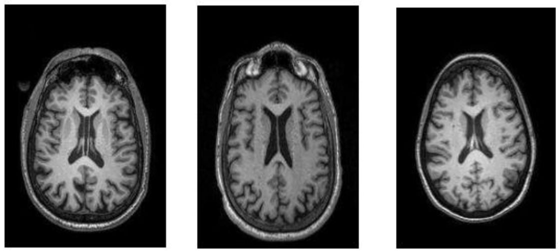
Part 2: Deep Learning(CNN)

```
def run_iter_epoch(data_train, data_val, data_test=0.0, crossval_fold=10):
    model, tuned_value = get_model_tuned_value(arg.fn, d)
    data_train = ADNI_dataset(arg.fn, arg.path, ratio_ds=arg.ratio_ds)
    data_val = ADNI_dataset(arg.fn, arg.path, ratio_ds=arg.ratio_ds)
    for itr in range(arg.iter_itr):
        loss_tr = train(model, train_loader)
        acc_tr = train(model, tuned_value)
        loss_val = eval(model)
        val_tr = eval(x_val1)

        if loss_val < loss_opt:
            save_model(model, loss_opt == 1)
            loss_opt = loss_val
```

Image Recognition of Layers for Alzheimer Detection, using Deep Learning(CNN)

- Here, I attempted to detect whether the scans had Alzheimer’s or not.
- The data used in this part of the project is from ADNI.
- The ADNI clinical dataset comprises clinical information about each subject including recruitment, demographics, physical examinations, and cognitive assessment data. The full set of clinical data may be downloaded in bulk as comma separated values (CSV) files.



(a) Alzheimer’s Disease (AD) (b) Mild Cognitive Impairment (MCI) (c) Normal Control (NC)

Identification Using CNN

A 3-D CNN Model was used with the following parameters:

- For each convolutional layer with 3x3x3 kernels of 64 layers.
- Different Models were tried, but the best was achieved with λ=0.01 and p= 0.2

  (λ= L2 regularization parameter, and p= dropout)

- Two images, which were obtained from the same patient, with a year between were taken as the input.
- Using subtraction of the two images, the decline in the gray matter in the brain is highlighted in red as the output image in the next slide.

Results

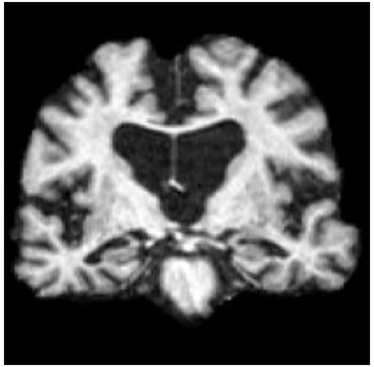
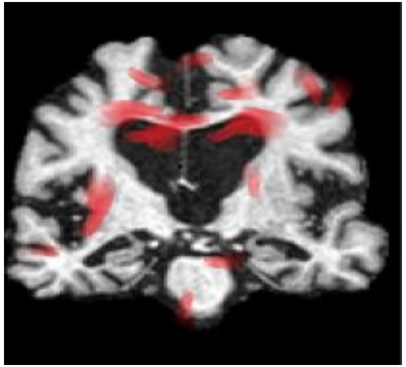
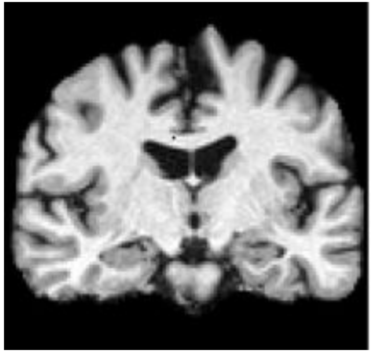
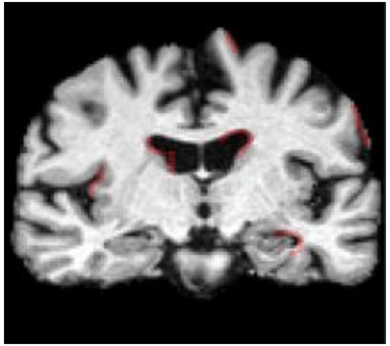


Image 1: Year 1



Subtracted Image: Red shows loss of gray matter





Note: The detection is not optimized due to lack of GPU and processing resources, due to which, the image was segmented, and parts of the image were used for comparison after subtraction. However, for the areas under comparison were detected successfully with an overall accuracy of 62% while the best validation accuracy was 64%.

Papers Referred:

1. **Estimating brain age based on a uniform healthy population with deep learning and structural MRI**Xinyang Feng,a Zachary C. Lipton,b Jie Yang,a Scott A. Small,c,d, and Frank A. Provenzano,, Alzheimer’s Disease Neuroimaging Initiative, Australian Imaging Biomarkers and Lifestyle flagship study of ageing, Frontotemporal Lobar Degeneration Neuroimaging Initiative
2. **Open Access Series of Imaging Studies (OASIS): Longitudinal MRI Data in Nondemented and Demented Older Adults**Daniel S. Marcus, Anthony F. Fotenos, John G. Csernansky, John C. Morris Randy L. Buckner
3. Elaheh Moradi, Antonietta Pepe, Christian Gaser, Heikki Huttunen, Jussi Tohka, Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects.