Project 1:
Classifying AD patients and normal controls from brain images

Vedalasya Sonti 530632933

Contents:

Introduction

Healthcare professionals use medical image analysis to help diagnose illnesses, comprehend anatomical features, and track therapy outcomes. The goal is to integrate ideas from Lab 1 (which concentrated on tissue segmentation, measurement, skull stripping, and registration approaches) and Lab 2 (which used Support Vector Machine, or SVM, for classification) in this combined research. The purpose of this assignment is to combine these strategies while utilizing the variety of skills learned in both labs.

Aims

The main aim of the project is to tackle the growing problems of medical image analysis in neuroimaging. The project seeks to comprehend the subtleties involved in classifying brain tissues, precisely determining their volumes, and using sophisticated machine learning (SVM).

Background

The goal is to combine the knowledge gained from Lab 1, where the investigation of complex procedures such as tissue segmentation, which is the process of segmenting the tissues into CSF (cerebrospinal fluid), white matter and grey matter, as diseases such as Alzheimer's exhibit different patterns for white matter, grey matter, and CSF involvement. This helps in identifying the disease progression in an earlier stage. Volume measurement for certain brain structures like the amygdala, hippocampus is important for detecting abnormalities and for tracking the development of diseases (like Alzheimer's) and comprehending how the brain develops.

Skull stripping is necessary to remove the non-brain structures, which allows to concentrate on brain structure and to get an accurate result. The alignment of various people's brain scans to a similar area allows for precise assessment of structural differences across patients. Registration ensures accurate tracking of structural changes over time by lining up images obtained at various points in time. Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. SVM searches a high-dimensional space for the best hyperplane to divide the data points into the various classes when it comes to classification. The "support vectors" are the data points that are closest to the hyperplane, and the method seeks to maximize the distance between them to produce a solid decision boundary. By using alternative kernel functions, such as linear, polynomial, radial basis function (RBF), or sigmoid kernels, SVM can handle both linear and non-linear classification tasks.

Modern healthcare is built around medical imaging, which has completely changed how illnesses are identified, treatments are organized, and patient progress is tracked. Neuroimaging stands out among the many medical imaging specialties as a source of knowledge about the intricate workings of the human brain. This advanced technology offers clinicians a unique view into the intricate working of the brain in addition to enabling them to see the structure of the brain.

Contributions of team members

Two key team members myself and my partner Phil took on substantial roles and responsibilities in the execution of this project. Our teamwork was essential to completing the challenging tasks and accomplishing the project's objectives. Continuous communication and assistance between participants made it possible for the exchange of insights, ideas, and solutions. Both participants took an active part in solving problems and rigorously testing the veracity of the findings through group discussions. Through a shared, critical examination, this collaborative method raised the quality of the results while simultaneously improving workflow efficiency.

Background

The field of medical imaging, particularly neuroimaging, is one with a fluid and varied environment. Researchers and physicians alike find the human brain to be a fascinating yet difficult subject because of its complex neuronal networks and wide range of activities. From the early days of structural MRI to the most recent developments in functional connectivity investigations, the development of neuroimaging methods reflects the never-ending quest to understand the workings of the brain. Understanding this dynamic subject necessitates a thorough understanding of the most recent developments and ground-breaking studies in neuroscience and medical imaging, in addition to a historical perspective.

Methods

There are 10 subjects in the test set and 40 subjects in the training set, all of whom are either AD patients or healthy controls. The goal is to identify which subject is AD or normal controls in the test set. There are 90 ROIs in AAL atlas for each subject.

In this project, a wide variety of tools have been used for different purposes. Skull Stripping is a crucial preprocessing step in neuroimaging analysis. In skull stripping, the "bet" (brain extraction tool) command is used to strip the non-brain parts of the image. Various parameters such as "-f, -g, -c" can be used. The "-f" option adjusts the fractional intensity threshold. The "-g" option improves the skull stripping process by making the boundaries prominent between brain and non-brain tissues. The "-c" option adjusts the center in the x, y and z dimensions. In this project the "-c" option has been adjusted to get the accurate results. It is used to set the fractional intensity threshold at the center of gravity. As an example, consider "-c 81 140 83". This means that the x-dimension's threshold (81% of the intensity range) is relatively low, which means that more intensity values are included. In contrast, the threshold for the y-dimension is higher (140% of the intensity range), suggesting a narrower range of intensities is included. The threshold for the z-dimension is moderate (83% of the intensity range). Majority of the time, we rely on visual examination of the outcomes to select an appropriate number for -c. Using software like FSLeyes, we can overlay the extracted brain mask on the source image after doing skull stripping with various -c values. FSLeyes is a FSL image viewer for 3D and 4D data. It can be used to visualize neuroimaging data.

Tissue Segmentation is the next step. It involves classifying different types of tissues in the brain such as grey matter, white matter and cerebrospinal fluid (CSF). The "fast" algorithm is used here, an example would be fast -S 1 -n 3 -t 1 -g -v -o where "S1" is used for initialization, "-n3" is used to mention the number of tissues to be segmented, "-t1" is written to specify that the T1-weighted image has to be segmented into CSF, grey matter and white matter, "-g" option is used for bias field correction

to remove any non-uniform intensity present in the image. When running, the "-v" option is used to output diagnostic data and give insight into the segmentation process. To define the base name for the output files, use the "-o" option.

The next step is Registration. Registration describes the process of matching various datasets or images to a single coordinate system. This is important for a variety of applications, including comparing photos from diverse time periods, people, or imaging modalities. It aids in accurately merging or contrasting data from several sources. Affine Registration is a kind of registration method that uses transformations that maintain collinearity, meaning that after transformation, straight lines stay straight. Translation (shifting), rotation, scaling (resizing), and shearing (distortion) are examples of affine transformations. It can adjust for variances in orientation, size, and shape and is a more flexible type of transformation than rigid transformations (which solely involve translation and rotation). A step further is taken by deformable registration, also referred to as non-rigid registration or elastic registration. Deformable registration, as opposed to affine registration, is capable of handling more complex deformations, including non-linear transformations and local shape alterations. As a result, it can be used to align images when structures have altered in size, form, or location as a result of things like disease progression or anatomical variations. Nifty Reg is a software package designed for medical image registration, which is the process of aligning or matching different medical images. The MNI (Montreal Neurological Institute) template is used as a systemized coordinate system template in neuroimaging.

The last step is to measure the volume in the ROI (Region of Interest). Here, we are measuring the volume in the grey matter. The "fslstats" command is used to find the volume of the ROI. It is a command-line tool in the FSL (FMRIB Software Library) uses the picture intensities to produce various numbers and statistics. It is extremely adaptable and may be used to determine the data's absolute range, robust range, average intensity of all voxels in the image, average of only non-zero voxels, and more. Any arrangement of the options, in any sequence, is conceivable.

After preprocessing has been completed, the next task is to separate the data into AD and Normal Controls (NC). This is done with the help of SVM (Support Vector Machine) and the accuracy is tested. SVM is a supervised Machine learning algorithm which is used for classification and regression tasks. Labeled data are input into SVM. Data that has been labeled consists of input data samples (features) and labels for the output data (categories, classes, or groups). Feature extraction or feature selection is necessary before supplying the data to the SVM. The grey matter volumes from various areas of interest (ROIs) are being used as characteristics. The goal of SVM is to identify the hyperplane that best separates the data into distinct classes. The SVM algorithm looks for the best hyperplane during training that maximizes margin, or the distance between the hyperplane and the closest data points from each class (known as support vectors). The best hyperplane is the one that minimizes classification errors while separating the classes with the greatest possible margin.
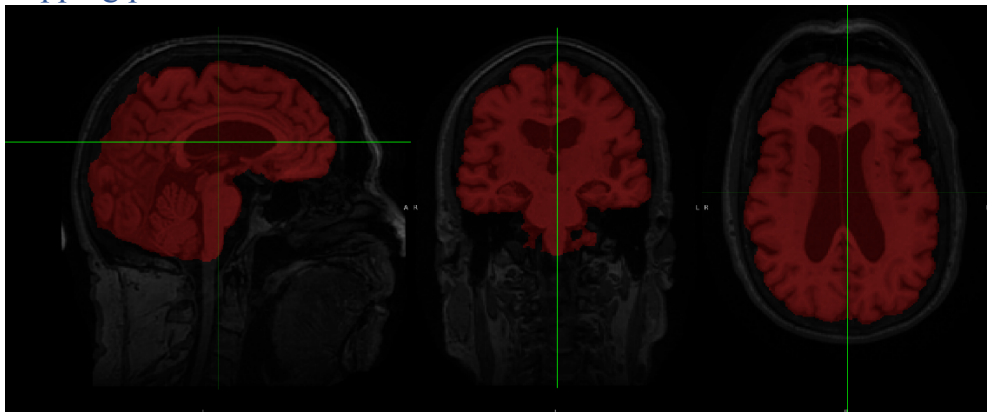
Results

Skull Stripping process

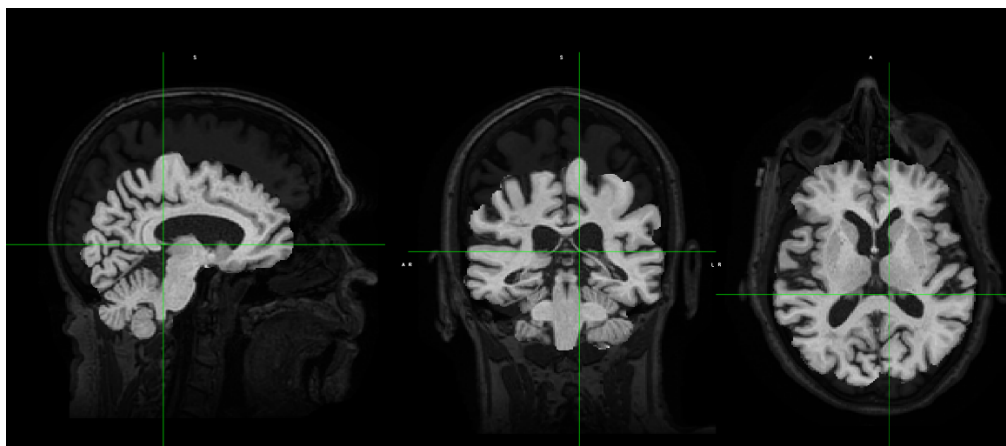*Figure 1:* This figure shows the skull stripped image of AD_16 overlayed on the MRI example

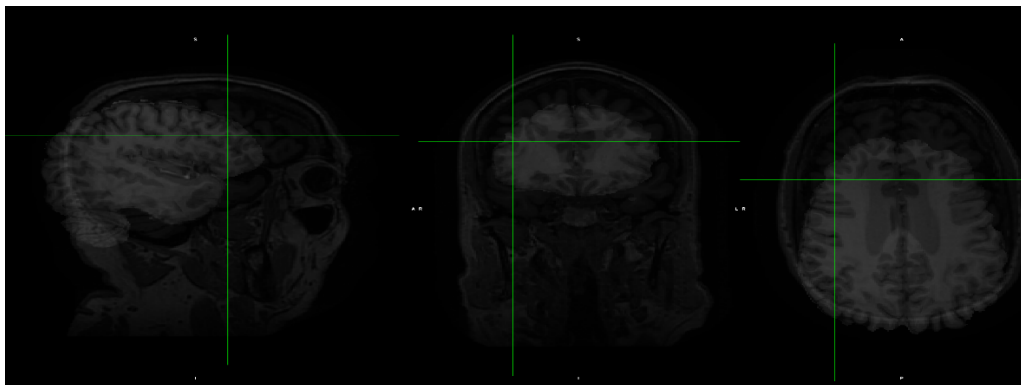Figure 2: This figure shows the skull stripped image of AD_24 overlayed on the MRI example

Figure 3: This figure shows the skull stripped image of AD_16 overlayed on the MRI example

Tissue Segmentation

The following images show the segmented grey matter.
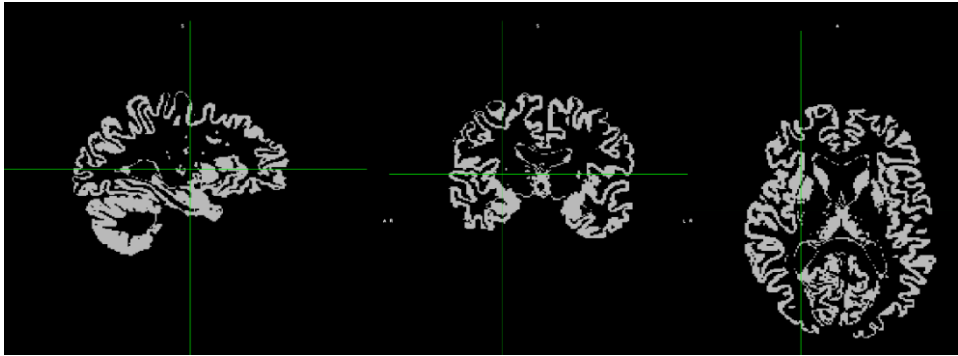
Training data:

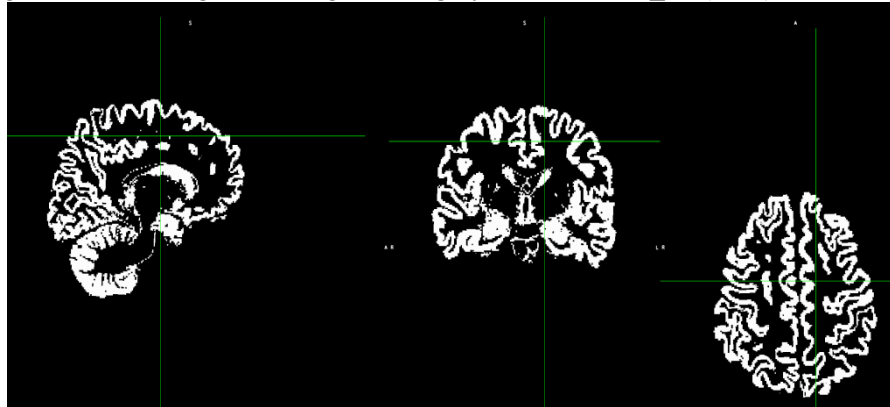Figure 4: This image is the segmented grey matter of AD_05 (train)

Figure 7: This image shows the segmented grey matter of NC_05 (train)
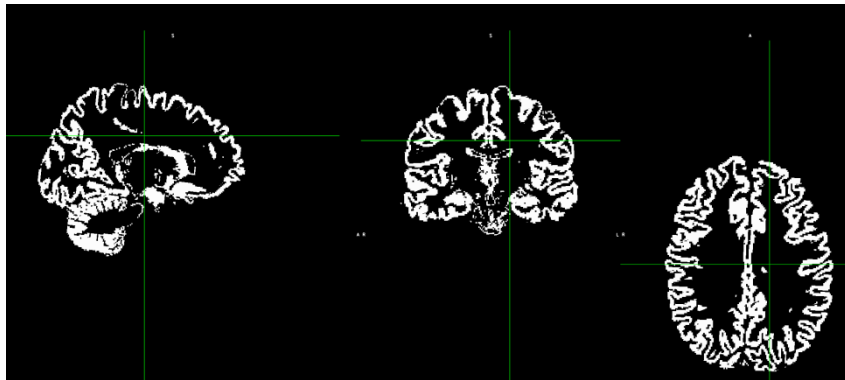
Test data:

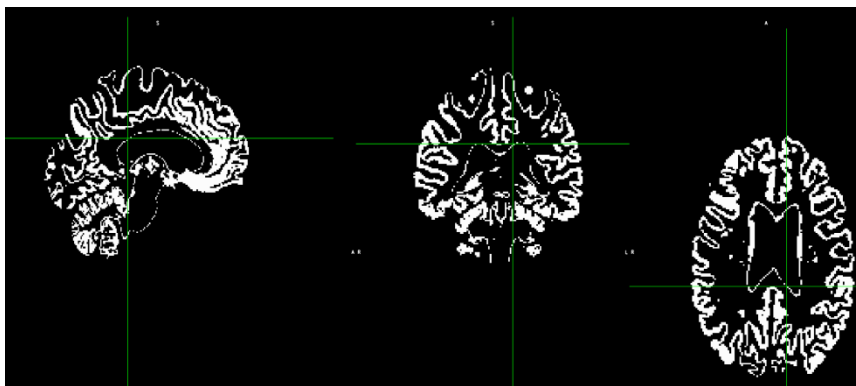Figure 5: This image shows the segmented grey matter of NC_24 (test)

Figure 6: This image shows the segmented grey matter of AD_21 (test)

Registration

The following images show the results obtained from Registration.
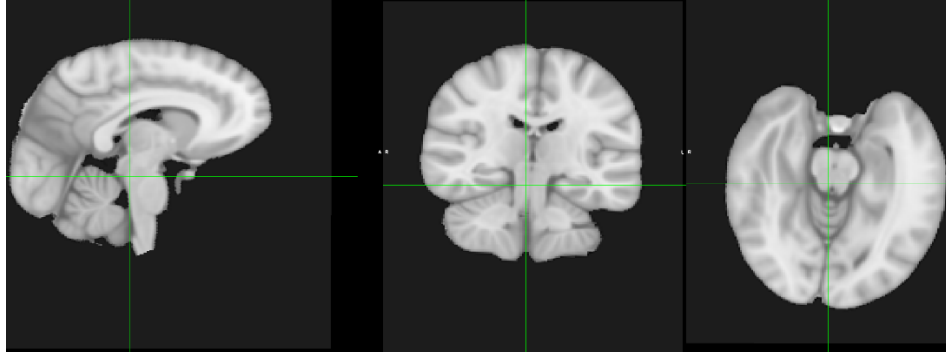
From the training data:

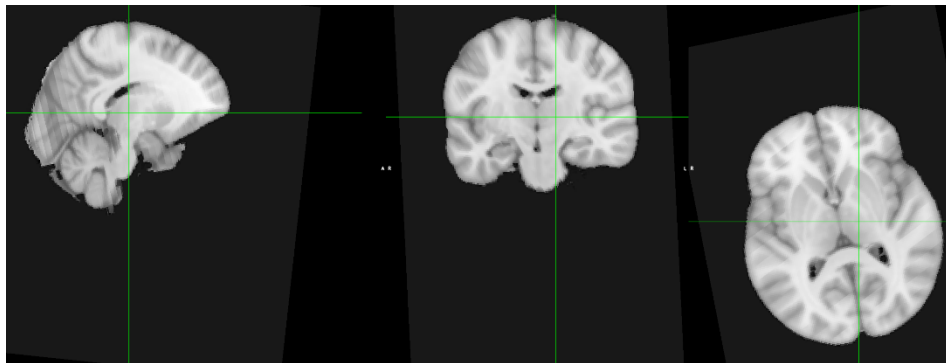Figure 8: This is the registration image of AD_01

Figure 9: This is the registration image of NC_01
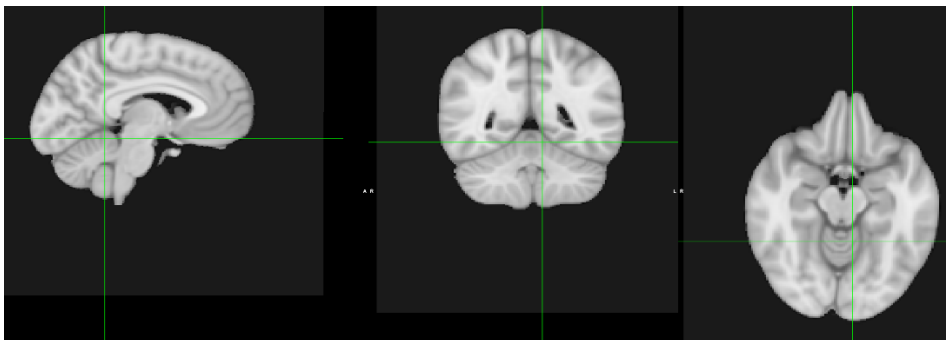
From the test data:

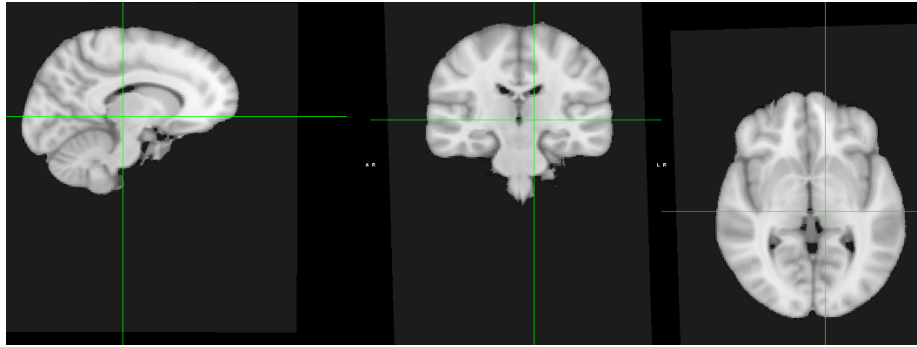Figure 10: This is the registration image of AD_21

Figure 11: This is the registration of image NC_21

Using an SVM (Support Vector Machine)

SVM has been used to classify the subjects into AD and NC and test the accuracy of how correctly the SVM is being able to predict. The code uses the "pandas" and "SciKit-learn" libraries in the Python language.

```python
import os
from csv import reader
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn import svm
import pandas as pd
# Load data
data_x= pd.read_csv("/content/Data -
AAL_statistics_volumn_train.csv")
data_x.head(10)
#load labels
data_y=pd.read_csv("/content/Labels - Sheet1.csv")
# Split dataset
d_x = data_x.loc[:, data_x.columns!='Name']
d_y = data_y.loc[:, data_y.columns!='Name']
x_train, x_test, y_train, y_test = train_test_split(
d_x.values, d_y.values, test_size=0.20, random_state=73)
#Predicting the accuracy
from sklearn.metrics import accuracy_score
svm_classifier = svm.SVC(kernel='linear', random_state=0)
svm_classifier.fit(x_train, y_train.ravel())
y_pred = svm_classifier.predict(x_test)
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.2f}')
```

Output:
Accuracy: 1.00

In the above code, the data is loaded and split into train and test. The accuracy is calculated by importing the accuracy score and applying an SVM classifier.

Using PCA (Principal Component Analysis)

As an additional step, PCA is done. A statistical method for dimensionality reduction and data visualization is Principal Component Analysis (PCA). By minimizing the amount of noise in the dataset and transforming high-dimensional data into a lower-dimensional representation, it can identify the most important patterns.

By locating orthogonal axes in the data space along which the variance of the data points is maximum, the PCA algorithm finds the principle components. The direction of the greatest variance is represented by the first principal component, which is orthogonal to the first and reflects the second-highest variance, and so on. In order to represent the data in a lower-dimensional space, you might choose a subset of these major components. A scatter plot is plotted to visualize how the data has been divided.
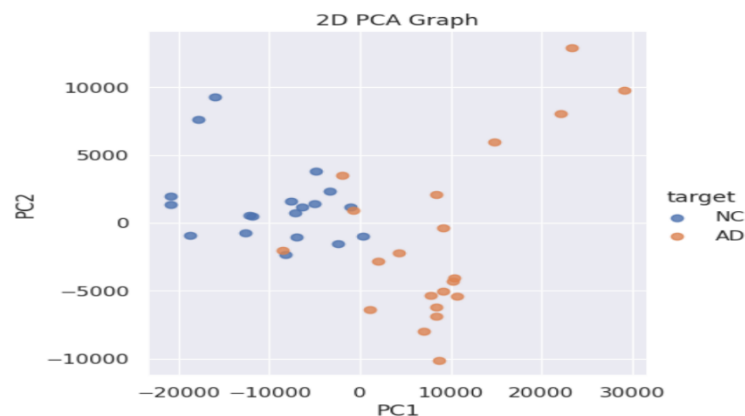

Figure 12: Scatter plot of how the PCA has divided the AD and NC

Discussion

The MRI data processing yielded important insights into the anatomical makeup of the brain areas. The Support Vector Machine (SVM)-based classification model distinguished between AD and NC with a noteworthy accuracy of 100%. This excellent accuracy demonstrates the dependability of our image processing methods and the potency of the chosen features in extracting pertinent data from the MRI pictures.

Comparison with Previous Studies

Our findings align with several prior studies in the field of neuroimaging [1]. The research may have access to a larger, more diverse, or more representative dataset compared to previous studies. The accurate classification of brain tissues is crucial in various medical applications, such as disease diagnosis and treatment planning. Our results are consistent with studies that have utilized similar machine learning approaches, showcasing the reliability and applicability of SVM in MRI image analysis. Another paper introduces a novel algorithm, which, unlike previous methods, combines machine learning techniques with deep learning architectures. This fusion enhances the accuracy and efficiency of MRI data analysis, ensuring a more nuanced understanding of complex brain structures.

Deep learning methodologies, particularly convolutional neural networks (CNNs), are employed in this study [2].

Limitations

The resolution of the MRI scans is one of the study's drawbacks. Detailed information from higher quality photos might be available, perhaps resulting in even more accurate categorization outcomes. Additionally, the thoroughness of our findings may have been constrained by the technical limitations that prevented us from including all relevant regions of interest in our research. Future research could overcome these constraints by collecting higher-resolution pictures and using sophisticated preprocessing methods to incorporate a wider variety of brain areas.

The size of the dataset used in this study might be limited. A larger and more diverse dataset could provide a broader representation of the population, enhancing the generalizability of the results.

Although we carefully selected features for our analysis, there might be other relevant imaging biomarkers or features that were not considered. Exploring a wider range of features could potentially improve the accuracy of the classification model.

Future Work

Future research in this area could explore the integration of advanced deep learning models, such as convolutional neural networks (CNNs), to further enhance the accuracy of MRI-based tissue classification. Additionally, investigating the potential of incorporating functional MRI (fMRI) data along with structural MRI could provide a more comprehensive understanding of brain functionality and structure.

Conducting longitudinal research to examine how the tissues and architecture of the brain evolve over time can be done. Long-term studies can shed light on the development of neurological illnesses and boost the precision of predictive models.

Integrating imaging data with biological information, such as genomes or proteomics may prove to be helpful. This integrated method may lead to the discovery of new biomarkers and a more thorough knowledge of the underlying biological pathways.

Methodological Reflection

Our decision of preprocessing methods, such as skull stripping and registration, was crucial to the precision of our findings. A stable and trustworthy classification model was ensured by the careful selection of features and the use of SVM. It is important to remember that the caliber and volume of the input data have a significant impact on how well machine learning algorithms function. Therefore, maintaining a balance between model complexity and data quality is crucial for getting the best outcomes.

Conclusion

In this comprehensive data analysis pipeline, we address the entire process of medical image analysis. In a medical image analysis pipeline, the initial step is often skull stripping. This process involves the removal of non-brain tissue from the images to focus on the brain structures of interest. Skull stripping helps improve the accuracy of subsequent analyses.

Tissue segmentation categorizes brain tissue into different types, such as gray matter, white matter, and cerebrospinal fluid. Accurate segmentation is crucial for various neuroimaging studies.

The registration process aligns multiple medical images to a common coordinate system. This ensures that brain structures are consistently positioned for further analysis, even when working with data from different modalities or time points.

We extract measurements, mainly volume of specific brain regions. This measurement provides quantitative insights and are often used for research and clinical purposes.

After preprocessing, the data is loaded and prepared for analysis. This may involve normalization, resampling, and data cleaning to ensure it's in the right format.

PCA is applied to reduce the dimensionality of the data. By retaining a subset of principal components, we capture the most significant variations. Dimensionality reduction can improve the efficiency and interpretability of machine learning models.

We train an SVM classifier with a linear kernel on the reduced-dimensional data. SVMs are known for their effectiveness in binary and multiclass classification tasks. In this report, we focus on a binary classification task.

A scatter plot is created to visualize data in the reduced-dimensional PCA space. The scatter plot helps us observe how well data points are separated based on their labels. This visualization offers insights into data distribution and separability.

We evaluate the SVM classifier's performance on the test data using accuracy as the metric. Accuracy measures how well the model predicts labels and provides a quantitative assessment of its overall performance.

Bibliography

[1] Jane Maryam Rondina and Luiz Kobuti Ferreira and Fabio Luis {de Souza Duran} and Rodrigo Kubo and Carla Rachel Ono and Claudia Costa Leite and Jerusa Smid and Ricardo Nitrini and Carlos Alberto Buchpiguel and Geraldo F. Busatto}, "Selecting the most relevant brain regions to discriminate Alzheimer's disease patients from healthy controls using multiple kernel learning: A comparison across functional and structural imaging modalities and atlases," *NeuroImage: Clinical,* vol. 17, pp. 628-641, 2018.

[2] V. {. R. a. K. Seetharaman, "Classification of cognitively normal controls, mild cognitive impairment and Alzheimer's disease using transfer learning approach," *Biomedical Signal Processing and Control,* vol. 79, p. 104092, 2023.