A comparison of classification accuracy between MRI and PET datasets in computer aided diagnosis of Alzheimer's disease

GUSTAV KJELLBERG
HENRIK KÄLVEGREN
En jämförelse av klassifieringssäkerhet mellan MRI- och PET-datamängder vid datorassisterad diagnostisering av Alzheimers sjukdom

GUSTAV KJELLBERG
HENRIK KÄLVEGREN
A comparison of classification accuracy between MRI and PET datasets in computer aided diagnosis of Alzheimer’s disease

GUSTAV KJELLBERG, HENRIK KÄLVEGREN

Date: June 5, 2018
Supervisor: Pawel Herman
Examiner: Örjan Ekeberg
Swedish title: En jämförelse av klassifieringssäkerhet mellan MRI- och PET-datamängder vid datorassisterad diagnostisering av Alzheimers sjukdom
School of Electrical Engineering and Computer Science
Abstract

The number of people suffering from Alzheimer’s disease (AD) is expected to increase rapidly in the coming years. Diagnosing the disease early is key to giving those affected a chance to maintain a higher quality of life. One of the most common ways to detect AD is to visually inspect scans of the patients’ brains. Computer aided diagnosis (CAD) can assist a physician’s judgement when searching for the disease in the brain scans, making the assessment more accurate. Progress has been made in this field throughout the years.

This paper compares the machine learning classification accuracy of AD on images from two different brain scanning procedures - Magnetic resonance imaging (MRI) and Positron emission tomography (PET). Both the MRI and PET datasets contained 60 images. 30 of the images were AD cases and 30 were normal cases in each of the datasets. The images were processed into 1-dimensional signals using Discrete wavelet transform (DWT). The classification accuracy of Support vector machine (SVM), Random forest (RF) and Naive bayes (NB) was obtained and then evaluated using 6-fold cross validation (CV).

This study showed that PET images are more suitable than MR images for diagnosing AD using machine learning classifiers. The highest accuracy for PET and MRI from 6-fold CV was 100% and 90% respectively. The lowest accuracy was 60% for PET and 40% for MRI.
Sammanfattning

Antalet människor drabbade av Alzheimers sjukdom (AD) förväntas öka kraftigt de kommande åren. Att kunna diagnostisera sjukdomen tidigt är nyckeln till att ge dem insjukna en chans att leva ett liv av högre kvalitet. Ett av de vanligaste sätten att upptäcka AD är att visuellt undersöka bilder från hjärnskanningar av patienter. Datorassisterad diagnostisering (CAD) kan hjälpa en läkare i sitt omdöme vid undersökning för sjukdomen i hjärnbilderna, vilket ökar omdömets pålitlighet. Framsteg har gjorts inom området genom åren.

Den här studien undersöker bedömningssäkerheten av maskinlärningsmetoder i klassifiering av AD på bilder från två olika skanningsmetoder för hjärnan - Magnetisk resonanstonografi (MRI) och Positronemissionstonografi (PET). Både PET- och MRI-datamängderna innehöll 60 bilder. 30 av bilderna var AD-fall och 30 var normala fall i varje datamängd. Bilderna processerades till endimensionella signaler med Discrete wavelet transform (DWT). Klassifieringssäkerheten av Support vector machine (SVM), Random forest (RF) och Naive bayes (NB) framtogs och utvärderades därefter med 6-delad korsvalidering (CV).

Studien visade att PET-bilder är att föredra vid diagnostisering av AD med maskinlärningsklassifierare framför MR-bilder. Den högsta bedömningssäkerheten för PET och MRI utifrån 6-delad CV var 100% och 90% respektive. Den lägsta säkerheten var 60% för PET and 40% för MRI.
Acknowledgement

Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. For up-to-date information, see www.adni-info.org.

We would also like to thank the supervisor for this project, Pawel Herman, who continuously helped and guided us throughout this study.
## Contents

1 Introduction .......................... 1
  1.1 Problem statement ..................... 2
  1.2 Outline .................................. 2

2 Background .......................... 3
  2.1 Computer aided diagnosis (CAD) ........... 3
  2.2 Magnetic resonance imaging (MRI) ........... 4
  2.3 Positron-emission tomography (PET) .......... 4
  2.4 Related work in CAD application to MRI and PET ........ 5
  2.5 Technical description & data processing ........ 6
      2.5.1 Discrete wavelet transform (DWT) ........... 6
      2.5.2 Support vector machine (SVM) ............... 7
      2.5.3 Random forest (RF) ......................... 8
      2.5.4 Naive Bayes (NB) ......................... 9

3 Methods .......................... 10
  3.1 Data ................................ 10
  3.2 Data processing ...................... 10
  3.3 Classifiers .......................... 12
  3.4 Validation .......................... 13

4 Results ................................ 14
  4.1 Classification accuracy .............. 14
      4.1.1 MRI CV accuracy ...................... 15
      4.1.2 PET CV accuracy ...................... 16
  4.2 Validation of results .................. 16

5 Discussion .......................... 19
  5.1 Discussion on data .............. 20
  5.2 Discussion on method ............. 21
6 Conclusion and future work 22

Bibliography 23
Chapter 1

Introduction

Since the first publication regarding CAD in 1967 the interest in the field has gradually grown. After the 1980 there has been a significant increase of interest in the field [9]. Today CAD can be used for a variety of areas, one of those is detection of AD. [10,23].

One way of identifying abnormalities within the human body, which may indicate a disease, is to look at biomarkers. Biomarkers are signs within the body that can be measured accurately and reproducibly. They have the advantage of giving away diseases in a very early stage [13]. There are three well-established biomarkers for AD: amyloid beta, tau and p-tau which are found in the cerebrospinal fluid (CSF). The quantity of these substances found in the CSF may reveal if the patient is affected by AD. However, the procedure is invasive, requiring a lumbar puncture to collect the fluid, which puts patients at discomfort [20].

According to the World Health Organisation (WHO) there were 50 million cases of dementia in 2017 of which 60-70 % is caused by AD. According to the same report the projected amount of cases in 2030 is 82 million [22]. CAD can be an aid for radiologists when conducting a diagnosis for a patient. CAD can help to identify potential AD and does not suffer from any of the human errors such as fatigue and distraction [9].

Visual diagnosis by means of inspecting brain images is one of the methods used to detect AD in patients. The most common neuroimaging technique is MRI due to its wide availability in hospitals. Another common technique is PET. Both MRI and PET are able to produce images with enough detail to be able to diagnose AD. However, a problem with brain images of elderly people is that it can be difficult to
differentiate between normal changes in the brain and changes caused by the disease, as the visual differences may be minimal. CAD is used to increase the detection of these abnormalities [12, 23].

1.1 Problem statement

AD is a growing health problem and is the only top ten cause of death in America that can not be cured [1]. CAD can be an aid for radiologists when conducting a diagnose for a patient. CAD can help to identify potential AD and does not suffer from any of the human errors such as fatigue and distraction [9]. Therefore, it is of interest to see which of the neuro-image types that yield the highest classification accuracy using machine learning (ML).

Which of PET and MR neuro-images give the best result in identifying AD using ML classifiers?

1.2 Outline

In Chapter 2 Background, background on AD, CAD, MRI and PET is presented followed by background regarding the used ML methods. In Chapter 3 Methods, information about the data is presented as well as a description of how the data is used with the three ML methods. In Chapter 4 Results, the result of the study is presented, showing the classification accuracy based on 6-fold CV for MRI and PET. In addition, an in depth comparison of the results is made using a two-way ANOVA test and Tukey’s HSD test. Chapter 5 Discussion concerns CAD in general, the approach of the study as well as data used. Finally, the conclusion and future work is presented in Chapter 6 Conclusion.
Chapter 2

Background

AD is a type of dementia, a syndrome of cognitive impairment severe enough to affect daily life. There are approximately 50 million people suffering from dementia of which AD is the most common form making out 60-70 % of the cases. There is no absolute way to determine if a person suffers from AD, nor does it exist a cure for AD [2, 3, 22].

One theory is that AD is caused by damage to brain cells, more specifically due to communication problems between the brain cells due to high levels of certain proteins inside and outside of the cells. It is not unusual that AD is assumed to be a condition caused by age, which is not correct. Though, it is an evolving disease and is usually most noticeable at higher ages. One of the earliest detectable symptoms are bad short term memory and memory loss, caused by damaged brain cells in the hippocampus, as this is one of the first places in which brain cells start to suffer from bad communication [2, 3, 22].

Detection of AD before symptoms occur is very hard due to the fact that it is very difficult to distinguish certain properties of AD in brain scans. Though, an early treatment of AD is crucial to achieve good patient care for affected persons, this making the field of study very important [2, 3, 22].

2.1 Computer aided diagnosis (CAD)

CAD is valuable for radiologists to serve as a second opinion in both diagnosis and identification. Human factors such as the visual system’s restrictions, i.e. distinguishing abnormalities with the human eye and fatigue restricts identification of abnormalities in medical im-
ages through visual inspection, therefore, CAD can be of help. CAD has developed significantly the past 20 years, it has been researched since the 1980’s and is used to support radiologists in many different fields of which AD is one.

There are two types of approaches, CADe and CADx. CADe is used to identify suspicious areas for the radiologists to inspect, classify and diagnose. CADx goes further and gives a likelihood for classification of such an area.

A CAD method said to have a certain specificity, sensitivity and classification accuracy is highly dependent on the database and dataset that have been used.

### 2.2 Magnetic resonance imaging (MRI)

MRI is a non-invasive method to obtain images of soft tissue within the human body. Instead of creating images using harmful x-rays like a normal computed tomography scan (CT), MRI uses strong magnets. By measuring the energy of the body’s protons within the magnetic field an image is created. MR images can be used for diagnosis and disease detection. Being good at separating white and grey matter in the brain, MR images can be used to identify AD, as one sign of AD is reduced volume of grey matter in the brain.

![Figure 2.1: MRI images showing an arbitrary slice of an AD subject.](image)

### 2.3 Positron-emission tomography (PET)

Contrary to MRI, PET is an invasive method that through injection or inhalation of a radiopharmasutical measures the energy of the positrons
emitted when the isotope decays. A series of scans are conducted as the isotope gradually decays, generating an approximate image of the brain [21]. Like MRI, PET has proven to work well in identifying AD [17, 21].

Figure 2.2: PET images showing an arbitrary slice of an AD subject.

2.4 Related work in CAD application to MRI and PET

There are many studies using either PET and MR images for CAD in with the objective to indentify and classify AD. Though, there are significantly fewer studies comparing the two image types for classification. In 2011 Illìan et al. [10] conducted a study using FDG PET images and SVMs to classify AD. The study is based on 401 ADNI patients and Illìan et al. obtained a classification accuracy of 88.64% [10]. Further on, Ramirez et al. [23] conducted a study in 2013 also using SVMs and PET images. Their results showed a classification accuracy of 90.38 % with sensitivity of 93.10 % and specificity of 86.96 % using RBF kernels.

Lahmiri et al. [14] in 2013 achieved better classification accuracy using MRI than Ramirez et al. obtained using FDG PET. Lahmiris et al. study used 93 MRIs obtained from ADNI and also used SVMs to classify AD. By exclusion of some atypical AD images they obtained an impressive classification accuracy of 100%, without discarding those images they obtained 99.18±0.01 [14]. Dukhart et al. [6] conducted a study using both MRIs and PET images in 2013. Using images obtained from two different databases (ADNI and Liepzieg) they did not only compare classification accuracy between the image types but also
between the databases. Also they compared separate classification accuracies, that being MRI and PET separate classification accuracy to the accuracy obtained from combining the image types. Dukhart et al. classified the data using many different processing techniques. The technique of interest to this study Volume of interest (VOI) based SVM classification yielded 87.5 %, 80.4 % and 85.7 % for PET, MRI and PET & MRI. Their conclusion was that independent classification of PET images or combined PET and MRIs yielded high accuracy while independent classification accuracy of MRIs was significantly lower [6].

In 2014 Matsunari et al. [19] conducted a study comparing the classification accuracy of AD detection in FDG PET and MRI respectively using SVMs. The study was based on 154 AD patients and 154 normal cases and Matsunari et al. concluded that FDG PET yielded higher classification accuracy than MRI.

2.5 Technical description & data processing

2.5.1 Discrete wavelet transform (DWT)

Discrete wavelet transform (DWT) is a function that transforms a signal into a more efficient representation of itself by extracting the details that provide the most significant changes. The function is defined as:

\[ X(a, b) = \int_{-\infty}^{+\infty} x(t)\Psi_{a,b}(t)dt \]

where \(a\) is a scale factor, \(b\) is a translation factor, \(x(t)\) describes the signal as a function of time and \(\psi(t)\) is an analyzing function defined by:

\[ \Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left( \frac{t - b}{a} \right) \]

In the above equation, \(\psi\) is a mother wavelet which definition depends on the type of DWT that is used. One important wavelet is the Haar wavelet. The Haar wavelet is commonly used because of its simplicity [5]. A wavelet is a waveform which has zero mean and has limited duration [18].
In the Haar DWT the wavelet used is defined as [24]:

\[
\psi(t) = \begin{cases} 
1, & \text{if } 0 \leq t < 1/2 \\
-1, & \text{if } 1/2 \leq t < 1 \\
0, & \text{otherwise}
\end{cases}
\]

2.5.2 Support vector machine (SVM)

SVM is a supervised ML method used for classifying data to belong to one of multiple classes (most commonly two). The idea is that each class in general contains some unique features that differentiate it from the other(s). Geometrically, SVM considers every feature of the input data as its own dimension in space. This means that the number of dimensions increases with the number of features.

Let \( m \) be the number of training data samples and let each sample \( i \) be represented by an \( n \)-dimensional vector \( x_i \in \mathbb{R}^n \), where \( n \) is the number of features. Then, a training dataset \( X \) consists of a set of \( m \) \( n \)-dimensional vectors:

\[
X : \{x_1, \ldots, x_m\}
\]

Every vector \( x_i, i = 1, \ldots, m \) belongs to one of the different classes and the task for SVM is to determine which one. Since the training data is labeled, SVM has the knowledge of which points in the \( n \)-dimensional space belong to each class. The objective of the model is to find a hyperplane which optimally segregates the classes in the space. Each class exists on their own side of the hyperplane. This is achieved by fitting the hyperplane in a way that maximizes its margin to the closest point, called support vectors, of either class [15].

The hyperplane is the threshold for the classification. Classification of previously unseen samples is achieved through observation of where the new sample vectors will point in space [15].

Linear separation of the classes is possible in certain cases, but not all; it depends on the spatial location of the points belonging to each class. Non-linear transformations, kernels, can be utilized to map the input space into higher dimensions where linear separation is more feasible [15].
Figure 2.3: This figure demonstrates the usage of a kernel. (a) shows a 2-dimensional input space, while (b) shows the same data after it has been transformed into 3 dimensions. This is the core principle of a kernel. After the data has been transformed it can become significantly easier to fit a hyperplane between the data points belonging to the different classes as the figure illustrates.

Commonly used kernels are [23]:

- Polynomial $K(x, y) = (\gamma(x \cdot y) + c)^d$
- Radial basis function (RBF) $K(x, y) = e^{-\gamma||x-y||^2}$
- Sigmoid $K(x, y) = \tanh(\gamma(x \cdot y) + c)$

where $\gamma$ and $c$ are constants and $d$ is the degree of a polynomial [27].

2.5.3 Random forest (RF)

RF is a method that uses a collection of Decision trees that are combined into a strong classifier. Working with RF involves two types of randomness: (1) generating bootstrap replicas of the data and (2) deciding predictor to split by. A tree that has a very strong predictor will without randomness likely split at the same predictor, making the collection of trees similar and thus the predictions similar. The method of randomness is referenced as decorrelation. The decorolated trees will be less variable and give better results than those obtained by i.e bagged trees [8, p.320,321].
2.5.4 **Naive Bayes (NB)**

NB is a machine learning method that classifies an observation to the most likely class based on conditional probability. For an observation \( x \) this is achieved by calculating the posterior of each class \( k_i \) in a set of given classes \( C \). This is achieved using Bayes’ *theorem* [7, 8].

\[
p(k|x) = \frac{p_k(x|k)p(k)}{\sum_{k' \in C} p(x|k')p(k')}
\]

Further on, calculations of the mean vector \( \mu_k \) and co-variance matrix \( \Sigma_k \) for classes \( k \) is calculated as; [7, 8]

\[
\mu_k = \frac{\sum_{i|c_i=k} x_i}{N_k}
\]

where \( N_k \) is the number of classes.

\[
\Sigma_k = \frac{1}{N_k} \sum_{i|c_i=k} (x_i - \mu_k)^T (x_i - \mu_k)
\]

Instead of calculating the entire Covariance matrix, which could be very large, is it possible to only calculate the diagonal. This is possible assuming that all feature dimensions are independent and there is no off diagonal co-variance elements. This is called *Naive Bayes’ assumption* and results in: [7, 8]

\[
\Sigma_k(m, m) = \frac{1}{N_k} \sum_{i|c_i=k} (x_i(m) - \mu_k(m))^2
\]
Chapter 3

Methods

3.1 Data

All images used for this thesis were obtained from ADNI. In order to have as consistent data as possible all images are of the same dimension and type. 60 MRI images were collected of which 30 were AD cases and 30 normal cases with dimension 192x192x160 pixels and slice thickness 1.2 mm. The images are all part of ADNIs project “ADNI1/GO Month 12 (M12)” and are of file format NIfTI (.nii). 60 PET images were collected, of which 30 were confirmed AD cases and 30 were confirmed normal cases. All images have the dimensions 160x160x96 pixels and a slice thickness of 2.0 mm. Each image is constructed from 96 DICOM (.dcm) files.

All tests were run on the same slice for each image type. To identify which slice to use all slices were tested and the one that generated the best classification accuracy was chosen. Therefore, the slice used differ between MRI and PET images, though, not in between the image type.

3.2 Data processing

The images were obtained from ADNI by manually selecting the images which have the format previously described. The same following procedure was used for both PET and MRI except for the usage of different slices and level of DWT. First, every image in a dataset was labeled as AD or NL depending on whether the image was a case of AD or a normal case. Then it was split into training and testing data
using 6-fold CV.

Second, 2-dimensional DWT Haar of level 4 was applied to vertical slice 82 of each PET image. For all MR images, DWT Haar of level 3 was applied to vertical slice 99 instead. Each image yielded a matrix of DWT approximation coefficients. Third, the obtained coefficient matrices from each image were individually unfolded row by row into one 1-dimensional column vector each. Then every column vector belonging to the training dataset was concatenated into one 2-dimensional training matrix, and the same was done to all column vectors belonging to the testing dataset into a testing matrix. Fourth, the training matrix was used as input data for the classifiers SVM, RF and NB to train them.

Finally, the testing matrix was used as input data for the classifiers to test their accuracies by comparing their predictions to the actual data labels. Accuracies were obtained using 6-fold CV.

Step 1: Apply 2-dimensional DWT Haar level 4 for PET slice 82 and level 3 for MRI slice 99 for each such image in the training dataset and the testing dataset separately to retrieve their approximation coefficient matrices.

Step 2: Unfold each coefficient matrix for each image into one 1-dimensional column vector.

Step 3: Concatenate the column vectors belonging to the training dataset into one matrix $M_{train}$, and concatenate the column vectors belonging to the testing dataset into another matrix $M_{test}$.

Step 4: Input $M_{train}$ and the image class labels to the classifiers to train the models.

Step 5: Input $M_{test}$ to the classifiers to predict the classes.
3.3 Classifiers

The expected output for each image from the classifiers SVM, RF and NB was a label AD or NL (normal), stating which class the classifiers predict the image belongs to. In this report all classifiers were modules from scikit-learn: SVM [27], RF [26] and NB [25]. The settings used for the classifiers are described in table 3.1.

<table>
<thead>
<tr>
<th></th>
<th>kernel</th>
<th>C</th>
<th>degree</th>
<th>n_estimators</th>
<th>distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>rbf</td>
<td>8.0</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>NB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gaussian</td>
</tr>
</tbody>
</table>

Table 3.1: This table shows the settings used for the classifiers which are available in the scikit-learn modules. Empty cells indicate that the setting is not applicable to the classifier.
3.4 Validation

In order to validate the implemented methods 6-fold CV was used. All images obtained from ADNI (60 AD and 60 Normal) were used in order to validate the result. First, images are processed as described in section 4.1.2 until step 4. A matrix consisting of column vectors where each column represent the 1-D signals for each image obtained from the DWT. 6-fold CV enables scoring of the classification, therefore, the validation is conducted for each classification method individually. Therefore there are 6 scores for each classifier and image type in the table below. Further on, a two-way ANOVA test was used described in section 5.3. If the difference between levels (rows), variables (columns) or in the interaction between levels and variables a post-hoc Tukey’s Honest Significant Difference (HSD) test was conducted to verify the difference.
Chapter 4

Results

This chapter presents the results obtained in this study - the classification accuracy obtained from the 6-fold CV followed by a validation of those results.

4.1 Classification accuracy

After processing the MR and PET images, each processed image was predicted to belong to one of two classes: AD or normal. The classification accuracies were obtained from the classifiers SVM, RF and NB using 6-fold CV. Tables 4.1 and 4.2 present six different classification accuracies per classifier and image type, which is the expected format of the output from 6-fold CV. Accuracy is defined as the percentage of times a classifier predicts an image to belong to the correct class.
### 4.1.1 MRI CV accuracy

<table>
<thead>
<tr>
<th>Classification accuracy (%)</th>
<th>Trail No. 2</th>
<th>Image type</th>
<th>No. of AD</th>
<th>No. of normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM 50.0</td>
<td>1</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>SVM 40.0</td>
<td>2</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>SVM 90.0</td>
<td>3</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>SVM 70.0</td>
<td>4</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>SVM 60.0</td>
<td>5</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>SVM 70.0</td>
<td>6</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF 60.0</td>
<td>1</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF 60.0</td>
<td>2</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF 70.0</td>
<td>3</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF 70.0</td>
<td>4</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF 60.0</td>
<td>5</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF 70.0</td>
<td>6</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB 70.0</td>
<td>1</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB 60.0</td>
<td>2</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB 90.0</td>
<td>3</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB 60.0</td>
<td>4</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB 60.0</td>
<td>5</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB 70.0</td>
<td>6</td>
<td>MRI</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 4.1: Accuracies obtained from 6-fold CV on the MRI data set for each classifier. Green coloured rows identifies the highest accuracy while red coloured rows identifies the worst accuracies.
4.1.2 PET CV accuracy

<table>
<thead>
<tr>
<th></th>
<th>Classification accuracy (%)</th>
<th>Trail No. 2</th>
<th>Image type</th>
<th>No. of AD</th>
<th>No. of normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>90.0</td>
<td>1</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>SVM</td>
<td>100.0</td>
<td>2</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>SVM</td>
<td>60.0</td>
<td>3</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>SVM</td>
<td>70.0</td>
<td>4</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>SVM</td>
<td>90.0</td>
<td>5</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>SVM</td>
<td>90.0</td>
<td>6</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF</td>
<td>90.0</td>
<td>1</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF</td>
<td>100.0</td>
<td>2</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF</td>
<td>60.0</td>
<td>3</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF</td>
<td>70.0</td>
<td>4</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF</td>
<td>90.0</td>
<td>5</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>RF</td>
<td>90.0</td>
<td>6</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB</td>
<td>90.0</td>
<td>1</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB</td>
<td>100.0</td>
<td>2</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB</td>
<td>60.0</td>
<td>3</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB</td>
<td>60.0</td>
<td>4</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB</td>
<td>90.0</td>
<td>5</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>NB</td>
<td>90.0</td>
<td>6</td>
<td>PET</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 4.2: Accuracies obtained from 6-fold CV on the PET dataset for each classifier. Green coloured rows identifies the highest accuracy while red coloured rows identifies the worst accuracies.

4.2 Validation of results

In order to conclude if there actually is a difference between using PET and MRIs to identify AD using ML-classifiers it is necessary to test the results. For this study the null-hypotheses says *There is no difference in classification accuracy of AD when using PET or MRIs*. Table 4.1 in section 4.1.1 and table 4.2 in section 4.1.2 present the classification accuracy obtained using 6-fold CV on the MRI and PET data sets respectively, for the three different classifiers; SVMs, RF and NB. A two-way ANOVA test yielded the following P-values; for image types: 0.0011, for classi-
fiers: 0.9600 and 0.8386 for interaction. As $P < 0.005$ for PET and MRI, this is considered significant, meaning that the null-hypotheses can be discarded [4]. Further on, the yielded $P$-value for classifiers is high which indicates that there is no significant difference in using different classifiers. Also, the $P$-value for interaction is high, indicating that there is no significant difference in any specific combination of classifier and image type. Therefore, the only feature causing a significant difference in classification accuracy is the image type.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image type</td>
<td>0.26694</td>
<td>1</td>
<td>0.26694</td>
<td>13.09</td>
<td>0.0011</td>
</tr>
<tr>
<td>Classifiers</td>
<td>0.00167</td>
<td>2</td>
<td>0.00083</td>
<td>0.04</td>
<td>0.96</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.00733</td>
<td>2</td>
<td>0.00361</td>
<td>0.18</td>
<td>0.8386</td>
</tr>
<tr>
<td>Error</td>
<td>0.61167</td>
<td>30</td>
<td>0.02039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.8875</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: The generated ANOVA table from the two-way ANOVA test. The column to the left shows source; Image type, Classifier, Etc. The right most column show the P-value obtained for each source.

As the yielded $P$-value implies that there is a significant difference between the two image types, a post-hoc test was conducted. The post-hoc test used was Tukey’s HSD. The test yielded $P$-value = 0.0011 for image types.
Figure 4.1: Confidence intervals obtained from the post-hoc Tukey’s HSD test. The figure shows how the confidence intervals do not overlap and therefore show that the two image types PET and MRI yield significant different results. The vertical axis is the group, 1 = MRI and 2 = PET. The horizontal axis represent the confidence interval.
Chapter 5

Discussion

The results used in this study were obtained using 6-fold CV. MR images resulted in relatively low classification accuracy, sometimes lower than random i.e. 50 % but in one case as high as 90 %, see table 1 in section 4.2. Comparing the result for AD classification in MR images with other studies this study fails to achieve results close to those of e.g. Lahmiri et al. [14], whom obtained 100 % classification accuracy using MR images and SVMs. Though, the highest overall accuracy this report obtained for MR images was from using a SVM classifier which yielded 90 %. This result is better to that obtained by Dukhart et al. [6] whom presented a classification accuracy of 87.5 % with SVMs.

On the other hand, PET images yielded significantly higher accuracies with some scores being as high as 100 %, see table one in section 4.2. Comparing the results with others such as Illian et al. [10] whom obtained 88.64 %, we find these results sufficient enough to be conclusive.

By simply inspecting the results that MRI and PET yielded, it is a significant difference between the accuracies. In order to conclude that there actually is a difference in using the different image types a two-way ANOVA test was conducted. The ANOVA test showed that there was a significant difference between classification accuracy obtained from PET and MRI. In order to verify this results further a post-hoc tested was conducted which yielded the same result.

Since the results obtained from MR images is much lower than what has been previously reached, there is a possibility that there are flaws in either the method or the data used in this study. Therefore, it can be debated if the results are too poor to be conclusive.
5.1 Discussion on data

ADNI had a recommended dataset to work with for MRI. All of the MRI data used in this study is from ADNI1/GO Month 12. The MRI data was already processed. Each image had the description MPR; GradWarp; B1 Correction; N3; Scaled or MPR-R; GradWarp; B1 Correction; N3; Scaled, describing how they were processed. The difference being that the first type is derived from MP-RAGE (MPR) and the second type is derived from MP-RAGE-REPEAT (MPR-R).

There did not exist any recommended dataset for PET in the ADNI database. There were no guidelines for what data that would be suitable. An arbitrary choice was made to pick images that there were a larger quantity of and that were non-blurry. The PET data chosen was already processed. Each image had the description AV45 Coreg, Avg, Standardized Image and Voxel Size, describing how they were processed. However, not all images had undergone completely the same procedure. The majority were derived from image type AV-45 (AC) AV45, whereas some were derived from ADNI2 AV45 (AC) and ADNI3 AV45 (AC). There did not exist 30 AD and 30 normal images that were acquired using exactly the same procedure. Therefore the decision had to be made to include such other images. The other images were manually compared to the AV-45 (AC) AV45 images and selected if they looked similar. This decision was justified because increasing the overall size of the dataset increased the general classification accuracy.

The reason why 30 AD images and 30 normal images were chosen both for MRI and PET is because there was a lack of data. This was a limitation that likely affected the performance of the classifiers.

Additional images could have been included, but it is not probable that they would have increased the classification accuracy, since those images were not similar enough to the data which was mainly used. Therefore those additional images would likely only have contributed with unnecessary variance, decreasing the importance of the pixel features of the other data.

One thing that might have affected the authenticity of the comparison between MRI and PET, is that the MRI scans and PET scans were not acquired from the same patient. This is because there were not enough patients who had undergone both the MRI and the PET procedures. Therefore, the datasets did not contain scans of the same brains. It could be argued that some of the patients’ brains naturally
would show a larger indication for AD independently from the scanning procedure used. In this case, it could have been true for the patients scanned with PET, because of the higher classification accuracy for PET.

5.2 Discussion on method

In order to conduct this study an initial method was planned based on other studies such as Lahmiri et al. [14] and Illian et al. [10]. Over time it became clear that the initial method had several flaws and yielded poor results. Therefore, several attempts to improve the results by changing the method were made. First, instead of working with an arbitrary slice or the entire brain, each slice was individually tested for PET- and MR images respectively. Using the vertical slice number 99 for MR images and number 82 for PET images gave a significant improvement. Moreover, an attempt to improve the results generated by a principal component analysis (PCA) were made. As no improvement showed when changing different settings it was removed entirely. This was a substantial change to the method resulting in classifications accuracy that were high enough to be conclusive, at least for PET images.

Further on, different settings were used when classifying PET- and MR images. One example of this is the DWT-level used. For MR images level 3 was used while level 4 was used when transforming PET images. This can be considered a flaw in our method. However, the choice to have different settings can be justified as it yielded better results for both image types.
Chapter 6

Conclusion and future work

Classification accuracy obtained from MRI images in this study is significantly lower than that obtained by Lahmiri et al. [14]. Though, we find the accuracy high enough to be used in a comparison with PET. Both the ANOVA test and the post-hoc Tukey’s HSD test indicated a significant statistical difference. Therefore, we conclude that there is a difference in classification accuracy when using PET and MRIs to identify AD. Our conclusion is similar to what previous research has shown [6] [19].

As previously stated, the classification accuracy obtained from MRIs was not as high as other studies have shown is possible [6] [14]. Therefore, future studies should focus on achieving state-of-the-art classification accuracy for both PET and MRI when comparing the image types for classification accuracy.
Bibliography


