Classification of Borderline Personality Disorder patients from fMRI scans

Antria Kkousii
Graduation Thesis, January 2019
Media Technology MSc program, Leiden University
Supervisors: Prof. Dr. Ir. Fons J. Verbeek and C.C. van Schie, MSc

Abstract

Borderline Personality Disorder (BPD) is a mental illness that affects many people. Some symptoms of the disorder are mood swings, negative self-image that are associated with self-harm and in some extreme cases to suicide. Some studies suggested that BPD is related to functionality and structure of specific brain areas. This is based on fMRI scans of patients with this disorder. The goal of our study is to examine whether a variety of machine learning models could classify different fMRI scans of control subjects and BPD patients. We have used two datasets with fMRI scans derived from two Leiden University studies focused on BPD. We explored different machine learning techniques like Support Vector Machines (SVM), Decision Tree algorithms (DTs) and Convolutional Neural Networks (CNN). As a part of our research, we also ran some experiments with the dataset itself so that it became possible to extract features from fMRI scan that can be recognized as the Grey Matter, the White Matter, and the Cerebrospinal fluid to examine whether the different brain areas can classify the BPD group from the control group. Deep learning models, such as CNN, usually require big datasets to be able to distinguish patterns and features and therefore categorize data. To overcome this limitation with our relatively small dataset we added noise, by artificially enlarged the dataset through copying the original dataset and adding noise. With this technique, it was possible to run different experiments with a variety of models and come to the conclusion that the best model for identifying a BPD patient through fMRI scans is a Convolutional Neural Network. The accuracy of such a model using the Cerebrospinal fluid dataset is 96%.

Introduction

A mental illness called Borderline personality disorder (BPD) affects about 1.3% of the population [3]. This disorder is characterized by mood swings,unstable interrelations with others and negative self-image. Furthermore, it is very frequent among individuals diagnosed with BPD to hurt them self or even commit

suicide for about 10% of the patients [16]. The abnormal behavior, in most cases, begins during the early adulthood [25] and can be comorbid with depression, eating disorders or drug abuse.

Until now, the causes of Borderline Personality Disorder are not completely defined, although the sources of the disorder, most probably, has a generic origin, the social environment of the person and the brain structure [16]. Many neuroimaging studies of BPD patients showed that some parts of the brain are smaller; i.e. the structures usually are involved in the emotions of the individual, like the hippocampus, the amygdala [5], the orbitofrontal cortex [24] and other [19]. The process of diagnosing a borderline personality disorder patient is done by a clinical psychologist and is based on signs and symptoms, such as: (a) excessive emotions (anger, terror, fear etc.) and also intense mood swings (b) impulsive behavior; self harm, substance abuse, excessive drinking or abnormal eating habits (c) unstable relationships; abandonment issues and low quality relationships.

BPD is quite complex and heterogeneous problem and it can be identified only with the help of a mental health professional. In order to get a better understanding of the disorder, many studies focused on analyzing the neuroimaging of BPD patients to get a different perspective of what processes might be underlying the symptoms in BPD. The benefit of these techniques over traditional analysis techniques is that more information can be processed at once and that there is an indication of how sensitive results are at distinguishing BPD from control group. A study that used functional magnetic resonance imaging (fMRI) showed abnormal activity in certain brain areas when the BPD patients had to watch different facial expressions [18, 10]. The findings of those studies showed that some patterns can be used to identify BPD. Researchers from another study extracted spectral power features from resting-state fMRIs using machine learning models. The most discriminant features to identify BPD were related to the areas of the left medial orbitofrontal cortex, the left thalamus, and the right rostral anterior cingulate cortex [27].

In this study we will focus on classifying fMRI scans of BPD patients and control group using machine learning and deep learning models. Machine learning on fMRI scans had lead to some interesting scientific results on different diseases like Alzheimer [22] , schizophrenia [13] and Attention deficit hyperactivity disorder (ADHD) [7]. Our goal is to examine whether machine learning can classify the BPD patients from the control group. The data we used were gathered from domain experts of Leiden University, faculty of Social and Behavioural Sciences, department of the Clinical Psychology. More specifically, the data we used as input in the different deep learning neural networks in this research, were extracted from two studies, focused on Borderline Personality Disorder. In both studies, the subjects belonged in one of the three groups; control group (subjects without BPD), low-self esteem group (people that were not diagnosed with personality disorder, but they have low self esteem) and the BPD group.

In this study we focused only on the first and third group, leaving the low self-esteem group out because it is more challenging to distinguish three groups instead of two and the main goal is differentiate BPD from the control group.

In order to be able to insert the data in the deep learning algorithms that were used, we had to process the raw fMRI datasets by extracting a variety of values and constructing new datasets that will fit better to our models. Before using deep learning algorithms we used some simple supervised learning classification algorithms, such as Support Vector Machine and (SVM) and Decision Trees (DTs) in order to examine whether simple algorithms can classify the datasets and continue on using deep learning models.

Support Vector Machines (SVM) are a category of supervised learning models, that can be use to classify data [4]. The SVM places the labeled data in an N-Dimensional space; where N is the length of each input. Based on the kernel of the support vector, a hyperplane divides the data in classes.

The Decision Tree algorithms (DTs) are another supervised learning model category. The DTs are mostly used for classification and regression of data [20]. These classification algorithms create a model of simple decision paths and anticipate the target outcome of a given input. Those decision paths are constructed by learning features that are deducted from past inputs. After receiving some results from the first experiments, we continued by feeding our datasets into deep learning networks. For that process we used a Convolutional Neural Network.

Convolutional Neural Networks (CNN) are deep learning artificial models and they were successfully applied in image analysis [15]. CNN have a feed-forward structure that was inspired by the connection of the biological neurons of the animal visual cortex [17]. These networks can identify patterns of the input images using convolution layers.

Our research question was whether deep learning models improve the classification of BPD patients using fMRI scans? This question led us into two other sub-questions. The first one was what features we will feed into the deep learning models. We could have selected different features from specific brain areas, but our approach was to examine whether a set of more generic features can classify the BPD patients. The second sub-question was how can we increase the size of a small dataset of fMRI scans to be able to be fed into a deep learning model? Deep learning algorithms require vast number of data to contemplate many examples of the control group and the BPD patients that we researched, so that they can identify with high certainty the results. This was the main difficulty for this research, because of the disorder itself in which it is hard to get many borderline personalty disorder patients to participate and because of the high cost of using MRI scanners. Given this obstacles we tried to find out solutions to work with the subjects that were available from our dataset.

In this paper we will explain our methods on how and which features we extracted and how we increased the size of our datasets. Furthermore, a small introduction and explanation on what machine learning models we used and we will close with the results and discussion.

Materials and Methods

Data

Neuroimaging is accomplished through functional Magnetic Resonance Imaging (fMRI). The fMRI scans detect the brain activity based on the blood oxygenation in different brain areas, which is a proxy for neural activation. [12]. The higher the blood oxygenation level in a brain area, its higher the activity. The recording process of the blood flow is called Blood-Oxygen Level Dependent (BOLD). The fMRI samples the volume in rectangular areas; the voxels. For each voxel, the image stores the level of oxygenation relative to a baseline in a specific area on a specific time. Thus, a BOLD fMRI scan keeps a time series of a 3D brain blood flow representation.

We use the data from two experiments from Leiden University in the department of clinical psychology that took place at the Leiden University Medical Center (LUMC); the Reliving Autobiographical Memories(RAM) Experiment and the Social Feedback experiment [23]. Both of them were focused on the borderline personality disorder. Three groups of participants were defined; control group (subject without BPD), people with low self-esteem (Insecure people group) and people that diagnosed with Borderline Personality Disorder (BPD group).

Social Feedback Task (SF)

In this experiment, the control group consisted of 37 participants, the Insecure group consisted of 23 and the BPD group consisted of 47. Before scanning, all participants were interviewed, by the domain experts for about 10 minutes. In addition, they had to fill in a questionnaire about their current state of self-esteem, tension and anger. While the subjects were in the scanner, stimuli words (negative, positive and neutral) were shown in front of them. This stimuli words were supposed to be the opinion of another participant for the subject, based on the personal interview. In reality, the feedback was standardized and all participants received the same feedback. After each stimuli word, the subjects were asked to evaluate their emotions, reflecting their mood on a scale of 1 (having negative emotions) to 4 (having positive emotions). This study focused on how positive and negative social feedback affects the human brain for different groups of people; control group, low self-esteem group and BPD group.

Reliving Autobiographical Memories task (RAM)

In this experiment, the control group consisted of 36 participants, the Insecure group consisted of 24 and the BPD group consisted of 42. All the subjects were asked to write down 8 memories, 4 neutral and 4 positive ones. After they were placed in the MRI scan, they were asked to recall this memories as clear as possible. For each memory the participant had to read the memory and then will relive it as best as possible for 30 seconds. This study was focused on recalling positive personal memories and the affect on peoples brain with variance of self-esteems levels. Additionally, this study also compared the three groups and investigated the role of self-esteem.

Feature Selection

Valuable data need to be retrieved from the images in order to make classification. To that end we used a software suite called FMRIB Software Library (FSL) [26], that handles the analysis and processing of fMRI scans.

Firstly, we used the FSL FEAT library, which was used to perform slice timing correction; the recording of one volume (one time the whole brain) takes 2.2 seconds. One volume consists of 38 slices which are recorded at slightly different time points within this 2.2 second time frame. The next step was to map the functional scan for each individual subject to a standard brain model in order to have comparable images between the subjects. For this task we used the FLIRT library of FSL.

Meanwhile, we used the FSL FAST library, to create masks for each individual brain, that separate the three different brain areas; Gray Matter (GM) which includes all the actual neural cells of the brain, White Matter (WM) which includes all the myelinated axons [1] and the Cerebrospinal fluid (CSF) which is a fluid to protect the brain from the skull. The fMRI is optimal to observe the blood oxygenation as we mention above. However, in the CSF there is no blood flow, thus the fMRI scans do not actually measure the CSF activation. The same applies to the White Matter, where there is activation, but the fMRI is not measuring this type of activation. Thus, the WM and CSF of the fMRI scans do not measure the activation of the cells, rather the noise that comes from movement, the heartbeat and the neighboring grey matter voxels. We use the T1-weighted scan, which is optimized for observing the brain structure of the subject, to create masks for the three categories (GM, WM and CSF). This process can be found in figure 1 on the first step. There were a lot of studies that focused on the grey matter as we mention in the introduction. Our experiments were focused also on the 'noisy' brain areas; white matter and Cerebrospinal fluid. By 'noisy' we refer to the parts of the brain that are not neurons. A study concluded that BPD patients may have small white matter damage at

the orbitofrontal areas [2], whereas a study found that some BPD patients may have lower levels of Cerebrospinal fluid[6]. Thus, our goal was to investigate whether different machine learning algorithms can detect some patterns and classify the subjects correctly. An example of those masks can be found in figure 2.

At that point we had a functional scan for each subject and three different masks for the three parts of the brain (GM, WM and CSF). We applied this masks to the functional scans in order to create functional scans of the Gray matter, the White matter and the Cerebrospinal fluid (Figure 1 step 1).

Each scan has 256 voxels on x-axis (from right side of the brain to the left), 256 voxels on y-axis(from the back side of the brain to the front) and 140 voxels on z-axis (from the bottom of the brain to the top). All of the voxels represent the level of blood oxygenation dependent response which is a proxy for neural activation, as we mention above. Thus, for each functional scan we had a 4-dimensional table of the size (256, 256, 140, t) where t represent the time points of the scan. For the ML models to learn patterns from the selected features we have reduces the data size. Therefore we move away from the original 4D fMRI data. Thus, we created chunks of different sizes (5, 10, 15, 20 voxels per chunk) along the z-axis (transverses plane). The use of chunks is a manner to deal with the large data size and reduce the size. The abstract logic behind the chunk division was that each time slot in the fMRI scan, had a 3D representation of the brain, as we mention above. For each 3D representation we created chunks along a specific axis and for each chunk we calculated the mean value of the values in the voxels. This process was repeated for all the time slots. More specifically, we use FSL ROI (Region Of Interest) command to achieve this division (Figure 1 step 2).

After we created the chunks in the z-axis, we calculated the mean value of each chunk for each time slot, using the FSL MEANTS (Mean Time Series) (Figure 1 step 3). After this command we managed to reduce the size of our data to one dimension array of size t for each chunk. The next step was to reconstruct all the averages into one table of size (t, chunkDivisions) where chunkDivisions is the number of chunks along z-axis based on the chunk size (5, 10, 15 and 20). Using this technique we got an overview of what is happening to Grey Matter, to the White Matter and the Cerebrospinal fluid for the duration of the scanning process. The smaller the chunk size, the more 'detailed' dataset we get, with more values.

Data Expansion

Having the prepossessed data for all the subjects' fMRIs, we could actually use the data in deep learning models. The number of subjects from both experiments (Social Feedback and Reliving Autobiographical Memories) was relatively

small for the deep learning models. A study focused on this issue found that deep learning models can generalize based on the noisy data instead of memorizing the noise [21]. Thus, we added noise to our datasets. Noise in the dataset, refers to a random generation of data (fMRI scan representations) using normal distribution with sigma = 10. This technique allowed us to generate a dataset 100 times bigger than the original one, thus be able to use deep learning networks. Figure 3 demonstrates this process.

Learning Models

Support Vector Machine (SVM)

Firstly, we used a Support Vector Machine (SVM) to investigate if our datasets were, at least, linear separable. In the figure 4 you can see a visualization of a simple usage of SVM. We ran the SVM using all the different datasets we have generated. For each dataset, we apply the SVM and use the 5-fold cross validation method to compute the accuracy of classification.

Decision Trees (DTs)

We apply Decision Trees to our datasets, in order to examine if they can classify them properly. We use the same cross validation tactic with the Support Vector Machines. The Decision Trees are no-parametric algorithms, which means, the algorithm does not assume that the data have a specific distribution. This algorithm will try to classify the data creating simple decision paths base the past input. More specifically, the model creates a binary tree. Each node checks a specific feature of the input and based on its value will redirect the input to the next node. This will continue until there is no other redirection and the last node contains the prediction (the class that the input belongs to)

Convolutional Neural Networks - CNN

Convolutional Neural Networks are deep learning artificial models and the most common usage is image analysis[15]. Deep learning models is a subcategory of machine learning models that can learn data representations. CNN have feed-forward structure that was inspired by the connection of the biological neurons of the animal visual cortex [17]. This networks can identify patterns in input images using convolution layers. The convolutional layers consist of a collection of filters (kernels) and they have a receptive field that scans the input across its height and width. The filter calculates the dot product of the receptive field

entry and creates a 2-dimensional activation map of that filter. This process imitates a visual stimuli of an single neuron.

We constructed mutable CNNs with different structures; with one, two, three and four convolutional layers. The basic structure of the CNN is convolutional layer(s) with Rectifier activation function (ReLU). ReLU can train the model faster than other activation functions, such as tahn and sigmoid, without a significant cost on the accuracy [14]. After each convolutional layer we added a max pooling layer. The pooling layers are responsible for merging the output of a set of neurons (convolutional filters) into the next layer [11], specifically it gathers the maximum values from the previous layer. The final layers of our network were a drop out layer and then a fully connected layer [9], with softmax activation function, that will classify the input into a control subject or a borderline personality disorder patient. In figure 5 you can find the network's diagram with more details. Our datasets represent a time series of features. The architecture of our CNN had to be able to classify this kind of data and we use similar structure from literature we found on classifying time series data using CNN [8, 28].

Results

Datasets

After prepossessing the fMRI scans from the Social Feedback dataset and the RAM dataset, we came up with multiple datasets. We use different sizes for the division of the brain on the Z-axis (chunk size = 5 or 10 or 15 or 20). Furthermore, we generated data using the existing data and added random noise with standard deviation (sigma=10) and repeated them for 100 times. Thus, we got for each chunk division the original set of data and the generated dataset. The table 1 lists all the datasets we used to run different deep learning networks.

Support Vector Machine - SVM

Firstly, we wanted to check whether our datasets were linearly separable. In machine learning, linear separability is one of the basic methods for classifying data. Support Vector Machines can examine if a dataset is linear separable by placing them in a higher dimension. Thus, we used a Support Vector Machine with a linear kernel. We ran each SVM using the cross validation method of 5 folds. We chose 5-fold cross validation to divide data into 80% training set and 20% testing set. As first experiment, we ran the SVM for all the original

datasets on table 1, to determine which datasets can be classified with high accuracy and the results can be found in table 2. The table shows the mean value of correct classification from all the folds and the standard deviation. On a general view on the original datasets, we cannot say that our datasets are linearly separable; we got accuracy close to 0.5 which is close to random guessing. Thus, the original dataset cannot be linearly classified.

Next experiment was to run the same SVM for the generated datasets. The results can be found on table 3. We can observe that in both cases, RAM and Social Feedback, we can observe that the smaller the chunk size is, the better accuracy we get. Hence, the more detailed datasets we feed into the SVM the better the classification. For the generated datasets we get a small increase in the accuracy, compared with the original datasets (Table 2) , however is not quite high for classification of the two classes(control group and BPD group). Nevertheless, for chunk size 5 in the RAM Dataset the accuracy is close to 75% and the highest accuracy is on the Gray Matter with 79%. To conclude, our dataset were not linearly separable, thus we continue exploring more machine learning techniques.

Before we continue on the next machine learning algorithm, there was another experiment we wanted to form. Given the fact that the number of our dataset is limited, we wanted to check whether we can combine the two datasets (RAM and Social Feedback) so we can increase the dataset size. Thus, we run a Linear Support Vector Machine to check if it can classify the RAM dataset from the Social Feedback dataset. We use the same SVM as before. Table 4 values the success mean value and the standard deviation of a Support Vector Machine with linear kernel. The white matter dataset has high accuracy, which means that the two datasets can be distinguished, therefore the White Matter behaves differently on the two datasets. The CSF datasets have accuracy close to 0.6 which means that they are not linearly separable and thus the machine cannot classify them properly. The gray matter dataset has accuracy close to 0.75 which is not the optimum, however the two datasets are not quite separable. Thus, we can use the combination of the two datasets for CSF because the SVM couldn't classify them and they were behaving similarly on both experiments (SF and RAM). The combination of both datasets for the WM would not be a good practice to be used in a deep learning model since the WM is behaving differently during the two experiments. For the GM we cannot be certain because the accuracy is close to 75%.

Decision Trees

The next approach we took, was to use decision trees to investigate if the algorithm can create a model with decision paths for classifying our datasets. We use again the cross validation method using 5-folds. Table 5 demonstrates

the result of the Decision Tree algorithms on the original datasets. The DTs couldn't classify the original datasets to BPD and control group. The highest accuracy is 63%, which is not optimum. The low accuracy percentage may cause by the size of our datasets. The DT algorithm did not have enough input in order the create the decision paths for such a big input from each subject, thus, our next experiment was to run DTs for the generated dataset. Similar results came from the SVM algorithm on the original datasets (Table 2). To conclude, our original datasets cannot be classified linearly (Table 2) with the Support Vector Machine and the size of our datasets is not big enough for the DTs algorithm to classify the data (Table 5).

Next step was to apply the Decision tree algorithm on the generated datasets and table 6 shows these results. Considering the table 5 with the original datasets, the table 6 has clearly better results. As we mention above, the datasets size of the original set may have caused the low accuracy, the result of table 6 supports this theory, where increasing the dataset lead to better results. The CSF datasets have the lowest accuracy compare with the other two kinds of datasets. The best mean accuracy came from the Social Feedback dataset of the white matter with chunk size equal to 5 (86% accuracy) and the second best from the RAM dataset of the white matter with again chunk size 5 (85% accuracy). Therefore, the Decision Tree algorithms could classify with relatively high accuracy the datasets of White Matter and Grey Matter for smaller chunk sizes.

The final experiment we did with the DTs was to examine whether it can classify the two datasets (RAM task and Social Feedback Task), same experiment we did with the SVM on table 4. Table 7 displays the results of the accuracy of the classification of all the original datasets to RAM task and Social Feedback Task. The Cerebrospinal Fluid, White Matter and Gray Matter datasets seems to not be able to classify whether the input belongs to the RAM or Social Feedback task. The accuracy does not overcome 56% in all of the datasets, which is contradictory with the SVM results (table 4); the decision tree algorithm cannot distinguish the two tasks for all the original datasets whereas the SVM can classify the White Matter dataset.

Convolutional Neural Network

After we run the Support Vector Machines and the Decision Trees, the next step was to run deep learning networks. For our experiments, we used Convolutional Neural Network (CNN). We chose the CNN model because this networks have the ability to detect patterns in images and it is one of the most powerful deep learning networks. We constructed multiple CNNs with different convolutional layers; one, two, three and four layers. Each CNN ran a 5-fold cross validation, for 50 epochs each. Tables 8, 9, 10 and 11 display the results of the CNN with

the different convolutional layers. On a general view, the best results derived from the CNN with two and three convolutional layers, whereas the CNN with one convolutional layer did not rise good results. The two and three layers lead to better results most probably because the one layer was not enough for the model to learn the training data whereas the four layers may lead to over fitting of the training dataset. In all experiments the White and Grey Matter datasets did not give any good results, where as the Cerebrospinal Fluid dataset gave some promising results. Furthermore, the combination of the datasets for all the experiments for the White Matter results are quite low that supports the findings from the SVM where the two datasets are separable on the White Matter features (table 4). Hence, the combination of them, troubled the learning process of the CNN because the BPD White matter from the Social Feedback has different behavior from the BPD White Matter of the RAM dataset, the same applies for the control group. The CNN presumably could not generalize the classification of the two groups (BPD and control group) for the combination of both datasets (SF and RAM). The same pattern appears on the Grey Matter for the combination of the two datasets, where all the CNNs cannot classify the BPD from the Control Group.

The first experiment we did with a CNN with one convolutional layer did not give good results, most probably because the one convolutional layer was not enough to detect features from the datasets. The CNNs with two and three convolutional layers had relatively good results only for the Cerebrospinal Fluid, whereas the White and Grey matter couldn't be classified by any CNN. The same pattern of results derived from the CNN with four convolutional layers. Comparing this results with Decision Trees (table 6), we were expecting to get better results from either White Matter or Grey Matter. Nevertheless, the best results were deducted from the Cerebrospinal fluid. More specifically, 95% accuracy was achieved in multiple cases; on the two layer CNN for the CSF for chunk size 15 on the SF dataset and the combination of both datasets (table 9) and for the tree layer CNN on the CSF features on the SF dataset (table 10). The best accuracy occurred on the three layer CNN network on the CSF features for the combination of both datasets with mean accuracy 96% and standard deviation 1%.

The figure 6 present the accuracy and the loss of a single run of the best 5-fold cross validation accuracy derived from the previous experiments; three layer CNN of CSF of SF and RAM datasets with chunk size 15. We run the CNN for 50 epochs with learning rate 0.001. The left graph expresses the mean accuracy per epoch for the training and test data (8:2 ration). The right graph expresses the mean square error for the train and test data for each epoch. It is clearly that the network is learning and the loss drops accordingly.

Division along Y-Axis

The CNN models could not classify the Gray Matter and the White matter datasets, and one possible reason for that is the chunk division along the Z-axis, which is not optimal for the Grey Matter cells; each chunk may averaging the signal from different brain areas. Thus, we created the same datasets but the chunks were created along the Y-Axis (coronal plane) to investigate whether this division can give better results. Based on the previous results, with the z-axis chunks, the Decision trees and the CNN classified with high accuracy some datasets. The decision trees classified the BPD from the control group for small chunk size in the RAM dataset (table 6). On the other hand, CNNs classified the Cerebrospinal Fluid features for higher chunk sizes (tables 9 and 10). One explanation for this pattern may be that for the Decision Trees the bigger chunks will create less decision paths because the input for each subject is smaller and this may lead to smaller success rates. When the chunk size is smaller then the input is bigger and the algorithm will create more decision paths and that may lead to better results. On the other hand, CNNs have the opposite behavior. The bigger chunks are providing less data points for each subject and it is easier for CNN to find distinct patterns between the two groups (control group and BPD group) compare with the smaller chunks. Thus, our approach was to run again this algorithms for the new features we extracted.

First, we ran the Decision Tree algorithms using the same methods with the previous experiments and the results can be found in table 12. The Decision Trees could classify with relatively high accuracy the White and Grey Matter (around 80% accuracy), whereas the Cerebrospinal Fluid features have lower accuracy. Furthermore, we see the same pattern as before; the smaller the chunk size is, the higher the accuracy.

We ran the CNN models with different convolutional layers to examine whether the division on the coronal plates can give a better classification on Gray Matter and White Matter. Tables 13 and 14 display the results of the Convolutional Neural Networks with 2 and 3 layers. In general the CNN cannot classify any of the categories (CSF, WM and GM). The y-axis division did not lead to any better results.

Discussion

Borderline Personality Disorder is a stable patterns of symptoms. Such symptoms are bad interpersonal relationships, mood swings, pessimistic self-image and others. Many BPD neuroimaging studies lead to some interesting findings regarding the brain structure and activity; some brain volumes have smaller size than a control subject that are responsible for the different emotions of

the individual [5, 24, 19] and abnormal brain activity after different stimuli [18, 10]. The aim of this study was to examine whether a deep learning model can classify fMRI scans of BPD patients from the control group.

They were two aspects we needed to take into consideration on using the fMRI scans in the deep learning models:

1. What features we will extract from the fMRI scans. We used datasets from two Leiden University studies focusing on BPD. Our approach was to extract generic features from the fMRIs. We divided the fMRIs in the Grey Matter, the White Matter and the Cerebrospinal Fluid and for each category we created chunks on the z-axis and y-axis where we calculated the mean value of all the voxels in each chunk. Our goal was to get an overview of the brain for each category (GM, WM and CSF) through time, where as similar studies used features from specific brain areas to feed into different machine learning algorithms. Although, the WM and the CSF features were noise due to movement and the heartbeat, we also fed them to different deep learning models to examine whether this noise can separate the control group from the BPD patients.

2. How we were going to increase the dataset size. The size of the datasets were relatively small to feed into a deep learning model; the Social Feedback had 84 scans and for the RAM dataset had 78 scans. We had to increase the size because deep learning models need a big dataset to learn from it. To achieve that, we replicated our dataset and altered them through additional noise. We use this technique because deep learning models can generalize the noise and not memorize it [21].

We form different experiments using different machine learning techniques, such as Support Vector Machine, Decision Tree Algorithms and Convolutional Neural Networks for all the datasets we have created (Table 1). We firstly use the Support Vector Machine algorithm to investigate whether our datasets are linearly separable and based on the results the original datasets were not (Table 2). Same behavior occurred from the Decision tree algorithms, where the original datasets were not able to be classify by the model (Table 5). We tested both algorithms on the generated datasets and there was some improvement (Tables 3 and 6). The increased size in the datasets lead to better results.

We extracted values from the Grey Matter, the White Matter and the Cerebrospinal Fluid and we examine them individually. Based on the results from the CNN experiments (Tables 8, 9, 10 and 11) the White Matter and Gray Matter were not classified from the model, in contradiction with the Cerebrospinal Fluid features were distinct between the BPD patients and the control.

The Cerebrospinal fluid lead to some interesting results. We were not expecting this results, given the fact that the fMRI data for the CSF were random noise.

Furthermore, the CSF is a fluid that flows around the brain, it does not have any specific activity based on its location. Extracting features that characterize an overview of the Cerebrospinal fluid prompted the CNN to find distinct patterns of it that classified a control subject to a BPD patient.

Considering the results that derived from the White Matter were not valuable. Given the fact that the WM are noisy data of the fMRI scans, we were expected this outcome. In addition, the White Matter contains all the connections between the neurons and with the same logic that the GM has abnormal activity in specific areas we could assume that the White Matter in the brain areas with the abnormal activity will have abnormal activity as well; a study found out that the white matter in the orbotfrontal area of BPD patients has a microstructural damage [2]. Thus, general overview of the Grey Matter and White Matter did not lead to any good results because the areas with the abnormal activities are canceled out from the generic features we extracted.

One possible justification for the bad results is that the division using the transverse plane is not optimal for the Grey Matter cells, whereas division using the coronal plane may have been in line with the functionality of the brain. Thus, we extracted the same features on this plane and tested the CNN models with 2 and 3 convolutional layers. As a result, the CNN could not classify any of the three categories (CSF, WM and GM). Thus, the generic feature extraction for the White and Grey Matter could not be categorized using Convolutional Neural Networks. That may occur because Grey Matter has specific behavior and functionality based on its location; brain regions work together to produce a certain function. With this logic, it would make more sense to extract features from specific areas that have abnormal activity in BPD patients, like the amygdala and the Anterior Cingulate Cortex (ACC)[18, 10].

Future Work

During this research, many decisions had to be made and thus, many other possibilities opened up that they were not study in this research. Firstly, the original datasets that we got from Leiden University had 3 groups; control group, low self-esteem group and BPD group. In our research we canceled out the second group because it could be more challenging to distinguish three categories. The results we got from the CNN from the Cerebrospinal Fluid features lead to some extremely good results. It will be interesting if the same CNNs can run the dataset with all three groups.

Furthermore, another choice we had to make was to create chunks on the z-axis and y-axis of the brain. Although, the CSF features gave some good results, the WM and GM features could not be classified by the CNN for the generic feature selection. In that end, a selection of features from specific areas may

lead to better classification accuracy where the model may find patterns that are distinct between the two groups. Also, in order to examine the actual activation of the White Matter, it would be optimum to use Diffusion Tensor Imaging (DTI), which is a different kind of MRI neuroimaging technique related to the White Matter cells.

Last, based on the CNN results (tables 9 and 10) the Cerebrospinal Fluid was distinct between the two groups. It would be interesting to have Cine MRI scans (exams the CSF of the brain) of BPD patients and control group and study whether there are different.

References

[1] Hal Blumenfeld. *Neuroanatomy Through Clinical Cases (Blumenfeld, Neuroanatomy through Clinical Cases)*. Sinauer Associates, 2002.

[2] José Luis Carrasco et al. "Microstructural white matter damage at orbitofrontal areas in borderline personality disorder". In: *Journal of Affective Disorders* 139.2 (2012), pp. 149–153.

[3] Jeremy Coid et al. "Prevalence and correlates of personality disorder in Great Britain". In: *The British Journal of Psychiatry* 188.5 (2006), pp. 423–431.

[4] Corinna Cortes and Vladimir Vapnik. "Support-vector networks". In: *Machine learning* 20.3 (1995), pp. 273–297.

[5] Martin Driessen et al. "Posttraumatic stress disorder and fMRI activation patterns of traumatic memory in patients with borderline personality disorder". In: *Biological psychiatry* 55.6 (2004), pp. 603–611.

[6] David L Gardner, Peter B Lucas, and Rex W Cowdry. "CSF metabolites in borderline personality disorder compared with normal controls". In: *Biological Psychiatry* 28.3 (1990), pp. 247–254.

[7] A Junyu Hao, B Lianghua He, and C Hongyi Yin. "Discrimination of ADHD children based on Deep Bayesian Network". In: (2015).

[8] Nima Hatami, Yann Gavet, and Johan Debayle. "Classification of time-series images using deep convolutional neural networks". In: *Tenth International Conference on Machine Vision (ICMV 2017)*. Vol. 10696. International Society for Optics and Photonics. 2018, 106960Y.

[9] Simon S Haykin et al. *Neural networks and learning machines*. Vol. 3. Pearson Upper Saddle River, 2009.

[10] Sabine C Herpertz et al. "Evidence of abnormal amygdala functioning in borderline personality disorder: a functional MRI study". In: *Biological psychiatry* 50.4 (2001), pp. 292–298.

[11] David H Hubel and Torsten N Wiesel. "Receptive fields and functional architecture of monkey striate cortex". In: *The Journal of physiology* 195.1 (1968), pp. 215–243.

[12] Scott A Huettel, Allen W Song, Gregory McCarthy, et al. *Functional magnetic resonance imaging*. Vol. 1. Sinauer Associates Sunderland, MA, 2004.

[13] Junghoe Kim et al. "Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia". In: *Neuroimage* 124 (2016), pp. 127–146.

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[15] Yann LeCun et al. "LeNet-5, convolutional neural networks". In: *URL: http://yann. lecun. com/exdb/lenet* (2015), p. 20.

[16] Klaus Lieb et al. "Borderline personality disorder". In: *The Lancet* 364.9432 (2004), pp. 453–461.

[17] Masakazu Matsugu et al. "Subject independent facial expression recognition with robust face detection using a convolutional neural network". In: *Neural Networks* 16.5-6 (2003), pp. 555–559.

[18] Michael J Minzenberg et al. "Fronto-limbic dysfunction in response to facial emotion in borderline personality disorder: an event-related fMRI study". In: *Psychiatry Research: Neuroimaging* 155.3 (2007), pp. 231–243.

[19] Aisling O'Neill and Thomas Frodl. "Brain structure and function in borderline personality disorder". In: *Brain Structure and Function* 217.4 (2012), pp. 767–782.

[20] J. Ross Quinlan. "Induction of decision trees". In: *Machine learning* 1.1 (1986), pp. 81–106.

[21] David Rolnick et al. "Deep learning is robust to massive label noise". In: *arXiv preprint arXiv:1705.10694* (2017).

[22] Saman Sarraf and Ghassem Tofighi. "Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks". In: *arXiv preprint arXiv:1603.08631* (2016).

[23] Charlotte C van Schie et al. "When compliments do not hit but critiques do: an fMRI study into self-esteem and self-knowledge in processing social feedback". In: *Social cognitive and affective neuroscience* 13.4 (2018), pp. 404–417.

[24] Christian G Schmahl et al. "Magnetic resonance imaging of hippocampal and amygdala volume in women with childhood abuse and borderline personality disorder". In: *Psychiatry Research: Neuroimaging* 122.3 (2003), pp. 193–198.

[25] Robert L Spitzer and Janet BW Williams. "Diagnostic and statistical manual of mental disorders". In: *American Psychiatric Association*. Citeseer. 1980.

[26] Mark W Woolrich et al. "Bayesian analysis of neuroimaging data in FSL". In: *Neuroimage* 45.1 (2009), S173–S186.

[27] Tingting Xu et al. "Classification of borderline personality disorder based on spectral power of resting-state fMRI". In: *Conference proceedings:... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*. Vol. 2014. NIH Public Access. 2014, p. 5036.

[28] Jianbo Yang et al. "Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition." In: *Ijcai*. Vol. 15. 2015, pp. 3995–4001.

Appendix

Tables

Social Feedback	RAM
ChunkSize = 5 - Original dataset	ChunkSize = 5 - Original dataset
ChunkSize = 5 - Generated dataset	ChunkSize = 5 - Generated dataset
ChunkSize = 10 - Original dataset	ChunkSize = 10 - Original dataset
ChunkSize = 10 - Generated dataset	ChunkSize = 10 - Generated dataset
ChunkSize = 15 - Original dataset	ChunkSize = 15 - Original dataset
ChunkSize = 15 - Generated dataset	ChunkSize = 15 - Generated dataset
ChunkSize = 20 - Original dataset	ChunkSize = 20 - Original dataset
ChunkSize = 20 - Generated dataset	ChunkSize = 20 - Generated dataset

Table 1: Datasets of the experiment

Chunk Size	CSF	WM	GM
	RAM Dataset		
5	0.38 (+/- 0.16)	0.56 (+/- 0.21)	0.49 (+/- 0.17)
10	0.44 (+/- 0.27)	0.59 (+/- 0.19)	0.49 (+/- 0.26)
15	0.48 (+/- 0.25)	0.58 (+/- 0.14)	0.46 (+/- 0.21)
20	0.50 (+/- 0.26)	0.53 (+/- 0.13)	0.46 (+/- 0.21)
	Social Feedback Dataset		
5	0.46 (+/- 0.18)	0.46 (+/- 0.16)	0.59 (+/- 0.17)
10	0.48 (+/- 0.18)	0.56 (+/- 0.27)	0.54 (+/- 0.11)
15	0.45 (+/- 0.21)	0.56 (+/- 0.20)	0.55 (+/- 0.07)
20	0.42 (+/- 0.18)	0.61 (+/- 0.25)	0.50 (+/- 0.24)

Table 2: Mean Accuracy and standard deviation of SVM with 5-fold cross validation using the original datasets

Chunk Size	CSF	WM	GM
	RAM Dataset		
5	0.74 (+/- 0.04)	0.75 (+/- 0.04)	0.79 (+/- 0.04)
10	0.60 (+/- 0.02)	0.63 (+/- 0.04)	0.64 (+/- 0.03)
15	0.58 (+/- 0.05)	0.59 (+/- 0.03)	0.59 (+/- 0.02)
20	0.56 (+/- 0.06)	0.58 (+/- 0.03)	0.60 (+/- 0.04)
	Social Feedback Dataset		
5	0.67 (+/- 0.04)	0.68 (+/- 0.05)	0.70 (+/- 0.03)
10	0.57 (+/- 0.04)	0.59 (+/- 0.05)	0.58 (+/- 0.01)
15	0.54 (+/- 0.04)	0.57 (+/- 0.01)	0.55 (+/- 0.03)
20	0.53 (+/- 0.03)	0.58 (+/- 0.03)	0.52 (+/- 0.02)

Table 3: Mean Accuracy and standard deviation of SVM with 5-fold cross validation using the generated datasets

Chunk Size	CSF	WM	GM
	Original Dataset		
5	0.55 (+/- 0.20)	0.84 (+/- 0.23)	0.76 (+/- 0.19)
10	0.65 (+/- 0.21)	0.86 (+/- 0.20)	0.75 (+/- 0.15)
15	0.64 (+/- 0.17)	0.89 (+/- 0.18)	0.74 (+/- 0.16)
20	0.63 (+/- 0.18)	0.88 (+/- 0.18)	0.76 (+/- 0.14)

Table 4: Mean Success Value and standard deviation of Linear Support Vector Machine with 5-fold cross validation for classification of RAM and Social feedback datasets

Chunk Size	CSF	WM	GM
	RAM Dataset		
5	0.58 (+/- 0.13)	0.55 (+/- 0.09)	0.62 (+/- 0.30)
10	0.44 (+/- 0.15)	0.53 (+/- 0.07)	0.63 (+/- 0.22)
15	0.57 (+/- 0.31)	0.53 (+/- 0.31)	0.61 (+/- 0.16)
20	0.51 (+/- 0.21)	0.59 (+/- 0.25)	0.62 (+/- 0.03)
	Social Feedback Dataset		
5	0.35 (+/- 0.10)	0.53 (+/- 0.31)	0.48 (+/- 0.43)
10	0.41 (+/- 0.19)	0.50 (+/- 0.27)	0.46 (+/- 0.28)
15	0.47 (+/- 0.27)	0.61 (+/- 0.29)	0.56 (+/- 0.29)
20	0.50 (+/- 0.27)	0.62 (+/- 0.26)	0.47 (+/- 0.35)

Table 5: Mean Accuracy and standard deviation of Decision Tree Model with 5-fold cross validation using the original datasets

Chunk Size	CSF	WM	GM
	RAM Dataset		
5	0.73 (+/- 0.03)	0.85 (+/- 0.02)	0.82 (+/- 0.03)
10	0.69 (+/- 0.03)	0.84 (+/- 0.02)	0.79 (+/- 0.02)
15	0.68 (+/- 0.03)	0.80 (+/- 0.01)	0.77 (+/- 0.02)
20	0.66 (+/- 0.03)	0.79 (+/- 0.02)	0.75 (+/- 0.02)
	Social Feedback Dataset		
5	0.69 (+/- 0.02)	0.86 (+/- 0.02)	0.82 (+/- 0.02)
10	0.67 (+/- 0.01)	0.83 (+/- 0.03)	0.79 (+/- 0.03)
15	0.66 (+/- 0.02)	0.81 (+/- 0.02)	0.76 (+/- 0.02)
20	0.65 (+/- 0.02)	0.79 (+/- 0.00)	0.74 (+/- 0.01)

Table 6: Mean Accuracy and standard deviation of Decision Tree Model with 5-fold cross validation using the generated datasets

Chunk Size	CSF	WM	GM
	Original Dataset		
5	0.52 (+/- 0.13)	0.54 (+/- 0.15)	0.48 (+/- 0.11)
10	0.48 (+/- 0.06)	0.45 (+/- 0.18)	0.48 (+/- 0.07)
15	0.49 (+/- 0.09)	0.50 (+/- 0.12)	0.49 (+/- 0.26)
20	0.54 (+/- 0.08)	0.49 (+/- 0.21)	0.56 (+/- 0.10)

Table 7: Mean Success Value and standard deviation of Decision Tree algorithm with 5-fold cross validation for classification of RAM and Social feedback datasets

Chunk Size	CSF	WM	GM
	RAM Dataset		
5	0.50 (+/- 0.04)	0.51 (+/- 0.02)	0.50 (+/- 0.06)
10	0.56 (+/- 0.26)	0.51 (+/- 0.02)	0.49 (+/- 0.05)
15	0.68 (+/- 0.32)	0.50 (+/- 0.02)	0.51 (+/- 0.06)
20	0.65 (+/- 0.17)	0.51 (+/- 0.01)	0.55 (+/- 0.15)
	Social Feedback Dataset		
5	0.50 (+/- 0.02)	0.50 (+/- 0.02)	0.50 (+/- 0.01)
10	0.51 (+/- 0.02)	0.51 (+/- 0.02)	0.50 (+/- 0.01)
15	0.73 (+/- 0.22)	0.51 (+/- 0.01)	0.56 (+/- 0.23)
20	0.65 (+/- 0.18)	0.51 (+/- 0.04)	0.55 (+/- 0.18)
	RAM + Social Feedback Dataset		
5	0.50 (+/- 0.01)	0.50 (+/- 0.01)	0.49 (+/- 0.03)
10	0.71 (+/- 0.36)	0.51 (+/- 0.01)	0.49 (+/- 0.02)
15	0.56 (+/- 0.24)	0.50 (+/- 0.02)	0.49 (+/- 0.01)
20	0.63 (+/- 0.21)	0.51 (+/- 0.02)	0.53 (+/- 0.18)

Table 8: Mean Accuracy and standard deviation of Convolutional Neural Network with 1 convolutional layer with 5-fold cross validation

Chunk Size	CSF	WM	GM
	RAM Dataset		
5	0.50 (+/- 0.04)	0.48 (+/- 0.02)	0.52 (+/- 0.05)
10	0.50 (+/- 0.02)	0.49 (+/- 0.02)	0.53 (+/- 0.01)
15	0.79 (+/- 0.31)	0.49 (+/- 0.01)	0.52 (+/- 0.05)
20	0.79 (+/- 0.11)	0.57 (+/- 0.27)	0.52 (+/- 0.05)
	Social Feedback Dataset		
5	0.50 (+/- 0.02)	0.50 (+/- 0.02)	0.50 (+/- 0.01)
10	0.79 (+/- 0.46)	0.50 (+/- 0.02)	0.50 (+/- 0.01)
15	0.95 (+/- 0.03)	0.51 (+/- 0.01)	0.50 (+/- 0.02)
20	0.79 (+/- 0.28)	0.52 (+/- 0.01)	0.64 (+/- 0.34)
	RAM + Social Feedback Dataset		
5	0.61 (+/- 0.39)	0.50 (+/- 0.01)	0.50 (+/- 0.03)
10	0.69 (+/- 0.46)	0.50 (+/- 0.02)	0.49 (+/- 0.03)
15	0.95 (+/- 0.02)	0.50 (+/- 0.02)	0.49 (+/- 0.03)
20	0.89 (+/- 0.02)	0.51 (+/- 0.02)	0.65 (+/- 0.40)

Table 9: Mean Accuracy and standard deviation of Convolutional Neural Network with 2 convolutional layers with 5-fold cross validation

Chunk Size	CSF	WM	GM
	RAM Dataset		
5	0.52 (+/- 0.02)	0.50 (+/- 0.04)	0.52 (+/- 0.05)
10	0.58 (+/- 0.35)	0.50 (+/- 0.02)	0.53 (+/- 0.01)
15	0.73 (+/- 0.38)	0.50 (+/- 0.02)	0.52 (+/- 0.05)
20	0.84 (+/- 0.05)	0.59 (+/- 0.34)	0.52 (+/- 0.05)
	Social Feedback Dataset		
5	0.50 (+/- 0.02)	0.50 (+/- 0.02)	0.50 (+/- 0.01)
10	0.77 (+/- 0.44)	0.50 (+/- 0.02)	0.50 (+/- 0.01)
15	0.95 (+/- 0.03)	0.51 (+/- 0.01)	0.50 (+/- 0.01)
20	0.81 (+/- 0.29)	0.52 (+/- 0.01)	0.57 (+/- 0.29)
	RAM + Social Feedback Dataset		
5	0.61 (+/- 0.39)	0.50 (+/- 0.01)	0.50 (+/- 0.03)
10	0.70 (+/- 0.48)	0.50 (+/- 0.02)	0.49 (+/- 0.03)
15	0.96 (+/- 0.01)	0.50 (+/- 0.02)	0.49 (+/- 0.03)
20	0.88 (+/- 0.04)	0.51 (+/- 0.02)	0.72 (+/- 0.40)

Table 10: Mean Accuracy and standard deviation of Convolutional Neural Network with 3 convolutional layers with 5-fold cross validation

Chunk Size	CSF	WM	GM
	RAM Dataset		
5	0.68 (+/- 0.82)	0.48 (+/- 0.90)	0.48 (+/- 0.90)
10	0.40 (+/- 0.98)	0.51 (+/- 0.90)	0.51 (+/- 0.90)
15	0.91 (+/- 0.08)	0.29 (+/- 0.79)	0.29 (+/- 0.79)
20	0.40 (+/- 0.67)	0.09 (+/- 0.37)	0.09 (+/- 0.37)
	Social Feedback Dataset		
5	0.29 (+/- 0.79)	0.09 (+/- 0.37)	0.09 (+/- 0.37)
10	0.59 (+/- 0.97)	0.11 (+/- 0.43)	0.11 (+/- 0.43)
15	0.77 (+/- 0.77)	0.32 (+/- 0.81)	0.32 (+/- 0.81)
20	0.40 (+/- 0.76)	0.33 (+/- 0.82)	0.33 (+/- 0.82)
	RAM + Social Feedback Dataset		
5	0.82 (+/- 0.32)	0.42 (+/- 0.69)	0.42 (+/- 0.69)
10	0.45 (+/- 0.89)	0.50 (+/- 0.73)	0.50 (+/- 0.73)
15	0.81 (+/- 0.46)	0.30 (+/- 0.60)	0.30 (+/- 0.60)
20	0.51 (+/- 0.71)	0.31 (+/- 0.59)	0.51 (+/- 0.71)

Table 11: Mean Accuracy and standard deviation of Convolutional Neural Network with 4 convolutional layers with 5-fold cross validation

Chunk Size	CSF	WM	GM
	RAM Dataset		
5	0.71 (+/- 0.01)	0.85 (+/- 0.03)	0.81 (+/- 0.02)
10	0.68 (+/- 0.01)	0.82 (+/- 0.02)	0.80 (+/- 0.02)
15	0.67 (+/- 0.01)	0.82 (+/- 0.01)	0.79 (+/- 0.03)
20	0.69 (+/- 0.03)	0.79 (+/- 0.02)	0.78 (+/- 0.03)
	Social Feedback Dataset		
5	0.70 (+/- 0.03)	0.88 (+/- 0.02)	0.81 (+/- 0.02)
10	0.68 (+/- 0.02)	0.85 (+/- 0.01)	0.80 (+/- 0.02)
15	0.67 (+/- 0.02)	0.84 (+/- 0.00)	0.78 (+/- 0.02)
20	0.65 (+/- 0.02)	0.83 (+/- 0.01)	0.76 (+/- 0.02)

Table 12: Mean Accuracy and standard deviation of Decision Tree Model with 5-fold cross validation using the generated datasets

Chunk Size	CSF	WM	GM
	RAM Dataset		
5	0.50 (+/- 0.04)	0.50 (+/- 0.04)	0.49 (+/- 0.04)
10	0.50 (+/- 0.02)	0.50 (+/- 0.02)	0.50 (+/- 0.02)
15	0.50 (+/- 0.02)	0.50 (+/- 0.02)	0.50 (+/- 0.02)
20	0.57 (+/- 0.27)	0.50 (+/- 0.01)	0.50 (+/- 0.01)
	Social Feedback Dataset		
5	0.50 (+/- 0.02)	0.51 (+/- 0.02)	0.49 (+/- 0.01)
10	0.50 (+/- 0.02)	0.50 (+/- 0.02)	0.49 (+/- 0.01)
15	0.61 (+/- 0.27)	0.51 (+/- 0.01)	0.51 (+/- 0.01)
20	0.75 (+/- 0.23)	0.57 (+/- 0.27)	0.51 (+/- 0.01)
	RAM + Social Feedback Dataset		
5	0.50 (+/- 0.01)	0.50 (+/- 0.01)	0.50 (+/- 0.01)
10	0.50 (+/- 0.00)	0.50 (+/- 0.01)	0.50 (+/- 0.01)
15	0.56 (+/- 0.24)	0.51 (+/- 0.01)	0.51 (+/- 0.01)
20	0.64 (+/- 0.32)	0.50 (+/- 0.02)	0.50 (+/- 0.01)

Table 13: Mean Accuracy and standard deviation of Convolutional Neural Network with 2 convolutional layers with 5-fold cross validation for the Y-Axis chunks

Chunk Size	CSF	WM	GM
	RAM Dataset		
5	0.50 (+/- 0.04)	0.49 (+/- 0.03)	0.52 (+/- 0.02)
10	0.51 (+/- 0.02)	0.50 (+/- 0.02)	0.50 (+/- 0.02)
15	0.56 (+/- 0.25)	0.50 (+/- 0.02)	0.50 (+/- 0.02)
20	0.60 (+/- 0.30)	0.50 (+/- 0.01)	0.50 (+/- 0.00)
	Social Feedback Dataset		
5	0.50 (+/- 0.02)	0.50 (+/- 0.02)	0.50 (+/- 0.02)
10	0.50 (+/- 0.02)	0.50 (+/- 0.02)	0.50 (+/- 0.02)
15	0.57 (+/- 0.26)	0.51 (+/- 0.01)	0.51 (+/- 0.01)
20	0.75 (+/- 0.24)	0.51 (+/- 0.01)	0.50 (+/- 0.02)
	RAM + Social Feedback Dataset		
5	0.50 (+/- 0.01)	0.50 (+/- 0.01)	0.50 (+/- 0.01)
10	0.50 (+/- 0.01)	0.50 (+/- 0.01)	0.50 (+/- 0.01)
15	0.57 (+/- 0.28)	0.51 (+/- 0.01)	0.51 (+/- 0.01)
20	0.77 (+/- 0.29)	0.50 (+/- 0.02)	0.50 (+/- 0.02)

Table 14: Mean Accuracy and standard deviation of Convolutional Neural Network with 3 convolutional layers with 5-fold cross validation for the Y-Axis chunks

Figures

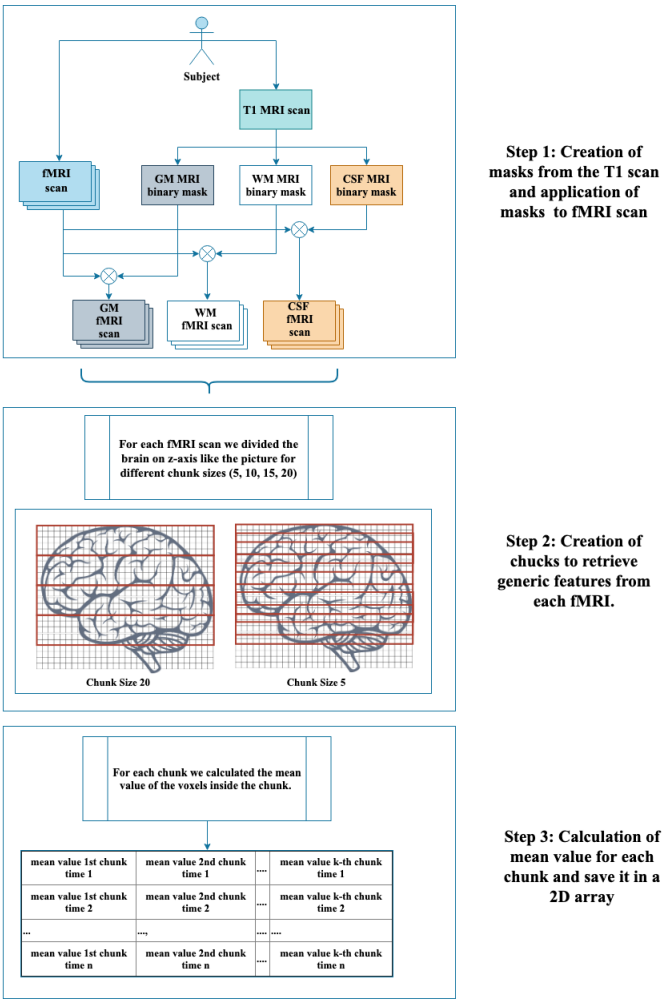

Figure 1: Feature extraction diagram

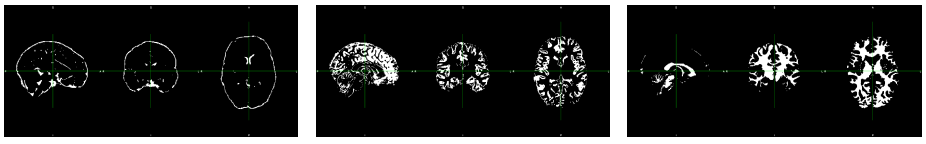

Figure 2: Examples of masks from left to right, Cerebrospinal fluid, Gray Matter and White Matter

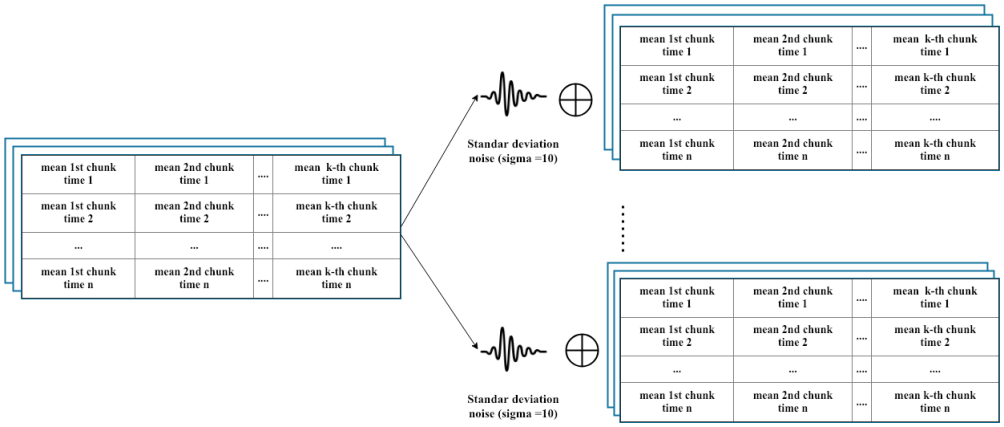

Figure 3: Data Expansion Process

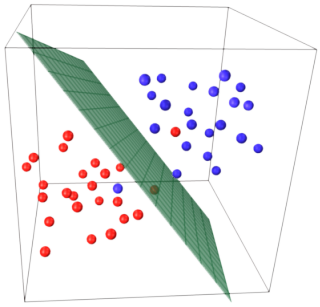

Figure 4: Support Vector Machine Illustration

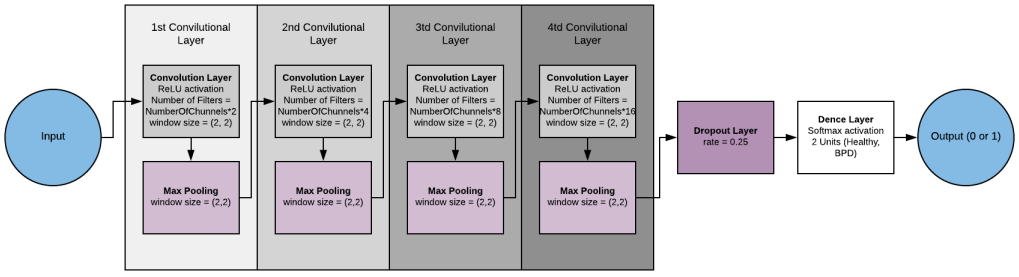

Figure 5: Convolutional Neural Network that we build for our experiments

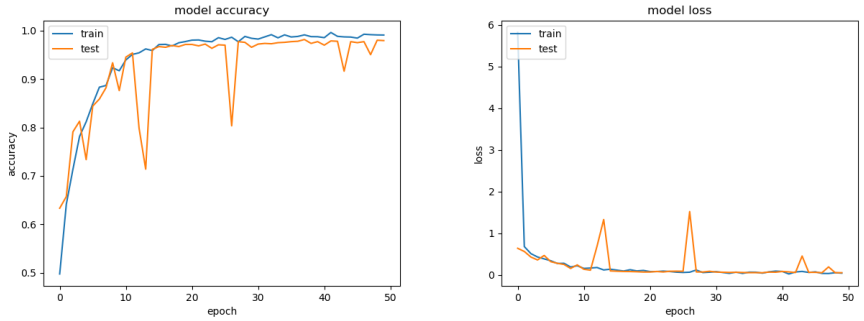

Figure 6: Accuracy and Loss of Three layers CNN of CSF for the RAM and SF Dataset with chunk size 15