UKRAINIAN CATHOLIC UNIVERSITY

BACHELOR THESIS

Diagnosis of neurological and psychiatric diseases based on whole-brain functional connectivity using Machine Learning techniques

Author:
Sofiia TATOSH

Supervisor:
Marco ZORZI

A thesis submitted in fulfillment of the requirements for the degree of Bachelor of Science

in the

Department of Computer Sciences
Faculty of Applied Sciences

Lviv 2022

Declaration of Authorship

I, Sofiia TATOSH, declare that this thesis titled, "Diagnosis of neurological and psychiatric diseases based on whole-brain functional connectivity using Machine Learning techniques" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"A real loser is someone who's so afraid of not winning he doesn't even try."

Grandpa from "Little Miss Sunshine"

UKRAINIAN CATHOLIC UNIVERSITY

Faculty of Applied Sciences

Bachelor of Science

Diagnosis of neurological and psychiatric diseases based on whole-brain functional connectivity using Machine Learning techniques

by Sofiia TATOSH

Abstract

The global problem this thesis aims to target is the inability of psychotherapists and psychiatrists always correctly to identify a presence of a mental illness. To give a constructive medical conclusion on a patient's state, usually, it is not enough to only rely on symptoms concluded from a therapeutic session. Moreover, the diagnosis of that kind could be biased from both a therapist and a patient's side. The former depends on the doctor's knowledge and experience, and the latter is based on an ability to communicate the mental state. Notably, the more researchers investigate the cause of psychiatric diseases, the more they make sure that mental illnesses are developed due to specific changes in one's brain. It could be the brain's structure, functionality, or damage, leading to changes in a person's behavior, thought process, interaction with other people, and sometimes difficulties in functioning as a healthy human being. It is believed that severe mental illnesses and neurological and developmental diseases result from abnormal connectivity in a brain network. That is why whole-brain functional connectivity is a significant source of information in this study. In simple words, it represents if and how the brain regions communicate with each other. This study presents generalized, usable, and reliable classification models to identify a specific neurological or psychiatric disease, Autism Spectrum Disorder and Schizophrenia, with 92.4% and 93.8% of accuracy respectfully, for further clinical application of the developed tool.

Acknowledgements

I am sincerely grateful to my supervisor, Marco Zorzi, for agreeing to cooperate with me on my Bachelor's thesis and for all of the recommendations, comments, and shared knowledge. It was a pleasure to work with prof. Zorzi, I gained much valuable experience, both technical and professional. I would also like to thank my parents, close friends, and my boyfriend, who were incredibly supportive during the process of thesis writing, during failures and hard times, and during the success at what I was doing. Thanks to the Ukrainian Catholic University, especially Applied Sciences Faculty, for providing me an opportunity to spend four years in a student-friendly environment. Special thanks to Yuliia Kleban for not giving up on IT and Business Analytics program students' progress on a thesis, keeping us up with all of the updates and deadlines, tracking every student's workflow, and being a great program coordinator in general.

The most important thing is to thank the Ukrainian army forces for letting Ukrainian students continue their studies and graduate from the universities by defending our country every day. Slava Ukraini!...

Contents

List of Figures

List of Tables

List of Abbreviations

ML	**M**achine **L**earning
HC	**H**ealthy **C**ontrol
ASD	**A**utism **S**pectrum **D**isorder
MDD	**M**ajor **D**epressive **D**isorder
BD	**B**ipolar **D**isorder
SCZ	**SC**hi**Z**ophrenia
fMRI	functional **M**agnetic **R**esonance **I**maging
sMRI	**s**tructural **M**agnetic **R**esonance **I**maging
ADHD	**A**ttention **D**eficit **H**yperactivity **D**isorder
FC	**F**unctional **C**onnectivity
AI	**A**rtificial **I**ntelligence
MRI	**M**agnetic **R**esonance **I**maging
MR	**M**agnetic **R**esonance
BOLD	**B**lood **O**xygenation **L**evel **D**ependent
rs	**r**esting **s**tate
LDA	**L**inear **D**iscriminant **A**nalysis
KNN	**K**-**N**earest **N**eighbors
SVM	**S**upport **V**ector **M**achine
ABIDE	**A**utism **B**rain **I**maging **D**ata **E**xchange
AUC	**A**rea **U**nder the **C**urve
ROC	**R**eciever **O**perator the **C**haracteristic
EEG	**E**lectro**E**ncephalo**G**ram
MNI	**M**ontreal **N**eurological **I**nstitute
ROI	**R**egion **O**f **I**nterest
PCA	**P**rincipal **C**omponent **A**nalysis
ICA	**I**ndependent **C**omponent **A**nalysis
NV	**N**uisance **V**ariable
CR	**C**onfound **R**egression
OLS	**O**rdinary **L**east **S**quares
ANN	**A**rtificial **N**eural **N**etwork
RMSprop	**R**oot **M**ean **S**quare propagation
DT	**D**ecision **T**ree

Dedicated to Orysia and Mykola…

Chapter 1

Introduction

Recent studies on the detection rates of severe psychiatric diseases show that more than a third of patients get misdiagnosed [4]. The most common ones are depression and bipolar disorder, following other disorders like ADHD and schizophrenia. Major Depressive Disorder shows a 65.9% of misdiagnosis rate when the bipolar disorder has around 92.7% [57]. In general, mental illnesses are hard to diagnose correctly, especially considering that there is no medical test to detect the presence of a particular disorder. The same works with neurological and developmental disorders, like autism spectrum disorder. Identifying ASD can be complicated because of the influence of many developmental factors outside of the ASD symptoms list. It could be an influence of age or language level, as some of the developmental factors [25].

To investigate the issue more profoundly, researchers have analyzed the correspondence between a mental disorder and its physical representation or symptoms on a physical level for the past several decades. According to the experimental results conducted during the past two decades, there is strong evidence that psychiatric and developmental disorders are dependent on the interactions between connected neural systems rather than some damage to a specific brain region [19]. It is also known that psychiatric diseases indicate problems in brain structure or its abnormal functioning. As a result, a supportive method of proper diagnosis is to ask specific questions about the patient's brain. Depending on a question, there are multiple techniques to use and non-invasive tools to investigate the brain. In this research, we will work with resting-state fMRI as a primary source of information. Functional MRI represents a brain's functioning throughout time. Unlike structural MRI, fMRI is used to determine the brain's metabolic function. We used functional connectivity derived from fMRI to understand the communication process among brain parts better. Functional connectivity is a term that is used to represent a map of inter-regional neural functional connections in the brain [3].

Studying the brain at rest, especially in analyzing the problems of psychiatric disease diagnosis, is more valuable than task-based fMRI. First, the brain consumes 20% of energy, where around $60 - 80\%$ is used to support communication between cells. Task-evoked activity account for 1% of increase in regional activity [18]. Thus, we can fully understand the brain's functional organization using rs-fMRI because the patterns that are aimed to support tasks are maintained at rest.

This work aims to develop an ASD and SCZ classifier based on whole-brain functional connectivity, generalized across demographic information and some of the other disorders. The classifiers could help to understand the reference among mental illnesses, identify a presence of an abnormality, and observe inter-regional relations responsible for proper differentiating between healthy controls (HC) and ones with a disorder. The main challenges we have to face are reducing data dimensionality caused by the number of features per subject, getting rid of unwanted

influence on the data, proper modeling and interpretation of classifiers, and gaining knowledge in brain anatomy to understand the problem from a biological point of view.

Chapter 2

Background

2.0.1 Human brain and functional connectivity

It is often said that the brain is the most powerful and complex network globally. It is still 30 times more potent than a supercomputer, even in the evolving world of AI. About 100 billion (10^{11}) of neurons are connected by around 100 trillion (10^{14}) synapses[1]. Moreover, it is not about the ability to calculate fast and not only about computational resources. It is about a lot more opportunities for a human brain to develop and a wide variety of options to degrade it.

Because a brain is a complex network, structurally separated regions are linked together and are constantly communicating with each other. In mathematical terms, it can be imagined as a graph with N nodes and E edges, where a single edge contains information about the connection between two structural regions of the brain. A graph itself is a $N * N$ adjacency matrix, in which zero and non-zero elements represent the absence and presence of a connection. The term for everything explained above is called connectome or connectivity.

In most cases, researchers use functional Magnetic Resonance Imaging to obtain any information about the functional relationships between regions (fMRI). It reflects the changes in signal as a response to particular experimental manipulation. An fMRI experiment consists of a sequence of individual MR images, where we can study oxygenation changes in the brain across time[18]. The most common approach toward fMRI uses the Blood Oxygenation Level Dependent (BOLD) contrast. The BOLD methodology enables us to identify which brain areas are more active than the others at a certain point in time. Technically, more active areas of the brain tend to receive higher levels of oxygenated blood.

Notably, MRI image contains information in a three-dimensional (3D) space. It is composed of voxels. Voxel is a 3D unit of the image with a single value, the same as a pixel for 2D[43]. Each voxel represents a spatial location and has an intensity associated with it. The BOLD signal is obtained from a single voxel through time. Two regions of the brain show functional connectivity if there is a statistical relationship between the measures of activity recorded for them[17]. With the BOLD signals, one can get a functional connectivity matrix, a correlation matrix of all the time series extracted.

[1]Synapse - the site of transmission of electric nerve impulses between two nerve cells (neurons)[18]. Britannica, The Editors of Encyclopaedia. "synapse". Encyclopedia Britannica, 18 Feb. 2011, https://www.britannica.com/science/synapse. Accessed 22 May 2022.

2.0.2 Mental health and functional connectivity

Modern neuroscience demonstrates the relation between mental diseases and brain abnormalities. Neuroscience also shows that severe psychiatric illnesses are associated with deviations in brain function, structure, and connectivity[30]. Functional connectivity is currently a widely used information source for dealing with problems related to mental health. Understanding the large-scale brain networks such as connectomes is gaining widespread acceptance. The questions of early diagnosis, and treatment reliability testing, were targeted mainly by looking into specific brain regions potentially responsible for a disease occurrence. Now it is known that psychiatric disorders are usually associated with abnormalities throughout the brain, which means it is not about a single region dysfunction[19].

With the development of the cognitive neuroscience field, which tries to answer the question of how the brain as an organ triggers the mind; and with the founding of new research fields like computational psychiatry, which aims to model the human brain in different diseases mathematically, an evolving interest of scientists to connectomic techniques to understand brain network abnormalities in various mental health issues occurs. This kind of curiosity around the topic has several reasons to exist. They are reliability, sustainability, and relative non-complexity of the data[65]. As the research topic gains a high level of excitement among scientists, more disorders are brought to the attention. The popular ones now are Parkinson's disease, ADHD, ASD, bipolar disorder, SCZ, MDD, and others. The ones we will concentrate on in this research are Autism Spectrum Disorder and Schizophrenia.

Autism Spectrum Disorder

ASD is a neurodevelopmental disorder that affects how a person behaves, socializes, and interacts with others and with the environment. The disease includes repetitive behaviors that are especially noticeable during infancy and adolescence. Even though there is no available cure for ASD, doctors develop intensive treatments that can lead to tremendous changes in one's behavior and life.

There were studies directed to investigate the development of a brain of a person with ASD throughout lifetime[23]. Analysis of a structural MRI showed that there are abnormalities in gray[2] and white[3] matter when comparing Healthy Controls (HCs) and patients with ASD. As was mentioned before, region-specific experiments demonstrate differences between HC and ASD, but it is also essential to include whole-brain network analysis to observe an entire picture[58]. Hence, recent research on that topic enlightens the ASD brain from the structural and functional connectivity that covers the entire cerebral cortex[4]. From an FC standpoint, subjects with ASD show increased FC in some of the areas[61], whereas there are also studies depicting an opposite finding suggesting a presence of so-called underconnectivity[29].

[2]Neural tissue, especially of the brain and spinal cord that contains nerve-cell bodies as well as nerve fibers and has a brownish-gray color.
"Gray matter." Merriam-Webster.com Dictionary, Merriam-Webster, https://www.merriam-webster.com/dictionary/gray%20matter. Accessed 24 May. 2022.

[3]Neural tissue, especially of the brain and spinal cord that consists mainly of myelinated nerve fibers bundled into tracts, has a whitish color and typically underlies the cortical gray matter.
"White matter." Merriam-Webster.com Dictionary, Merriam-Webster, https://www.merriam-webster.com/dictionary/white%20matter. Accessed 24 May. 2022.

[4]Outermost layer of the brain consisting of convoluted gray matter.

Schizophrenia

Schizophrenia is one of the severe mental illnesses which affects how a person thinks and feels. The most common associations with SHZ are loss of reality and hallucinations that result in constant misunderstanding among family and friends. SHZ is considered a neurodevelopmental disorder and ASD, but they still have some differences. Unlike ASD, SHZ is usually diagnosed between the ages of 16 and 30, when the first episode of psychosis[5] comes into play. The symptoms are not only present during episodes but throughout daily life. These include loss of motivation, withdrawal from social life, and difficulty showing emotions[40].

The investigation of structural brain changes in schizophrenia exposes the ability to detect the shifts in gray and white matter. A breakthrough is that the structural changes can occur even prior to the onset of clinical symptoms, which can lead to early detection and treatment and prevent a disease from progressing[16]. Concerning FC analysis, people with schizophrenia possess less firmly integrated functional connectivity compared with HCs[36, 34].

2.1 Literature review

2.1.1 ASD classification related works

Today, there are already many works dedicated to mental disorders diagnosis. Different research teams work with different data. The datasets vary from answers to a questionnaire and demographical data as a relevant information resource to non-invasively gathered signal data from the brain. The most popular ones today are EEG and fMRI data.

To approach the question of ASD diagnosis prediction, Altay and Ulaş [1] research team developed classification models, specifically Linear Discriminant Analysis Classifier (LDA) and K-Nearest Neighbors (KNN), basing their learning on the response data obtained from a mobile application for ASD diagnosis. The results obtained have 90.8% and 88.5% of accuracy based on the situational questions[1]. The data of that type was also used along with the Support Vector Machine (SVM) classifier, getting valuable accuracy results as well[6]. Even though the classifying models may be accurate, well-projected, generalized, and adequately validated, this assessment is more valid in terms of the behavioral part of human life. However, it does not analyze the physical core of a disease.

One of the most recent studies on ASD classification using resting-state functional connectivity[63] developed a kernel SVM based on Autism Brain Imaging Data Exchange (ABIDE) fMRI data with gaining 69.43% of accuracy as the most optimal result. It is important to note that the team performed a reasonable amount of testing throughout their development pipeline, starting from choosing brain parcellation[6] technique up until Machine Learning (ML) model selection.

The other reference ASD classification study involves the application of Deep Learning techniques. They also worked on the ABIDE dataset and achieved 90.39%

[5] Conditions that affect the mind, where there has been some loss of contact with reality.
National Institute of Mental Health (2021). Understanding Psychosis. Retrieved May 24, 2022, from https://www.nimh.nih.gov/sites/default/files/documents/health/publications/understanding-psychosis/20-mh-8110-understandpsychosis.pdf.

[6] Defines distinct partitions in the brain, be they areas or networks that comprise multiple discontinuous but closely interacting regions.
Eickhoff, S.B., Yeo, B.T.T. Genon, S. Imaging-based parcellations of the human brain. Nat Rev Neurosci 19, 672–686 (2018). https://doi.org/10.1038/s41583-018-0071-7

of accuracy and an area under the curve (AUC) of 0.9738. Comparing the results to other works on the same dataset, their model achieved the best results almost in all of the metrics[32].

Finally, the research done on the dataset used in this thesis presented an ASD classifier performing with 85% of accuracy and covering 0.93 AUC. The classifier was developed to generalize it with correspondence to demographical data, such as age, sex, site[7], and others. The generalization was also tested regarding other psychiatric diseases, like SCZ, MDD, and ADHD. They discovered that the classifier could differentiate SCZ from HC but performs worse on MDD and ADHD[62].

2.1.2 Schizophrenia classification related works

There are much fewer potentially dependent research papers on SCZ classification based on functional connectivity compared to ASD. However, some of them provide essential information on the possibilities of approaching such problems. One of the latest works related to SCZ classification was working with EEG functional connectivity [5]. The feature selection process was based on graph metrics like degree[8], strength[9], and others. A Logistic Classifier with 10-fold cross-validation was tested and compared at different frequency bands of a signal.

Works on discriminative analysis of functional connectivity of fMRI and its application in SCZ diagnosis were brought to attention in Shen et al. [51]. The team implemented a classifier based on C-Means clustering [10] obtaining experimental results of 93.75% and 75% of accuracy for SCZ patients and healthy controls respectively[51]. The feature selection part used a correlation coefficient method to extract highly discriminative regions that reduced the number of features up to only 2.24% of all FCs. One of the significant concerns regarding the work mentioned above is the number of subjects used for training and prediction. It consists of 32 patients and 20 healthy controls. A small dataset usually leads to an inability to generalize the model and project it onto real-life problems.

One of the latest research on SCZ detection based on rs-FC deals with the multi-site problem and presents a Deep Learning classifier with the accuracy of approximately 85%[64]. A research team used multi-atlas for brain region separation in functional connectivity extraction. As a result, they measured the FC of the same image in different spaces of several atlases. The final classifier was constructed with three hidden layers and 100 nodes in each of them. The study also provides comparison results with other classifiers' performance.

[7] Refers to a location where fMRI scan was used to measure the brain activity of a subject.

[8] Number of edges connected to a node.

[9] A weighted variation of degree is the sum of all the neighboring link weights for each node.

[10] Algorithm is creating k numbers of clusters, assigning each data point to each cluster, and defining how strongly the data belongs to that cluster based on the distance to the cluster center.

Chapter 3

Methodology

3.1 Data

3.1.1 Dataset description

The dataset we are working with in this work is a multi-site, multi-disorder resting-state Magnetic Resonance Image database[54]. The database consists of rs-fMRI and structural images of the brain of 993 patients and 1,421 healthy controls, along with demographic information like age, sex, handedness, and specific to the subject's clinical assessments. There are four major datasets with different data restrictions and consistency levels. The one we used in our study is the SRPBS Multi-disorder Connectivity Dataset. The data were collected from 15 different scanners and eight sites. The data itself is resting-state functional connectivity of each subject and participant demographic information.

3.1.2 Functional Connectivity from raw fMRI

Preprocessing of raw fMRI data and rs-FC calculation was performed in the scope of SRPBS Multi-disorder Dataset creation. rs-fMRI data was preprocessed using SPM8[20], software implemented in MATLAB. It is an academic software for the analysis of functional imaging data[55]. As described in the paper, SPM8 preprocessing steps included slice-timing correction, realignment, co-registration, segmentation of T1-weighted structural images, normalization to Montreal Neurological Institute (MNI) space, and spatial smoothing with an isotropic Gaussian kernel of 6mm full width at half maximum (FWHM).

After the standard approach to rs-fMRI preprocessing and getting rid of the noise, they split the brain into 140 regions of interest (ROIs). ROIs were defined anatomically by Brainvisa Sulci Atlas[14] and three sub-regions of the cerebellum[1] (the left and right cerebellum, and the vermis[2]). The brain parcellation is depicted in figure 3.1. A full region naming is present in Appendix C (C.2, C.1).

Then BOLD signals extracted from those regions were put through a bandpass filter ($0.008 - 0.1Hz$). Notably, to generate a BOLD time series per region, the BOLD signal was extracted from each voxel and then averaged as by brain segmentation

[1] A large dorsally projecting part of the brain concerned especially with the coordination of muscles and the maintenance of bodily equilibrium.
"Cerebellum." Merriam-Webster.com Dictionary, Merriam-Webster, https://www.merriam-webster.com/dictionary/cerebellum. Accessed 31 May. 2022.

[2] The constricted median lobe of the cerebellum that connects the two lateral lobes.
"Vermis." Merriam-Webster.com Dictionary, Merriam-Webster, https://www.merriam-webster.com/dictionary/vermis. Accessed 31 May. 2022

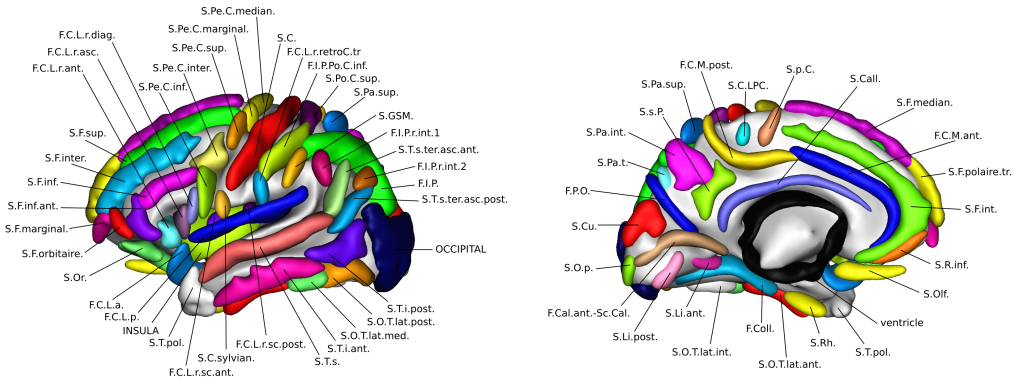


FIGURE 3.1: BrainVisa Atlas[14].

procedure applied. The next step was to regress out confounds which included temporal fluctuations of the white matter, the cerebrospinal fluid, and head motion parameters.

As a result, each subject was attached to 140 ROIs signals that were then pairwise correlated to each other using Pearson correlation:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

The output correlation matrix was vectorized to form a lower-triangular matrix consisting of 9,730 functional connections. Based on the previous research[45], they calculated the framewise displacement [3] and removed volumes with FD > 0.5mm.

Finally, the outcome of the preprocessing pipeline was flattened FC matrices with $m = 9,730$ features per subject.

3.1.3 SCZ and ASD functional connectivity data

To analyze SCZ and ASD data, we retrieved two main sub-datasets from the general one to balance out HCs and patients in amount of data points. We performed a similarity search to balance the data across subjects regarding demographic data like age and gender. It was made by calculating a pair-wise distance between vectors using normalized Euclidean distance. Normalized Euclidean distance helped deal with features having different variations, meaning one of the features (sex) was a binary one, and the other (age) was continuous.

$$NED = \sqrt{\sum_s \left(\frac{A_s}{|A|} - \frac{B_s}{|B|} \right)^2}$$

In the above equation, s denotes the subject's number, and A_s and B_s represent the age and sex of each subject, respectively. Similar healthy control was selected for each ASD or SCZ patient depending on the measure. The following datasets were obtained.

[3] This measure indexes the movement of the head from one volume to the next and is calculated as the sum of the absolute values of the differentiated realignment estimates at every timepoint.[44]

TABLE 3.1: SCZ dataset demographics.

SCZ / HC	**Sex**	**Age**	**Site**
HC	male 85 female 61	42.5±24.5	KyotoU 98 HiroshimaU 33 OsakaU 3 ShowaU 1 CiNet 7 UTO 4
SCZ	male 85 female 61	41.5±25.5	KyotoU 92 ShowaU 19 UTO 35

TABLE 3.2: ASD dataset demographics.

ASD / HC	**Sex**	**Age**	**Site**
HC	male 109 female 16	36.5±16.5	HiroshimaU 15 KyotoU 85 OsakaU 2 ATR 8 ShowaU 8 UTO 4 CiNet 3
ASD	male 109 female 16	37±17	ShowaU 115 UTO 10

3.2 FC selection for the classifier

Feature selection is the next major step in our data preparation. The FC dataset used in this study presents 9,730 functional connections per subject. We also have to consider the number of subjects we have, much lower than the number of features per subject. The problem is dealing with a so-called curse of dimensionality where $n << m$, where n is a number of subjects, and m is a number of features. In most cases, Machine Learning problems assume $n >> m$ because otherwise, it can cause overfitting. That is because some of the predictors are noise variables, so the model cannot be validated correctly in future dataset[67]. We applied and validated several dimensionality reductions and feature selection techniques to tackle this problem.

3.2.1 Dimensionality reduction

Reducing amount of variables can be split into two main approaches. One of them keeps the most relevant features given in the dataset but does not modify an input. This approach is called feature selection. The other one is selecting a smaller set of features that contain the same information as input data, usually using a linear combination of the given variables. It is called dimensionality reduction.

Dimensionality reduction, especially in neuroimaging data, is one of the essential steps to successfully building a robust and non-overfitting ML model. According to the recent study on feature extraction methods comparison in resting-state functional connectivity, the best dimensionality reduction methods were Principal Component Analysis (PCA) and Independent Component Analysis (ICA). The validity

of the methods was tested on rs-FC of stroke patients predicting ability of neuropsychological scores[9].

PCA

The main idea behind Principal Component Analysis (PCA) is to reduce the dimensionality of the data and, at the same time, keep as much variability present in the dataset as possible. PCA algorithm linearly transforms the dataset into a smaller amount of variables, principal components (PCs), that are uncorrelated and sorted in the way that the first ones retain most of the variance.

The main intention in PCA is to perform matrix decomposition of m by n matrix X with m features and n subjects, such that

$$X = USV^T$$

where U is m by m matrix and $U^TU = I$, V is n by n matrix and $V^TV = I$, and S is an m by n singular-value matrix. V is a matrix containing the eigenvectors[4] of X^TX. To decompose matrix X, we need to find the eigenvalues[5] and eigenvectors of X^TX. Applying singular value decomposition, we get

$$X^TX = VS^TU^TUSV^T = VS^TSV^T$$

where S^TS results into n by n zero-valued matrix except having eigenvalues of X^TX on its diagonal.

$$X^TXV = VS^TSV^TV = VS^TS$$

V is a matrix of eigenvectors, or in other words, principal components. Hence, we can conclude that

$$X^TXv_k = \lambda_k v_k$$

where λ_k is k-th eigenvalue and v_k is an eigenvector of X^TX. The components are sorted by importance, that is why the highest value λ_k has, the more important v_k is for describing the data variance[42].

In our work, we performed the PCA reduction method in the way that at least 80% of variance is explained by the reduced data. The maximum we tried to sustain was 95% of information described by the output from PCA. We iteratively increased the percentage by 5% to find the most convenient choice of variables. Before the PCA application, the data was centered on a zero mean.

ICA

Another one of the most widely used dimensionality reduction techniques is Independent Component Analysis (ICA). An important difference between PCA and

[4] A nonzero vector that is mapped by a given linear transformation of a vector space onto a vector that is the product of a scalar multiplied by the original vector.
"Eigenvector." Merriam-Webster.com Dictionary, Merriam-Webster, https://www.merriam-webster.com/dictionary/eigenvector. Accessed 27 May. 2022.

[5] A scalar associated with a given linear transformation of a vector space.
"Eigenvalue." Merriam-Webster.com Dictionary, Merriam-Webster, https://www.merriam-webster.com/dictionary/eigenvalue. Accessed 27 May. 2022.

ICA is that PCA transforms the data into a set of uncorrelated[6] variables, when ICA looks for independent[7] factors.

To understand the process of ICA, we can look at the statistical latent variables[8] model. Let's assume that we observe $x_1, x_2, ..., x_n$ such that

$$x_j = a_{j1}s_1 + a_{j2}s_2 + ...a_{jn}s_n$$

The x_j variables are observable, when s_k are independent components. Let's assume $\mathbf{x}$ is a random vector of elements $x_1, x_2, ..., x_n$, and $\mathbf{s}$ is a random vector of elements $s_1, s_2, ..., s_n$. Let's also introduce a matrix $\mathbf{A}$ which elements are a_{ij}. Then the ICA statistical model can be represented as

$$\mathbf{x} = \mathbf{As}$$

The ICA model is a generative model, which means that it explains how the observed data are generated by mixing the components s_i. As was mentioned before, the independent components obtained are latent variables, which means they cannot be observed. At the same time, matrix $\mathbf{A}$ is unknown. That is why the main goal is to estimate the matrix and independent components vector. As soon as we estimate the matrix, we can get an independent components set[26].

$$\mathbf{s} = \mathbf{A^{-1}x}$$

Before the ICA application, the input dataset was centered and whitened[9][10]. After that, we iteratively estimated the results of choosing from 10 to 30 independent components with 5 components step.

3.2.2 Feature selection

Random Forest feature selection

Random Forest ML algorithm is popular in applying supervised regression and classification problems. It is one of the most widely used algorithms because of its simplicity, low overfitting rates, and good performance. In general, the algorithm uses many decision trees built based on random subset selection from the dataset. For classification problems, it then uses the majority vote to decide on the test observation result. An important fact is that the features used to train Random Forest are split randomly into particular decision trees, which means not all trees get the same feature set. Notably, each tree is a particular sequence of yes-no questions based on selected features.

As decision trees build, one has to decide which feature will be considered next in the data subset on each step. To select a feature for splitting, Gini Index is used in classification models. It measures an impurity of the split. A pure split is getting either "yes" or "no". Gini Index can be calculated as follows:

$$GI = 1 - \sum_{i=1}^{n} (P_i)^2 = 1 - \left[(P_+)^2 + (P_-)^2\right] \tag{3.1}$$

[6]There is no linear relation between the variables.

[7]Not dependent on other variables.

[8]A variable that cannot be observed.

[9]Data transformation to make it have a covariance matrix as an identity matrix (1 on the diagonal, 0 otherwise)

Where P_+ is the probability of a positive class and P_- is the probability of a negative class[47]. This reflects how much impurity a node has.

To calculate the node's importance, we use

$$ni_j = w_j C_j - w_{left(i)} C_{left(j)} - w_{right(j)} C_{right(j)}$$

where ni_j is the importance of node j, w_j is a weighted number of samples reaching node j, C_j reflects the impurity value of node j, $left(j)$ and $right(j)$ is a child node from left and right split on node j respectively.

Feature importance on a single decision tree is calculated using

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k \in all nodes} ni_k}$$

Then each feature importance value is normalized by dividing the calculated fi_i by the sum of all feature importance values and is equal to $normfi_i$. To generalize the feature importance score across all of the decision trees involved in a Random Forest algorithm, the last step is to average the fi_i score among all of the trees. This is done as

$$RFfi_i = \frac{\sum_{j \in all\ trees} normfi_{ij}}{T}$$

where $normfi_{ij}$ is a normalized feature importance for i in tree j and T equals a total number of trees[31].

Backward feature elimination

Backward feature elimination is much simpler from a mathematical standpoint in comparison to the ones discussed above. The first step is to take all m features from the input dataset and train a model on them. The model we have chosen to deal with is Logistic Regression (see Logistic Regression explanation in 3.4.1), one of the standard classifier models. Then we calculate the performance of the model. After that, we recursively take out one feature, so there would be one feature less than during the previous run at each iteration. During the process of elimination, we identify the variable that did not influence the model's performance after its removal and drop the variable. We repeat this process until the variables can no longer be removed. That would mean that we reached our desired number of components[24].

Forward feature selection

The forward feature selection method is the opposite of the backward feature elimination approach. The naming says by itself that in forward feature selection, we select new features, and in backward feature elimination, we get rid of them.

The algorithm starts with training a model with a single feature. Our model is trained m times during the first step. The variable that gives the best performance is selected as a starting point for adding one more variable. The feature that influences a model's performance is selected to be kept in further steps. The process is repeated until there is no significant change in a model's performance[24].

3.3 Confound Regression

3.3.1 A problem of nuisance variables

Nuisance variables, or confounding variables, are some factors that influence the independent variable within an experimental study but are not of research interest. It introduces bias to analyzing data. Usually, when dealing with medical data, confounding variables are demographic, like age, gender, and ethnicity. Sometimes they also include the presence of a treatment or tool specifications' bias that was used to perform clinical tests. In fMRI data, the confound variables are also the scanner parameters, experiment setup, and imaging conditions.

3.3.2 Confound regression approach

One of the most common ways of controlling confounds in the dataset is to remove the variance explained by a confound. That is called confound regression[52], which means that we regress out the influence of an unwanted variable. This method consists of fitting a linear regression model on each feature of the data (FC) with the confounds as predictors.

We constructed a confound matrix C to identify the relationships between demographic data and each feature from the FC matrix. The variables we consider to nuisance are age, sex and site. The number of columns in matrix C is 11, that is $p = 11$. First column represents age, which is a continuous variable. The next column is sex containing 1 as male and 0 as female. The next 8 columns contain information about the site in a form of [1 0 0 0 0 0 0 0 0] for site 0, [0 1 0 0 0 0 0 0 0] for site 1, [0 0 1 0 0 0 0 0 0] for site 2 and so on. The last column is either 1 (ASD / SCZ) or 0 (HC). As a result, each feature is modelled as a linear function of confounding variables in the following way:

$$X_i = C\beta_i + \epsilon$$

where X_i is a single feature vector for n subjects, C is a n by p confound matrix, where p is an amount of confound variables we consider.

To estimate the parameters $\hat{\beta}_i$ for a feature X_i we use ordinary list squares as follows:

$$\hat{\beta}_i = \left(C^T C\right)^{-1} C^T X_i$$

To regress the confounds from the data, we subtract the variance associated with a confound from the original data.

$$X_{i,corr} = X_i - C\hat{\beta}_i$$

$X_{i,corr}$ represents a feature vector X_i where the variance associated with confounds was removed. It is important to note that the influence of a disease label is not considered in regressing out the variance, which means the $\hat{\beta_{i,disease}}$ is automatically set to 0.

3.4 Machine Learning models

3.4.1 Mathematics behind ML models

Logistic Regression

Logistic Regression is one of the most common classifying ML models. It predicts a binary outcome based on a probability of an observation appearing either in the first or second group. Because Logistic Regression is a linear classifier, it uses a linear function called logit.

$$f(\mathbf{x}) = b_0 + b_1 x_1 + ... + b_m x_m$$

A core of Logistic Regression is a sigmoid function, which predicts the probability.

$$sigmoid(\mathbf{x}) = \frac{1}{1 + e^{-f(\mathbf{x})}} \tag{3.2}$$

Model training is based on maximizing the log-likelihood function for all observed data.

$$LLF = \sum_i (y_i log(sigmoid(\mathbf{x}_i)) + (1 - y_i) log(1 - sigmoid(1 - \mathbf{x}_i))$$

In ML, it is conventional to work with log-loss minimization using gradient descent rather than maximizing an objective function. The negative value of LLF is a cost function we want to minimize. Mathematically

$$-LLF = -\sum_i (y_i log(sigmoid(\mathbf{x}_i)) + (1 - y_i) log(1 - sigmoid(1 - \mathbf{x}_i))$$

where y_i represents a class and $log(sigmoid(\mathbf{x}_i))$ represents a probability measure[2].

Ridge

In general, Ridge, Lasso, and Elastic Net models are called Regularized Regression models. They are developed to eliminate the sparsity and overfitting problem when the model performs well on train data but fails to do as well on unseen data. A regression vector is sparse of only some of its components are nonzero while the rest is set to zero[7]. This type of model aims to reduce variance by introducing some bias. Ridge, for instance, reaches regularization by reducing the importance of some features, whereas Lasso completely blocks the importance given to some of them[41]. In Ridge regression, the OLS loss function penalizes the size of parameter estimates and is also called the L2 regularization model[39].

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^{n} (y_i - x_i^{'}\hat{\beta})^2 + \lambda \sum_{j=1}^{m} \hat{\beta}_j^2 = ||y - \mathbf{X}\hat{\beta}||^2 + \lambda ||\hat{\beta}||^2$$

Here λ factor is controlling the strength of the penalty. This parameter is often tuned using the cross-validation technique. The main impact on the model's better performance is caused by adding a shrinkage estimator, meaning all of the coefficients are shrunk by the same factor so that none of them are dropped.

Elastic Net

Elastic Net model is a combination of Ridge and Lasso ones, where the penalty term is added to a loss function from both of them. The main advantage of Elastic Net is that because of combination of two different penalties it allows to cover up limitations of both methods[68]. Therefore, Elastic Net minimizes the following function:

$$L_{enet}(\hat{\beta}) = \frac{\sum_{i-1}^{n}(y_i - x_i^{'}\hat{\beta})^2}{2n} + \lambda\left(\frac{1-\alpha}{2}\sum_{j-1}^{m}\hat{\beta}_j^2 + \alpha\sum_{j-1}^{m}|\hat{\beta}_j|\right)$$

where λ and α are parameters regularizing the penalization strength.

Decision Tree

A decision tree algorithm is a supervised model operating on a tree datatype. The root node is the beginning of a decision tree, and it is the first feature that starts splitting the data. Decision nodes are responsible for preserving conditional information about the further split. The next split would be based on the decision node. A leaf node is the one that makes a final decision, and no splitting then is possible[12]. As was mentioned before, the selection of features is based on a purity measure (3.1).

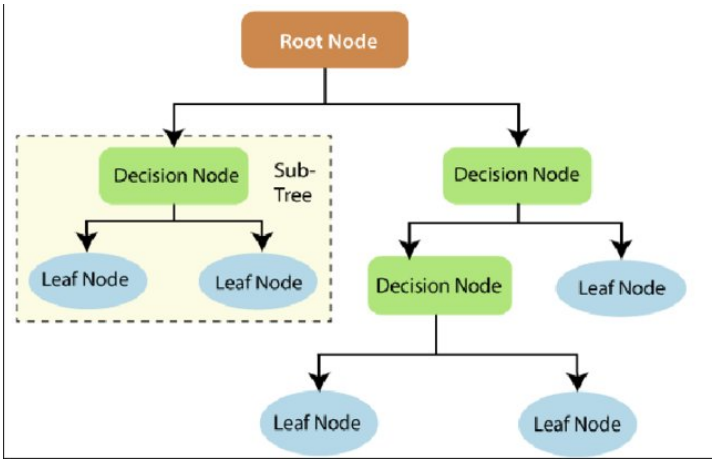


FIGURE 3.2: Decision Tree template[12].

Support Vector Machine

Support Vector Machine (SVM) is a linear ML algorithm primarily used for solving classification problems; that is also known as Support Vector Classification. In the case of binary classification, we refer to SVM as a problem of finding a hyper-plane so that it divides the n-dimensional data into two separate groups. The goal of the SVM algorithm is to choose an optimal hyper-plane, and here support vectors come into play.

A black line in the above figure is an optimal hyper-plane, whereas the dotted lines are two other planes that pass through the nearest data points to the optimal one. The area with no points (the one between two dotted hyper-planes) is called the margin, and the closest points to an optimal hyper-plane are known as support vectors. The equation for a hyperplane is

$$H : \mathbf{w}^T(\mathbf{x}) + \mathbf{b} = 0$$

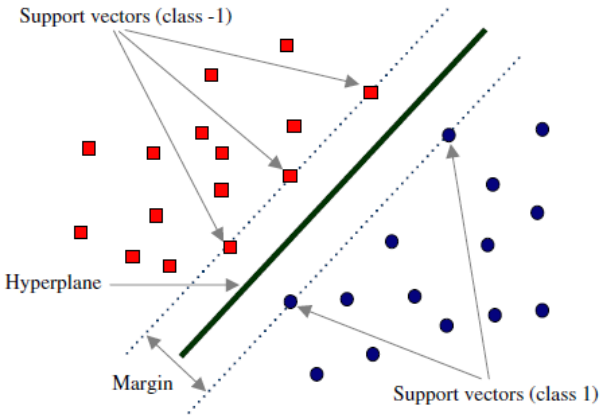


FIGURE 3.3: SVM linearly-separable example[56].

The primary goal of SVM is to maximize margin so that the model can be generalized well on the training dataset and perform better on unseen data. For that, we would like to present a distance measure from a given point vector to a hyperplane as

$$d_H(\phi(x_0)) = \frac{|w^T(\phi(x_0)) + b|}{||w||_2}$$

where $\phi(x_0)$ is a point vector and $||w||_2$ is a Euclidean norm given by

$$||w||_2 = \sqrt{w_1^2 + w_2^2 + ... + w_n^2}$$

To find a hyperplane with the maximum margin, we must maximize the minimum distance to the closest points (support vectors). The objective function becomes

$$w^* = arg_w max[min_n d_H(\phi(x_n))] = arg_w max \frac{1}{||w||_2}, s.t.\ min_n y_n[w^T\phi(x_n) + b] = 1 \tag{3.3}$$

In most cases working with real-life data, the data points are not linearly separable. Making non-linearly separable data into linearly separable ones is a kernel trick. The kernel is a function responsible for projecting data to a higher dimensional space where linear separation becomes possible. It can be expressed as an inner product in another space. The kernel is usually defined as a feature map:

$$\omega : \phi \rightarrow \psi$$

that satisfies

$$k(x, x^{'}) = (\omega(x), \omega(x^{'}))\psi$$

Kernel plays an important role in so-called dual form of the objective function. Dual form differs from the primal form (3.3) by using different variables to serve the same purpose[15].

K-nearest Neighbors

KNN algorithm is developed in the way that it finds the closest point in a given dataset to the input point. Using Euclidean distance measure, it determines the neighbors and the input point distance to them. The distance metric is

$$d(x, x^{'}) = \sqrt{(x_1 - x_1^{'})^2 + ... + (x_n - x_n^{'})^2} \tag{3.4}$$

The input point gets assigned to a class by calculating the probability of appearing in that class.

$$P(y = j|X = x) = \frac{1}{K}\sum_i I(y^{(i)} = j)$$

A value K is a hyperparameter that denotes an amount of nearest neighbors to which the distance is measured during the process of input point class assignment[66].

Gaussian Naive Bayes

Naive Bayes is a classification algorithm that is based on Bayes' theorem. It calculates the probability of an entry observation being assigned to each class, and the one with the highest probability is a predicted class. The Bayes rule states

$$P(C = i|\mathbf{X} = \mathbf{x}) = \frac{P(\mathbf{X} = \mathbf{x}|C = i)P(C = i)}{P(\mathbf{X} = \mathbf{x})}$$

$P(\mathbf{X} = \mathbf{x})$ is identical for all the classes, so that it can be omitted. Hence, the Bayes classifier

$$h^*(\mathbf{x}) = arg\, max_i P(\mathbf{X} = \mathbf{x}|C = i)P(C = i)$$

that finds the maximum posterior probability given an observation $\mathbf{x}$. Usually, $P(\mathbf{X} = \mathbf{x}|C = i)$ is hard to calculate because of the high dimensionality of the data, which is why it is rather approximated than directly estimated. It is performed by assuming that all of the features are independent in a given feature vector[48]. That is why the objective function to maximize is

$$f_i^{NB}(\mathbf{x}) = \prod_{j=1}^{n} P(X_j = x_j|C = i)P(C = i)$$

XGBoost classifier

Gradient Tree Boosting is one of the techniques in ML that show excellent results in a lot of different applications. At every iteration, a base learner[10] is fit to the negative gradient performing gradient descent on the loss function. XGBoost is a model that consists of a collection of base learners, and the model is trained in an additive manner[13]. The loss-function at iteration t we want to minimize is

[10]Is usually generated from training data by a base learning algorithm which can be a decision tree, neural network, or other kinds of ML algorithms.
Zhou ZH. (2009) Ensemble Learning. In: Li S.Z., Jain A. (eds) Encyclopedia of Biometrics. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-73003-5_293

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \Omega(f_t)$$

where

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda ||w||^2$$

Here y_i is a true label from the training dataset, and Ω is responsible for penalizing the complexity of the model. The objective function cannot be optimized using traditional methods in Euclidean space[13], so the Taylor expansion to calculate the value of a loss function for the base learners. Briefly, Taylor Series is used to approximate a function based on first, second, third, and so on, derivatives, and the factorial.

Artificial Neural Networks

Artificial Neural Networks (ANNs) are computational models whose structure and logic were inspired by the brain's biological structure. That is why ANNs consist of many connected nodes, or in other words, neurons that perform a particular mathematical operation. It also looks like directed graphs, where each node is responsible for a particular calculation, the output of which then is passed to the next connected node. This process is called an activation. It is usually represented as $f(z)$, where z is an aggregation of all the input. The activation functions are either linear or nonlinear, where linear follows the concept of $f(z) = z$, and others perform a special transformation.

The most popular activation functions are Rectified Linear Units (ReLU), sigmoid activation (3.2) and *tanh*. With ReLU, we make sure that the output is not going below zero, meaning

$$f_{ReLU}(z) = max(0, z) \tag{3.5}$$

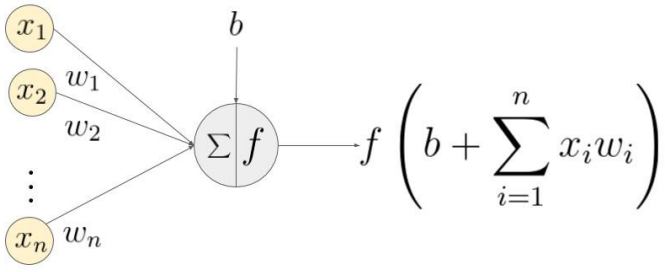


FIGURE 3.4: Artificial neuron example[50].

In the above figure there is a function

$$f\left(b + \sum_{i=1}^{n} x_i w_i\right) \tag{3.6}$$

where b is a bias term, x is an input to a neuron, w are weights, n is the number of inputs from the incoming layer.

ANNs learn based on two major concepts - backpropagation and optimization. Backpropagation computes the gradient of the loss function with respect to the weights. Weights are connections between neurons that carry a value, and the higher the value, the more importance is dedicated to a neuron on the input side. Generally, cross-entropy is used as a loss function for solving classification problems with ANNs. On the other hand, optimization is the selection of the best element from a given set of variables' alternations. Mainly it is about choosing the best weights for the model.

3.4.2 Evaluation metrics

Accuracy

Accuracy is the fraction of predictions our model got right. For binary classification problems, it is calculated in terms of positives and negatives[37].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.7}$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

Recall and precision

Precision and recall are the measures that use the same concept as accuracy, but those metrics work well on imbalanced datasets.

$$Precision = \frac{TP}{TP + FP} \tag{3.8}$$

$$Precision = \frac{TP}{TP + FN} \tag{3.9}$$

Precision shows the proportion of positive identifications that were actually correct. Recall depicts the proportion of actual positives that were identified correctly[37].

F-1 score

In binary classification, the F-1 score measures test accuracy calculated from the precision and recall. The score represents a harmonic mean of precision and recall[37].

$$F_1 = \frac{2}{recall^{-1} + precision^{-1}} = 2\frac{precision \cdot recall}{precision + recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{3.10}$$

Pseudo R-Squared

For binary classification problems, OLS R-squared is not suitable because of dealing only with continuous predictions. Classification is otherwise a discrete categorical one; there are several ways of measuring so-called pseudo-R-Squared. Efron's R-Squared is one of the most used ones, especially for Logistic Regression model evaluation.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{\pi}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{3.11}$$

where $\hat{\pi}$ are model predicted probabilities. Efron's R-Squared explains variability, as the denominator can be considered as the total variability in the dependent variable, and the numerator of the ration can be considered as the variability in the dependent variable that is not predicted by the model. The more variability explained by the model, the better it is[8].

AUC-ROC curve

The Receiver Operator Characteristic (ROC) curve is an evaluation matric for binary classification models. It is a probability curve that plots recall against precision at various threshold values. The Area Under the Curve (AUC) is a measure that shows the ability of a classifier to identify the class of an input observation. The performance of the model is good if the AUC is high[28].

Log-loss

Log-loss is also one of the major metrics for evaluating classification models. It is known that the binary classification first predicts the probability of an observation to be related to class 1. This prediction probability is a value log-loss is dependent on. Log-loss indicates how close the prediction probability is to the actual value. The higher the log-loss value, the more the predicted probability is different from a true class. Log-loss value is calculated as

$$Logloss_i = -[y_i ln(p_i) + (1 - y_i) ln(1 - p_i)]$$

The log-loss score for a classifier is an average of all of the observations' log-loss values.

$$Logloss = -\frac{1}{N}\sum_{i=1}^{N}[y_i ln(p_i) + (1 - y_i) ln(1 - p_i)]$$

Chapter 4

Experimental results

4.1 Confound Regression results

While almost all of the nuisance variables were regressed from BOLD signals, as mentioned in the data preprocessing part, the demographic features are still included in functional connectivity. The biggest concern in working with the datasets we have obtained is the potential influence of sites in identifying disease. Both ASD and SCZ have an unbalanced distribution of data across sites, leading to the model's learning specifications about a site but not about a disease per Se. To check the possibility of an influence of site bias, we tried to perform feature selection on FC data using the approaches described below, selected the best one by a performance of Logistic Regression, and used those 20 features to classify the sites, but not the disease. Because this is a multi-class problem, we developed a CatBoost multi-class classifier model to check the model's ability to distinguish among given classes. We performed classification on the ASD and SCZ datasets along with healthy controls and disease patients separately for both sets.

CatBoost is a boosted decision tree machine learning algorithm. It is a gradient boosting algorithm that can be used both in the form of regression and classification. Briefly, a series of N classifiers are trained, and weights are updated in a way that the following classifiers will pay more attention to a training set of features that were misclassified by a previous classifier[46].

After the confound regression, we repeated the steps of control runs for site classification with the help of the CatBoost model (See table 4.1). The performance results were obtained by 5-fold cross-validation, and accuracy results were calculated on the test set only.

The baseline accuracy was calculated using Zero Rule Algorithm. It implies that the ML model predicts only the values of the major class so that the baseline accuracy would be a number of subjects of a prevalent class divided by a number of all subjects.

TABLE 4.1: Sites classification results

Data	**% accuracy before CR**	**% accuracy after CR**	**Baseline % accuracy**
ASD	99%	88%	92%
HC from ASD set	70%	64%	68%
HC + ASD	72%	58%	49.2%
SCZ	72%	65.3%	63%
HC from SCZ set	68.3%	63.3%	67%
HC + SCZ	68.8%	62.3%	65%

4.2 Feature selection results

To decide which features work best in explaining the difference between ASD / SCZ and healthy controls, we applied feature selection algorithms to regressed FCs data. All of the feature selection approaches followed the same pipeline. We set up a range of features to try out from 10 up to 30 with a 5 features step. Except PCA had to deal with a percentage of explained variance from 80% to 95%. At each iteration, a 5-fold cross-validation was performed to evaluate models' (Logistic Regression) prediction success. The best performing feature selection algorithm turned out to be Backward Feature Elimination with 20 features. Test accuracy, an area under the curve, and log-loss were the main model metrics. We also considered a number of features, especially in the case of the PCA approach. Thus, according to the metrics we obtained, even though PCA performed slightly better in accuracy, it underperformed the number of features. The least features chosen by PCA were no less than 100 in both cases. By the overall performance comparison, the Backward Feature Elimination algorithm with 20 FCs turned out to be the most suitable one. The results are presented in Appendix B (B.2, B.1).

4.3 Experiments' results

We trained Machine Learning models on regressed ASD and SCZ datasets as the features were selected. The models included were Elastic Net, Ridge, Decision Tree Classifier, SVC, KNN, Gaussian NB, XGBoost, and ANN.

Almost all of the developed models went through a similar grid search parameters tuning process, where a range of values was proposed for some of the major ones. For instance, the Elastic Net classifier was trained and cross-validated with a $0 < l1\_ratio < 1$ range of L1 and L2 penalty combination with a 0.1 step. The exact process was performed for the Ridge classifier, changing the L2 penalty value. The Decision Tree Classifier was set to work with Gini Index to measure the quality of the split. For SVC γ value was selected as $\frac{1}{n}$ with all of the other parameters set to default.

As for ANN, we developed a four-layer Feedforward Neural Network. The activation functions used on hidden layers were ReLU; the output layer had a sigmoid function because of the classification problem. Grid search was performed on batch size[1], number of epochs[2], optimizer type and amount of neurons.

RMSprop optimizer was selected as the best option considering the data and other model parameters. RMSprop is a gradient-based optimizer developed for mini-batch learning. It works by using a moving average of squared gradients to normalize the gradient. It simply means that RMSprop is an adaptive learning approach, and the learning rate changes over time[49].

Models were cross-validated with a 5-fold Stratified Cross-Validation technique. Stratified cross-validation makes sure that training and test sets have the same proportion of the feature of interest as in the original dataset. 80/20 rule was applied to split the data into subsets. In this way, we ensure the generalization of metrics calculated. As soon as we split the dataset into train and test, we start the confound regression algorithm on the train set, obtaining β coefficients from OLS regression. Those coefficients are saved for future application to test data to ensure that test observations are as unknown to the model as possible.

[1] Number of samples processed before the model is updated.
[2] The number of complete passes through the training dataset.

We applied several metrics to evaluate models' performance, including accuracy, recall and precision scores, F-1 score, pseudo R-squared and log-loss. Based on the metrics' results, we concluded the best performing model, Gaussian Naive Bayes for ASD data and NuSVC for SCZ data. The best test accuracy for the ASD dataset is 92.4% with a ROC-AUC score equal to 0.95. For the SCZ dataset, accuracy results were up to 93.8% and a ROC-AUC score of 0.98. The results obtained from all of the models are presented below.

TABLE 4.2: Classification results on ASD dataset.

Model	**Acc. %**	**Efron's R^2**	**F-1**	**Recall**	**Precision**	**AUC**	**Log-loss**
Log. Reg.	90.8	0.63	0.90	0.90	0.91	0.95	0.26
Elastic Net	91.2	0.65	0.91	0.90	0.92	0.95	0.24
Ridge	91.2	0.65	0.91	0.91	0.91	0.94	0.2
DT	73.2	-0.07	0.73	0.71	0.76	0.73	0.71
SVC	90.8	0.63	0.92	0.94	0.89	0.93	0.23
NuSVC	88.4	0.54	0.89	0.90	0.88	0.93	0.20
KNN	87.6	0.50	0.88	0.90	0.87	0.93	0.27
Gauss. NB	92.4	0.70	0.92	0.92	0.92	0.95	0.16
XGB	84.8	0.39	0.85	0.86	0.84	0.91	0.24
ANN	91.2	0.65	0.91	0.90	0.91	0.92	0.23

TABLE 4.3: Classification results on SCZ dataset.

Model	**Acc. %**	**Efron's R^2**	**F-1**	**Recall**	**Precision**	**AUC**	**Log-loss**
Log. Reg.	93.5	0.74	0.94	0.95	0.93	0.98	0.24
Elastic Net	92.1	0.69	0.92	0.92	0.92	0.98	0.24
Ridge	92.5	0.70	0.93	0.92	0.94	0.98	0.12
DT	69.9	-0.2	0.68	0.65	0.72	0.70	0.18
SVC	92.8	0.71	0.93	0.91	0.95	0.98	0.14
NuSVC	93.8	0.75	0.94	0.92	0.95	0.99	0.20
KNN	88.7	0.55	0.88	0.86	0.92	0.95	0.31
Gauss. NB	92.8	0.71	0.93	0.93	0.93	0.98	0.24
XGB	83.2	0.33	0.82	0.81	0.85	0.94	0.25
ANN	92.8	0.73	0.93	0.95	0.93	0.99	0.21

As we can observe in Table 4.3, the best performing model based on almost all of the metrics is NuSVC. NuSVC is the same model as SVC mathematically, but the only difference is that NuSVC uses parameter ν to control the number of support vectors. Its value represents an upper bound on the fraction of margin errors and a lower bound on the support vector fraction. As training is done on roughly the same dataset (features were selected using the same algorithm), we can compare the NuSVC model and Gaussian Naive Bayes model on SCZ data. We can see that the difference in metrics variates by ± 0.05, which is not a huge gap. So, even though a NuSVC classifier performs slightly better than Gaussian NB, Gaussian NB can still be a reliable model for implementation. It means the pipeline of building a classifier can be generalized without considering a disease.

4.3.1 Models generalization

One of the goals we set in this work is to develop a generalized classifier, which will look for differences between healthy and mentally ill patients without considering any other influences of demographic data. Key points were not to base the model on only one site but try to cover as many as possible; regress out the influence of demographic data; check whether classifiers can be generalized to other diseases, and find the similarities between them on a brain level, if present.

Our datasets contain data from around seven different sites; the influence of fMRI scanners and sites per Se, along with age and sex, were omitted from the FC. After developing the classifiers, we examined the ability to predict patients with other psychiatric diseases. We used accuracy results and ROC-AUC scores as significant generalizability identifiers.

We attempted to use a set of diseases for testing a hypothesis, including ASD, SCZ, bipolar disorder, and Major Depressive Disorder (MDD). Our prior assumption is that if accuracy and ROC-AUC scores are around 0.5, the model cannot distinguish between patients with the disorder and healthy controls. On the other hand, if the values are closer to 1, the classifier can differentiate one having mental problems from the others. Below are the results we obtained from the generalizability tests.

TABLE 4.4: Generalizability measures of ASD and SCZ classifiers on other psychiatric disorders.

Data	ASD		Data	SCZ	
	Acc. (%)	ROC-AUC		Acc. (%)	ROC-AUC
SCZ	51.2%	0.51	ASD	60%	0.65
BD	83.6%	0.87	BD	72.8%	0.81
MDD	48.8%	0.52	MDD	46.3%	0.44

As we can see, both of the classifiers did well in generalizing to bipolar disorder, having a 0.8+ AUC score. That tells us that there is some relationship between ASD, SCZ, and BD, which helps differentiate them from healthy controls. ASD classifier did not show any correspondence to SCZ patients, even though one of the recent works on ASD classifier generalization[62] shows that SCZ has a 0.65 ROC-AUC score and is the most relatable disease to ASD among SCZ, ADHD, and MDD. Even though the ASD classifier did not manage to generalize to SCZ, it happened vice versa. SCZ classifier performs slightly better on ASD data. The reason could be either feature selected for an ASD classifier that does not suit SCZ data or because the SCZ dataset had more observations and hence, a bigger train set compared to the ASD dataset. From the above results (4.4 we can conclude that there exists a possible relationship between ASD, SCZ, and BD on the brain functional level, which can be explained by functional connectivity.

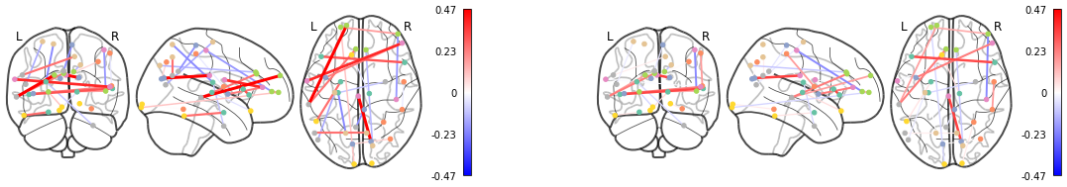
The relationship between ASD, schizophrenia, and bipolar disorder was investigated. It included both symptomatic and external similarities, as well as physiological ones. Autistic-like traits were present in people with bipolar disorder and schizophrenia. According to the investigation[38], there is shared pathophysiology among ASD, SCZ, and BD. It is also proven that there exists an overlap between ASD and psychotic illnesses, meaning patients with ASD can potentially experience schizophrenia and bipolar disorder at the same time[33]. The studies on physiological similarities among those disorders also show a potential overlap. One of the studies provides evidence of diseases' genetic factors representing common biological overlap[11]. At the molecular level, gene expressions in the brain share some

physical characteristics, implying a direct relationship between ASD, SCZ, and BD. It depicts a potential causal genetic component responsible for risk factors, formation, and development of those illnesses[21].

4.3.2 Selected features from biological standpoint

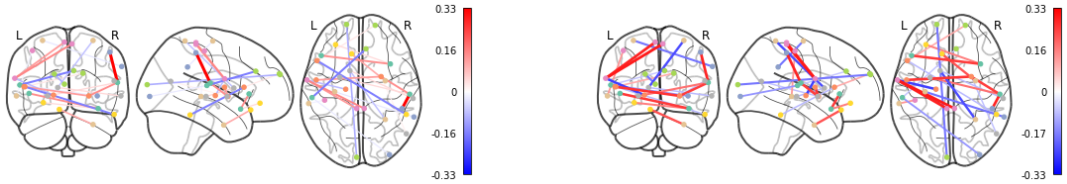
As we performed feature selection for each model, it is important to look at the functional connectivity features selected and the same brain regions responsible for it. We would also like to check the FC patterns' hyper- or under- connectivity presence.

On figures 4.1 and 4.2 we depicted the FCs of selected features to visually compare intensiveness of the connectivity among subjects. The FCs plotted are the average among all of the HCs, ASD, and SCZ patients used in the research so that it is possible to cover not only a single case but also see the whole picture.



(A) ASD (B) HC

FIGURE 4.1: ASD vs HC Functional Connectivity.



(A) SCZ (B) HC

FIGURE 4.2: SCZ vs HC Functional Connectivity.

From the above figures, we clearly can see that ASD connectivity seems to be much higher than the one for a control group (4.1). The opposite, however, corresponds to SCZ and controls, where the intensity of FCs is higher for HC in comparison to SCZ subjects (4.2). It can was mentioned in Mental Health and FCs chapter (2.0.2), that Schizophrenia tends to possess less strongly integrated functional connectivity when comparing to HCs[36, 34]. Moreover, ASD patients tend to have hyper-connectivity patterns[61].

Figures 4.3 and 4.4 show a closer look on a relationship among the brain regions, the connection values, and a type of connection. By the type of connection, we mean whether the connection is more intense (black line) or less intense (dotted line) than healthy controls. We depicted the connectivity pattern between brain regions selected as the distinguishing ones in the form of a graph, where nodes are brain regions, edges are connections, and the FC value is captured between the two. Nodes naming is present in Appendix C (C.1, C.2).

According to the studies on hyper- and hypo-connectivity of ASD and SCZ, ASD is believed to be prone to hyper-connectivity between primary sensory and paralimbic, primary sensory and association, and paralimbic and association sybsystems[53]. The paralimbic subsystem consists of the insula, anterior cingulate cortex,

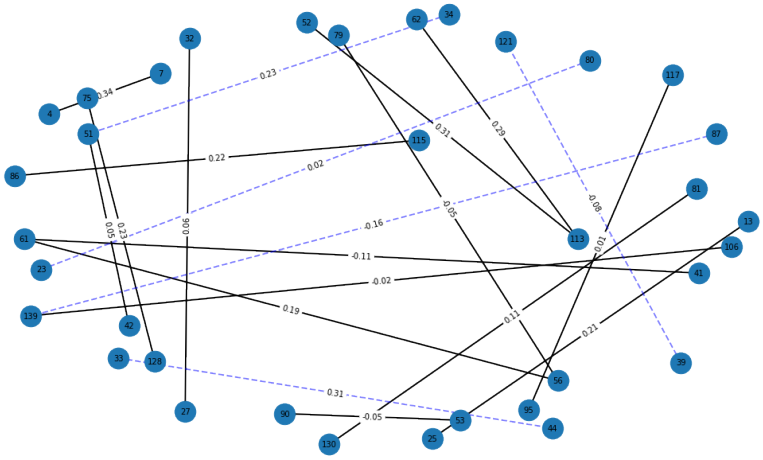


FIGURE 4.3: Graph representation of selected ASD FCs and it's values as edges. Dotted line means $\mathbf{fc}_{ASD} \leq \mathbf{fc}_{HC}$, thick black line corresponds to $\mathbf{fc}_{ASD} > \mathbf{fc}_{HC}$.

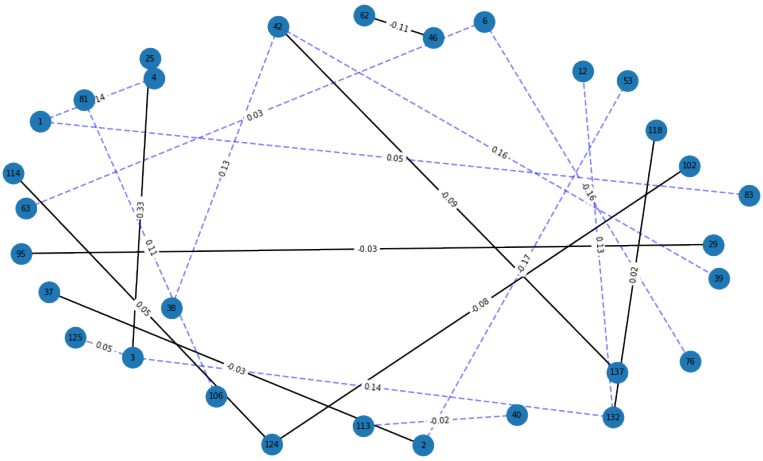


FIGURE 4.4: Graph representation of selected SCZ FCs and it's values as edges. Dotted line means $\mathbf{fc}_{SCZ} \leq \mathbf{fc}_{HC}$, thick black line corresponds to $\mathbf{fc}_{SCZ} > \mathbf{fc}_{HC}$.

posterior cingulate cortex, and the orbitofrontal cortex, while association areas include the lateral frontal and parietal cortices. Other works state that there is an obvious hyper-connectivity between the thalamus and cortical regions like the temporoparietal junction and posterior cingulate cortices, or amygdala and some other cortical regions[27]. On Figure 4.3 we observe many functional connections values higher in ASD than in HC. Comparing the regions involved in those intense connections, we can conclude that the results somewhat relate to the Supekar et al. [53] investigation. We have noticed that some of the sub-regions of the primary cortices correspond to the ones mentioned in the research([53], [27]). For instance, there is a strong connection between the occipitotemporal lateral sulcus and amygdala (ref. C.2), where the sulcus relates to the os-called Lingual Gyrus (marked yellow on Figure 4.5). According to Iidaka et al. [27], the hyper-connectivity is present within the

lingual gyrus and amygdala.

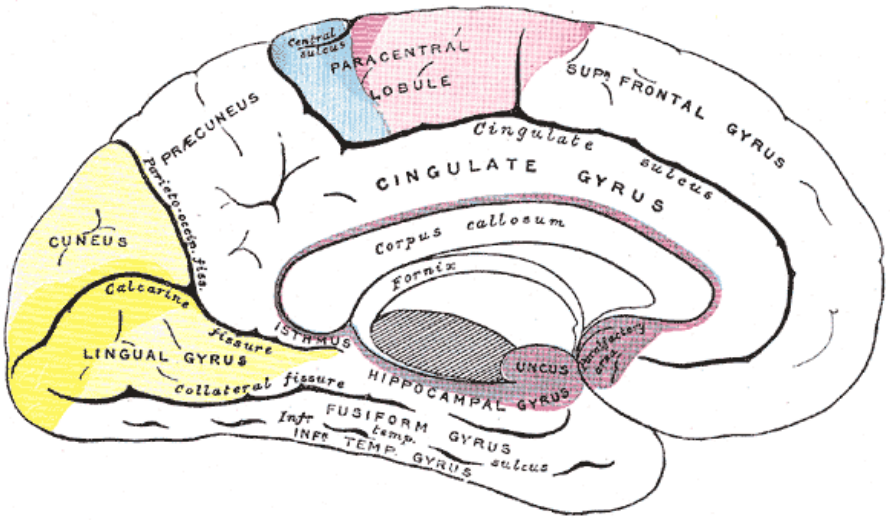


FIGURE 4.5: Areas of localization on medial surface of hemisphere. Motor area in red. Area of general sensations in blue. Visual area in yellow. Olfactory area in purple. The psychic portions are in lighter tints[22].

As for schizophrenia, it was characterized by hyper-connectivity in motor cortices to the thalamus, motor cortices to the cerebellum, and prefrontal cortex to subthalamic nucleus[59]. We have a strong relationship between the central sylvian sulcus and a ventricle among the selected functional connections. The central sylvian sulcus is related to the motor cortex, and a ventricle is one of four ventricles in the cerebellum. That again proves a correspondence between the reliability of our research and other available works.

Chapter 5

Applications

5.1 Market overview

Artificial Intelligence plays a priceless role in every human's life. It surrounds us everywhere; it can entertain us, help us do our work, and save our lives.

Since the pandemic started, the demand for psychological therapy has increased, leading to higher workloads among professionals.

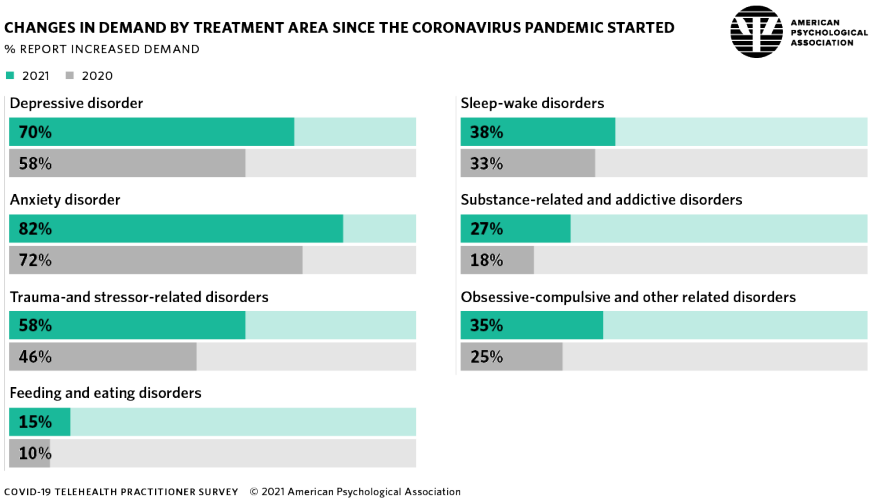


FIGURE 5.1: Changes in demand by treatment area since the coronavirus pandemic started[60].

Such an increase in demand results from various factors that are now present in almost every human-caused by a rapid change in the style of living, habits, and relationships with other people.

According to a recent interview with Dr. Ellen Lee, assistant professor of psychiatry, University of California San Diego, it is quite an issue in the psychiatry field not to have specific biomarkers of diseases to conclude a diagnosis based on biological pathologies[35]. Because of an absence of medical tests that could clearly state the patient's problems, doctors rely only on the symptoms provided by a patient. Usually, it does not cover that different people can have different symptoms but still be diagnosed with the same disorder.

As an observed tendency in demand for mental care increases, many improvements in patients processing speed, the accuracy of diagnosis, and preventive treatment, could be applied. There are many different ML solutions for mental care support that cover various issues, from short-term support or symptoms treatment to

proper therapy type selection. One of those tools is Ieso Digital Health. They provided an AI solution to their platform to help therapists and patients be on the same page with tracking the insights from sessions and predictions of patients' current state. It integrates cognitive behavioral therapy online to support mental health care provision clinicians.

There are some other startups on the market directed to patients as users. They assist patients through their mental illness process (Wysa), provide access to real-world interactions with AI coaches/therapists (Happify), or provide therapy chatbots to track one's mood (Woebot).

5.2 Use cases

The use cases of the developed classifiers can vary a lot depending on a primary goal. If appropriately modeled, they can play a significant role not only in a diagnosis of disease but bio-markers identification, personalized treatment, supporting mechanism to track the process of a disease, and hence a possibility to change treatment depending on the current state.

The most direct option is to integrate a model into a more complex system, following the concept of Ieso Digital Health. The diagnostic tool, available to psychotherapists, with an ability to obtain a hint on potential diagnosis and develop therapy sessions based on the results from medical data. The diagnosis results can not only be used for figuring out whether there is any mental illness but can help in choosing the best therapist match. As soon as the therapist's choice and treatment are personalized, it is time to track the progress, where the classifiers could also be helpful. With the additional support of other personalized tools, it also can provide information on the direction in treatment, explanation of symptoms, and its relation to the diagnosis given by a model. Thus, the whole mental illness journey has the potential to be much more successful from the starting point, covering the duration up till the patient is no longer having issues because of the disease's influence.

Chapter 6

Conclusion and future work

As the result of this thesis, we developed generalized, applicable, and reliable classification models for Autism Spectrum Disorder and Schizophrenia based on resting-state fMRI functional connectivity data. To develop a pipeline for model training, we tested several feature selection algorithms validated on each of the used datasets. We also applied the confound regression approach to limit the effect of bias coming from variations of fMRI scanners and sites.

Consequently, we trained and validated different classical Machine Learning algorithms for classification. The selected models were properly evaluated and managed to gain 92.4% and 93.8% of accuracy, respectively. Other metrics to evaluate performance were applied and considered in a final decision.

The classifiers were able to generalize to bipolar disorder, well-meaning an existence of a relationship between ASD, SCZ, and bipolar disorder. The hypothesis was supported by some of the works related to investigating this interrelationship. We also managed to overview the functional connectivity and brain regions playing a significant role in driving our models.

One of the limitations of this research is data. It would be great to work with the data collected from all over the world, not only from one country (in our case - Japan). It may be possible that styles of living, the influence of society, and genetics are also important factors to consider in future work. By broadening the data range, we will cover more cases with more variability in demographics; thus, the amount of observation will also expand.

The whole work was a journey into new topics, both from data science and medical points of view. Much new information and interesting results were gained from other scientists' work and a possibility to combine various approaches to similar problems.

The future steps for the work include extending the subjects' amount, data from other countries' institutions, and revisiting the developed models with new data. It would also be an excellent point to dive deeper into a psychiatric and medical spectrum of the problem to understand better the processes behind brain functions and their dependencies on mental illness specifications.

Appendix A

Appendix A

TABLE A.1: Consortium Sites Data.

Site	Scanner	Participants	Number of data
ATR	Siemens	Obsessive compulsive disorder (OCD) Healthy controls (HC)	OCD 4 HC 108
University of Tokyo	GE, Philips	Autism spectrum disorder (ASD) Major depressive disorder (MDD) Schizophrenia (SCZ) Bipolar disorder (BPD) Dysthymia Other psychiatric disorders (Other) Healthy controls (HC)	ASD 10 MDD 62 SCZ 35 BPD 41 Dysthymia 4 Other 28 HC 170
Osaka University	Siemens	Intractable neuropathic pain (INP) Post-stroke without pain (PSP) Healthy controls (HC)	INP 43 PSP 10 HC 29
Showa University	Siemens	Autism spectrum disorder (ASD) Schizophrenia (SCZ) Healthy controls (HC)	ASD 115 SCZ 19 HC 101
Kyoto University	Siemens	Schizophrenia (SCZ) Depression (DEP) Healthy controls (HC)	OSCZ 92 DEP 16 HC 234
Hiroshima University	GE, Siemens	Depression (DEP) Healthy controls (HC)	DEP 173 HC 261

Appendix B

Appendix B

TABLE B.1: Feature selection results on SCZ dataset.

FS alg.	No. of features	Var. expl.	Acc. %	AUC	Log-loss
PCA features	147	0.8	0.77	0.87	0.43
PCA features	172	0.85	0.79	0.88	0.55
PCA features	200	0.9	0.78	0.87	0.55
PCA features	235	0.95	0.81	0.88	0.52
ICA	10	-	0.79	0.89	0.55
ICA	15	-	0.81	0.88	0.55
ICA	20	-	0.81	0.87	0.41
ICA	25	-	0.79	0.88	0.51
ICA	30	-	0.79	0.89	0.51
Random Forest	10	-	0.83	0.92	0.30
Random Forest	15	-	0.85	0.93	0.31
Random Forest	20	-	0.86	0.95	0.32
Forward FS	10	-	0.72	0.80	0.26
Forward FS	15	-	0.72	0.79	0.26
Forward FS	20	-	0.73	0.79	0.26
Backward FE	10	-	0.90	0.97	0.33
Backward FE	15	-	0.93	0.99	0.31
Backward FE	20	-	0.96	0.99	0.27

TABLE B.2: Feature Selection results on ASD dataset.

FS alg.	**No. of features**	**Var. expl.**	**Acc. %**	**AUC**	**Log-loss**
PCA features	128	0.8	0.92	0.97	0.39
PCA features	149	0.85	0.93	0.98	0.41
PCA features	174	0.9	0.95	0.98	0.41
PCA features	203	0.95	0.96	0.99	0.28
ICA	10	-	0.97	0.99	0.28
ICA	15	-	0.99	0.99	0.24
ICA	20	-	0.99	0.99	0.23
ICA	25	-	0.99	0.99	0.22
ICA	30	-	0.99	0.99	0.22
Random Forest	10	-	0.90	0.97	0.26
Random Forest	15	-	0.93	0.98	0.28
Random Forest	20	-	0.94	0.99	0.28
Backward FE	10	-	0.97	0.99	0.22
Backward FE	15	-	0.98	0.99	0.21
Backward FE	20	-	0.99	0.98	0.17
Forward FS	10	-	0.90	0.98	0.31
Forward FS	15	-	0.92	0.98	0.30
Forward FS	20	-	0.94	0.99	0.30

Appendix C

Appendix C

TABLE C.1: BrainVisa parcellation regions for SCZ dataset.

Index	BrainVisa acronym	Label
1	F.C.L.a._right	Right anterior lateral fissure
2	F.C.L.p._left	Left posterior lateral fissure
3	F.C.L.p._right	Right posterior lateral fissure
4	F.C.L.r.ant._left	Left anterior ramus of the lateral fissure
6	F.C.L.r.asc._left	Left ascending ramus of the lateral fissure
12	F.C.L.r.sc.ant._left	Left anterior sub-central ramus of the lateral ...
25	F.I.P.Po.C.inf._right	Right superior postcentral intraparietal superi...
29	F.I.P.r.int.1_right	Right primary intermediate ramus of the intrapa...
37	OCCIPITAL_right	Right lobe occipital
38	S.C.LPC._left	Left paracentral lobule central sulcus
39	S.C.LPC._right	Right paracentral lobule central sulcus
40	S.C._left	Left central sulcus
42	S.C.sylvian._left	Left central sylvian sulcus
46	S.Cu._left	Left cuneal sulcus
53	S.F.int._right	Right internal frontal sulcus
62	S.F.polaire.tr._left	Left polar frontal sulcus
63	S.F.polaire.tr._right	Right polar frontal sulcus
76	S.O.T.lat.med._right	Right median occipito-temporal lateral sulcus
81	S.Olf._left	Left olfactory sulcus
83	S.Or._left	Left orbital sulcus
95	S.Pe.C.marginal._left	Left marginal precentral sulcus
102	S.Po.C.sup._right	Right superior postcentral sulcus
106	S.Rh._right	Right rhinal sulcus
113	S.T.s._left	Left superior temporal sulcus
114	S.T.s._right	Right superior temporal sulcus
118	S.T.s.ter.asc.post._right	Right posterior terminal ascending branch of th...
124	caudate._left	Left Caudate
125	putamen._left	Left Putamen
132	putamen._right	Right Putamen
137	ventricle._left	Left ventricle

TABLE C.2: BrainVisa parcellation regions for ASD dataset.

Index	BrainVisa acronym	Label
4	F.C.L.r.ant._left	Left anterior ramus of the lateral fissure
7	F.C.L.r.asc._right	Right ascending ramus of the lateral fissure
13	F.C.L.r.sc.ant._right	Right anterior sub-central ramus of the lateral...
23	F.Coll._right	Right collateral fissure
25	F.I.P.Po.C.inf._right	Right superior postcentral intraparietal superi...
27	F.I.P._right	Right intraparietal sulcus
32	F.P.O._left	Left parieto-occipital fissure
33	F.P.O._right	Right parieto-occipital fissure
34	INSULA_left	Left insula
39	S.C.LPC._right	Right paracentral lobule central sulcus
41	S.C._right	Right central sulcus
42	S.C.sylvian._left	Left central sylvian sulcus
44	S.Call._left	Left subcallosal sulcus
51	S.F.inf.ant._right	Right anterior inferior frontal sulcus
52	S.F.int._left	Left internal frontal sulcus
53	S.F.int._right	Right internal frontal sulcus
56	S.F.marginal._left	Left marginal frontal sulcus
61	S.F.orbitaire._right	Right orbital frontal sulcus
62	S.F.polaire.tr._left	Left polar frontal sulcus
75	S.O.T.lat.med._left	Left median occipito-temporal lateral sulcus
79	S.O.p._left	Left occipito-polar sulcus
80	S.O.p._right	Right occipito-polar sulcus
81	S.Olf._left	Left olfactory sulcus
86	S.Pa.int._right	Right internal parietal sulcus
87	S.Pa.sup._left	Left superior parietal sulcus
90	S.Pa.t._right	Right transverse parietal sulcus
95	S.Pe.C.marginal._left	Left marginal precentral sulcus
106	S.Rh._right	Right rhinal sulcus
113	S.T.s._left	Left superior temporal sulcus
115	S.T.s.ter.asc.ant._left	Left anterior terminal ascending branch of the ...
117	S.T.s.ter.asc.post._left	Left posterior terminal ascending branch of the...
121	S.s.P._left	Left sub-parietal sulcus
128	amygdala._left	Left Amygdala
130	thalamus._right	Right Thalamus
139	cerebral_white Matter._left	Left Cerebral White Matter

Bibliography

[1] Osman Altay and Mustafa Ulaş. "Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children". In: Mar. 2018, pp. 1–4. DOI: 10.1109/ISDFS.2018.8355354.

[2] Antônio Alves et al. "Read this paper if you want to learn logistic regression". In: *Revista de Sociologia e Política* 28 (Dec. 2020), p. 1. DOI: 10.1590/1678-987320287406en.

[3] Dirk Jan Ardesch, Lianne H. Scholtens, and Martijn P. van den Heuvel. "Chapter 6 - The human connectome from an evolutionary perspective". In: *Evolution of the Human Brain: From Matter to Mind*. Ed. by Michel A. Hofman. Vol. 250. Progress in Brain Research. Elsevier, 2019, pp. 129–151. DOI: https://doi.org/10.1016/bs.pbr.2019.05.004. URL: https://www.sciencedirect.com/science/article/pii/S0079612319301323.

[4] Getinet Ayano et al. "Misdiagnosis, detection rate, and associated factors of severe psychiatric disorders in specialized psychiatry centers in Ethiopia". In: *Annals of General Psychiatry* 20 (Feb. 2021), pp. 1–10. DOI: 10.1186/s12991-021-00333-7.

[5] Sima Azizi, Daniel B. Hier, and Donald C. Wunsch. "Schizophrenia Classification Using Resting State EEG Functional Connectivity: Source Level Outperforms Sensor Level". In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*. 2021, pp. 1770–1773. DOI: 10.1109/EMBC46164.2021.9630713.

[6] Mousumi Bala et al. "Efficient Machine Learning Models for Early Stage Detection of Autism Spectrum Disorder". In: *Algorithms* 15.5 (2022). ISSN: 1999-4893. DOI: 10.3390/a15050166. URL: https://www.mdpi.com/1999-4893/15/5/166.

[7] Lea Bottmer, Christophe Croux, and Ines Wilms. "Sparse regression for large data sets with outliers". In: *European Journal of Operational Research* 297.2 (2022), pp. 782–794. ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2021.05.049. URL: https://www.sciencedirect.com/science/article/pii/S037722172100477X.

[8] J. Bruin. *FAQ: WHAT ARE PSEUDO R-SQUAREDS?* Feb. 2011. URL: https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/.

[9] Federico Calesella et al. "A comparison of feature extraction methods for prediction of neuropsychological scores from functional connectivity data of stroke patients". In: *Brain Informatics* 8 (Dec. 2021). DOI: 10.1186/s40708-021-00129-1.

[10] Federico Calesella et al. "A comparison of feature extraction methods for prediction of neuropsychological scores from functional connectivity data of stroke patients". In: *Brain Informatics* 8 (Dec. 2021). DOI: 10.1186/s40708-021-00129-1.

[11] Liam Carroll and Michael Owen. "Genetic overlap between autism, schizophrenia and bipolar disorder". In: *Genome medicine* 1 (Oct. 2009), p. 102. DOI: 10.1186/gm102.

[12] Bahzad Charbuty and Adnan Mohsin Abdulazeez. "Classification Based on Decision Tree Algorithm for Machine Learning". In: 2021.

[13] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: Aug. 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.

[14] Yann Cointepas et al. "BrainVISA: Software platform for visualization and analysis of multi-modality brain data". In: vol. 13. June 2001, S98. DOI: 10.1016/S1053-8119(01)91441-7.

[15] Nello Cristianini and Elisa Ricci. "Support Vector Machines". In: *Encyclopedia of Algorithms*. Ed. by Ming-Yang Kao. Boston, MA: Springer US, 2008, pp. 928–932. ISBN: 978-0-387-30162-4. DOI: 10.1007/978-0-387-30162-4_415. URL: https://doi.org/10.1007/978-0-387-30162-4_415.

[16] Lynn Delisi et al. "Understanding structural brain changes in schizophrenia". In: *Dialogues in clinical neuroscience* 8 (Feb. 2006), pp. 71–8.

[17] S.B. Eickhoff and V.I. Müller. "Functional Connectivity". In: *Brain Mapping*. Ed. by Arthur W. Toga. Waltham: Academic Press, 2015, pp. 187–201. ISBN: 978-0-12-397316-0. DOI: https://doi.org/10.1016/B978-0-12-397025-1.00212-8. URL: https://www.sciencedirect.com/science/article/pii/B9780123970251002128.

[18] A. Fornito, A. Zalesky, and Edward Bullmore. *Fundamentals of Brain Network Analysis*. Mar. 2016, pp. 1–476.

[19] Alex Fornito and Ben Harrison. "Brain Connectivity and Mental Illness". In: *Frontiers in Psychiatry* 3 (2012). ISSN: 1664-0640. DOI: 10.3389/fpsyt.2012.00072. URL: https://www.frontiersin.org/article/10.3389/fpsyt.2012.00072.

[20] K. J. Friston et al. "Statistical parametric maps in functional imaging: A general linear approach". In: *Human Brain Mapping* 2.4 (1994), pp. 189–210. DOI: https://doi.org/10.1002/hbm.460020402. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.460020402. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.460020402.

[21] Michael Gandal et al. "Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap". In: *Science* 359 (Feb. 2018), pp. 693–697. DOI: 10.1126/science.aad6469.

[22] Henry Gray. *Anatomy of the human body*. Vol. 8. Lea & Febiger, 1878.

[23] Sungji Ha et al. "Characteristics of Brains in Autism Spectrum Disorder: Structure, Function and Connectivity across the Lifespan". In: *Experimental Neurobiology* 24 (Dec. 2015), p. 273. DOI: 10.5607/en.2015.24.4.273.

[24] Xuan Huang, Lei Wu, and Yinsong Ye. "A Review on Dimensionality Reduction Techniques". In: *International Journal of Pattern Recognition and Artificial Intelligence* 33.10 (2019), p. 1950017. DOI: 10.1142/S0218001419500174. eprint: https://doi.org/10.1142/S0218001419500174. URL: https://doi.org/10.1142/S0218001419500174.

[25] Marisela Huerta and Catherine Lord. "Diagnostic Evaluation of Autism Spectrum Disorders". In: *Pediatric clinics of North America* 59 (Feb. 2012), pp. 103–11, xi. DOI: 10.1016/j.pcl.2011.10.018.

[26] A. Hyvärinen and E. Oja. "Independent component analysis: algorithms and applications". In: *Neural Networks* 13.4 (2000), pp. 411–430. ISSN: 0893-6080. DOI: https://doi.org/10.1016/S0893-6080(00)00026-5. URL: https://www.sciencedirect.com/science/article/pii/S0893608000000265.

[27] Tetsuya Iidaka et al. "Thalamocortical Hyperconnectivity and Amygdala-Cortical Hypoconnectivity in Male Patients With Autism Spectrum Disorder". In: *Frontiers in Psychiatry* 10 (2019). ISSN: 1664-0640. DOI: 10.3389/fpsyt.2019.00252. URL: https://www.frontiersin.org/article/10.3389/fpsyt.2019.00252.

[28] Alberto Jiménez-Valverde. "Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling". In: *Global Ecology and Biogeography* 21.4 (2012), pp. 498–507. DOI: https://doi.org/10.1111/j.1466-8238.2011.00683.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1466-8238.2011.00683.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1466-8238.2011.00683.x.

[29] Marcel Just et al. "Autism as a neural systems disorder: A theory of frontal-posterior underconnectivity". In: *Neuroscience and biobehavioral reviews* 36 (Feb. 2012), pp. 1292–313. DOI: 10.1016/j.neubiorev.2012.02.007.

[30] K. Kiehl. *The Psychopath Whisperer: Inside the Minds of Those Without a Conscience*. Oneworld Publications, 2014. ISBN: 9781780745404. URL: https://books.google.ch/books?id=jhy9DwAAQBAJ.

[31] M. Kirk. *Thoughtful Machine Learning with Python: A Test-driven Approach*. O'Reilly Media, Incorporated, 2017. ISBN: 9781491924129. URL: https://books.google.ch/books?id=DNtmtAEACAAJ.

[32] Yazhou Kong et al. "Classification of Autism Spectrum Disorder by Combining Brain Connectivity and Deep Neural Network Classifier". In: *Neurocomputing* 324 (May 2018). DOI: 10.1016/j.neucom.2018.04.080.

[33] Felicity Larson et al. "Psychosis in autism: comparison of the features of both conditions in a dually-affected cohort". In: (Jan. 2016). DOI: 10.17863/CAM.4652.

[34] Siyi Li et al. "Dysconnectivity of Multiple Brain Networks in Schizophrenia: A Meta-Analysis of Resting-State Functional Connectivity". In: *Frontiers in Psychiatry* 10 (2019). ISSN: 1664-0640. DOI: 10.3389/fpsyt.2019.00482. URL: https://www.frontiersin.org/article/10.3389/fpsyt.2019.00482.

[35] Heather Lindsey. *AI's push to understand psychiatry research has the potential to tackle mental illness*. URL: https://www.businessinsider.com/how-ai-is-advancing-psychiatry-medicine-to-tackle-mental-illness-2021-9?r=US&IR=T.

[36] Mary-Ellen Lynall et al. "Functional Connectivity and Brain Networks in Schizophrenia". In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30 (July 2010), pp. 9477–87. DOI: 10.1523/JNEUROSCI.0333-10.2010.

[37] *Machine Learning Crash Course*. URL: https://developers.google.com/machine-learning/crash-course.

[38] Junko Matsuo et al. "Autistic-Like Traits in Adult Patients with Mood Disorders and Schizophrenia". In: *PLoS ONE* (Apr. 2015). DOI: 10.1371/journal.pone.0122711.

[39] Gary C. McDonald. "Ridge regression". In: *WIREs Computational Statistics* 1.1 (2009), pp. 93–100. DOI: https://doi.org/10.1002/wics.14. URL: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wics.14.

[40] National Institute of Mental Health. *Schizophrenia*. URL: https://www.nimh.nih.gov/health/topics/schizophrenia.

[41] Michał Oleszak. *Regularization Tutorial: Ridge, Lasso and Elastic Net*. URL: https://www.datacamp.com/tutorial/tutorial-ridge-lasso-elastic-net.

[42] Nick Patterson, Alkes L Price, and David Reich. "Population Structure and Eigenanalysis". In: *PLOS Genetics* 2.12 (Dec. 2006), pp. 1–20. DOI: 10.1371/journal.pgen.0020190. URL: https://doi.org/10.1371/journal.pgen.0020190.

[43] Jonathan Peelle. *Introduction to MRI. Some basic terminology.* URL: http://jpeelle.net/mri/general/preliminaries.html.

[44] Jonathan Power et al. "Methods to detect, characterize, and remove motion artifact in resting state fMRI". In: *NeuroImage* 84 (Aug. 2013). DOI: 10.1016/j.neuroimage.2013.08.048.

[45] Jonathan D. Power et al. "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion". In: *NeuroImage* 59.3 (2012), pp. 2142–2154. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2011.10.018. URL: https://www.sciencedirect.com/science/article/pii/S1053811911011815.

[46] Liudmila Prokhorenkova et al. "CatBoost: Unbiased Boosting with Categorical Features". In: NIPS'18. Montréal, Canada, 2018, 6639–6649.

[47] Laura Elena Raileanu and Kilian Stoffel. "Theoretical comparison between the gini index and information gain criteria". In: *Annals of Mathematics and Artificial Intelligence* 41.1 (2004), pp. 77–93.

[48] Irina Rish. "An Empirical Study of the Naïve Bayes Classifier". In: *IJCAI 2001 Work Empir Methods Artif Intell* 3 (Jan. 2001).

[49] Rajit Sanghvi. *A Complete Guide to Adam and RMSprop Optimizer*. URL: https://medium.com/analytics-vidhya/a-complete-guide-to-adam-and-rmsprop-optimizer-75f4502d83be.

[50] Aditya Sharma. *Understanding Activation Functions in Deep Learning*. URL: https://learnopencv.com/understanding-activation-functions-in-deep-learning/.

[51] Hui Shen et al. "Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI". In: *NeuroImage* 49 (Nov. 2009), pp. 3110–21. DOI: 10.1016/j.neuroimage.2009.11.011.

[52] Lukas Snoek, Steven Miletić, and H. Scholte. "How to control for confounds in decoding analyses of neuroimaging data". In: *NeuroImage* 184 (Sept. 2018). DOI: 10.1016/j.neuroimage.2018.09.074.

[53] Kaustubh Supekar et al. "Brain Hyperconnectivity in Children with Autism and its Links to Social Deficits". In: *Cell reports* 5 (Nov. 2013). DOI: 10.1016/j.celrep.2013.10.001.

[54] Saori Tanaka et al. "A multi-site, multi-disorder resting-state magnetic resonance image database". In: *Scientific Data* 8 (Aug. 2021), p. 227. DOI: 10.1038/s41597-021-01004-8.

[55] Saori Tanaka et al. "A multi-site, multi-disorder resting-state magnetic resonance image database". In: *Scientific Data* 8 (Aug. 2021), p. 227. DOI: 10.1038/s41597-021-01004-8.

[56] Cagdas Ulas. "INCORPORATION OF A LANGUAGE MODEL INTO A BRAIN COMPUTER INTERFACE BASED SPELLER". PhD thesis. Aug. 2013. DOI: 10.13140/2.1.2017.6326.

[57] Monica Vermani, Madalyn Marcus, and Martin Katzman. "Rates of Detection of Mood and Anxiety Disorders in Primary Care: A Descriptive, Cross-Sectional Study". In: *The primary care companion to CNS disorders* 13 (Apr. 2011). DOI: 10.4088/PCC.10m01013.

[58] Marlies Vissers, Michael Cohen, and Hilde Geurts. "Brain connectivity and high functioning Autism: A promising path of research that needs refined models, methodological convergence, and stronger behavioral links". In: *Neuroscience and biobehavioral reviews* 36 (Sept. 2011), pp. 604–25. DOI: 10.1016/j.neubiorev.2011.09.003.

[59] Sebastian Walther et al. "Aberrant Hyperconnectivity in the Motor System at Rest Is Linked to Motor Abnormalities in Schizophrenia Spectrum Disorders". In: *Schizophrenia Bulletin* 43.5 (July 2017), pp. 982–992. ISSN: 0586-7614. DOI: 10.1093/schbul/sbx091. eprint: https://academic.oup.com/schizophreniabulletin/article-pdf/43/5/982/19682927/sbx091.pdf. URL: https://doi.org/10.1093/schbul/sbx091.

[60] *Worsening mental health crisis pressures psychologist workforce*. URL: https://www.apa.org/pubs/reports/practitioner/covid-19-2021.

[61] Shoujun Xu et al. "Altered Functional Connectivity in Children With Low-Function Autism Spectrum Disorders". In: *Frontiers in Neuroscience* 13 (2019). ISSN: 1662-453X. DOI: 10.3389/fnins.2019.00806. URL: https://www.frontiersin.org/article/10.3389/fnins.2019.00806.

[62] Noriaki Yahata et al. "A small number of abnormal brain connections predicts adult autism spectrum disorder". In: *Nature Communications* 7 (Apr. 2016), p. 11254. DOI: 10.1038/ncomms11254.

[63] Xin Yang, Ning Zhang, and Paul Schrader. "A study of brain networks for autism spectrum disorder classification using resting-state functional connectivity". In: *Machine Learning with Applications* 8 (2022), p. 100290. ISSN: 2666-8270. DOI: https://doi.org/10.1016/j.mlwa.2022.100290. URL: https://www.sciencedirect.com/science/article/pii/S2666827022000226.

[64] Ling-Li Zeng et al. "Multi-Site Diagnostic Classification of Schizophrenia Using Discriminant Deep Learning with Functional Connectivity MRI". In: *EBioMedicine* 30 (2018), pp. 74–85. ISSN: 2352-3964. DOI: https://doi.org/10.1016/j.ebiom.2018.03.017. URL: https://www.sciencedirect.com/science/article/pii/S2352396418301014.

[65] Jiahe Zhang et al. "What have we really learned from functional connectivity in clinical populations?" In: *NeuroImage* 242 (2021), p. 118466. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2021.118466. URL: https://www.sciencedirect.com/science/article/pii/S1053811921007394.

[66] Zhongheng Zhang. "Introduction to machine learning: K-nearest neighbors". In: *Annals of Translational Medicine* 4 (June 2016), pp. 218–218. DOI: 10.21037/atm.2016.03.37.

[67] Zhongheng Zhang. "Too much covariates in a multivariable model may cause the problem of overfitting". In: *Journal of thoracic disease* 6 (Sept. 2014), E196–E197. DOI: 10.3978/j.issn.2072-1439.2014.08.33.

[68] Hui Zou and Trevor Hastie. "Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Statist Soc B. 2005;67(2):301-20". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (Apr. 2005), pp. 301 –320. DOI: 10.1111/j.1467-9868.2005.00503.x.