UNIVERSITA' DEGLI STUDI DI MILANO

Doctoral School of Computer Science
Department of Computer Science
Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science (31^{st} cycle)

Machine-Learning based analysis and computer aided classification of neuropsychiatric-disorders using neuro-imaging

INF/01

Doctoral dissertation of:
Tewodros Mulugeta Dagnew

Supervisor:
Prof. Roberto Sassi

Co-advisor:
Dr. Letizia Squarcina

Doctorate school's director:
Prof. Paolo Boldi

Oct. 2015 - Oct. 2018

This thesis work has been done in collaboration with Prof. Paolo Brambilla's medical team from Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, University of Milan.

ABSTRACT

Machine learning (ML) based analysis of neuroimages in neuropsychiatry context are advancing the understanding of neurobiological profiles and the pathological bases of neuropsychiatric disorders. Computational analysis and investigations on features derived from structural magnetic resonance imaging (sMRI) of the brain are used to quantify morphological or anatomical characteristics of the different regions of the brain that have role in several distinct brain functions. This helps in the realization of anatomical underpinnings of those disorders that cause brain atrophy. Structural neuroimaging data acquired from schizophrenia (SCZ), bipolar disorder (BD) patients and people who experienced psychosis for the first time, are used for the experiments presented in this thesis.

The cerebral cortex (i.e., gray matter) of the brain is one of the most studied anatomical part using 'cortical-average-thickness' distribution feature in the literature. This helps in the realization of the anatomical underpinning of those mental illnesses that cause brain atrophy. To this regard, based on statistical background, 'cortical-skewness' feature, a novel digital imaging-derived neuroanatomical biomarker that could potentially assist in the differentiation of healthy control (HC) and patient groups is proposed and tested in this thesis.

The core theme of machine intelligence relies in extracting and learning patterns of input data from experience. Classification is one of the task. In a basic set up, ML algorithms are trained using exemplary multivariate data features and its associated class labels, so that they could be able to create models and do predictive classification and other tasks. Considering the conundrum nature of psychiatric disorders, researchers in the field, could benefit from ML based analysis of complex brain patterns. Out of many, one task is computer aided classification (CAC). This is achieved by training the algorithms, these complex brain patterns and their corresponding diagnostic statistics manual (DSM) based clinical gold standard labels. Indeed, in the literature, supervised learning methods such as support vector machines (SVM) which follow inductive learning strategy are widely exploited and achieved interesting results. Observing this and due to the fact that the most widely available relevant anatomical features of the cortex such as thickness and volume values, could not be considered satisfactory features because of the heterogeneous nature of the human brain anatomy due to differences in age, gender etc., a contextual similarity based learning is proposed. This learning uses a transductive learning mechanism (i.e, learn a specific function for the problem at hand) instead of learning a general function to solve a specific problem.

Based on this, it is adopted, a formulation of a semi supervised graph transduction (label propagation) algorithm based on the notions of game theory, where the consistent labeling is represented with Nash equilibrium, to tackle the problem of learning from neuroimages with subtle microscopic difference among different clinical groups.

However, since such kind of algorithms heavily rely on the graph structure of the extracted features, we extended the classification procedure by introducing a pre-training phase based on a distance metric learning strategy with the aim of enhancing the contextual similarity of the images by providing a 'must belong in the same class' and 'must not belong in the same class' constraint from the available training data. This would result to increase intra-class similarity and decrease inter-class similarity.

The proposed classification pipeline is used for searching anatomical biomarkers. With the goal of identifying potential neuroanatomical markers of a psychiatric disorder, it is aimed to develop a feature selection strategy taking into consideration the widely exploited cortical thickness and the proposed skewness feature, with the objective of searching a combination of features from all cortical regions of the brain that could maximize the possible differentiation among the different clinical groups

Considering Research Domain Criteria (RDoC) framework developed by National Institute of Mental Health (NIMH) with the aim of developing biologically valid perspective of mental disorders by integrating multimodal sources, clinical interview scores and neuroimaging data are used with ML methods to tackle the challenging problem of differential classification of BD vs. SCZ.

Finally, as deep learning methods are emerging with remarkable results in several application domains, we adopted this class of methods especially convolutional neural networks (CNNs) with a 3D approach, to extract volumetric neuroanatomical markers. CAC of first episode psychosis (FEP) is performed by exploiting the 3D complex spatial structure of the brain to identify key regions of the brain associated with the pathophysiology of FEP. Testing of individualized predictions with big dataset of 855 structural scans to identify possible markers of the disease is performed.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION AND THE STATE-OF-THE-ART

> The beginning is the most important part of the work.
>
> Plato

The brain is a vital organ that controls all the activities of the body, interprets information and stimulus from the outside world, furthermore, it plans reaction to the outside world. It governs much more complex tasks. Anatomically speaking it has 3 layers based on tissue types: the gray matter where mostly the cell bodies of neurons reside, the white matter formed by axon of the neurons and the cerebral fluid. The brain is localized into four main regions based on specialized functions: frontal, temporal, parietal and occipital. Unfortunately, the brain is vulnerable to be affected by diseases that have negative impact on the physiology and anatomy that would eventually cause psychiatric disorders. A psychiatric disorder is a mental illness that could be associated with anatomical and/or physiological alteration of the brain. The alteration of physiology may manifest into anatomical variation as a course of time.

Considering anatomical variations between healthy and patient subjects, in this thesis, it is explored how anatomical feature variation/classification from imaging data can be a vital source of information in understanding the diseases and facilitate the diagnosis process of patients from computational analysis perspective. Many psychiatric disorders share a common attribute called brain (cortical) atrophy, a degeneration of neuron cells in the brain that would cause cortical thinning in many areas, at the end, affecting the normal functioning of the brain. The atrophy in a specific region of the brain causes a decline of the associated function. For example, the degeneration of neurons in the occipital region (an area responsible for vision processing function), results in the impairment, disorder or corruption of visual information processing (i.e., hallucination) [1].

Among the major psychiatric disorders, schizophrenia (SCZ), bipolar disorder (BD) and psychosis are found. These mental illnesses still remain to be conundrum in the neuropsychiatry community. Taking this into consideration, in this thesis, structural brain imaging scan data of healthy controls, first episode psychosis (FEP), BD and SCZ patients using magnetic resonance imaging (MRI) machines for experimental analysis are considered.

BD is known to be a mental illness affecting around 1-2% of the population [2], characterized by oscillation of mood, with alternating depressive and manic episodes. It can lead to chronic disability, if there is delay in treatment [3, 4]. SCZ is known to be among the most prevalent mental disorders and affects about 1% of the population worldwide [5]. This heterogeneous disease is usually characterized by disintegration in reality perception and plenty of cognitive problems [6]. Psychosis is a phenomenon that refers to detachment from reality, in which people have trouble distinguishing between what is real and what is not. When this occurs, it is called a psychotic episode while the first episode of psychosis (FEP) is the first time a person experiences a psychotic episode. It has been observed in the literature that interventions at the earlier stages, when the illness is no yet chronic, may prevent irreversible neuro-biological and social changes [7].

In this regard, the invention of structural Magnetic Resonance Imaging (MRI) machines makes possible to investigate and visualize the anatomy and structure of the brain in this exciting time of brain research. Using the acquired anatomical images one can investigate neuroanatomical differences between healthy brain and ill brain. As an input data, MRI is the gold standard technique to explore the anatomical and functional underpinnings of psychiatric disorders. Furthermore, this imaging modality, helps in the process of characterizing the particular neuroanatomical profiles of those disorders. Computational methods that aim to quantify and differentiate subtle anatomical differences due to brain atrophy induced psychiatric disorders would definitely advance the understanding of the biological underpinning of the pathology and help in the realization of a better mental health care systems. Currently, there is an increasing need to imaging derived neuroanatomical biomarkers for diagnosis and prognosis of neuropsychiatric disorders, however, in order to reach a justifiable conclusion, it is required a fair amount of brain imaging samples and automated methods that can accurately analyze those imaging data [8, 9, 10]

In this regard, machine learning (ML) algorithms come to the rescue by providing computational solutions with the aim of identifying the needed objective biomarkers for the following three reasons. First, machine learning algorithms allow predictions at an individual subject level and therefore able to facilitate individualized clinical decisions [11, 12]. Second, machine learning frameworks can naturally handle 'multivariate' features that means they are capable of analyzing multiple biological measurements simultaneously as opposed to traditional 'univariate' statistical methods which are only able to analyze single measurement at a time [13]. Third, they are evaluated by a robust 'cross-validation' methods to ensure generalization of results by 'testing' the algorithm using previously 'unseen' observations [14, 15, 16, 17]. These virtues make them an ideal computational tool to be used in neuroimaging based psychiatric studies. For a

detailed overview of ML approaches mainly used so far in psychiatric neuroimaging, it is advised to refer [13, 14, 15].

The core theme of machine intelligence relies in the process of extracting and learning patterns of input data from experience. The learning part is achieved by trainable algorithms that can be classified into three groups: 1) Supervised learning algorithms aim to find a mapping function of data features x_i to a set of labels y_i. The input labeled training examples are provided in the form of feature-label pairs $\{x_i, y_i\}$. 2) Unsupervised learning algorithms aim to learn the underlying structure of the data so as to group features into clusters. The learning process aims to find the structure or pattern of the input data $\{x_i\}$ without having label information provided. 3) Semi-supervised approach incorporates both the labeled and unlabeled part of the data in the learning process.

ML based computer aided classification (CAC) pipelines are associated with one or more of the following 3 tasks: 1) feature extraction: involves quantifying representation of an object, 2) feature selection: involves in selecting optimal relevant features for data/object representation and 3) classification: involves in assigning the selected/extracted features into a predefined category known as label assignment. When the predefined categories or classes are two, the problem is casted as binary classification, otherwise the problem becomes a multi-class one.

Quantitative objective measurements of psychiatric disorders using MRI are believed to facilitate the journey in the understanding of the cause [18]. Neuroimaging increasingly exploits ML techniques in an attempt to achieve clinically relevant single-subject predictions. In this regard, the two important issues are: brain image analysis methodology that is capable of extracting the most relevant information and a pattern classification method that is designed to process the extracted information in order to estimate or predict the likelihood of the psychiatric disorder under investigation from the imaging data [19].

Structural neuroimaging is mainly used for studying and analyzing the anatomy of the brain and in understanding which brain parts are affected by atrophies (or in general structural alterations) with the aim of finding morphological biomarkers for the diagnosis of neuro-psychiatric disorders. Diagnosis and prognosis of neuropsychiatric disorders is still a challenging task to this date. In this regard computer aided diagnosis (CAD) systems try to facilitate the understanding of these illnesses so as to improve health care systems and the quality of human life, in the era where the holistic understanding of the brain is still limited. One of the goal of ML methods in psychiatric neuroimaging studies is to propose objective (neuroanatomical) biomarkers. In the best case scenario, these markers should quantify some structural characteristics which can help in differentiation of healhy brain vs. patient brain. Also, they can be of help in the identification of the neurological and biological basis of psychiatric disorders [20, 21, 22]. Moreover, automatic classification methods could lead to the definition of biomarkers representing the vulnerability to a specific psychiatric disorder of individuals with a familial high risk [23], which represents an issue which is difficult to deal with based only on the clinical practice.

Below, we will briefly overview the essence of feature extraction, feature selection and classification from pure machine learning, neuroimage processing and computer aided classification (CAC) of psychiatric disorders perspective respectively.

Feature extraction is under the umbrella of pattern recognition which is a science and art that uses statistics and mathematics to develop an algorithm and program a computer to recognize patterns in a dataset. In the field of medical science, pattern recognition is the basis of computer aided diagnosis (CAD) systems. It allows computers to handle new situations by means of previous experience, analysis and training [24]. The performance of pattern recognition systems heavily relies in the quality of features to be analyzed. The set of features extracted from an object can be considered as a signature which describes the object. The more succinct and optimal the better. Brain feature extraction is aimed at characterizing the image in terms of properties or features, such as cortical thickness, cortical volume, gyrification, shape and texture. Good features are those that, when belonging to subjects of the same category or class, are very similar, on the contrary, they should be statistically significantly different from subjects in different categories. A good feature vector should be able to discriminate objects belonging to different classes. Most anatomical investigations follow either of the following two approaches while extracting anatomical information from structural neuroimages with the objective of representing complex brain patterns using features: 1) surface reconstruction based: where initially the raw input volumetric brain image is reconstructed based on the voxel intensities. This involves computing distances between surfaces (i.e., the boundary surface between white matter and gray matter and the pial surface that separates gray matter from CSF). It is usually used to extract cortical informations such as thickness and volume 2) voxel-intensity based: where the statistical information of the voxels are analyzed either for the whole brain tissues or focusing only in some region of interests (ROIs). In this thesis we used both approaches. Features are usually organized in the so-called feature vector, a vector of arbitrary length which collects all the properties that are considered useful in order to describe the objects under analysis. Recently, after the emergence of deep learning algorithms that uses convolutional neural networks, the feature extraction process is embedded automatically in the learning pipeline avoiding hand crafting features with the aim of learning a representative feature trained from example.

After features are extracted, the process of feature selection is of great importance in the application of automatic algorithms to classify data. In fact, especially when dealing with imaging data, there is a risk of over fitting, due to the fact that the number of features is usually large [15]. A single subject scan has millions of voxels. Thus, feature reduction is an important preprocessing step, and the process itself can give insight into which are the most important characteristics of the datasets, related to the problem of interest. In our case, it gives insight into which brain region and its corresponding feature can be a potential imaging-derived morphological biomarker. In neuroimage based studies a huge amount of features from cardinally many voxels from the volumetric brain imaging modality are extracted with the aim of characterizing anatomical and physiological profiles. In practice all the extracted features does not hold equally relevant information in understanding the attributes among different classes so a fea-

ture reduction step as part of feature selection process may be employed. For example, in [25], the authors employed this method to reduce the dimensionality of the features derived from functional MRI in depression without affecting the classification performances. Other feature selection methods refer to the statistical properties of the features themselves such as correlation between features or t-test comparing features across groups, this methods have been used in neuroimaging studies in psychiatry: in the context of schizophrenia [26], cognitive impairment [27], Alzheimer's disease etc. A common way of selecting the most relevant features is to rank the features on the basis of their classification performances. Considering the multivariate neuroanatomical pattern, a relevant feature selection approach has been used in studies on psychosis [28], BD and SCZ [29] for succinct representation of the biological profiles and their corresponding localization of the pathologies.

ML classification algorithms aim at automatically assigning a set of class labels to a new observation. The commonly used term 'learn' signifies finding statistical regularities on a set of data [24]. In particular, features in the data are classified into different predefined classes. This is done on the basis of a set of data containing observations of which the class label is known, called training set.

In the context of computer aided neuropsychiatric disorder classification schemes, neuroimages intended for training a classification algorithms come with their corresponding class labels based on diagnostic statistical manual (DSM) and/or international classification of disease (ICD) based gold standard class labels assigned to the subjects after a diagnosis is performed by psychiatrists. Then the learning algorithm finds a function for mapping of the extracted brain features to the labels. A validation set may be used to optimize hyper parameters of the learning algorithm. Finally, after the training procedure, follows a testing procedure to evaluate the generalization capability of the learned model in assigning a new unseen brain feature to a correct diagnostic class label. The predictions are compared with the ground truth DSM/ICD based labels to evaluate the performance in terms of accuracy, sensitivity and specificity using cross validation routines. The capability of the learned model, how accurately classified those unseen data will serve as the reliability evaluation of the model on how well it can generalize the phenomena (i.e., in our case cortex atrophy). The main motivation for the need of adopting complex reasoning systems is that there cannot be fixed threshold values of thickness/volume that can let the doctor know a normal or pathological brain. In other words, it is unreliable to say, less than 2.5 mm of cortical thickness is a brain affected by psychiatric disorder, so it is needed a mechanism that compares those thickness features based on exemplary data from healthy control and pathological groups and analyze the similarity of a test image which is effective and proven to characterize or recognize the subtle changes.

Considering the absence of standard clinical test for schizophrenia, there is a growing interest in machine learning based diagnosis of schizophrenia using neuroimaging features from brain scans acquired using MRI machine. Basing on the studies that SCZ is highly correlated with brain atrophy, the ultimate goal of ML classifiers in processing the brain images is to quantify and classify the subtle and diffused changes in the

anatomy of the brain. These helps in deriving imaging based biomarkers that can help individualized prediction of subjects.

The application of automatic classification methods to imaging-derived features has also shown promising results so far in BD, reaching classification accuracies of around 70% [30]. Using average of the cortical thickness distribution as a feature in the discrimination of BD patients from patients with major depression [31] resulting accuracies of around 75% is reported, considering fronto-temporal and parieto-temporal regions. However, results are not always in agreement, and several studies did not reach completely satisfactory results, with poor accuracies not reaching 60% [32, 33, 34].

When using machine learning analysis on the features extracted from neuroimaging data, it is important to take into account that supervised ML methods consider only previously labeled data, where each sample is associated to an apriori known label defining the class which it belongs to. The use of such methods in the psychiatry field has led to some criticism, especially because the diagnosis is subjective and not unaffected by uncertainty. Therefore, the objectivity of biomarkers defined with supervised classification methods has been questioned [35]. In contrast, semi-supervised learning (SSL), which is a contextual-based learning framework in which the learning process exploits information from both the labeled and unlabeled parts of the dataset, might overcome some of the aforementioned criticalities, furthermore it may facilitate discovery of phenotypes. SSL methods employ user-provided labels to bias the process of discovering structure or patterns in the data as supervised learning methods and, in addition, the learning process regroups subjects with similar characteristics without the use of the labels, achieving theoretically a more robust prediction in the presence of label uncertainty. In the literature, the majority of ML applications to psychiatric data are purely supervised methods that learn only from labeled data, with promising and interesting results [36, 22, 37]. However, while these findings have been received with great optimism within the neuropsychiatric community, a major criticism has been that these algorithms are ordinarily 'trained' to categorize patients based on a symptom-based diagnosis. As such, there are inevitable uncertainty in the gold standard labels due to the heterogeneous nature of psychiatric disorders. Learning from the unlabeled data seems a possibility to mitigate the problem. In these situations, classification performances might improve when the learning process incorporates unlabeled data. Moreover, semi-supervised and unsupervised schemes could provide a better phenotype identification and classification of diseases [38].

As for whole brain analysis, it is not always the best way to analyze changes in brain regions, since misleading significant correlations may exist in some brain regions that are not involved in the brain disease under diagnosis. Some studies were driven by a priori hypothesis basing on domain knowledge and consistently detected specific structural markers [39, 40].

ML algorithms have been used in SCZ studies [24] with the aim of detecting sets of features which could be discriminative in the diagnosis.

Some of the ML algorithms exploited for diagnosis of psychosis, BD and SCZ include: support vector machines (SVM), classification trees, linear discriminant analysis, quadratic discriminant analysis, neural networks, generalized linear models, near-

est neighbor classifiers etc. They are all based on different algorithms, whose aim is to decide how new brain images should be classified [41]. SVM [42] a supervised machine learning classifer has been extensively employed in neuroimaging studies [24, 20, 43, 19]. Using structural and/or functional neuroimaging data as input to ML algorithms represents a valid diagnostic aid for classifying major neurological and psychiatric illnesses, allowing inferences at individual level, rather than at group level [13] . During the past few years an increasing number of studies have used SVM in order to investigate the presence of potential neuroanatomical biomarkers of neurological and psychiatric disorders [44, 39, 45, 46, 47, 22]. Specifically, in schizophrenia patients, structural and functional brain abnormalities have been demonstrated [48].

Regarding the reliability of machine learning based diagnosis in BD, a study in [33] reported 61% prediction accuracy in discriminating BD patients from healthy controls using structural neuroimaging scans from a total of 132 subjects. A study using 80 subjects reported prediction accuracies around 72.5% [30]. Although many studies in this direction tried to show the ML algorithms predictive validity of structural brain images, there remains some questions still needed to be answered such as diagnostic accuracies have not been established using large samples and the employment of these algorithms in supporting clinically relevant applications (e.g. patient stratification or clinical staging) has not been fully explored.

As an example, studies using sMRI for investigating anatomical and biological underpinning of BD have extensively reported reductions in gray matter density in orbitofrontal cortex, superior temporal gyrus and insula regions [49, 50, 51]. It has consistently been shown that brain structure is broadly affected in BD, especially reductions in gray matter volume [52, 53], ventricular enlargement [54, 55] and damage to corpus callosum [56, 57, 58, 59]. Studies investigating cortical thickness typically found reduced gray matter thickness in cingulate, temporal and frontal cortices [60], regions which are known to be involved with emotions processing and regulation [61, 62], the parietal lobe as well as the fronto-parietal network, involved in cognition [63].

Considering the state of the art approaches and their limitations, in this thesis, with the objective of designing a neuropsychiatric disorder classification pipeline using sMRI, it is proposed a natural approach of classification by considering contextual similarity among the brain images in order to incorporate both labeled and unlabeled brain imaging data. For this purpose graph based semi-supervised transductive learning is exploited. The multivariate nature of the proposed frame work makes it possible to analyze anatomical informations from different regions simultaneously unlike the traditional univariate analysis approach. This make it suitable to address the challenging phenomena that the diseases causes dispersed cortex thinning. The contextuall similarity based analysis is a good fit to the nature of neuroimages that have subtle difference among classes. The transduction (or label propagation) problem is formulated based on the notion of game theory where the feature derived from anatomical images act as players and the DSM based labels as choices. The game is formulated as a pair wise polymatrix game where the whole game is represented by a simultaneous two player game. The reward matrix is calculated from the pairwise feature similarity. The final consistent labeling is considered the Nash equilibria of the game. The

non parametric feature of the classifier makes it easier to use. The approach is similar to KNN in which a nearest neighbor is assigned similar label but with radically different perspective based on the notions of game theory having a solid mathematical root. However, since the reliability of this classifier heavily relies on the graph structure in the space (where the nodes represent brain features and their edges represent the similarity among them) we introduced a pre-training phase before the final learning of class label assignment takes place. To do so, a distance metric learning is adopted with the aim of increasing the intra-neuroimage feature similarity and decreasing the inter-neuroimage feature similarity. The process transform the graph with a new structure. Since the type of features matter a lot in ML based neuropsychiatry, it is also proposed a novel diagnostic metric called cortical-skewness that measures the symmetricity of the cortex. Furthermore, the proposed feature is used in the identification of neuroanatomical biomarker for FEP and BD using the proposed classification pipeline together with other establish ML algorithms such as: SVM and logistic regression. Especially, the biomarkers identified for FEP will be a great input for predictive analysis on longitudinal studies. As part of this thesis, we also explored a graph-based semi-supervised learning method to identify and select the most discriminant regions of the brain between BD patients and controls, using MRI-derived cortical thickness, in terms of average and skewness of the thickness distribution. Most studies consider only average thickness in their analyses, which can be seen as a global thickness feature for each region of interest (ROI). Imaging derived biomarker 'cortical skewness' that we proposed it as a novel structural brain imaging derived anatomical-diagnostic-metric for psychiatric disorders, instead, provides a measure of the asymmetry of a distribution around its mean. We chose this quantity because we hypothesize that the pathology could modify cortical thickness not only in its average value, but also in its distribution shape. In particular, a distribution with a positive skew (i.e. asymmetrical shape with more values on the left of the mean) could reflect a progressive shrinkage of cerebral cortex which, in turn, could not be so evident considering only the average values. Following the motive dictated by research domain criteria (RDoC) that encourages the use of multi modality, a combination of structural neuroimaging data and clinical interview score data are tested for classification. Using this set up the challenging problem of differential diagnosis of BD vs SCZ is also addressed. Finally, using a big data of FEP dataset, we applied a 3D convolutional neural network for classifying subject with first episode psychosis. In the process, we aim to extract a 3D spatial marker for the disease using both supervised and unsupervised approach (using 3D convolutional auto encoders). It is also proposed for the first time the usage of a 3D residual network in volumetric structural brain imaging.

However, even though all this achieved results and insights, these techniques are still not available for clinical use, due to the complexity and heterogeneousness of the diseases and their lack of objective morphological biomarkers. Exploring imaging derived biomarkers or features other than the traditional parameters such as cortical thickness and volume seems the right direction in this regard.

1.1 CONTRIBUTIONS OF THE THESIS

The Contribution of this thesis can be summarized as follows:

- The designing and testing of a machine learning based neuro-image processing and analysis pipeline and its application in computer aided classification of neuro-psychiatric disorders. To explore and exploit the intra-heterogeneous features of neuro-images further more, we employed a contextual similarity based semi-supervised learning approach. Furthermore, we introduced a data driven 'similarity metric learning' strategy to enhance the learning process of neuroimaging features.
- Employing a new diagnostic metric called 'cortical-skewness feature' to characterize the gyrification and cortical folding that can help the diagnosis process of psychiatric disorders. The metric has been tested on the context of bipolar-disorder.
- Employing an automated feature selection method in structural brain imaging without relying on priori hypothesis from the domain knowledge with the aim of identifying imaging-derived biomarkers and identification of which brain regions can be potential biomarkers. In this experiment, we tested the proposed diagnostic metric: cortical-skewness feature.
- Employing the recent state of art machine learning algorithms known as deep learning (specifically 3D convolutional neural networks) to learn the 3D spatial feature from brain imaging, with the aim of deriving a 3D anatomical biomarker in diagnosis of first episode psychosis and avoiding hand crafting structural marker features. We used a large population in the experiment.

By designing different experimental set ups, we showed the application of the designed pipeline for computer-aided-diagnosis of psychiatric disorders. In the mean time, we tried to answer biological questions associated with the biological underpinning of the pathologies by tasks such as searching for specific regions and associated biomarker.

1.2 THESIS OVERVIEW

In the following, the chapters of this thesis are briefly described.

- Chapter 2: In this chapter, a novel neuro-imaging derived diagnostic metric that can be potentially used in the computer aided classification of neuropsychiatric disorders (that affect the brain by resulting brain atrophy or cortical shrinkage) is introduced.
- Chapter 3: In this chapter, it is described a contextual similarity based neuro-image processing and classification system in the context of computer aided diagnosis/classification of neuropsychiatric disorders using a semi-supervised

learning approach. The experiments are carried on classification of brain scans of healthy control, BD and SCZ patients.

- Chapter 4: In this chapter, a proposed system for identification of neuroanatomical biomarkers for chronic BD and first episode patients is described.
- Chapter 5: In this chapter, we describe about multi-modality where we applied the designed pipeline to learn from multi-modal information by combining neuroimage features with clinical interview scores assessing cognitive abilities. It also discusses the challenging problem of the differential diagnosis of SCZ vs. BD.
- Chapter 6: Narrates about deep learning application in computer aided neuroimage classification in the context of diagnosis of psychiatric disorders, where it is tested on brain scan of healthy controls and first episode psychosis patients. It focuses on a 3D spatial feature of the brain and their classification outcome.
- Chapter 7: Finally this chapter concludes all the proposed approaches and methods and the results of this thesis and explains the future works.

2

CORTICAL "SKEWNESS" FEATURE AS A DIGITAL BIOMARKER: A NOVEL STRUCTURAL BRAIN IMAGING DERIVED ANATOMICAL-DIAGNOSTIC-METRIC FOR PSYCHIATRIC DISORDERS

> All truths are easy to understand once they are discovered; the point is to discover them.
>
> Galileo Galilei

In ML based neuropsychiatric studies, investigating neuroanatomical differences among different diagnostic clinical groups using cerebral cortex (or gray matter) data, mostly use direct brain measurement such as cortical thickness, volume, density, area etc. While they inform biologically relevant phenomenon, still other indirect measures (or features) with comparable or better performance could open another perspective in the understanding of the neurobiological profiles and characterization of psychiatric disorders. Based on this, a new sMRI based neuroanatomical marker useful in the computer aided classification of psychiatric disorders is proposed and tested for its discriminative ability in the differentiation of structural changes among BD and FEP patients vs. healthy controls. This metric is based on a skewness measurement of the cortical thickness distribution of the brain and it measures the asymmetric/symmetricity of gray matter distribution in a particular region of interest.

In the medical domain, skewness is used together with related metrics such as kurtosis and percentiles as biomarkers of tumour heterogeneity in the context of cancer study using MRI [64]. According to the best of the author's knowledge, this is the first investigation using skewness measurement applied on MRI based computer aided diagnosis of psychiatric disorders. While the most widely used diagnostic metric of gray matter measurement is the average of the thickness distribution, skewness is a mea-

sure of symmetry and in this context it is used to measure the gray matter thickness distribution symmetry.

2.1 STATISTICAL BACKGROUND OF SKEWNESS

In probability theory and statistics, histograms are one of the mainly used data and scenario representations. It can also be used to represent and measure geometrical shapes. The histogram can give a general idea/global feature of a shape. From it, one could calculate measures of shape that give a more precise evaluation such as skewness, that tells the amount and direction of skew (departure from horizontal symmetry), relative to a standard bell curve. In other words, it is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive, negative or zero.

It has applications such as testing for normality: many statistics inferences require that a distribution be normal or nearly normal. A normal distribution has zero skewness, so if a data distribution is close to this value then it is probably close to normal.

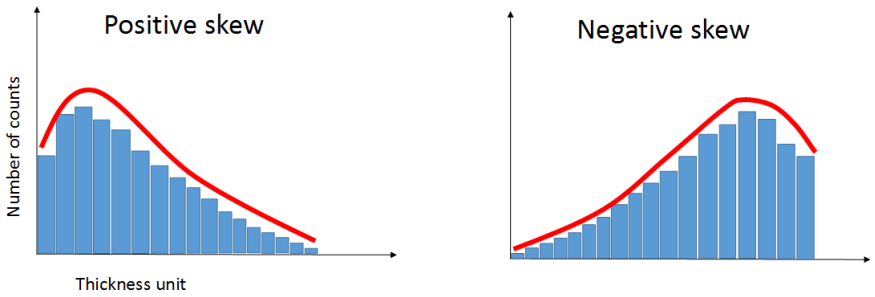


Figure 2.1: Interpretation of the skewness of data distribution histogram.

Skewness is the function of a mean (or average) as expressed by the following formula.

$$Skewness = \sqrt{n}\frac{\sum_{i=1}^{n}(v_i - v_{avg})^3}{(\sum_{i=1}^{n}(v_i - v_{avg})^2)^{3/2}}$$

Where n is the number of vertices used to reconstruct the 2 surfaces (i.e, the distance between the white matter surface and the pial surface) using meshes. In the context of 3D computer graphics and geometrical modeling, a mesh is a collection of vertices (v_i), edges and faces. v_i is an individual vector where the magnitude represents the distance between the two surfaces (i.e., gray matter thickness) at a particular spatial location. v_{avg} is the mean of the v_is.

When the distribution is symmetrical then the value of coefficient of skewness is zero because the mean, median and mode coincide. If the coefficient of skewness is a positive value then the distribution is positively skewed and when it is a negative value, then the distribution is negatively skewed.

2.2 APPLICATION TO CORTICAL THICKNESS MEASUREMENT

Most neuroimaging studies that investigate cortical thickness use mean thickness measure. The work reported here is quantifying another metric that is potentially useful as anatomical biomarker for the diagnosis of psychiatric disorders that would result in brain atrophy (or neurodegenesis also refered as cortex shrinkage in some literatures).

Other than the most widely used average gray matter thickness, some studies in the literature employed innovative metrics to quantify cortex features such as cortical gyrification, folding index etc. [65, 66, 67].

Aiming from analyzing the brain atrophies that can result in the subtle change of brain anatomy caused by psychiatric disorders, skewness might capture those subtle differences in some regions of the brain in addition to the insight hold on the average thickness measurement. The measurement is done on parcellated brain with a region of interest (ROI) based approach.

The average and skewness feature extraction from the brain cortex is depicted in figure 2.2.

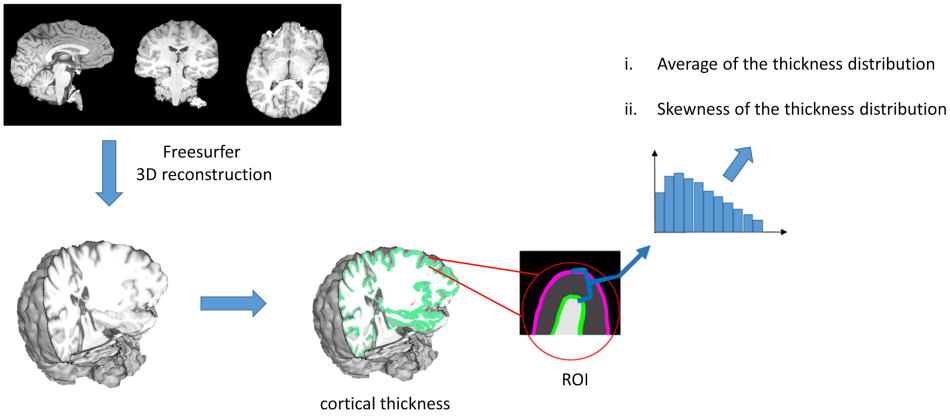


Figure 2.2: Cortical thickness feature extraction illustration. We quantified average and skewness of the distribution of cortical thickness values for each of the 58 ROIs (29 from each hemispheres).

2.3 EXPERIMENT

2.3.1 DATASET

Forty-one patients (48.40 $\pm$ 8.12, 25 females) with DSM-IV BD and 34 healthy controls (HC) (39.55 $\pm$ 12.17, 17 females) were recruited. Group and gender proportions were not significantly different ($p > 0.05$), during statistical hypothesis testing of proportion. The average age between the two populations was instead significantly different (t-test

$p < 0.05$). The recruitment of patients was performed using the South-Verona Psychiatric Care Register (PCR) (Tansella and Burti, 2003), a community-based mental health register. Diagnoses for BD were obtained using the Item Group Checklist of the Schedule for Clinical Assessment in Neuropsychiatry (IGC-SCAN) (Wing, J.K., World Health Organization, 1992) and confirmed by the clinical consensus of two staff psychiatrists. All patients with other Axis I disorders (i.e., all psychological diagnostic categories except mental retardation and personality disorder), history of alcohol or substance abuse, history of traumatic head injury with loss of consciousness, epilepsy or other neurological or medical diseases, including hypertension and diabetes, were excluded from the study. HC were recruited through word of mouth and advertisements in the local communities and had no history of head injury or psychiatric disorders and no psychiatric disorders among first-degree relatives. The absence of psychiatric disorders was tested using an interview modified from the non-patient version of the Structured Clinical Interview for DSM-IV Axis IV Disorders (SCID-IV). All the procedures were approved by the Biomedical Ethics Committee of the Azienda Ospedaliera di Verona, Italy, and were in accordance with the Helsinki Declaration of 1975. All subjects signed a written informed consent to the protocol. Demographic and clinical details of participants are reported in Table 1.

T1 images were acquired with a 1.5 T Siemens Symphony (Siemens Healthcare, Erlangen, Germany). Parameters for the T1 MRI sequence were set as follows: 384 x 512 x 144 voxels, 0.45 x 0.45 x 1.25 mm3, TR = 2140 ms, TE = 3.9 ms, flip angle 15 Â°.

2.3.2 MEASUREMENT OF CORTICAL THICKNESS DISTRIBUTION

In table 2.1 and 2.2 are reported the calculated mean and skewness metric of cortical thickness of the brain of healthy controls and bipolar disorder patients. Further more we try to see if this anatomical imaging derived feature has a potential to be a digital bio-marker by having a discriminative capability between the two clinical groups.

The cortical thickness distribution have been obtained using FreeSurfer software first by reconstructing the brain surfaces (i.e., the surface between white matter & gray matter and between gray matter & the cerebral fluid known as the pial surface), then calculating the distance between those surfaces (which is refereed as cortical thickness): one value per vertex. To obtain the specific region thickness distribution, Desikan-Killiany brain atlas is used to mask.

The feature extraction procedure is illustrated in figure 2.2. To take into account possible differences in brain structure due to age difference of the considered subjects (which could alter the results), for the nuisance effect of the age mean thickness and skewness values are corrected using a linear model, following the suggestion of [21]. Specifically, age is considered as predictor, and mean thickness and skewness as responses. Then, once removed what was predicted by the model, only residuals were further employed. Equal proportion of gender did not require any correction of cortical thickness based on sex.

Table 2.1: Cortical thickness and skewness features of HC and BD groups.
Lh-Ba: Left hemisphere Banks of the Superior Temporal Sulcus; Rh-Ba: Right hemisphere-Banks of the Superior Temporal Sulcus; Lh-Cu: Left hemisphere-Cuneus; Rh-Cu: Right hemisphere-Cuneus; Lh-IP: Left hemisphere-Inferior Parietal; Rh-LO: Right hemisphere-Lateral Occipital; Lh-Mt: Left hemisphere-MiddleTemporal; Lh-Pe: Left hemisphere-Pericalcarine; Lh-Sm: Left hemisphere-Supramarginal; Lh-OL: Left hemisphere-Orbito Lateral; Lh-SFG: Left hemisphere-Superior Frontal Gyrus; Rh-In: Right hemisphere-Inferior Temporal; Lh-Tr: Left hemisphere-Triangularis; Lh-FP: Left hemisphere-Frontal Pole; Rh-En: Right hemisphere- Enthorinal; Rh-TL: Right hemisphere- Temporal Lobe; Rh-PC: Right hemisphere- Post Central ; Rh-SP: Right hemisphere-Superior Parietal; Lh-Op: Left hemishpere-Opercularis.

	HC		BD	
Cortical features / Brain regions	Mean thickness (± std)	Mean skewness (± std)	Mean thickness (± std)	Mean skewness (± std)
Lh-Ba	2.47(±0.17)	0.11(±0.25)	2.31(±0.18)	0.10(±0.26)
Rh-Ba	2.55(±0.16)	0.18(±0.29)	2.37(±0.16)	0.10(±0.26)
Lh-Cu	1.64(±0.13)	0.72(±0.22)	1.53(±0.10)	0.92(±0.28)
Rh-Cu	1.66(±0.16)	0.72(±0.35)	1.57(±0.11)	0.93(±0.37)
Lh-IP	2.40(±0.15)	0.19(±0.12)	2.27(±0.13)	0.32(±0.14)
Rh-LO	2.14(±0.12)	0.61(±0.27)	2.03(±0.12)	0.71(±0.21)
Lh-MT	2.88(±0.16)	−0.02(±0.15)	2.75(±0.18)	0.12(±0.22)
Lh-Pe	1.42(±0.11)	0.84(±0.23)	1.34(±0.10)	1.12(±0.42)
Lh-Sm	2.55(±0.18)	0.14(±0.13)	2.38(±0.15)	0.21(±0.17)
Lh-OL	2.72(±0.19)	−0.01(±0.18)	2.57(±0.17)	0.13(±0.19)
Lh-SFG	2.80(±0.16)	−0.18(±0.15)	2.61(±0.16)	−0.02(±0.17)
Rh-InT	2.86(±0.17)	0.08(±0.13)	2.73(±0.15)	0.14(±0.15)
Lh-Tr	2.76(±0.27)	−0.16(±0.34)	2.57(±0.26)	−0.04(±0.33)
Lh-FP	2.90(±0.35)	−0.26(±0.41)	2.66(±0.39)	−0.11(±0.41)
Rh-En	3.55(±0.33)	−0.19(±0.37)	3.37(±0.36)	−0.25(±0.27)
Rh-TL	3.86(±0.27)	−0.51(±0.38)	3.63(±0.28)	−0.37(±0.35)
Rh-PC	2.00(±0.14)	0.54(±0.20)	1.89(±0.13)	0.69(±0.18)
Rh-SP	2.05(±0.11)	0.43(±0.15)	1.95(±0.13)	0.55(±0.18)
Lh-Op	2.61(±0.19)	−0.03(±0.14)	2.45(±0.18)	0.03(±0.25)

2.3.3 RESULTS AND DISCUSSIONS

Increased skewness values are observed in the left inferior parietal gyrus of BD patients, which implies a reduction in thickness in some portions of these regions with respect to HC. The mean thickness, while affected, captures less effectively this variation.

Table 2.2: Age-corrected cortical thickness and skewness features of HC and BD groups (here, the intercept of the linear model was re-added to the residuals).

	HC		BD	
Cortical features / Brain regions	Mean thickness (± std)	Mean skewness (± std)	Mean thickness (± std)	Mean skewness (± std)
Lh-Ba	2.78(±0.15)	0.14(±0.25)	2.70(±0.18)	0.14(±0.26)
Rh-Ba	2.90(±0.13)	0.25(±0.29)	2.81(±0.16)	0.19(±0.26)
Lh-Cu	1.69(±0.13)	0.53(±0.22)	1.60(±0.10)	0.68(±0.28)
Rh-Cu	1.69(±0.16)	0.51(±0.35)	1.61(±0.11)	0.66(±0.37)
Lh-IP	2.66(±0.13)	0.07(±0.12)	2.60(±0.13)	0.16(±0.15)
Rh-LO	2.27(±0.12)	0.33(±0.25)	2.19(±0.13)	0.36(±0.21)
Lh-MT	3.20(±0.13)	−0.28(±0.14)	3.14(±0.17)	−0.21(±0.21)
Lh-Pe	1.48(±0.11)	0.60(±0.26)	1.42(±0.10)	0.82(±0.42)
Lh-Sm	2.91(±0.14)	0(±0.11)	2.83(±0.16)	0.03(±0.18)
Lh-OL	2.96(±0.18)	−0.32(±0.16)	2.88(±0.18)	−0.25(±0.18)
Lh-SFG	3.17(±0.12)	−0.44(±0.14)	3.06(±0.17)	−0.34(±0.18)
Rh-InT	3.06(±0.15)	−0.04(±0.12)	2.98(±0.16)	−0.01(±0.16)
Lh-Tr	3.22(±0.22)	−0.41(±0.33)	3.15(±0.26)	−0.35(±0.33)
Lh-FP	3.43(±0.32)	−0.68(±0.41)	3.32(±0.38)	−0.62(±0.39)
Rh-En	3.80(±0.31)	0.01(±0.37)	3.68(±0.37)	0.01(±0.27)
Rh-TL	4.22(±0.27)	−0.42(±0.38)	4.09(±0.27)	−0.26(±0.34)
Rh-PC	2.24(±0.12)	0.21(±0.17)	2.19(±0.13)	0.27(±0.18)
Rh-SP	2.20(±0.10)	0.27(±0.14)	2.15(±0.13)	0.35(±0.18)
Lh-Op	2.96(±0.15)	−0.14(±0.14)	2.88(±0.19)	−0.10(±0.24)

The mean and age of the 34 healthy controls was 39.15(± 12.30). The mean and age of the 41 BD patients was 48.73(± 8.17). The difference in age between the groups was statistically significant (p = 0.00014).

The discriminatory capability of the biomarkers for BD (i.e., mean cortical thickness or skewness) is indicated using AUCs and ROC curves in figure 2.3. The values in the bracket are the p-values between the correlated ROC's AUC values of thickness and skewness estimated using a non parametric statistical test [68]. The result indicates that the proposed diagnostic metric show similar performance with the bench mark metric (i.e., mean thickness).

2.4 SUMMARY

We proposed skewness as a feature to measure the cortical thickness distribution that can potentially serve as a digital imaging derived neuro-anatomical biomarker.

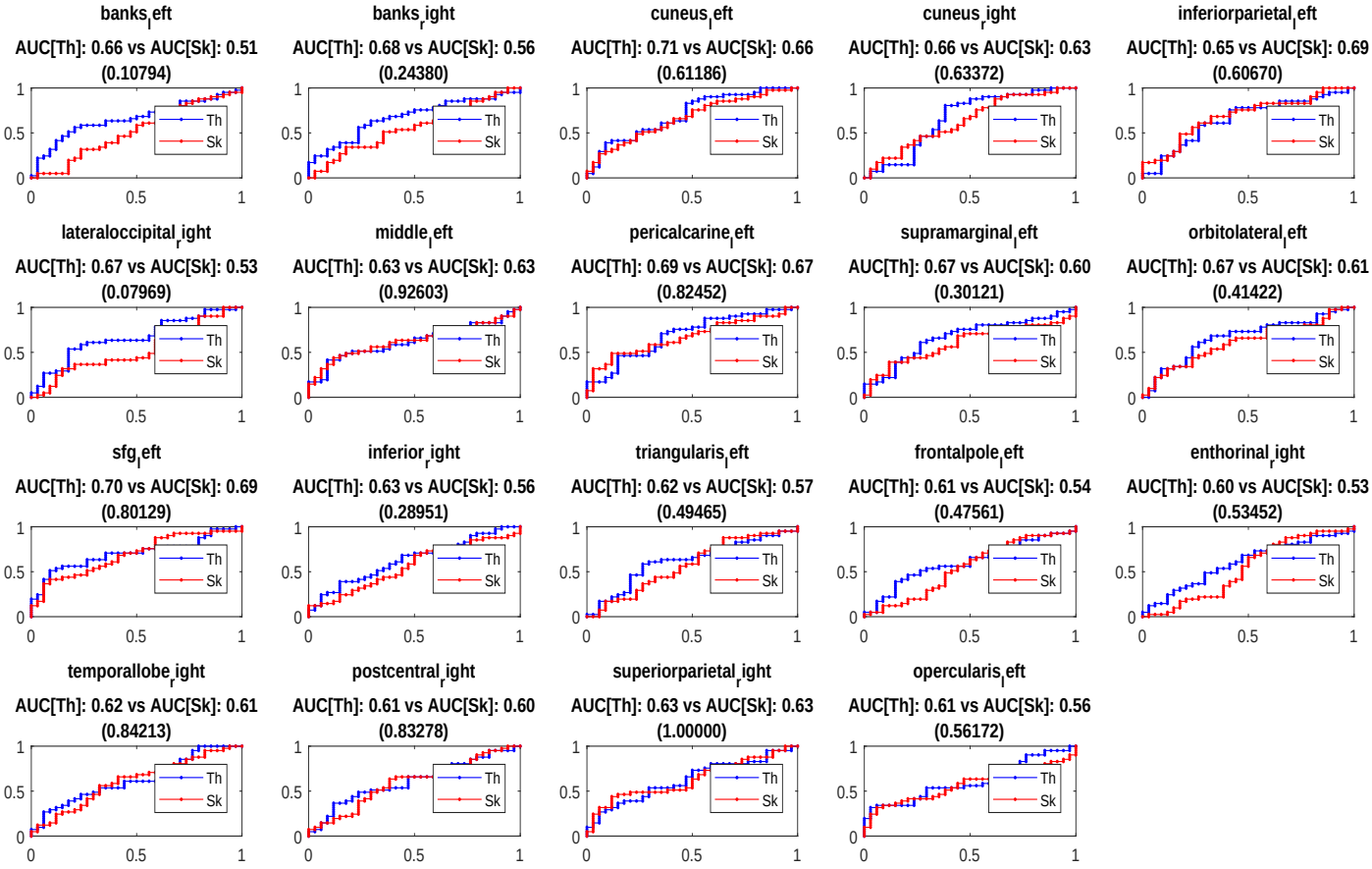


Figure 2.3: The AUC under roc curve of cortical thickness and skewness diagnostic metrics in differentiating HC vs. BD. The values in the bracket are the p-valuevbetween the correlated AUC values of thickness and skewness.

In the context of anatomical variation measurement among BD patients and healthy controls, skewness has been shown to be a better biomarker on inferior parietal region of the left hemisphere than average thickness (table 2.3).

According to the best of the author's knowledge, this is the first investigation of using skewness measurement applied on MRI based computer aided diagnosis of psychiatric disorders. While the most widely used diagnostic metric of gray matter measurement is the average of the thickness distribution, skewness is a measure of symmetry and in this context it is used to measure the gray matter thickness distribution symmetry.

Table 2.3: a univariate statistical analysis results on group level classification i.e, HC vs. BD

	P-value of ave. thickness	P-value of skewness	AUC of mean-thickness	AUC of skewness
Lh-Ba	1 (0.03982)	0 (0.97921)	0.66	0.51
Rh-Ba	1 (0.00859)	0 (0.33737)	0.68	0.56
Lh-Cu	1 (0.00076)	1 (0.01213)	0.71	0.66
Rh-Cu	1 (0.00913)	0 (0.07311)	0.66	0.63
Lh-IP	1 (0.04191)	1 (0.00292)	0.65	0.69
Rh-LO	1 (0.00962)	0 (0.56018)	0.67	0.53
Lh-MT	0 (0.11403)	0 (0.08496)	0.63	0.63
Lh-Pe	1 (0.00574)	1 (0.00810)	0.69	0.67
Lh-Sm	1 (0.02070)	0 (0.31740)	0.67	0.60
Lh-OL	1 (0.03555)	0 (0.11390)	0.67	0.61
Lh-SFG	1 (0.00273)	1 (0.00766)	0.70	0.69
Rh-IT	1 (0.03961)	0 (0.31435)	0.63	0.56
Lh-Tr	0 (0.15888)	0 (0.44885)	0.62	0.57
Lh-FP	0 (0.18789)	0 (0.57393)	0.61	0.54
Rh-En	0 (0.15648)	0 (0.97795)	0.60	0.53
Rh-TL	1 (0.03447)	0 (0.05995)	0.62	0.61
Rh-PC	0 (0.10396)	0 (0.13540)	0.61	0.60
Rh-SP	1 (0.04034)	1 (0.03823)	0.63	0.63
Lh-Op	1 (0.04859)	0 (0.45557)	0.61	0.56

3

CONTEXTUAL-SIMILARITY BASED CLASSIFICATION OF NEUROIMAGES: A SEMI-SUPERVISED LEARNING APPROACH

> Prediction is the essence of intelligence. Differentiation is the essence of learning.
>
> Yann LeCun

Structural (anatomical) neuroimages to be used in computer aided classification of neuropsychiatric disorders have very subtle differences between different classes (i.e 'healthy control' & 'pathological' class). In psychiatric neuroimaging, usually, one does not expect macroscopic differences between images of healthy individuals and psychiatric patients.

Considering this, we hypothesize, transductive learning (trying to learn a specific function for the problem at hand) could easily exploit well this scenario easier than inductive learning (trying to learn a general function to solve a specific problem). As indicated in [69] transductive learning is easier than inductive learning.

Observing that the neuroimage literature is dominated with supervised machine learning algorithms, we wondered to see how a semi-supervised transductive learning would result. Supervised models such as SVM learns hyperplane parameters as decision boundaries between data from different classes. This does not tell about the underlying interconnection between the data and we tried to explore if the contextual similarity of the data at hand could hold some information that could help the computer aided classification process of neuropsychiatric disorders.

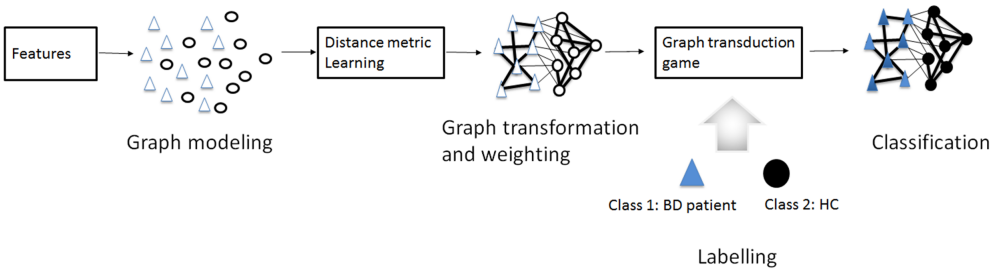


Figure 3.1: Pipeline of the proposed extended semi-supervised scheme for classification of HC and BD patients.

3.1 TRANSDUCTIVE LEARNING

Transductive learning aims at classifying only unlabeled objects exploiting the information derived from labeled ones, rather than finding a general rule for classifying future examples. Usually, the objects are represented by graphical models that use graph to represent data. The nodes in the graph correspond to labeled and unlabeled points while the edges encode the pairwise similarity among each pair of nodes.

Graph transduction algorithms uses graphical models of the data for the the learning process. Basically, graph transduction is a semi supervised learning technique that aims at estimating a classification function defined over a graph of labeled and unlabeled data points. This technique works propagating the label information from labeled nodes to unlabeled, exploiting the graph structure.

The problem is formulated by a set of labeled objects (x_i, y_i) $(i = 1, 2, ..., l)$, where $x_i \in \mathbb{R}^n$ is a real-valued vector describing the object i, and $y_i \in (1, ..., c)$ its label, and a set of u unlabeled objects $(x_{l+1}, ..., x_{l+u})$.

In this framework, the geometry of the data is represented using a weighted graph [70] formally expressed as $G = \{V, E, W\}$, where $V = (v_1, v_2, ..., v_n)$ indicates the available (n) nodes, E the available edges and W the edge weights.

For achieving transductive learning that we wanted to apply to contextual similarity based neuro-image classification for solving a computer aided diagnosis of neuropsychiaty disorders problem, we adopted the approach of Erdem and Pelillo [71] where the graph transduction task is interpreted as a non-cooperative multi-player game. We chose to use this framework because it has a solid mathematical foundation rooted in game theory, plus it does not impose any constraint on the pairwise similarity function that computes the weight of the graph. Additionally, we can exploit the multi-class nature of the framework which could suit well for multiple simultaneous differential diagnosis of neuropsychiatric disorders that can help in understanding of the biological underpinning of the pathologies.

One of the important assumptions of graph tansduction framework is, adjacent nodes often have same labels known as the principle of 'homophily' [72]. The similarity information among data points is exploited to realize this cluster assumption. In the following sections, it is shown that this information can be easily embedded in a

game-theoretical framework as part of the payoff function. In the next section, for the sake of clarity, a brief introduction about basic concepts of game theory will be given.

3.2 GAME THEORY

The concept of game theory was introduced by Von Neumann and Morgenstern in 1944 and later popularized by John F. Nash. The aim was developing a mathematical framework that can model the essentials of decision making in interactive situations. When we see it from the perspective of the problem in our hand, classification can be seen as a decision making process. The concept is that there are a finite set of players $P = \{p_1, p_2, ..., p_n\}$, a set of pure choices $CH = \{1, 2, ..., c\}$ for each players, and a reward function ($reward : ch_1 \times ch_2 ... \times ch_n \rightarrow \mathbb{R}$), that assigns reward values to chosen choices by the players. Here, an assertive assumption is that, players always try to maximize their reward. The key concept of this interactive game is called Nash Equilibrium (NE). It represents those choice profiles in which each choice is the best response to the choice of the co-player and in which no player has the motive to change its decision. This phenomena signify, the players are in equilibrium state. The NE choice profile could be either 'pure-choice' or 'mixed-choice'. In a pure-choice NE, each player choose only one choice, while in mixed strategy NE the players choice is a probability distribution among the possible options. The mixed-choice of a player lies in a standard simplex $\triangle^c$, where c is the number of available decision choices. The mixed-choice of a player is denoted with a vector: $ch = [ch^1, ..., ch^c] \in \triangle^c$ and each component ch^h denotes the probability that a particular player chooses its h^{th} pure-choice.

The standard simplex is, mathematically represented by:

$$\triangle^c = \{ch \in \mathbb{R} : \sum_{h=1}^{c} ch^h = 1, ch^h \geqslant 0, \forall h\} \tag{3.1}$$

Now let us see how the expected reward is computed when two players play a game having a pair of choice profiles represented as (ch_i, ch_j), where each players decision lies in the standard simplex (i.e., $ch_i \in \triangle^c$ and $ch_j \in \triangle^c$). It will be defined as follows:

$$\begin{aligned} reward(ch_i, ch_j) &= ch_i^T R_{ij} ch_j \quad ; \\ reward(ch_j, ch_i) &= ch_j^T R_{ji} ch_i \quad . \end{aligned} \tag{3.2}$$

R_{ij} represents the $c \times c$ reward matrix, holding information about the quantified values of reward when player i chooses decision on the rows, while player j chooses decisions on the columns.

The reward for a player i choosing the h^{th} pure-choice is computed using:

$$reward(ch_i^h) = \sum_{j=1}^{n} (R_{ij} ch_j)^h \tag{3.3}$$

while the expected reward for its entire mixed-choice is computed as follows:

$$reward(ch_i) = \sum_{j=1}^{n} ch_i R_{ij} ch_j \tag{3.4}$$

where n is the number of players that i plays with. Finally, using the terms of eq. 3.3 and 3.4, we can determine the NE state of the game using a concept from evolutionary game theory [73] inspired from the behavior of animals conflict for survival by updating their view of the environment for making decisions. When it is related to our case, since the game is played repeatedly, in every iteration the fittest choice survives. To obtain those states, a discrete time version of multi population replicator dynamics [74] is used, described by the following equation:

$$ch_i^h(t+1) = ch_i^h(t)\frac{reward(ch_i^h)}{reward(ch_i)}, \forall h \in S \tag{3.5}$$

At each iteration, for each player, the probability of its all available decisions will be updated individually and the solution brings a NE state of the game that is being played. Later, it will be described how this NE state is interpreted, as a consistent labeling assignment in the context of classification using graph models.

3.3 CREATING CLASSIFICATION CONTEXT (ENHANCING CONTEXTUAL FEATURE SIMILARITY) BY LEARNING A PROPER FEATURE-DISTANCE METRIC

Distance metric learning (DML) represents a useful technique widely exploited in pattern recognition, which aims to find a metric that maximizes the distance between features belonging to different classes (and vice versa, minimizes the distance between features belonging to the same class). With this aim, linear and non-linear metrics had been investigated. On one hand, linear metrics can be computationally less expensive, but often provide lower performances. On the other hand, non-linear algorithms might perform better but they are computationally expensive and application-dependent.

In the linear domain, DML remaps features using a linear combination carried out by the transformation matrix L, as follows:

$$\bar{x}' = L\bar{x},$$

where $\bar{x}$ is the input feature vector and $\bar{x}'$ is the transformed feature vector. If the matrix L is full rank, it is possible to show that the Euclidean distance between two elements in the transformed space,

$$D(\bar{x}_i, \bar{x}_j) = ||L(\bar{x}_i - \bar{x}_j)||_2 \quad ,$$

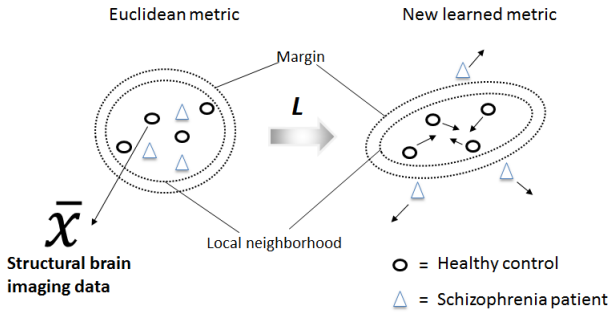


Figure 3.2: Illustration of feature context enhancement by means of large margin nearest neighbor (LMNN) distance metric learning. Before training (left) and after training (right).

represents a valid metric. Furthermore, the Euclidean distance can be rewritten using a matrix notation which becomes the so-called *Mahalanobis* distance. Such distance is defined as

$$D_M(x_i, x_j) = \sqrt{(\bar{x}_i - \bar{x}_j)^\top M(\bar{x}_i - \bar{x}_j)} \quad ,$$

being $M = L^T L$ the Mahalanobis positive semidefinite matrix. The effect of such transformation is shown in fig. 2.1. When L is the identity matrix, the Mahalanobis distance becomes the standard Euclidean distance.

In this thesis, we used a linear DML to modify the pre-existing neighbouring structure of MRI data before feeding it to GT, aiming to achieve classification improvements. In order to determine the transformation matrix L, we used the *Large Margin Nearest Neighbor* DML method described in [75]. The algorithm makes use of the following equations

$$\mathrm{pullpush}(L) = (1-\mu)\ \mathrm{pull}(L) + \mu\ \mathrm{push}(L) \tag{3.6}$$

with

$$\mathrm{pull}(L) = \sum_{i,j \to i} \|L(\bar{x}_i - \bar{x}_j)\|^2$$

$$\mathrm{push}(L) = \sum_{i,j \to i} \sum_{k} (1 - \delta_{ik})[1 + \|L(\bar{x}_i - \bar{x}_j)\|^2 - \|L(\bar{x}_i - \bar{x}_k)\|^2]_+$$

where y_i is the class to which $\bar{x}_i$ belongs and $\delta_{ik} = 1$ if $y_i = y_k$ or $\delta_{ik} = 0$ otherwise. $[f]_+$ implies a hinge-loss such that $[f]_+ = \max(0, f)$. The term $j \to i$ in the equation implies that j belongs to the same class where i belongs too. Finally, the parameter μ sets the trade-off between the pulling and pushing objectives and was set to 0.5 as suggested in [75].

The process of getting the transformation metric L involves minimizing the overall objective function in eq. (3.6). The first term pulls subjects with the same class label closer in terms of the Mahalanobis distance. The second term pushes away differently-

labeled instances by a large margin, so that they are located further apart in the transformed space (fig. 2.1).

As stated in [75], it is worth noting that Eq. (3.6) does not define a convex optimization problem in terms of L. However, it can be rephrased in a convex fashion using a semi-definitive programming approach by determining M instead of L. Then, L can be computed using matrix factorization of M.

3.4 COMPUTER AIDED CLASSIFICATION (CAC) OF STRUCTURAL MRI AS A GRAPH TRANSDUCTION GAME

In this section, the formulation of game of neuroimages will be explained. Following the path proposed by Erdem and Pelillo, we aim to achieve a transductive learning on neuroimage features based on the notions of game theory to solve the problem of CAC of neuropsychiatric disorders. We will treat the problem as a binary classification.

First, by casting the problem as a graph transduction problem, we allow any features extracted from structural MRI to be represented using a graph $G = \{V, E, W\}$, where $V = (v_1, v_2, ..., v_n)$ indicates the available (n) nodes representing features, E as the edges among nodes and W as similarity weights (i.e., edge weight) as depicted in figure 3.3.

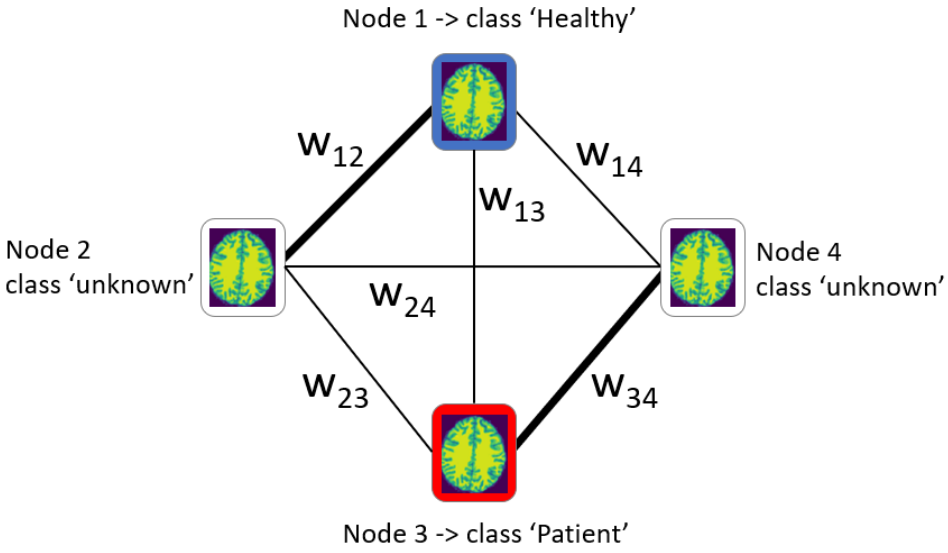


Figure 3.3: A hypothetical weighted graph of brain images with labeled (v1 & v3) and unlabeled nodes (v2 & v4). The bold edges signify higher similarity between the Node features.

When the problem is faced from game theoretic perspective, the nodes of the graph representing the imaging features will be players and the available DSM based categorical labels assigned by psychiatrists (i.e., healthy and patient) will be treated as choices. The nodes are divided in two disjoint sets namely labeled and unlabeled nodes. The labeled players have knowledge of their class membership and they do not need to maximize their reward since they always play their already chosen h^{th} pure choice, where $h = 1, ..., c$, as the unlabeled ones do not have any idea about their membership

at the beginning of the game. The transduction game can be easily reduced to a game with only unlabeled players that need to find their mixed choice decisions $ch_i \in \triangle^c$ and the fixed choices of labeled players act as bias over the choices of unlabeled players. From a machine learning perspective, the labeled nodes are used to train the model and the unlabeled nodes will be treated as a test set. For example, as shown in the figure 3.3, node 1 and 3 are labeled where as node 2 and 4 are unlabeled. As a result, our aim is to propagate the label information from nodes 1 and 3 to nodes 2 and 4. In other words, we will determine the classes of node 2 and 4.

The hypothetical graph presented above has the following properties:

- The weighted adjacency matrix, representing the similarity between the features of the brain images, looks like the following, where each entry is the quantification of the edge weight.

	Node 1	Node 2	Node 3	Node 4
Node 1	w_{11}	w_{12}	w_{13}	w_{14}
Node 2	w_{21}	w_{22}	w_{23}	w_{24}
Node 3	w_{31}	w_{32}	w_{33}	w_{34}
Node 4	w_{41}	w_{42}	w_{43}	w_{44}

- Node 1 and Node 3 are labeled (colored as blue and red respectively), belonging to class 'Healthy' and 'Patient' (i.e they know their class membership), in other words player 1 and player 3 already know their decisions as a result they are playing the game with their pure choice always. Here notice that, the terms 'Node' and 'Player' are equivalent and can be used interchangeably.

- Node 2 and Node 4 are unlabeled (i.e player 2 and player 4 did not decide their class choice which can yield them a higher reward yet, out of their available choice profile which is either belonging to 'class Healthy 'or'class Patient'). In the beginning (i.e., at time t=0, when the iteration of the game begins) they do not know their class membership.

- Initial choice probability distribution matrix ch(0) i.e., at time t = 0, will be initialized as described in figure 3.4. $ch(0) \in \mathbb{R}^{n \times c}$, where n is the number of players and c is the number of labels (class) existing in the current configuration. In this specific toy example, because there are four nodes playing to maximize their reward by choosing a decision or selecting a best choice from the available two choices, n will be four and c will be two. The goal is to update the entries of this ch matrix for the unlabeled node. As a result, final consistent labeling assignment expressed as Nash Equilibrium (NE) will be achieved. ch_i is normalized such that it add up to 1 ($ch_i^h = \frac{ch_i^h}{\sum_{h=1}^{c} ch_i^h}$) so that the choice of each player remains in the standard simplex $\triangle^c$.

As explained above, the rule of the game is to have players choose their best choice that could maximize a reward. Following this concept, we represent the neuroimage

	Class Healthy	Class Patient	
Node 1	1	0	→ Labeled
Node 2	0.5	0.5	→ Unlabeled
Node 3	0	1	→ Labeled
Node 4	0.5	0.5	→ Unlabeled

Figure 3.4: The initial class (or choice) probability distribution ch(0) of each node.

scan of individuals as players and their DSM based diagnostic category as choices. The reward matrix is computed by considering the similarities between the features of the images (i.e., nodes of the graph) exploiting the structure of the weighted graph. In a classical way, the weight is computed using the equation given below:

$$w_{ij} = \exp\left[-\frac{d(\bar{x}_i, \bar{x}_j)}{2\sigma}\right] \tag{3.7}$$

where d(.) can be any distance function.

Unfortunately, label propagation using these kind of algorithms heavily relies on the pre-existing underlying structure of the graph, so in here, it is proposed an extension on top of the semi-supervised contextual similarity based learning framework (i.e., graph transduction game) by modifying the structure of the graph using distance metric learning strategy by pre-training using the available label information on the training sets in the form of a 'must be in the same class' and 'must not be in the same class'. The aim is to increase the similarity weight between data points belonging to similar classes and decreasing the similarity weight belonging to different classes. This procedure as explained in section 3.3 derive a (feature space) transformation matrix that increases separation of nodes belonging to different classes while increasing the proximity of the features belonging to the same class. From the game theoretic perspective, since the weight of the graph are being modified, this procedure results in the modification of the reward matrix that the players look up to. As a result the equation is modified 3.8 as follows:

$$w'_{ij} = \exp\left[-\frac{d(L\bar{x}_i, L\bar{x}_j)}{2\sigma}\right] \tag{3.8}$$

Where L is the newly learnt distance metric on the pre-training phase. Briefly stating the above concept, what is done is that, it is introduced a supervised pre-training phase before the final training so as to learn a new graphical representation of the data. Then the final training is done using the new graph representation.

The effect of DML is illustrated on figure 3.5. The new weights (W') will be the ones used to quantify the reward matrix when the game takes place.

Being the solution considered as the equilibrium in a non-cooperative game, the adjacency matrix W is used to compute the reward value between players. It is important

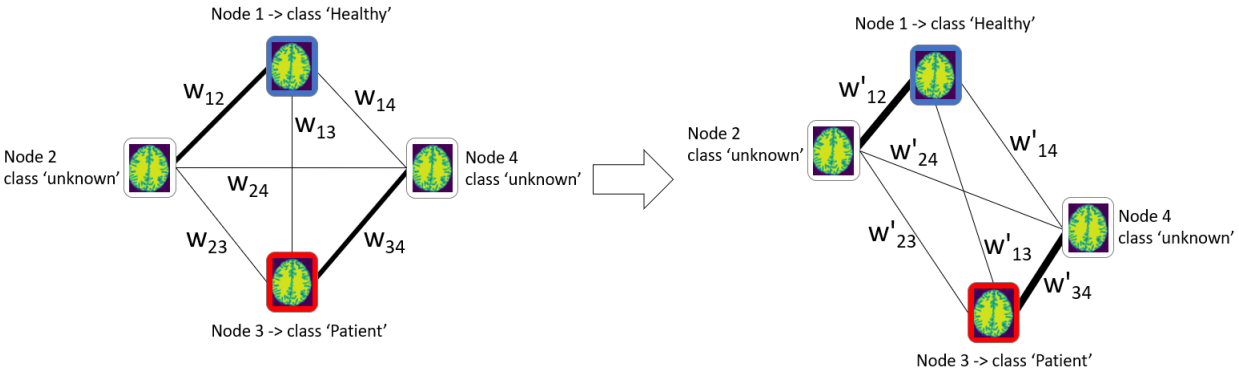


Figure 3.5: The effect of data driven distance metric learning.

to notice that the game taking place is the so called sub class of multi-player games known as polymatrix game. Polymatrix games are games such that the whole interaction of the game is understood as the combination of pairwise interaction in which the edge between two nodes (or players) is seen as a single game, and the reward values for these two players will be a function of the edge weight. Even though it is a multi-player game, the polymatrix game rule states that at a specific moment one player can only interact with one of its neighbors only, not with all its neighbors at the same time. So the final payoff for a player P_i will be the summation of the rewards that it has got from the individual pairwise games that it played with its neighbors. So being the game considered as an instance belonging to the class of polymatrix games [76, 77] where players are nodes of a graph and every edge denote a two-player game between corresponding pair of players, the partial reward matrix between two players i and j and their choice ch_i and ch_j respectively, is computed as $R_{ij} = w_{ij} \times I_c$ where I_c is the identity matrix of size c.

As explained in section 3.2, the reward values are computed as:

$reward(ch_i^h) = \sum_{j=1}^{n}(R_{ij}ch_j)^h$

$reward(ch_i) = \sum_{j=1}^{n} ch_i R_{ij} ch_j$

The Nash equilibria, thus the labeling for unlabeled neuroimage features, is computed using the evolutionary approach [78, 79] as the following,

$$ch_i^h(t+1) = ch_i^h(t)\frac{reward(ch_i^h)}{reward(ch_i)}, \forall h \in S \tag{3.9}$$

The dynamic interpretation of Nash equilibria through the evolutionary approach imagines that the game is played repeatedly, generation after generation, during which a selection process acts on the multi-population of decision choices, thereby resulting in the evolution of the fittest choice. The particular class of dynamics used in this are the so called replicator dynamics.

For the Nash equilibrium theorem [80] the graph tansduction game always has equilibrium in mixed choices (not necessarily converging to pure choices) that corresponds to a steady state where each player's choice could yield the highest reward, and it provides a globally consistent labeling of the data set. Once an equilibrium is reached,

	Class Healthy	Class Patient	
Node 1	1	0	
Node 2	1	0	→ New information
Node 3	0	1	
Node 4	0	1	→ New information

Figure 3.6: The final class probability distribution of the subjects status to be predicted, i.e. the final consistent labeling assignment ch(t).

the label of a data point (player) i is simply given by the strategy with the highest probability in the equilibrium mixed choice of player i as:

$ch_i = \text{argmax}_{h=1,..,c}\, ch_{ih}$

and is interpreted as the final output of the player's class membership (or final probability distribution or the final consistent labeling assignment.

The cost complexity to compute Nash equilibrium of the described graph transduction game using the discrete version of replicator dynamic is $O(IcP^2)$ where I is the number of iteration for the algorithm to finish, c indicates the number of classes (i.e the available pure strategies for the players to choose) and P corresponds to the number of the participant players (i.e the nodes of the graph).

3.5 CLASSIFICATION OF SCHIZOPHRENIA: INCORPORATING 'DOMAIN-SPECIFIC KNOWLEDGE' APRIORI HYPOTHESIS

In this experiment, we applied the method described to classify healthy indivuduals vs schizophrenia patients considering gray matter thickness information extracted from their MRI scan. We incorporated a domain specific apriori hypothesis from schizophrenia studies and we considered features only from some selected regions of the brain that are believed to be affected by the disease (i.e., that would result in cortex shrinkage).

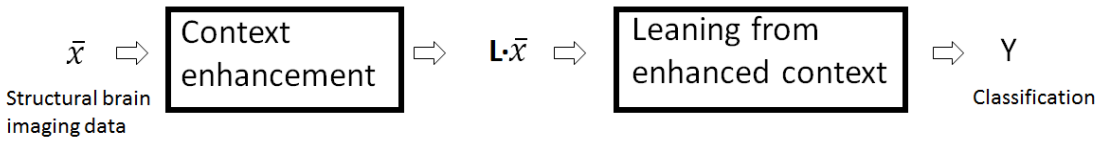


Figure 3.7: The proposed schizophrenia classification scheme using structural brain imaging data.

The proposed classification scheme exploits contextual brain anatomical similarities of subjects from MR images, so as to differentiate healthy controls from SCZ patients. A set of features, characterizing the anatomy of the brain, was obtained from the MR images of every single subject. Then, we used a DML technique, specifically the one proposed in [75], to enhance the CS of the input MRI data and apply the GT algorithm

[71] on top of this new metric space to learn from the newly enhanced context. The overall scheme is depicted in fig. 3.7 and described step-by-step in the next sections.

As explained above, the aim of GT is to address the problem of consistent labeling, with the aim of predicting or propagating class membership to unlabeled data by exploiting learning both from the labeled and unlabeled samples. Such methodology involves three different areas: i) graph theory; ii) evolutionary game theory; and iii) dynamical systems and optimization.

The main idea behind GT is to consider the samples of the dataset as nodes of a graph, and to propagate class labels to unlabeled nodes, by considering the CS among the samples. In particular, it exploits CS among data features to perform label propagation in a consistent way, relying on a common *a priori* assumption known as the "cluster assumption" (a reminiscent of the homophily principle used in social network analysis): nodes that are close to each other, in the same cluster or on the same manifold are expected to have the same label. Each node is then a feature vector $\in \mathbb{R}^d$ (with d being the number of features). Moreover, each node can select a strategy, *i.e.*, class membership, that maximizes its CS. Finally, the output labeling corresponds to the Nash equilibrium of the game.

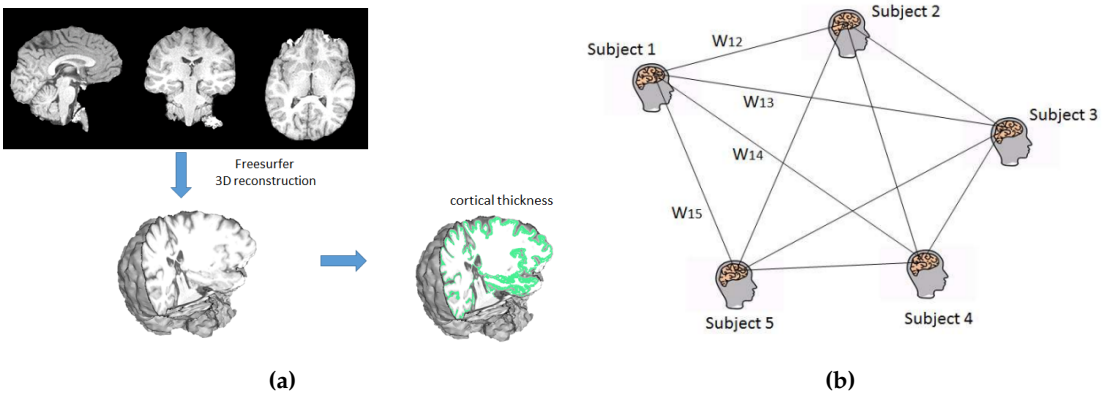
Input features are represented with graph nodes $\mathcal{G} = (\mathcal{V})$, where the vertex set $\mathcal{V}$ is composed of $n = l + u$ elements $\in \mathbb{R}^d$ and consists of a first labeled set $\{(x_1, y_1), ..., (x_l, y_l)\}$ of l elements and a second unlabeled set $\{(x_{l+1}, ..., (x_{l+u})\}$ of u elements. Then, the similarity matrix $\mathcal{E}$ between pairs of nodes is computed, after having selected a similarity metric. A simple and effective optimization algorithm to propagate the labels through the graph is given by the so-called replicator dynamics, developed and studied in evolutionary game theory, which has proven to be effective in many applications [81, 82].

In practice, labeled examples in the form of "must-be-in-the-same-class" and "must-not-be-in-the-same-class" pairs of subjects are provided to the DML framework, to learn the best feature space transformation matrix L using Eq. (3.6). Afterwards, the class label propagation occurs on such transformed feature space (*i.e.*, $L\bar{x}$) by constructing the fully connected graph $\mathcal{G} = (\mathcal{V})$, where $\mathcal{V}$ is now the set of graph nodes representing the transformed feature vectors, and $\mathcal{E}$ encodes the brain anatomy similarity between subjects by means of the edge weights (similarity matrix) as depicted in fig. 3.8b. $\mathcal{E}$ is constructed in the following manner (for simplicity we show how an edge is constructed between two transformed feature vectors):

$$\mathcal{E}_{ij} = \exp\left[-\frac{d(L\bar{x}_i, L\bar{x}_j)^2}{2\sigma^2}\right] \tag{3.10}$$

where $d(L\bar{x}_i, L\bar{x}_j)$ is the Euclidean distance. For estimating σ, which is a critical parameter of the graph's ability in representing the CS between data points, we adopted an automatic self tuning method as proposed in [83].

We performed two series of comparisons to assess the performances of the proposed classification scheme in differentiating healthy controls from SCZ subjects. First, we verified whether learning from CS (from both labeled and unlabeled data) might provide better classification results than just learning from labeled data. Second, we tested if the enhancement of CS by DML might provide further improvements. To do so,



(a) **(b)**

Figure 3.8: (a) ROI and cortical thickness feature extraction from brain images. (b) Representation of brain anatomy similarity between subjects.

we compared the proposed classification scheme (DML+GT) with both GT [71] and KNN, with and without metric learning (KNN, DML+KNN [75]), linear SVM and our implementation of DML+SVM. We evaluated the classification performances by using accuracy (Acc), sensitivity (Se), specificity (Sp) and Cohen's kappa (Ck) coefficients. Sensitivity refers to the true recognition of SCZ patients.

We considered first 70%, then 80% of the data from each class for training and input labeling of GT, while the rest of data was left to be predicted. In fact, GT was found to perform sufficiently well when the labeled data were just a small fraction of the dataset [71]. However, given the small size of our dataset, we considered labeling 70% and 80% of the data at disposal. We repeated this procedure by randomly sampling the dataset 100 times and computed the average performance. In all the experiments, we avoided the risk of circularity [84]. For KNN, we chose $K = 3$, for limiting the possible overfitting due to the relatively small sample size.

3.5.1 DATASET AND REPRESENTATION

The dataset consisted in T1-weigthed MR images of 20 healthy control subjects (35.8±13, 8 males) and 20 SCZ patients (37.9±11, 13 males). The size of this dataset is in line with the dimensionality of datasets used in academic works aimed at medical applications [85, 86, 87]: in particular, it is not straightforward to obtain consistent MRI data of psychiatric patients, due to difficulties in recruitment and feasibility of MRI acquisitions in this kind of patients. The data were collected at the Psychiatric department of Ospedale di Verona (Verona, Italy). All involved subjects signed an informed consent, following the principle of the Helsinki's declaration.

The T1-weigthed images were preprocessed using the software FreeSurfer[1] as depicted in fig. 3.8a. Based on prior knowledge on schizophrenia [88, 89], we considered the average cortical grey matter thickness of frontal and temporal regions (namely: *caudal middle frontal*, *inferior temporal*, *middle temporal*, *rostral middle frontal* and *superior*

1 http://surfer.nmr.mgh.harvard.edu/

ROI	HC (mean± std)	SCZ (mean± std)
Caudal-middle-frontal	2.56 ± 0.21	2.44 ± 0.18
Inferior-temporal	2.85 ± 0.16	2.64 ± 0.17
Middle-temporal	2.89 ± 0.17	2.73 ± 0.18
Rostral-middle-frontal	2.47 ± 0.19	2.34 ± 0.17
Superior-frontal	2.82 ± 0.18	2.65 ± 0.20

Table 3.1: Left hemisphere gray matter cortical thickness of ROIs (in mm) of healthy controls and schizophrenia patients.

frontal of the left hemisphere) as features in the classification task. The ROI thickness measurement of the subjects is reported in table 3.1. Also, in order to take into account the effect of age on the cortical thickness, we corrected all the data for age differences using a generalized linear model [90].

3.5.2 CORTICAL THICKNESS FEATURE EXTRACTION

Grey matter thickness values are obtained using Freesurfer. Since the brain images aquired from MRI machines have low signal to noise ratio, FreeSurfer reconstructs brain surfaces in a deterministic approach so as to have a standard comparable images between individuals. Usually, it takes several hours to process a single scan and several days to analyze a dataset.

Using a T1 weighted anatomical image with a good contrast between the white matter and the gray matter as an initial input, the skull from the volume is removed as a first step. Then the volume is segmented to cortex (i.e., gray matter), white matter and sub cortical structures. More specifically, the boundaries between the tissues are computed (i.e., between white matter tissue and grey matter tissue), then between gray matter tissue and cerebral fluid (known as pial surface). Finally, the distance between the white matter surface and the pial surface i.e., the cortical (gray matter) thickness distribution is extracted as a feature.

3.5.3 RESULTS

The average and standard error of the classification performance for DML+GT (proposed scheme), DML+SVM, DML+KNN, GT, SVM and KNN (used for comparison), when 70% and 80% of the samples in each class are labeled are reported in table 3.2 and fig. 3.9.

As expected, performance got better when using higher percentage of labeled data on small datasets for DML+GT and GT. Moreover, in our proposed scheme, sensitivity was always lower than specificity (fig. 3.9c and fig. 3.9d), meaning that some subjects with schizophrenia were classified as healthy, regardless the labeled sample size. In addition, the increase of training data provided different relative improvements between sensitivity and specificity (fig. 3.9c and fig. 3.9d). This means that these methodologies,

under the settings we considered, are capable of recognizing the healthy subjects more easily than the schizophrenia patients.

GT was more affected by the training set's size (fourth bar in each plot of fig. 3.9) than the other methods. However, when DML was applied before GT, we obtained a drastic classification improvement of all measures except the sensitivity, even with a smaller training set. Furthermore, the use of DML resulted in a higher performance in all the cases except sensitivity (fig. 3.9).

Finally, when 80% of the data is used as training, CS learning, *i.e.*, learning from unlabeled data as well, enhanced with DML outperformed both SVM and KNN with DML (first bar vs second and third bar).

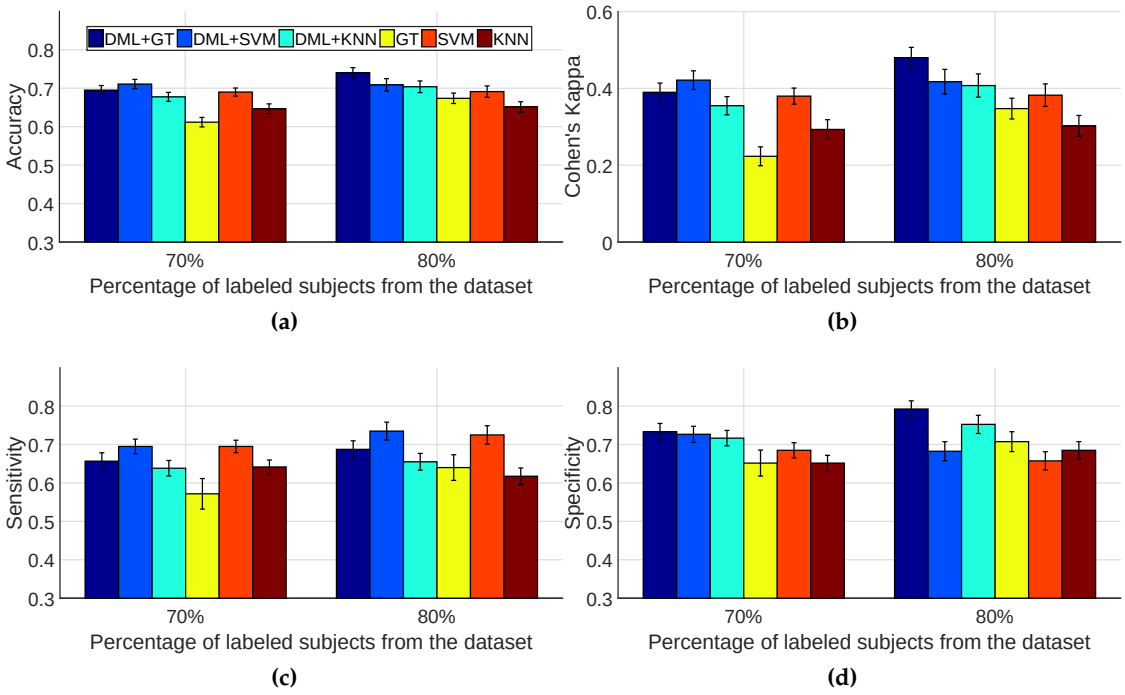


Figure 3.9: Classification results for healthy controls vs. schizophrenia patients. Average performances and standard errors of the mean are reported.

3.5.4 DISCUSSION

This work supports the finding that DML+KNN is better than KNN (*i.e.*, with respect to every evaluation metric considered), as found by other authors [75]. In particular, we showed that this finding holds true when applied to thickness features extracted from MRI data.

Moreover, GT is consistently improved by the proposed scheme (DML+GT), which suggests that contextual similarity (CS) enhancement of MRI data coupled with learning from unlabeled samples, can result in a better performance of classifying schizophrenia. This result is also supported in [91], within the computer vision domain (object recognition and scene classification).

Table 3.2: Average test-set classification performance (± standard deviation across subjects) on brain sMRI data features using 70% and 80% of the data for training.

Methods		70%	80%
DML+GT	Acc	0.70 ± 0.01	0.74 ± 0.01
	Se	0.66 ± 0.02	0.69 ± 0.02
	Sp	0.73 ± 0.02	0.79 ± 0.02
	Ck	0.39 ± 0.02	0.48 ± 0.03
DML+SVM	Acc	0.71 ± 0.01	0.71 ± 0.02
	Se	0.70 ± 0.02	0.74 ± 0.02
	Sp	0.73 ± 0.02	0.68 ± 0.02
	Ck	0.42 ± 0.02	0.42 ± 0.03
DML+KNN	Acc	0.68 ± 0.01	0.70 ± 0.02
	Se	0.64 ± 0.02	0.66 ± 0.02
	Sp	0.72 ± 0.02	0.75 ± 0.02
	Ck	0.36 ± 0.02	0.41 ± 0.03
GT	Acc	0.61 ± 0.01	0.67 ± 0.01
	Se	0.57 ± 0.04	0.64 ± 0.03
	Sp	0.65 ± 0.03	0.71 ± 0.03
	Ck	0.22 ± 0.02	0.35 ± 0.03
SVM	Acc	0.69 ± 0.01	0.69 ± 0.01
	Se	0.70 ± 0.02	0.73 ± 0.02
	Sp	0.69 ± 0.02	0.66 ± 0.02
	Ck	0.38 ± 0.02	0.38 ± 0.03
KNN	Acc	0.65 ± 0.01	0.65 ± 0.01
	Se	0.64 ± 0.02	0.62 ± 0.02
	Sp	0.65 ± 0.02	0.69 ± 0.02
	Ck	0.29 ± 0.03	0.30 ± 0.03

Finally, DML+GT performances are higher likely due to the additional information obtained from the unlabeled MRI data features. This confirms that DML and CS has the potential to improve schizophrenia classification.

The results obtained are comparable to the state-of-the-art in classification of schizophrenia. For example, in [92] using functional MRI (fMRI), they obtained an average classification accuracy of 0.59 and 0.84 using both static and dynamic resting-state functional network connectivity approach respectively and linear SVM. In [93] they obtained up to 0.75 accuracy (combining ROI thickness features) using 1.5 T sMRI and covariate multiple kernel learning approach using SVM. In [40], they achieved 0.75 accuracy considering the left hemisphere.

3.6 SUMMARY

In this study, we designed a classification scheme to discriminate healthy controls from schizophrenia patients using MR images-derived data as features. We believe that learning from contextual anatomical similarity of subjects (*i.e.*, learning both from labeled and unlabeled MRI data features) has a great potential in dealing with schizophrenia, due to the nature and complexity of the disease and its associated diagnostic uncertainty (and further more, due to a very subtle difference in the MRI images of healthy controls and patients). So transductive learning is an easier and it could learn using subtle anatomical differences better than inductive approaches by using a specific model for a specific problem instead of deriving a generalized model for a general problem [69, 94]. Also considering the limited dataset in the field, it can be an advantage to exploit this learning approach. It is meant to describe the learning approach by the algorithms, for example in neural neural network (NN) training, the common approach is that the NN is trained to model a certain representation of a data, and that model would be considered a general model. But the proposed extended learning algorithm in chapter 3, it learns the feature-class association from similarity of features by exploiting the graph structure. The features are represented by graph nodes and edges denoting their distance-based similarity. So, the learning approach is learning from contextual similarity instead of updating a learning model parameter after each single visit of data in example-based training. This is useful because it is effective method when there are not enough training data or enough phenomena found in the sample training data, for building a general model.

Furthermore, we showed that enhancing the CS improved the classification performances of the label propagation algorithm (semi-supervised context learning). We demonstrated that the combination of metric learning and graph transduction (DML+GT) is useful to learn a meaningful underlying pattern from MRI data by exploiting contextual information, resulting in better classification performances.

In the future, we would like to test a non-linear metric for context enhancing to assess if it can further improve the classification results. Also, GT could be improved by using another anatomical feature (dis)similarity measurement instead of the symmetric Euclidean distance of eq. (3.10), since it can handle asymmetric (dis)similarities also.

4

TOWARDS AUTOMATIC IDENTIFICATION OF MRI-DERIVED ANATOMICAL BIOMARKERS

> Nothing in the world is ever completely wrong. Even a stopped clock is right twice a day.
>
> Paulo Coelho

In this chapter of the thesis, it is described the attempt for finding imaging derived anatomical biomarker(s) for psychiatric disorders. For this task, we chose to consider the case of chronic stage BD and first episode psychosis (FEP).

As mentioned in the introduction section, BD is believed to cause brain atrophy in several cortical regions of the brain, however, until now there is not definitive atrophy quantification and identification of regions that can serve as markers. In the literature some results overlap and some not.

By taking this into consideration, we divide the brain regions using the Desikan-killiany gyral based brain atlas [95], then we extracted cortical thickness distribution features from each region of interest (ROI) independently. This thickness distribution is treated as a diagnostic metric. After that, it is aimed to find which region (or combination of regions) has the most contextual difference between healthy and the pathological groups. Finally, by applying our proposed method described in chapter 3, we chose to consider the results as a potential marker for the considered psychiatric disorder. Specifically, by exploiting a signal processing and machine learning strategy, we tried to answer, the question of 'which ROIs could result in more accurate differentiation of patients from health controls?'. To do so, skewness and mean thickness of the cortical thickness distribution is considered to asses the brain atrophy phenomenon. The problem is casted as a relevant feature selection problem.

Trivially, one can try several combination of the features from the 68 regions obtained from the atlas and aim to find which region (or combination of regions) would yield

the maximum differentiation between groups (i.e., higher classification performance), but eventually one would face explosion of combinations which is a NP hard problem. As a consequence, we chose to use heuristics to avoid this phenomenon.

The approach in this chapter is the extension of the approach from the previous experiment (in chapter 3, sub section 3.5). In here, we considered all the features derived from whole brain regions in a region of interest (ROI) based manner i.e., without considering apriori domain specific knowledge, unlike the previous experiment where we considered only five regions that are known to be affected by the psychiatric disorder, neglecting features from other ROIs. Furthermore, in here, the novel brain feature proposed in chapter 2 (i.e., cortical skewness) is used in addition to the mean thickness feature which is most widely used in the literature.

4.1 DATASET

- **BD dataset**: comprising 75 sMRI scans grouped into 34 healthy controls and 41 chronic bipolar-disorder patients, which we used for this experiment is described in chapter 2, section 2.3.1. The scans are collected from Verona center and diagnosis labels are assigned by an experienced psychiatrist.
- **FEP dataset**: comprising 174 sMRI scans grouped into 93 healthy controls and 81 subjects who experienced first episode psychosis. The scans are collected from Verona center and diagnosis labels are assigned by an experienced psychiatrist.

4.2 NEUROANATOMICAL BIOMARKER FEATURE EXTRACTION

Cortical thickness distributions of 58 ROIs were obtained from 1.5 and 3T MR images using FreeSurfer [1], 29 located in the right and 29 in the left brain hemisphere, identified with the Desikan-Killiany atlas [95]. For each ROI, two scalar values are quantified to be used as the features of interest: the mean and skewness of the distribution of the gray matter thickness of the cerebral cortex. Skewness is a measure of asymmetry of the distribution around its mean and can assume a negative, zero or positive value. Overall, each subject was represented with a 116-dimensional feature vector composed of 58 values of mean thickness and 58 values of skewness. We referred all this 116 features as candidate anatomical biomarkers in campaign where one (or a specific combination of them) are going to be selected as a relevant potential biomarkers which can facilitate the computer aided diagnosis (or classification) of bipolar disorder.

The feature extraction procedure is illustrated in figure 2.2 of chapter 2. To take into account possible differences in brain structure due to age difference of the considered subjects (which could alter the results), we corrected for age the mean thickness and skewness values using a linear model, following the suggestion of [21]. Specifically, we considered age as predictor, and mean thickness and skewness as responses. Then, once removed what was predicted by the model, only residuals were further employed. Equal proportion of gender did not require any correction of cortical thickness based

1 version 4.3.1, http://surfer.nmr.mgh.harvard.edu/

on sex. After age-correction, features were normalized (by subtracting the average and dividing by their standard deviation).

4.3 NEUROANATOMICAL FEATURE CLASSIFICATION

We applied a graph-based semi-supervised learning algorithm [71] to classify our data. As in any semi-supervised method, labels are not necessarily available for each subject, and classification information is derived from the labels that are present. Graph-based semi-supervised learning operates as follows. In a first step, the subjects' feature vectors (both labeled and unlabeled) are represented as graph nodes and the similarity weight between them is calculated using the Euclidean distance and an exponential kernel (Equation 4.1, where L is the identity matrix). Then, class-membership labels propagate from the labeled nodes to the unlabeled ones based on proximity in feature space (transductive algorithms). Unfortunately, label propagation using these algorithms heavily relies on the pre-existing underlying structure of the graph. A possible strategy to improve classification performances is to derive a (feature space) transformation matrix that increases separation of nodes belonging to different classes (a procedure called "distance metric learning" that is described in 3.3). Specifically, we implemented the large margin nearest neighbor (LMNN) method [75].

The aim of this procedure is to derive a transformation matrix considering only labeled data (training set), which was used to transform the entire dataset by increasing the intra-similarity and decreasing inter-similarity of the two classes (healthy subjects and bipolar disorder patients). That is, the transformation matrix L causes a displacement of the nodes (feature vectors) in space, such that nodes belonging to the same class became closer to each other as illustrated in figure 3.2 and 3.5. The displacement depends on the amount of neighboring nodes considered (a parameter of the procedure). Given the limited dimension of the dataset, we considered 24 nodes during feature selection and 32 during comparison between selected feature sets (that is, the largest possible number of neighboring nodes). The similarity weight on the transformed graph between subject i and subject j as explained in chapter 3, was computed by:

$$w_{ij} = \exp\left[-\frac{d(L\bar{x}_i, L\bar{x}_j)}{2\sigma}\right] \tag{4.1}$$

where L is the learned transformation matrix, computed using LMNN, x_i and $x_j$are the anatomical feature vectors of subject i and j respectively, σ is the scaling parameter (or kernel bandwidth, estimated automatically from data using the self-tuning procedure proposed in [83]) and d is the Euclidean distance. Finally, the label propagation on the graph was executed using a class of dynamical systems that solves a quadratic optimization problem known as replicator dynamics [74]. The processing scheme which is used also for the previous experiment is depicted in Figure 3.1.

4.4 A GAME OF COLLABORATIVE CAMPAIGN OF BIOMARKERS

As discussed in the introduction of this thesis, most studies of neurological and psychiatric disease consider only mean thickness in their analyses as an MRI derived imaging biomarker as a diagnostic metric, and often characterizing a single region of the brain at a time. In this experiment, we considered a novel "skewness" metric.

The procedure is termed as a game of collaborative campaign of biomarkers in the sense that the term game is used because the classifier used is based on the notions of game theory and the features used are neuroanatomcal markers of disease. Finally, the term collaborative campaign is used because of the fact that, combinations of features that could yield in better differentiation of the two classes are searched and finally the best combination is voted based on the classification performance.

A feature selection procedure (intending as 'feature' either the mean or the skewness of the values of cortex thickness in one of the 58 regions of interest) was set up to identify those combinations of features that were more effective in the classification problem considered. First, we empirically decided to limit the selection to combinations of at most five features (around 5% of the features, as done in the work by Taylor and Kim (2011)). We believe that five features could be a suitable choice to reduce the risk of over-fitting considering the dataset size, at the same time preserving a sufficient descriptive capability. To ensure generalization of our results, feature selection was performed, using a weighted-vote-based algorithm, repeatedly splitting the dataset into training and validation set. The procedure was performed iteratively keeping each subject out (i.e., test subject) using a jackknife procedure. A training set was obtained by randomly sampling 25 HC and 25 BD subjects, while the remaining 24 subjects of each class (not including the test subject) were used for the validation set. Votes were assigned to features accordingly to a metric computed on the validation set. The metric we employed was the geometric mean (GMEAN) between sensitivity (Se, here true recognition rate of BD) and specificity (Sp), i.e., the square root of the product of Se and Sp. GMEAN tends to be large only when Se and Sp have similar values, thus avoiding a bias towards either of them.

In detail, the feature selection algorithm was defined as follows. First, the training and validation sets were used to determine a preliminary smaller set (S1) of features less correlated among them. In particular, for each pair of features that were moderately correlated between each other ($|\text{Pearson's correlation coefficient}| > 0.6$), we retained the one with the highest validation GMEAN. The latter was estimated by setting the cut-off value corresponding to the left-upper corner of the ROC curve (which ensures similar values of Se and Sp).

Second, we further reduced the feature set S1 and created the feature set S2 with a Greedy Forward Feature Selection (GFFS) approach, using the algorithm introduced in the Classification section. GFFS iteratively adds up to five features, one at a time, from S1 to S2, initially empty, to maximize the validation GMEAN. The GFFS procedure was stopped when the addition of any extra feature was going to reduce the GMEAN. Due to statistical fluctuations, the features in S2 might depend on the specific random sample used to build the training set. To improve repeatability and generalization,

feature selection, as just described, was repeated 100 times, and a weighted-voting procedure was used to decide among the possible 100 different S2 sets. A histogram of votes was created, where each S2 set was weighted by its GMEAN value (regardless of the order with which the features entered the set). For example, with the S2 sets: (F1, F3, 0.9), (F1, F2, F3, 0.8), (F3, F1, 0.5), where the last number in each group is the maximum GMEAN value, then, the combination (F1, F3) would have obtained a vote of 0.9+0.5=1.4 and (F1, F2, F3) of 0.8. Given the fact that we obtained a voting histogram for each test subject, a comprehensive voting histogram was built simply summing the 75 different ones, as in the standard jackknife. The whole procedure is depicted in Figure 4. Finally, from the comprehensive voting histogram, we took, as the outcome of our feature selection procedure, the most voted combination of M features, with M ranging from 1 to 5 (please notice that each S2 set which built the histogram was composed of up to 5 features).

4.5 CONTRIBUTION OF MEAN THICKNESS AND SKEWNESS ALONE

The feature selection procedure just described was started from a large feature set composed of 58 features describing the mean thickness and other 58 characterizing the skewness of different regions of the brain (i.e., modality 1). To verify if features based on mean thickness or skewness alone were more effective in classifying HC and BD subjects, we repeated the feature selection: i) considering only mean thickness (i.e., modality 2); ii) considering only thickness skewness (i.e., modality 3). These extra two modalities led to 5 final combinations of features each.

4.6 CLASSIFICATION ASSESSMENT

The classification accuracy for each of the final combinations of features was assessed, as already indicated, on the test subject, using a so-called Leave-One Subject-Out Cross-Validation approach (LOSOCV, see Figure 4.1). We also evaluated whether the classification accuracies obtained for each combination were statistically different from those of a (simple) majority classifier, by means of a McNemar's test adjusted for multiple comparisons using the Holm-Bonferroni's correction (significance level at 0.05). All the analyses were performed using MATLAB 2016 (Mathworks, Natick, MA, USA).

4.7 RESULTS

4.7.1 NEUROANATOMICAL BIOMARKERS OF CHRONIC BD

In any modality considered, the largest GMEAN was achieved when more than just one feature were selected (last column of Table 4.1, 4.2, 4.3). In particular, when considering the mean thickness alone (modality 2), or in combination with the skewness (modality 1), GMEAN generally increased when including in the feature set up to 3 elements (Table 4.1 and 4.2). On the other hand, with skewness-derived features (modality

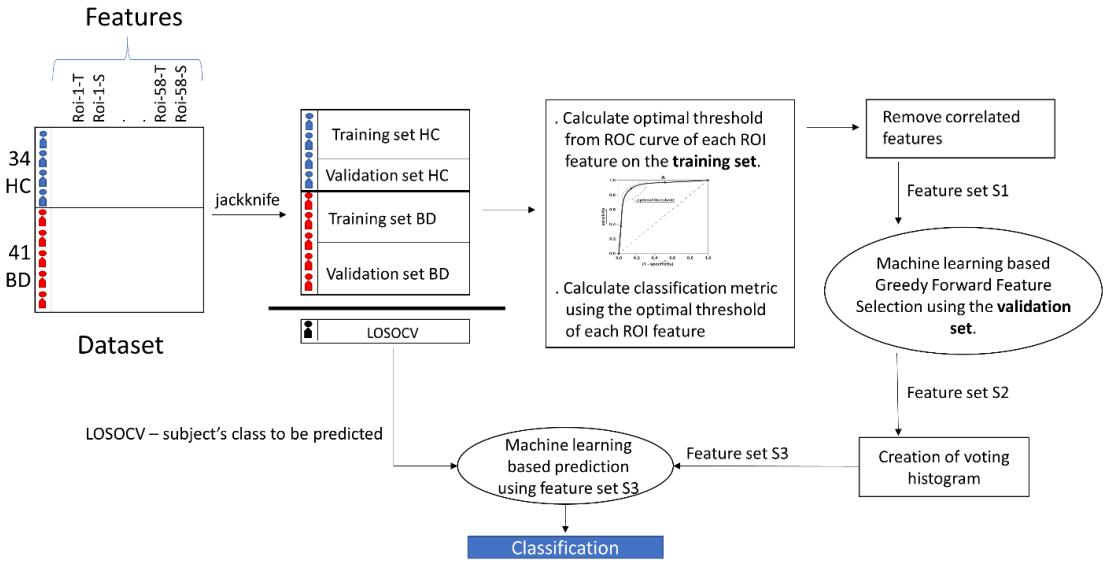


Figure 4.1: Illustration of the feature selection process and subject's class prediction. First, the dataset is divided into training and validation sets after taking a subject out with the jackknife technique. These two sets are used to select relevant features by means of a combination of ROC analysis and Greedy Forward Feature Selection algorithm, repeated 100 times to produce the voting histogram. Then, the average voting histogram is computed, as in the standard jackknife procedure, and the most-voted not-correlated features are retained. The semi-supervised learning algorithm is represented in the oval block and described in Figure 3.1. After selecting the most voted features, a LOSOCV procedure is used to determine the validation accuracy.

3) up to 5 features were required to maximize the overall validation GMEAN (Table 4.3). As expected, with lower feature set sizes, the percentage of votes of the combination selected was significantly larger than the average of vote received by the other combinations (up to about 19 times the average, third column of Table 4.1, 4.2 and 4.3). This is likely due to the large number of possible combinations of features that can be encountered when the feature set size increases. Interestingly, in modality 3, features sets of size two displayed a percentage of votes larger than those of size one. Accuracies ranged from 60% to 75% for all the three modalities (second column of Table 4.1, 4.2 and 4.3). However, only combinations of features that obtained an accuracy higher than about 70% were statistically significantly different from the one obtained by a majority classifier. Interestingly, when considering the mean thickness (alone or in combination with the skewness), the accuracy obtained with a single feature did not differ from the one achieved with many features (first row vs last row of the second column of Table 4.1 and 4.2). However, it seemed that the addition of other regions might set a different trade-off between sensitivity and specificity (second column of Table 4.1 and 4.2). On the contrary, when using only skewness, we found an increasing trend in terms of performance from dimension 1 to 5 (second column in Table 4.3). Of note, the combination of 5 skewness features (last row of Table 4.3) obtained the highest specificity (76%) for modality 3 and the second highest among all the 15 combinations within S3 (78% obtained by the right-hemisphere lateral occipital mean thickness; first row of table 4.1).

The regions involved in the 15 combinations were from the frontal lobe (superior frontal, pars triangularis, pars opercularis, lateral orbitofrontal, frontal pole), from the temporal lobe (entorhinal, banks of the superior temporal sulcus, temporal pole), from the parietal lobe (superior, inferior and middle parietal, postcentral) and from the occipital lobe (lateral occipital, pericalcarine). Among the mean thickness features, the banks of the superior temporal sulcus, part of the temporal lobe, were selected the highest amount of times (3 in modality 1 and 7 in modality 2), and pericalcarine (part of the occipital lobe) ranked second (1 in modality 1 and 3 in modality 2). On the other hand, for the skewness-related features, inferior parietal was chosen five times (3 in modality 1 and 2 in modality 3) while middle temporal, pericalcarine and superior frontal gyrus were selected 3 times. Regarding modality 1, mean thickness and skewness were selected 9 and 6 times, respectively. In particular, mean thickness was predominant for the first 2 dimensions (first and second row of Table 4.1), while skewness came into play starting from feature sets of size 3 (third row of Table 4.1). Also, when 5 features were considered, 3 were skewness-based (last row of Table 4.1). The left hemisphere was selected more often than the right one for any modality (10 vs 5 for modality 1, 9 vs 6 for modality 2 and 11 vs 4 for the modality 3).

Table 4.1: Feature and region selected, along with their classification performance (Acc, Se and Sp) for modality 1, i.e., mean thickness and skewness combinations in S3. $*$ refers to $p<0.05$ and ** for $p<0.05$ adjusted with the Holm-Bonferroni's correction.

Features	Acc/Se/Sp[%]	% of vote/average of the others (ratio)	% of times a specific combination dimension has been selected
Rh-LO-T	71/73/78*	27.09/1.43 (18.95)	7.79
Rh-Cu-T + Lh-Ba-T	64/71/56	1.51/0.13 (11.46)	21.09
Lh-OL-T + Lh-IP-S + Lh-Ba-T	67/68/65	0.32/0.06 (5.41)	25.43
Lh-SFG-S + Lh-Pe-T + Lh-IP-S + Rh-In-T	75/83/65**	0.13/0.06 (2.22)	23.73
Lh-Pe-S + Lh-OL-T + Lh-IP-S + Rh-Ba-T + Rh-Ba-S	73/80/65**	0.12/0.06 (2.02)	21.96

Lh: Left hemisphere; Rh: Right hemisphere; T: mean thickness; S: skewness; Ba: Banks of the Superior Temporal Sulcus; Cu: Cuneus; IP: Inferior Parietal; LO: Lateral Occipital; OL: Orbito Lateral; Pe: Pericalcarine; SFG: Superior Frontal Gyrus; In: Inferior Temporal.

Finally, all the regions and features selected had normally distributed residuals (Kolmogorov-Smirnov; $p<0.05$), suggesting a proper correction for the age. In the result section of chapter 2, it is reported the values of raw and age-corrected mean thickness and skewness, for all the regions selected by the algorithm.

Table 4.2: Feature and region selected, along with their classification performance (Acc, Se and Sp) for modality 2, i.e., mean thickness combinations in S3. * refers to p<0.05 and ** for p<0.05 adjusted with the Holm-Bonferroni's correction.

Identified Neuroanatomical biomarkers	Acc/Se/Sp [%]	% of vote/average % of the others (ratio)	% of times a specific combination dimension has been selected
Rh-LO	71/73/68*	25.49/1.66 (15.39)	15.11
Rh-Cu + Lh-Ba	64/71/56	2.92/0.18 (16.60)	24.53
Lh-Pe + Rh-Ba + Lh-Ba	69/76/62	0.83/0.08 (10.91)	22.79
Lh-Tr + Lh-Pe + Rh-Ba + Lh-Ba	72/76/68*	0.34/0.07 (4.76)	19.72
Lh-Pe + Lh-FP + Rh-En + Rh-Ba + Lh-Ba	71/78/62*	0.22/0.08 (2.96)	17.85

Lh: Left hemisphere; Rh: Right hemisphere; Cu: Cuneus; Pe: Pericalcarine; LO: Lateral Occipital; Ba: Banks of the Superior Temporal Sulcus; Ba: Left hemisphere Banks of the Superior Temporal Sulcus; Tr: Triangularis; En: Enthorinal; FP: Frontal Pole.

Table 4.3: Feature and region selected, along with their classification performance (Acc, Se and Sp) for modality 3, i.e., skewness combinations in S3. * refers to p<0.05 and ** for p<0.05 adjusted with the Holm-Bonferroni's correction.

Identified Neuroanatomical biomarkers	Acc/Se/Sp [%]	% of vote/average % of the others (ratio)	% of times a specific combination dimension has been selected
Lh-Sm	59/68/47	18.89/3.00 (6.29)	3.88
Lh-MT + Lh-IP	63/66/59	3.98/0.22 (18.12)	14.73
Rh-TL + Lh-SFG + Lh-IP	72/78/65*	0.73/0.07 (10.35)	25.09
Rh-TL + Rh-PC + Lh-Pe + Lh-MT	61/66/56	0.26/0.05 (4.84)	26.24
Rh-SP + Lh-SFG + Lh-Pe + Lh-Op + Lh-MT	75/73/76*	0.10/0.04 (2.14)	30.05

Lh: Left hemisphere; Rh: Right hemisphere; Sm: Supramarginal; MT: Middle Temporal; IP: Inferior Parietal; TL: Temporal Lobe; PC: Post Central; Pe: Pericalcarine; SP: Superior Parietal; SFG: Superior Frontal Gyrus; Op: Opercularis.

4.7.2 NEUROANATOMICAL BIOMARKERS OF FIRST EPISODE PSYCHOSIS

The whole result with rich information is presented in table 4.4. When both mean and skewness features are considered (i.e., modality 1) aiming to find 2 best neuroanatomical markers, 3 of the ML algorithms (including the one proposed in chapter 3) identified left temporal sulcus bank and right cuneus mean thickness.

DML+GT (the classification pipeline proposed in chapter 3), within modality 1 setting, highlighted the mean thickness of the lateral occipital region as important biomarker with 71% accuracy, 73% sensitivity and 68% specificity. The same result is exhibited in the setting of modality 2, followed by another highlight, two combination of biomarkers (i.e., right hemisphere cuneus and left temporal sulcus bank) with 67% accuracy, 71% sensitivity and 62% specificity. Finally, under modality 3 setting, when only skewness metric is employed, the combination of right hemisphere temporal lobe, left hemisphere superior frontal gyrus and left hemisphere inferior parietal, is selected as a good combination of markers of FEP with 72% accuracy, 80% sensitivity and 62% specificity.

The highest promising biomarkers are highlighted when SVM is used under the setting of modality 1 (i.e., mean thickness of left hemisphere SFG & right hemisphere temporal sulcus bank, and skewness of left hemisphere inferior parietal & middle temporal) with 76% accuracy, 73% sensitivity and 79% specificity.

Finally, the biomarkers identified for chronic BD is tested on the FEP data (table 4.5). The most voted biomarkers are the skewness measurements from the left hemisphere of inferior parietal, middle temporal, pericalcarine and supramarginal regions of the brain, with 69% accuracy, 80% sensitivity and 56% specificity.

Table 4.4: Identified neuroanatomical biomarkers when considering both candidate features (i.e., mean and skewness of cortex thickness distribution) from 58 regions.

ML methods	Identified neuroanatomical biomarkers	Acc/Se/Sp [%]
DML+GTG	**Considering both Candidate features**	
	LO-Rh-T	71/73/68
	Cu-Rh-T + Ba-Lh-T	67/71/62
	Ol-Lh-T + IP-Lh-S + Ba-Lh-T	68/73/62
	SFG-Lh-S + Pe-Lh-T + IP-Lh-S + IT-Rh-T	71/88/50
	Pe-Lh-S + Ol-Lh-T + IP-Lh-S + Ba-Rh-S + Ba-Rh-T	67/83/47
	Considering only Cortical mean-thickness Features	
	LO-Rh	71/73/68
	Cu-Rh + Ba-Lh	67/71/62
	Pe-Lh + Ba-Rh + Ba-Lh	67/76/56
	Tr-Lh + Pe-Lh + Ba-Rh + Ba-Lh	65/78/50

	Pe-Lh + FP-Lh + Ent-Rh + Ba-Rh + Ba-Lh	61/78/41
	Considering only Cortical skewness Features	
	Sm-Lh	59/68/47
	MT-Lh + IP-Lh	67/73/59
	TL-Rh + SFG-Lh + IP-Lh	72/80/62
	TL-Rh + PC-Rh + Pe-Lh + MT-Lh	65/85/41
	SP-Rh + SFG-Lh + Pe-Lh + Op-Lh + MT-Lh	68/76/59
SVM	**Considering both Candidate features**	
	Cu-Rh-T	65/83/44
	Cu-Rh-T + Ba-Lh-T	69/71/68
	Pe-Lh-S + Pe-Lh-T + Cu-Lh-T	64/78/47
	SFG-Lh-T + MT-Lh-S + IP-Lh-S + Ba-Rh-T	76/73/79
	Pe-Lh-S + Pe-Lh-T + IP-Lh-S + IT-Lh-S + Ba-Lh-T	73/76/71
	Considering only Cortical mean-thickness Features	
	Cu-Rh	65/83/44
	Cu-Rh + Ba-Lh	69/71/68
	Or-Lh + Ba-Rh + Ba-Lh	67/71/62
	OL-Rh + LO-Rh + Ba-Rh + Ba-Lh	67/73/59
	PCu-Lh + Or-Lh + Li-Rh + En-Rh + Ba-Rh	65/71/59
	Considering only Cortical skewness Features	
	SFG-Rh	57/76/35
	SFG-Rh + Li-Rh	57/76/35
	SFG-Lh + Pe-Lh + IP-Lh	71/73/68
	SP-Rh + PC-Rh + Ol-Rh + MT-Lh	71/71/71
	Om-Lh + RMT-Rh + RMT-Lh + IP-Rh + IP-Lh	72/73/71
Log-Reg	**Considering both Candidate features**	
	OL-Lh-T	60/78/38
	Cu-Rh-T + Ba-Lh-T	68/71/65
	MT-Lh-S + IP-Lh-S + Ba-Rh-T	71/80/59
	Pe-Lh-T + IP-Lh-S + IT-Rh-T + Ba-Lh-T	67/76/56
	Tr-Rh-T + PCu-Rh-S + IP-Rh-S + Ba-Rh-T + Ba-Lh-T	63/68/56
	Considering only Cortical mean-thickness Features	
	Ol-Lh	60/78/38
	SFG-Lh + Ba-Rh	67/73/59
	Pe-Lh + IT-Rh + En-Rh	68/76/59
	TT-Rh + Pe-Lh + Ba-Rh + Ba-Lh	73/73/74
	Pe-Lh + FP-Lh + En-Rh + Cu-Rh + Ba-Lh	65/71/59

	Considering only Cortical skewness Features	
	IP-Rh	48/73/18
	SFG-Lh + IP-Lh	71/76/65
	SFG-Lh + IP-Lh + Cu-Lh	71/76/65
	Ol-Lh + MT-Lh + IP-Rh + IP-Lh	65/71/59
	TL-Rh + SFG-Lh + Pe-Lh + IP-Lh + Cu-Lh	75/80/68

Table 4.5: Using neuroanatomical biomarkers identified from chronic BD to test the discriminative performance in FEP.

DML+GTG	**Considering both Candidate features**	
	Pe-Lh-T	65/71/59
	Pe-Lh-T + IP-Lh-S	60/71/47
	Pe-Lh-T + IP-Lh-S+ Ba-Rh-T	65/76/53
	Pe-Lh-T + IP-Lh-S + Ba-Rh-T + LO-Rh-T	67/78/53
	Pe-Lh-T + IP-Lh-S + Ba-Rh-T + LO-Rh-T + Cu-Rh-T	64/78/47
	Considering only Cortical mean-thickness Features	
	Pe-Lh	65/71/ 59
	Pe-Lh + Ba-Rh	61/68/53
	Pe-Lh + Ba-Rh + Ba-Lh	67/76/56
	Pe-Lh + Ba-Rh + Ba-Lh + LO-Rh	65/78/50
	Pe-Lh + Ba-Rh + Ba-Lh + LO-Rh + Cu-Rh	64/80/44
	Considering only Cortical skewness Features	
	IP-Lh	59/63/53
	IP-Lh + MT-Lh	67/73/59
	IP-Lh + MT-Lh + Pe-Lh	65/73/56
	IP-Lh + MT-Lh + Pe-Lh + Sm-Lh	69/80/56
	IP-Lh + MT-Lh + Pe-Lh + Sm-Lh + Op-Lh	69/83/53

4.8 DISCUSSION

In this work, we aimed at automatically identifying the gray matter regions most involved with BD, using a SSL method. The learning phase was not entirely driven by the subjects' labels: we also took advantage of the natural groupings in our data. The possibility of using non-labelled data might avoid critical issues typical of fully supervised methods, as their vulnerability to target labels affected by uncertainty. In other words, in the context of psychiatry, with our approach, a diagnosis which can be considered, to a certain extent, subjective, is most likely to drive the definition of objective biomarkers. We considered 58 brain gray matter ROIs (29 from left and 29 from right

hemisphere), obtained from the Desikan atlas (Desikan et al., 2006), using the parcellation method included in FreeSurfer. With our proposed feature selection method, we identified the most discriminative regions, between HC and BD patients, with respect to gray matter mean thickness and skewness. The results show that the selected regions were located in particular in the temporal, parietal and occipital areas. The temporal lobe, which is involved in emotion processing and regulation, as well as behavioral control, has already been found to be affected in BD and implicated in its pathogenesis, even if results are somewhat contrasting (Heng et al, 2010). In a recent study (Altamura et al., 2017) temporal abnormalities, both in structure and metabolism, were found to be a neurobiological marker of BD and of the disruption of the implicated brain neural network. Some studies found significant thinning of temporal cortex (e.g., Elvsashagen et al., 2012), while others found increased volumes (Adler et al., 2007), or no differences in temporal structures between patients and controls (Brambilla et al., 2003). Temporal lobe white matter microstructure has also been found related to genetic risk for BD and cognitive deficits (Mahon et al., 2010; Mahon et al., 2013). The parietal lobe is involved in cognition, especially attention and memory (Ferro et al., 2017), which are heavily affected in BD, and its involvement in mental disorders, for example in schizophrenia, has been shown (Bellani et al., 2010). In a recent study (Ferro et al., 2017), it has been demonstrated that white and gray matter in parietal lobe of BD patients are reduced in comparison with healthy individuals. The gray matter volume of parietal lobe was also found reduced in BD patients with recurring mood episodes in respect to remitted patients and to HC (Kozicky et al., 2016). Altered parietal gray matter density, which could indicate tissue disruption, was also detected (Ha et al., 2009; Li et al., 2011). Moreover, the fronto-parietal circuitry has been found to be altered in BD during attention (Cabeza et al., 2000) and memory tasks (Townsend et al., 2010). An implication of parietal lobe in BD has been found also employing diffusion weighted imaging (Bellani et al., 2012), implying an involvement of parietal tissue microstructure in the disease. Our method also selected the occipital lobe thickness as discriminative between BD patients and controls. This is in agreement with previous literature: Bruno and colleagues (2008) found white matter abnormalities in temporal and occipital areas, indicating perturbations in the affect processing networks, possibly leading to mood and cognition deterioration. White matter tracts pertaining to occipital regions were found disrupted also in other studies (Benedetti et al., 2011; Sui et al., 2011). Alterations in this area have been found also in brain metabolism (Bhagwagar et al., 2007) and blood flow (Liang et al., 2013).

Regarding FEP, left hemisphere temporal sulcus bank, right hemisphere cuneus, lateral occipital, left temporal sulcus bank, right hemisphere temporal lobe, left hemisphere superior frontal gyrus, left hemisphere inferior parietal, left hemisphere SFG, right hemisphere temporal sulcus bank, left hemisphere inferior parietal, middle temporal, the left hemisphere inferior parietal, middle temporal, pericalcarine and supramarginal regions of the brain were highlighted as important biomarkers. Furthermore, the skewness measurement highlighted the anatomical variation in parietal and temporal regions.

4.9 LIMITATIONS

Some limitations should be taken into account when considering the result of this particular experiment. We recruited BD chronic patients, thus under the effects of medications, which might play a role in our findings. Also, we excluded subjects with a history of alcohol or substance abuse, a frequent co-morbidity for BD. While these choices could lead to a selection bias, we believe that the exclusion of such patients is methodologically appropriate, because it removes potential confounding effects. The size or our dataset is relatively limited, but in line with other recent machine learning works in the field (Valli et al., 2016; Mwangi et al., 2014b). Our HC and BD groups did not match for age. This was however compensated by correcting mean thickness and skewness values with a linear model. Finally, the results of our analysis were derived by using a jackknife strategy. An independent test set might have been possibly preferable. However, the main goal of our analysis was to identify important regions along with their corresponding features and it was impossible, in our view, to further split the dataset because of the limited size. Regarding FEP patients, even though there is a less confounding factors in the dataset, to confirm the markers identified, it still needs to be validated using very big multi center scans.

4.10 SUMMARY

Our results from the second experiment, demonstrate that machine learning approaches have the potential of discerning the most meaningful characteristics that differentiate patients from healthy controls. Automatically selecting the regions which discriminate BD patients and healthy subjects can be of great importance when dealing with the pathogenesis of the disorder. Our method selected regions which are known to be involved with BD, indicating that damage to the identified areas can be considered as a marker of disease, consistently differentiating brain characteristics of patients when compared with healthy. We observed increased skewness values in the left inferior parietal gyrus of BD patients, which implies a reduction in thickness in some portions of these regions with respect to HC. The mean thickness, while affected, captures less effectively this variation. As a future extension of this research, it would be interesting to explore other ways of characterizing the histogram representation of the cortical thickness distribution, such as entropy and kurtosis, while substantially increasing the population size. It would also be interesting to verify the findings in BD patients at the early phases of the illness.

en

5

MULTI-MODAL APPROACH IN COMPUTER AIDED DIFFERENTIAL CLASSIFICATION OF NEUROPSYCHIATRIC DISORDERS

> ... the patients often try to starve themselves, to hang themselves, to cut their arteries; they beg that they may be burned, buried alive, driven out into the woods and there allowed to die. One of my patients struck his neck so often on the edge of a chisel fixed on the ground that all the soft parts were cut through to the vertebrae ...
> ... there exists ... a vast area of transition where we are merely dealing with the estimation of differences in degree, so that it often depends upon the discretion and viewpoint of the observer whether the range of mental disease is wide or narrow ...
>
> Emil Kraepelin

Schizophrenia (SCZ) and bipolar disorder (BD) are among the major neuropsychiatric disorders. SCZ affects around 0.4% of the population [5] while BD is known with approximate lifetime prevalence of 4-5% in the general population [43]. Their differential diagnosis is still a very challenging task and it is not trivial, since they share a vast amount of symptoms. In fact, the Kraepelinian dichotomy [96] between SCZ and BD is currently under discussion: the hypothesis that they are not two completely distinct disease is supported by increasing evidence [97].

In these regard, anatomical neuroimaging studies try to investigate anatomical differences between those disorders [98, 97]

With the same aim of understanding the differences, psychiatrists investigate using clinical interviews, those tests asses the manifestation of the working of the cognition of the brain (i.e., attention, memory, decision making etc.).

In this regard, with the goal of understanding the nature of mental health and illness with a holistic approach in terms of varying degrees of dysfunctions in general psychological/biological systems, Recent Research Domain Criteria (RDoC) [99], which

is a research framework for new approaches to investigate mental disorders, suggests a multi-modal analysis and understanding of different data such as psychiatric interview scores, neuroimaging and genetics. Our experiment in this section, can be seen as a partial realization of this framework.

In this chapter of the thesis, the aim is to investigate the power of psychiatric interview scores in differentiating BD and SCZ, which is prevalently used by clinicians and at the same time a less costly modality, following a machine learning approach. The problem is viewed from the machine learning perspective.

Using Functional connectivity of the brain, authors in [92] tried to classify BD and SCZ patients.

With the same aim, authors of [100] studied three psychiatric interview scores to address the challenging task of differential diagnosis of SCZ and BD. Determining whether motivational, cognitive and response selection component processes of Iowa Gambling Task (IGT) [101] performance are differentially affected in SCZ and BD.

The Iowa Gambling Task [101], a simulated gambling task, is a prototypical tool for investigating the processes underpinning incentive decision making. The IGT requires participants to select cards arranged in four decks. Each card has a monetary value that is revealed only after it has been selected, and can either be a gain or a loss. The participants' aim is to optimize their net gains across trials. Two of the decks have high rewards (gains) but also higher punishments (losses), resulting in monetary loss over time and they are therefore disadvantageous. The other two decks have lower rewards but also lower punishments, making them advantageous in the long term. Participants are not told about the distribution of gains and losses associated with each deck but must deduce this from experience during the task. Over several trials, healthy individuals learn to favour the advantageous cards.

Psychiatric interviews, in the context of disease diagnosis, aim to asses and quantify the cognition abilities such as attention & psychomotor, executive function, memory and emotional & social cognition of an individual.

However, in this work, with the aim of identifying the cognition difference between controls and patients from self-reports, we focus only on analyzing nine interview response scores namely: TIB-score, Raven-score, Wisconsin Card Sorting test (WCST), memory score (Nback_1, Nback_2 & Nback_3), Motivational parameter, Expectancy/Learning parameter and Response consistency parameter. We used the interview scores to represent the subjects, i.e., we used the interview scores as features, and train an algorithm so that in the future it can predict the state of an unseen individual (i.e., healthy control, BD or SCZ) based on the interview scores. Using those nine scalar values, we represented an individual with 9-dimensional feature vector.

We used a logistic regression [102] that we used in chapter 4 for the task of neuroanatomical identification of FEP and a semi-supervised learning (SSL) algorithm [71] which works by using a very few labeled examples (minimum supervision) for training. Furthermore, we enhanced the SSL algorithm by a so called metric learning strategy [75]. The proposed application may help psychiatrists as a decision support system while diagnosing patients. Furthermore, this study may help in understanding how cognition (such as decision making) is affected by those psychiatric disorders and

shed light on the biological underpinning of the pathology by identifying the association between the interview result and the brain region of the subject.

5.1 MATERIALS AND METHODS

5.1.1 DATASET

For our analysis we used 35 healthy controls, 41 Bipolar disorder patients and 29 Schizophrenic patients. The procedure was approved by ethical committees and it is in accordance with the Helsinki declaration.

5.1.2 DATA NORMALIZATION

We normalized the data, in the following manner: (Feature-Minimum value)/(Maximum value-Minimum value). The min and max values are computed from the data.

5.1.3 STATISTICAL ANALYSIS

The t-test with Bonferroni correction was used to preliminarily compare the discriminative power of each feature between groups. We further evaluated the discriminative power of each feature using the area under the curve (AUC) of the ROC (receiver operating characteristic). Finally, we checked if the features are correlated with each other using the Pearson's correlation coefficient. During the training of the machine learning algorithms, we split the data in 80%/20%, training and test set, randomly. And this procedure was repeated 100 times to obtain an estimate of the variability for the evaluation metrics of the classification.

5.1.4 LEARNING FROM CLINICAL INTERVIEW SCORES

To learn and then predict the mental state of the subjects from their clinical interview scores, we used machine learning algorithms namely: logistic regression and a graph transduction (see chapter 3) classifiers. Logistic Regression estimates the chance, for an input, to belong to any one of the various classes we have. It assigns a weight to the 9 individual cognitive features along with a bias. The graph transduction algorithm considers the similarity between the subjects' features to give class label to unlabeled data. Even though it is a standalone method, we utilized a distance metric learning method to enhance the feature similarity weights by giving a 'must be in the same group' and 'must not be in the same group' training examples (this procedure increases the intra-cluster similarity and decreases the inter-cluster similarities). In fact, graph transduction algorithms in general heavily rely on the pre-existing graph structure of the features to classify unseen data, in this context the metric-learning procedure displaces the graph nodes (i.e., the cognitive feature represented as a graph node) with

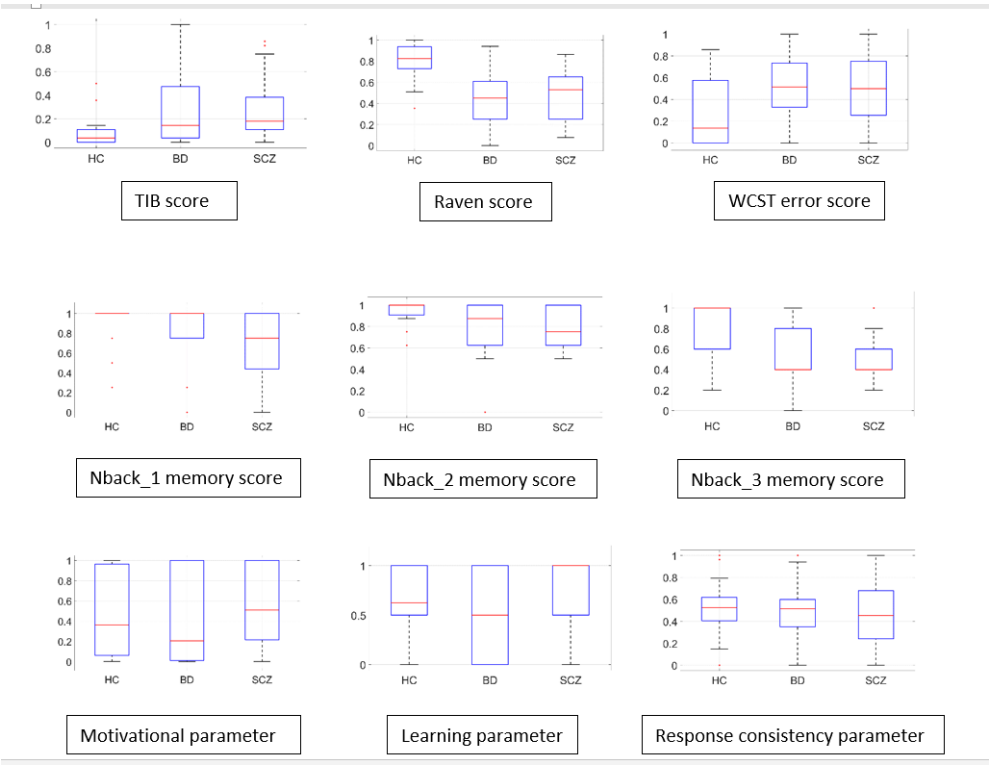


Figure 5.1: Box plots of the psychological test scores (or cognitive features).

the aim of realizing the cluster assumption (i.e., feature sets with similar class labels lie closer to each other and vice versa in N-dimensional space).

5.1.5 CLASSIFICATION AND CROSS-VALIDATION

We performed 3 independent experiments. In particular, we determined whether the machine learning algorithms mentioned in the previous subsection, might be able to correctly classify the status of a subject. Therefore, we tackled three classification problems: i) Healthy controls vs Bipolar disorder patients; ii) Healthy controls vs Schizophrenic patients; and iii) Bipolar disorder patients vs Schizophrenic patients. In the first two cases of the classification, we set apart 5 subjects from each class in a random manner (i.e., 10 subjects in total) as a test set and we used the rest for training purpose. In the third classification problem, due to a statistically significant ($p<0.05$) class imbalance in the training data, we included in the training set only 24 randomly selected bipolar-disorder patients out of the 36. Still, 5 subjects from each group were set apart as independent test set.

5.2 RESULTS AND DISCUSSIONS

5.2.1 STATISTICAL RESULTS

In figure 5.1, the box plot of the features used to represent the individuals are reported.

As it can be seen in Table 5.1, where the discriminative power of the features (quantified by p-values using t-test) are reported. Regarding HC vs BD, the Raven and Nback3 memory scores are shown to have high discriminative power followed by Wcst error score. In the case of HC vs SCZ, the Raven score, all the 3 memory scores, Wcst error score and TIB scores are shown to be statistically significantly different between the groups. But in the case of BD vs SCZ, none of the features showed statistically significant difference between the groups (p-value= 0.017, after Bonferroni correction).

Table 5.1: P-values (t-test) of the cognitive features in distinguishing between different clinical groups. P-value -> 0.05/3 = 0.017 (Bonferroni correction). Significant results are highlighted in bold.

T-test b/n groups / Cognitive features	HC vs. BD	HC vs. SCZ	BD vs. SCZ
TIB score	**0.001**	**0.000**	0.931
Raven score	**0.000**	**0.000**	0.751
WCST error score	**0.002**	**0.015**	0.739
Nback_1 memory score	**0.011**	**0.000**	0.187
Nback_2 memory score	**0.001**	**0.002**	0.693
Nback_3 memory score	**0.000**	**0.000**	0.699
Motivational parameter	0.649	0.262	0.129
Learning parameter	0.091	0.404	0.022
Response consistency parameter	0.587	0.382	0.664

The following pair of features are found to have a Pearson correlation statistically significant (i.e., p <0.05/36) after Bonferroni correction: 1. TIB score vs Raven score: -0.60, 2. TIB score vs Nback_1 memory score: -0.41, 3. TIB score vs Nback_2 memory score: -0.35, 4. Raven score vs Nback_1 memory score: 0.50, 5. Raven score vs Nback_2 memory score: 0.51, 6. Raven score vs Nback_3 memory score: 0.47, 7. Nback_1 memory score vs Nback_2 memory score: 0.51, 8. Nback_2 memory score vs Nback_3 memory score: 0.36, 9. Motivational parameter vs Learning parameter: 0.47, 10. Motivational parameter vs Response consistency parameter: -0.35 and finally, 11. Learning parameter vs Response consistency parameter: -0.49.

In table 5.2 and figure 5.2, the area under ROC curve (AUC) of the interview scores (cognitive features) are reported. Regarding HC vs SCZ, the highest AUC, that is 0.88, is obtained by Raven score followed by Nback3 memory score (0.78). Regarding HC vs BD, still those two features resulted the highest AUC and TIB is also shown to have a discriminative power (i.e., 0.81). Interestingly, regarding BD vs SCZ, the learning

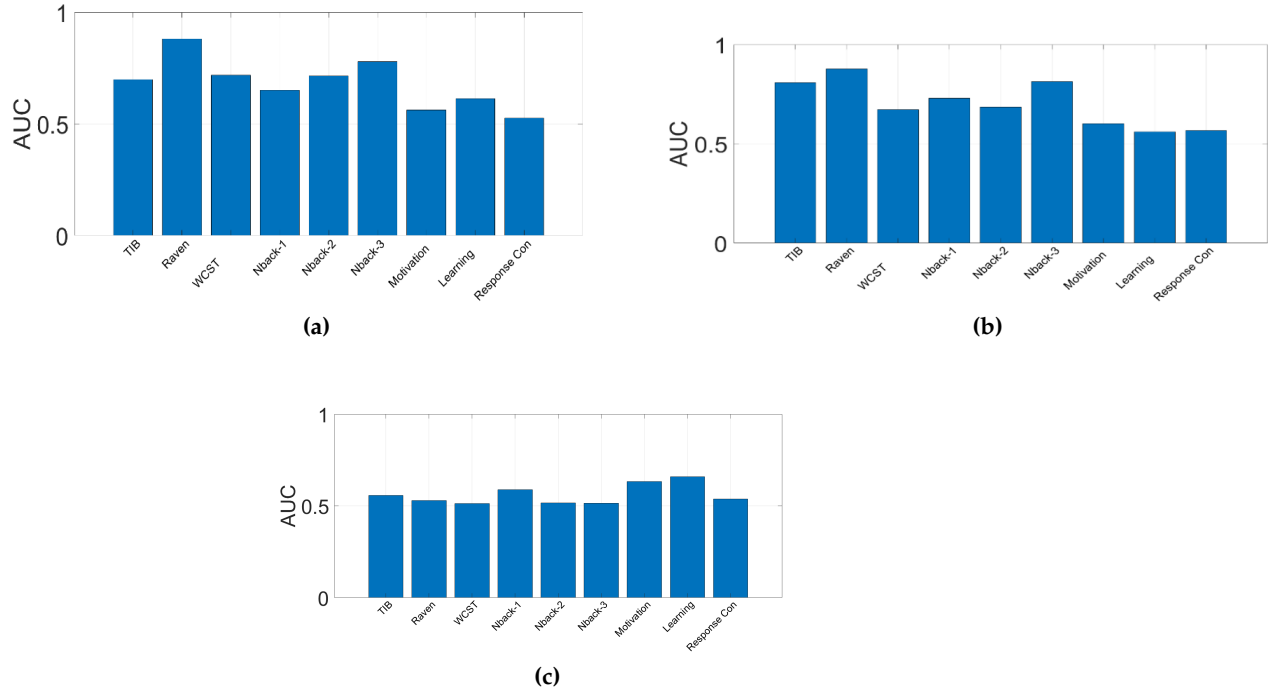


Figure 5.2: Area under the curve (AUC) of the ROC (receiver operating characteristic) for HC vs BD (a), HC vs SCZ (b) and BD vs SCZ (c).

parameter, the Motivational parameter and the response consistency parameters are shown to have the highest discriminative power when compared with the other cognitive features. And this result is in line with the study [100] where the authors reported associative learning underlying the representation of expectancies was disrupted in SCZ whereas BD was associated with increased incentive salience of gains.

5.2.2 ML CLASSIFICATION RESULTS

In table 5.3, the classification Accuracy, Sensitivity, Specificity and G-mean of the machine learning algorithms employed on the interview scores (cognitive features concatenated) are reported. In all the 3 classification problems, the sensitivity referred to the correct recognition of the second class, whereas, the specificity referred to the correct recognition of the first class.

Using the semi-supervised classifier (enhanced by the distance metric strategy) using the whole 9 features, 80% accuracy, 80% sensitivity, 80% specificity and 79% geometric mean is achieved for the case of HC vs BD. For the case of HC vs SCZ, 74% accuracy, 61% sensitivity, 87% specificity and 70% geometric mean is achieved. Finally, for the case of BD vs SCZ, it is reported a lower classification result i.e., 54% accuracy, 62% sensitivity, 45% specificity and 47% geometric mean. Using the Logistic regression classification, 78% accuracy, 78% sensitivity, 78% specificity and 76% geometric mean is achieved for the case of HC vs BD. For the case of HC vs SCZ, 77% accuracy, 70% sensitivity, 84% specificity and 75% geometric mean is achieved. Finally, for the case of

Table 5.2: AUC of the 9 cognitive features (representing a subject) during three different classification tasks.

	HC vs. BD	HC vs. SCZ	BD vs. SCZ
TIB score	0.70	0.81	0.56
Raven score	0.88	0.88	0.53
WCST error score	0.72	0.67	0.51
Nback_1 memory score	0.65	0.73	0.59
Nback_2 memory score	0.72	0.69	0.52
Nback_3 memory score	0.78	0.81	0.51
Motivational parameter	0.56	0.60	0.63
Learning parameter	0.61	0.56	0.66
Response consistency parameter	0.53	0.57	0.54

Table 5.3: Classification accuracy using the combined 9 psychiatric interview features (i.e., cognitive features) using Logistic regression and distance metric learning and semi-supervised learning algorithm (GT).

	Method	Acc	Se	Sp	Gmean
HC vs BD	DML+GT	0.80 $\pm$ 0.12	0.80 $\pm$ 0.16	0.80 $\pm$ 0.19	0.79 $\pm$ 0.13
	LR	0.78 $\pm$ 0.12	0.78 $\pm$ 0.18	0.78 $\pm$ 0.19	0.76 $\pm$ 0.13
HC vs SCZ	DML+GT	0.74 $\pm$ 0.13	0.61 $\pm$ 0.25	0.87 $\pm$ 0.15	0.70 $\pm$ 0.18
	LR	0.77 $\pm$ 0.13	0.70 $\pm$ 0.22	0.84 $\pm$ 0.17	0.75 $\pm$ 0.15
BD vs SCZ	DML+GT	0.54 $\pm$ 0.14	0.62 $\pm$ 0.23	0.45 $\pm$ 0.24	0.47 $\pm$ 0.20
	LR	0.53 $\pm$ 0.16	0.55 $\pm$ 0.22	0.51 $\pm$ 0.24	0.49 $\pm$ 0.20

BD vs SCZ, it is reported a lower classification result as in the previous case i.e., 53% accuracy, 55% sensitivity, 51% specificity and 49% geometric mean.

Table 5.4: Classification performance evaluation result using each interviews separately in a univariate approach for differentiating HC vs BD.

	Method	Acc	Se	Sp	Gmean
TIB score	DML+GT	0.68 (0.14)	0.66 (0.25)	0.70 (0.20)	0.66 (0.16)
	LR	0.65 (0.13)	0.56 (0.23)	0.74 (0.18)	0.62 (0.18)
Raven Score	DML+GT	0.79 (0.11)	0.78 (0.18)	0.81 (0.18)	0.78 (0.12)
	LR	0.80 (0.11)	0.80 (0.18)	0.79 (0.18)	0.78 (0.12)
WCST	DML+GT	0.76 (0.11)	0.89 (0.15)	0.62 (0.19)	0.73 (0.13)
	LR	0.70 (0.13)	0.77 (0.18)	0.63 (0.20)	0.677 (0.15)
Nback-1	DML+GT	0.64 (0.13)	0.42 (0.22)	0.86 (0.15)	0.57 (0.20)
	LR	0.64 (0.13)	0.42 (0.22)	0.86 (0.15)	0.57 (0.20)
Nback-2	DML+GT	0.70 (0.16)	0.65 (0.21)	0.74 (0.18)	0.69 (0.16)
	LR	0.70 (0.16)	0.65 (0.21)	0.74 (0.18)	0.69 (0.16)
Nback-3	DML+GT	0.67 (0.13)	0.73 (0.21)	0.61 (0.22)	0.65 (0.14)
	LR	0.73 (0.14)	0.714 (0.19)	0.74 (0.18)	0.716 (0.14)
MP	DML+GT	0.51 (0.13)	0.63 (0.22)	0.38 (0.21)	0.44 (0.19)
	LR	0.44 (0.09)	0.82 (0.25)	0.07 (0.11)	0.10 (0.15)
LP	DML+GT	0.54 (0.12)	0.67 (0.27)	0.41 (0.22)	0.45 (0.19)
	LR	0.50 (0.14)	0.60 (0.20)	0.41 (0.20)	0.46 (0.17)
RCP	DML+GT	0.372 (0.12)	0.54 (0.25)	0.20 (0.19)	0.23 (0.19)
	LR	0.47 (0.07)	0.87 (0.16)	0.06 (0.11)	0.11 (0.19)

Table 5.5: Classification performance evaluation result using each interviews separately in a univariate approach for differentiating HC vs SCZ.

	Method	Acc	Se	Sp	Gmean
TIB score	DML+GT	0.78 (0.13)	0.73 (0.183)	0.84 (0.17)	0.77 (0.15)
	LR	0.70 (0.12)	0.478 (0.21)	0.93 (0.12)	0.64 (0.18)
Raven Score	DML+GT	0.78 (0.12)	0.75 (0.18)	0.82 (0.18)	0.77 (0.13)
	LR	0.80 (0.12)	0.72 (0.19)	0.88 (0.15)	0.78 (0.13)
WCST	DML+GT	0.71 (0.13)	0.62 (0.18)	0.80 (0.18)	0.69 (0.14)
	LR	0.62 (0.14)	0.50 (0.21)	0.75 (0.19)	0.59 (0.15)
Nback-1	DML+GT	0.73 (0.136)	0.60 (0.20)	0.85 (0.14)	0.71 (0.15)
	LR	0.73 (0.14)	0.61 (0.20)	0.85 (0.14)	0.70 (0.14)
Nback-2	DML+GT	0.63 (0.13)	0.50 (0.19)	0.77 (0.17)	0.60 (0.14)
	LR	0.68 (0.13)	0.49 (0.19)	0.85 (0.14)	0.64 (0.14)
Nback-3	DML+GT	0.76 (0.13)	0.79 (0.20)	0.73 (0.18)	0.74 (0.15)
	LR	0.73 (0.12)	0.69 (0.24)	0.77 (0.20)	0.70 (0.15)
MP	DML+GT	0.47 (0.11)	0.41 (0.23)	0.53 (0.25)	0.40 (0.17)
	LR	0.50 (0.09)	0.28 (0.19)	0.72 (0.22)	0.37 (0.21)
LP	DML+GT	0.52 (0.14)	0.57 (0.26)	0.47 (0.24)	0.45 (0.21)
	LR	0.49 (0.08)	0.15 (0.22)	0.83 (0.25)	0.16 (0.23)
RCP	DML+GT	0.57 (0.14)	0.47 (0.22)	0.67 (0.22)	0.52 (0.18)
	LR	0.53 (0.08)	0.18 (0.16)	0.88 (0.15)	0.30 (0.23)

Table 5.6: Classification performance evaluation result using each interviews separately in a univariate approach for differentiating BD vs SCZ.

	Method	Acc	Se	Sp	Gmean
TIB score	DML+GT	0.54 (0.15)	0.57 (0.28)	0.51 (0.26)	0.47 (0.21)
	LR	0.41 (0.14)	0.41 (0.23)	0.42 (0.25)	0.34 (0.12)
Raven Score	DML+GT	0.43 (0.13)	0.44 (0.25)	0.42 (0.23)	0.35 (0.19)
	LR	0.43 (0.13)	0.42 (0.20)	0.44 (0.22)	0.39 (0.18)
WCST	DML+GT	0.53 (0.13)	0.40 (0.22)	0.65 (0.24)	0.45 (0.20)
	LR	0.43 (0.14)	0.41 (0.25)	0.46 (0.22)	0.37 (0.18)
Nback-1	DML+GT	0.50 (0.11)	0.23 (0.22)	0.77 (0.20)	0.32 (0.24)
	LR	0.55 (0.13)	0.39 (0.28)	0.72 (0.24)	0.43 (0.24)
Nback-2	DML+GT	0.43 (0.15)	0.50 (0.31)	0.36 (0.23)	0.33 (0.20)
	LR	0.40 (0.16)	0.39 (0.22)	0.41 (0.23)	0.35 (0.20)
Nback-3	DML+GT	0.47 (0.14)	0.52 (0.27)	0.42 (0.23)	0.39 (0.21)
	LR	0.45 (0.13)	0.47 (0.22)	0.42 (0.21)	0.39 (0.18)
MP	DML+GT	0.58 (0.13)	0.73 (0.27)	0.43 (0.24)	0.49 (0.20)
	LR	0.57 (0.16)	0.53 (0.24)	0.60 (0.21)	0.54 (0.18)
LP	DML+GT	0.65 (0.13)	0.85 (0.18)	0.44 (0.22)	0.58 (0.20)
	LR	0.61 (0.15)	0.61 (0.21)	0.60 (0.20)	0.59 (0.16)
RCP	DML+GT	0.54 (0.15)	0.52 (0.27)	0.56 (0.24)	0.48 (0.21)
	LR	0.49 (0.17)	0.45 (0.22)	0.54 (0.24)	0.46 (0.20)

Table 5.7: Classification of the 3 interviews for BD vs SCZ

	Method	Acc	Se	Sp	Gmean
MP+LP+RCP	DML+GT	0.584 (0.153)	0.70 (0.27)	0.46 (0.25)	0.51 (0.21)
	LR	0.56 (0.14)	0.66 (0.25)	0.47 (0.23)	0.51 (0.19)
MP+LP	DML+GT	0.59 (0.14)	0.71 (0.26)	0.48 (0.27)	0.51 (0.23)
	LR	0.57 (0.14)	0.58 (0.20)	0.56 (0.22)	0.55 (0.15)
MP+RCP	DML+GT	0.55 (0.13)	0.58 (0.29)	0.52 (0.29)	0.45 (0.21)
	LR	0.496 (0.131)	0.44 (0.21)	0.55 (0.213)	0.45 (0.16)
LP+RCP	DML+GT	0.63 (0.12)	0.75 (0.20)	0.53 (0.25)	0.59 (0.17)
	LR	0.61 (0.14)	0.66 (0.25)	0.56 (0.24)	0.56 (0.18)

Table 5.8: Multimodal differential classification result of BD vs SCZ using semi-supervised method.

	SSL			
	ACC	SE	SP	G-mean
Learning parameter (LP)	0.60 (0.12)	0.76 (0.23)	0.44 (0.20)	0.53 (0.18)
Motivation parameter (MP)	0.60 (0.11)	0.79 (0.18)	0.41 (0.22)	0.52 (0.19)
Response consistency parameter (RCP)	0.58 (0.15)	0.57 (0.23)	0.58 (0.22)	0.55 (0.16)
Left hemisphere Thalamus volume (LT)	0.50 (0.13)	0.55 (0.25)	0.45 (0.25)	0.43 (0.19)
Right hemisphere Thalamus volume (RT)	0.59 (0.14)	0.57 (0.26)	0.61 (0.23)	0.54 (0.20)
Right hemisphere Insula cortical thickness (RI)	0.61 (0.13)	0.59 (0.22)	0.63 (0.25)	0.56 (0.18)
Left Insula cortical thickness (LI)	0.61 (0.14)	0.45 (0.26)	0.76 (0.24)	0.52 (0.22)
LP+MP+RCP	0.59 (0.15)	0.69 (0.27)	0.49 (0.24)	0.52 (0.22)
LT+RT+RI+LI	0.60 (0.14)	0.50 (0.20)	0.70 (0.26)	0.55 (0.18)
LT+RT	0.55 (0.14)	0.51 (0.25)	0.59 (0.25)	0.50 (0.19)
RI+LI	0.62 (0.12)	0.45 (0.23)	0.80 (0.20)	0.55 (0.18)
LP+MP+RCP+LT+RT	0.57 (0.15)	0.69 (0.25)	0.45 (0.27)	0.48 (0.24)
LP+MP+RCP+RI+LI	0.59 (0.14)	0.66 (0.26)	0.51 (0.25)	0.53 (0.17)
all	0.59 (0.14)	0.63 (0.28)	0.55 (0.24)	0.53 (0.19)

Table 5.9: Multimodal differential classification result of BD vs SCZ using logistic regression.

	LR			
	ACC	SE	SP	G-mean
Learning parameter (LP)	0.57 (0.15)	0.56 (0.24)	0.59 (0.20)	0.54 (0.18)
Motivation parameter (MP)	0.62 (0.13)	0.59 (0.20)	0.64 (0.18)	0.60 (0.13)
Response consistency parameter (RCP)	0.52 (0.16)	0.51 (0.24)	0.52 (0.23)	0.47 (0.20)
Left hemisphere Thalamus volume (LT)	0.56 (0.16)	0.57 (0.21)	0.55 (0.24)	0.53 (0.1812)
Right hemisphere Thalamus volume (RT)	0.53 (0.14)	0.58 (0.22)	0.48 (0.20)	0.49 (0.17)
Right hemisphere Insula cortical thickness (RI)	0.55 (0.15)	0.61 (0.21)	0.48 (0.22)	0.51 (0.17)
Left Insula cortical thickness (LI)	0.65 (0.13)	0.66 (0.21)	0.63 (0.23)	0.61 (0.17)
LP+MP+RCP	0.57 (0.15)	0.64 (0.22)	0.51 (0.22)	0.54 (0.17)
LT+RT+RI+LI	0.56 (0.14)	0.54 (0.19)	0.57 (0.22)	0.53 (0.16)
LT+RT	0.53(0.15)	0.55 (0.21)	0.52 (0.22)	0.50 (0.18)
RI+LI	0.59 (0.14)	0.60 (0.21)	0.58 (0.23)	0.56 (0.16)
LP+MP+RCP+LT+RT	0.55 (0.15)	0.59 (0.22)	0.51 (0.24)	0.50 (0.19)
LP+MP+RCP+RI+LI	0.58 (0.14)	0.61 (0.23)	0.55 (0.22)	0.55 (0.17)
all	0.59 (0.15)	0.59 (0.23)	0.59 (0.23)	0.56 (0.17)

6

DEEP LEARNING BASED CLASSIFICATION OF NEUROIMAGES

> You never change things by fighting the existing reality. To change something, build a new model that makes the existing model obsolete.
>
> Buckminster Fuller

While the utility of machine learning (ML) in this decade has been enormously increasing in the medical imaging community, an emerging sub-field of ML called deep learning (DL) is resulting remarkable results in tasks such as computer-aided classifications (CAC).

As we keep on checking computer-assisted analysis of images in the field of medical imaging, recent advances in machine learning, especially with regard to deep learning, are helping to identify, classify, and quantify patterns in medical images. It is expected that ML with image input will be the mainstream area in the field of medical imaging in the next few decades [103]. The field of computer assisted neuropsychiatry is also beneficiary of this progress.

DL methods learn vital representations of input raw data through consecutive non-linear transformations unlike traditional ML methods based on learning from hand crafted or engineered features. In DL framework, the representations in the form of hierarchical features are learned from the raw input data. At the core of these advances is the ability to exploit hierarchical, latent and invariant feature representations learned solely from data using mathematical convolution operations, over neural network architectures, an idea inspired from the working of information processing in the visual cortex of primates [104].

This characteristics of DL frameworks are vital in understanding discriminative brain features that are helpful in the comprehension of psychiatric disorders and their effect on the brain. Usually, the use of ML algorithms in neuroimaging based psychiatric

studies, in the context of anatomical investigation of psychiatric disorders, is based on analyzing human insight based features such as quantified gray matter characteristics (i.e., cortical thickness, volume, gyrification index etc.), quantified white matter characteristics and related quantifyable metrics. The motivation of this chapter is therefore, exploring deep learning based classification using the brain image's intensity in the context of computer aided neuropsychiatric disorder diagnosis.

6.1 RELATED WORKS

Deep learning models can learn a hierarchy of features, in which high-level features are built upon low-level image features layer-by-layer. CNN [105, 106] is a useful deep learning tool, when trained with appropriate regularizations, CNN has been shown with superior performance on both visual object recognition and image classification tasks (e.g., [106]).

In [107], authors proposed a frame work by combining sparse regression and deep learning concepts for Alzheimer's disease/mild cognitive impairment diagnosis and prognosis. First, they represented the subjects using gray matter tissue volumes of 93 ROIs using Kabani atlas [108] i.e., 93-dimensional features from an MR image. Then, they trained ten sparse regression models with 10 different values of regularization parameters (equally spaced between 0.01 and 0.3), as a result selecting different feature subsets from the original feature set thereby having different powers to predict the response values (i.e., clinical label: healthy or patient and clinical scores: MMSE (mini mental state examination) and ADAS-Cog (Alzheimer's Disease Assessment Scale - Cognition)). Finally, by considering the response values from the sparse regression models as target-level representations (i.e., 10x4 feature matrix), they built a deep convolutional neural network for final classification (two convolution layers followed by one max-pooling layer, two fully connected layers, and one output layer). In this study a total of 805 subjects are used (i.e., 186 AD, 226 HC and 393 MCI), from the Alzheimer's Disease National Initiative (ADNI) cohort dataset which is publicly available.

In [109], authors described how to characterize differences in brain morphometry in schizophrenia on 258 subjects in total (i.e., 143 SCZ, 83 HC and 32 FEP), with the main objective of investigating the most abstract features that are invariant across patients with schizophrenia, learned using a multivariate analysis of the latent features of the deep belief network (DBN). First, utilizing FreeSurfer, they represented the subjects using 68 ROI cortical thickness and 45 anatomical structure volumes using the Desikan-Killiany atlas [95]. Then, they trained a deep neural network, pre-trained by a deep belief network, i.e., by adding a last softmax layer to get the final class label predictions. The training of a DBN is performed in two steps: DBN pre-training and supervised fine-tuning. The smaller networks converged and were then used as initializations for the larger, deeper networks - this process is called pre-training.

In a recent paper on ADNI data classification [110], the authors proposed to use a 3D convolutional neural network for feature extraction from MRIs. To be more specific, the authors used the Deeply Supervised Adaptive 3D-CNN (DSA-3D-CNN) which was initialized by training convolutional autoencoders for feature extraction and fine tun-

ing the network for classification on different domain images. They showed impressive performance compared to the other approaches, with ROC AUC over 0.96 on binary classification setting. They tested their methods on 30 subjects from CADementia as a source domain and 210 subjects from ADNI as target domain. Similarly, In [111], authors pretrained their T1 MRI data using sparse auto encoder and used the learned filters as an initializer for their 2D and 3D CNN. They reported a slightly better improvement using 3D CNN in classifying Healthy controls, Mild cognitive impairment and Alzheimer. They used 2265 scans which is almost more than double when compared to our data. In [112], authors implemented the 3D version of VGG and Resnet neural network architectures on 3D structural MRI brain scans and demonstrated the performance of the proposed approach for classification of Alzheimer's disease versus mild cognitive impairment and normal controls on the ADNI dataset. In [113], authors presented a classification scheme or automatic diagnosis for Attention Deficit Hyperactivity Disorder (ADHD) using 3D CNN and Functional and Structural MRI. In [114], the authors proposed a 3D multi-view convolutional neural networks for lung nodule classification.

Regarding the usage of deep convolutional networks in medical domain, [115] discussed CNN architectures, dataset characteristics and transfer learning. [116] discussed about whether end to end training from scratch or fine tuning is better in Medical domain, considering the limited amount of available clean data, and they highlighted that fine tuning has its promises. In [117], the editors presented an overview and future promise of DL in medical imaging. We also noticed that in the literature of medical image analysis domain, DL has been applied mainly for segmentation tasks.

6.2 CLASSIFICATION OF FIRST EPISODE PSYCHOSIS (FEP) USING 3D CONVOLUTIONAL NEURAL NETWORK

The alteration in physiology and functional activity of neurons due to FEP, as generally in neuropsychiatry disorders, is manifested in the anatomy of the brain. These structural changes (or neuroanatomical variations) has been observed and validated using structural magnetic resonance imagining (MRI) studies [20]. In this regard, the analysis of neuroimaging data using machine learning techniques is offering promising results, both biologically and methodologically sound. Classification algorithms that can learn an intrinsic input data patterns from MRI of clinical groups (i.e., healthy controls and patients) are playing an important role in the diagnosis process of a disease and individualized prediction. Recently, deep learning techniques are achieving state of the art results in classification tasks in several applicative domains. The algorithms, being trained with examples of MRI images from healthy subjects and patients, are able to classify a new unseen subjects' data [104].

Motivated from the points mentioned above, currently we are performing a research with the aim of finding/validating digital neuroanatomical biomarker (i.e., features extracted from MRI) for diagnosis and individualized prediction of FEP, on a novel big dataset of about 855 multi-center human brain (structural MRI) scans of subjects. We use machine learning techniques, specifically convolutional neural network (CNN)

and related approaches to learn the hierarchical representations (features) of the brain structure (it might be from a particular region of interest, ROI) and capture or discover the anatomical variation between healthy controls class and FEP class.

Instead of extracting features from each slices of MRI (2D approaches), we aim to extract 3D features from the input volumetric 3D MRI data by employing 3D-CNN, 3D-autoencoders and other tools necessary to extract more reliable anatomical features that represent the spatial structure of the brain more accurately. We believe these extracted features should have a discriminative capability between the classes (i.e., healthy controls class vs FEP class), when used in the right context. Even though this approach is computationally expensive, it is easily interpretable for the medical community unlike other studies with 2D approaches.

To study this and obtain a meaningful valid result, having quality data is a must, since CNN algorithms need large amounts of quality training example data in order to learn intrinsic patterns. In this regard, to the best of our knowledge, this will be the first study with a large population of novel FEP data i.e., 855 scans in a class balanced manner, which are acquired as a result of the collaboration with 7 different medical centers.

The contributions of this part of our work is: 1. The creation of a multi-center based neuroimaging pipeline for diagnosis and individualized prediction of FEP that takes into consideration age, gender and scan-center biases in the input data. 2. The contribution of knowledge to the neuroimaging field of FEP studies with a holistic approach (i.e., using clinical, cognitive, immunological, neuroimaging and genetic data). 3. The detection of a set of digital neuroanatomical-biomarkers for diagnosis and individualized prediction of FEP (the same approach could then be adapted for diagnosis process of other neuropsychiatry disorders). 4. The validation of results of medical studies, where region of interests (ROI) of the brain are identified as regions affected by FEP (e.g., the frontal lobe is found to be highly affected). This part of our work will be validated by our medical team (medical doctor, neuropsychiatry expert and biomedical engineer). We believe our results will advance the understanding of the biological underpinning of the pathology by exploring a whole brain approach and a ROI based approach using already existing anatomical atlases.

Figure 6.1 illustrates the convolution operation that is mathematically expressed using the following formula: $op_{1,1,1} = \sum_{x=1}^{3} \sum_{y=1}^{3} \sum_{z=1}^{3} w_{x,y,z}.in_{x,y,z} + b$ where b is the bias value to be added to the out put of the activation fired by a neuron. The activation would come after the convolution operation.

6.3 EXPERIMENTS AND RESULTS

In this section the experiments is carried on with the aim of identifying neuroanatomical biomarker for diagnosis and individualized-prediction of first episode psychosis using sMRI data and DL methods.

The CNN is used for extracting representative features (diagnosis/prognosis metrics) that can distinguish healthy brain from FEP affected brain. The stochastic back-

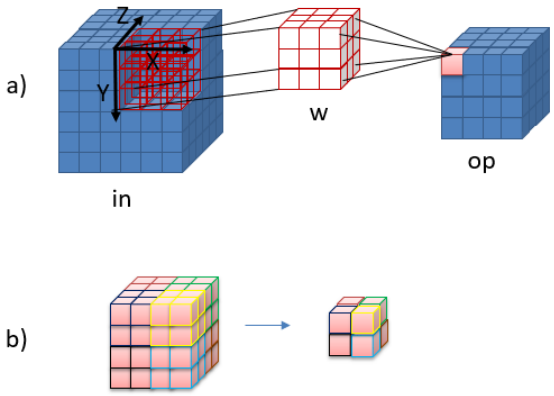


Figure 6.1: Illustration of 3D convolution operation (a) and 3D maximum pooling operation (b).

propagation algorithm is used for training. A 3D version of CNN is exploited, following a natural approach, to learn from 3D spatial data (i.e., brain MRI).

6.3.1 DATASET

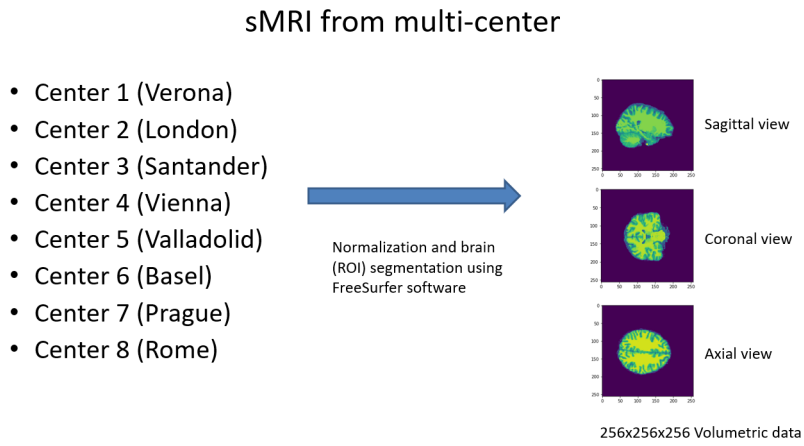


Figure 6.2: First episode psychosis (FEP) dataset from eight centers.

174 scans (93 hc, 81 fep) from Verona center, 218 scans (91 hc, 127 fep) from London center, 50 scans (25 hc, 25 fep) from Santander center, 134 scans (101 hc, 33 fep) from Vienna center, 43 scans (27 hc, 16 fep) from Valladolid center, 123 scans (44 hc, 79 fep) from Basel center, 63 scans (30 hc, 33 fep) from Prague center, 50 scans (25 hc, 25 fep) from Rome center.

In total 855 scans were collected which is a high numerosity when compared to state of the art FEP studies in the literature.

6.3.2 EXPERIMENTAL SETUP

Having a total of 855 structural scans, the training and testing procedure is validated using a 10 fold cross validation resulting, in each iteration, 770 training and 85 test samples. Furthermore, 20% of the training sample is used as a validation set. Performance of the classification is evaluated by accuracy (ACC), sensitivity (SE) and specificity (SP) metrics calculated as follows:

$$\text{Accuracy} = \frac{\text{true positive} + \text{true negative}}{\text{true positive} + \text{true negative} + \text{false positive} + \text{false negative}}$$

$$\text{Sensitivity} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}$$

$$\text{Specificity} = \frac{\text{true negative}}{\text{true negative} + \text{false positive}}$$

where 'true positive' indicates the number of correctly classified MRI scans from FEP class, 'true negative' indicates the number of correctly classified MRI scans from HC class, 'false positive' indicates the number of misclassified scans from HC class and 'false negative' indicates the number of misclassified scans from FEP class.

6.3.3 CAC USING SIMPLE 3D CNN ARCHITECTURE

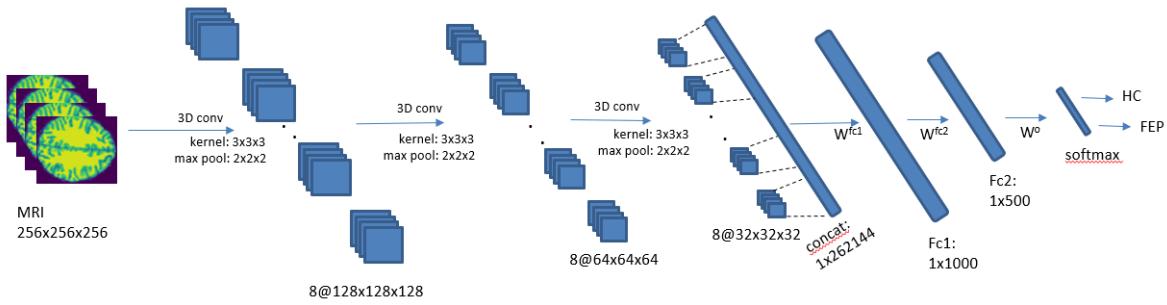


Figure 6.3: 3D CNN for individualized predictive classification of FEP vs. HC.

The first tested simple 3D CNN architecture is composed of three conventional layers with max pooling operations and two fully connected upper layers with softmax output.

In the first convolution layer, eight 3D filters having $3 \times 3 \times 3$ dimension (initialized with zero) scan and convolve the input volumetric MRI (with dimension of $256 \times 256 \times 256$), followed by a max pooling operation, outputting eight convolved cubes

(with dimension of $128 \times 128 \times 128$). The max pooling operation squeezes $2 \times 2 \times 2$ voxels into $1 \times 1 \times 1$ voxel by selecting the maximum value from the target cube. In the second layer and third layer of the convolution, same operation is performed. At this step, the eight convoluted cubes (each having size of $32 \times 32 \times 32$) are stretched to form a vector of 262144 dimension that are going to be fed to the fully connected upper layers. The first and second layers of the fully connected upper layer consist of 1000 and 500 neurons respectively. Finally the last layer is composed of 2 neurons because we are dealing with binary classification and the diagnostic category of the input MRI is determined by a softmax operation. The architecture is depicted in figure 6.3.

Regarding the training procedure, adadelta optimization is used for minimizing binary cross entropy loss.

6.3.4 CAC USING PRE-TRAINED 3D CNN ARCHITECTURE

This second architecture looks quite similar to the simple architecture proposed above, except that the first three convolution layer filters are not initialized with zero, instead they were substituted with the filters learned by an independent and unsupervisedly trained 3D convolutional auto encoder (CAE). The training part with 3D CAE is served as a pre-training for the final classifier.

The 3D CAE was composed of encoding and decoding layers as depicted in figure 6.4 used to reconstruct the volumetric structural MR image. Then, the filters, learned in the encoding part, are used to initialize the first 3 layers of the final classifier. In this method, with the aim of extracting features in unsupervised manner, 3D stacked convolutional auto encoder is used. Then, the decoded information is used as features. This learned (or extracted) features are fed to the fully connected upper layers of the CNN. The reconstruction is done by minimizing the Root mean square (RMS) metric.

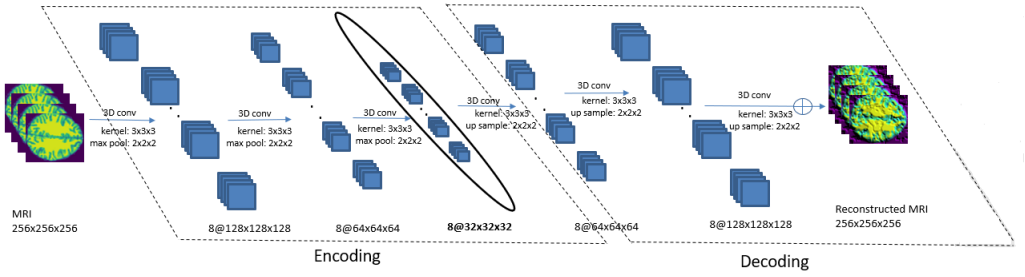


Figure 6.4: Reconstruction of volumetric structural brain image data using 3D auto encoder.

6.3.5 CAC USING 3D RESIDUAL NETWORK (RESNET) ARCHITECTURE

The third neural network architecture tested is an 18 layers 3D resnet with 33,171,842 trainable parameters as depicted in 6.7.

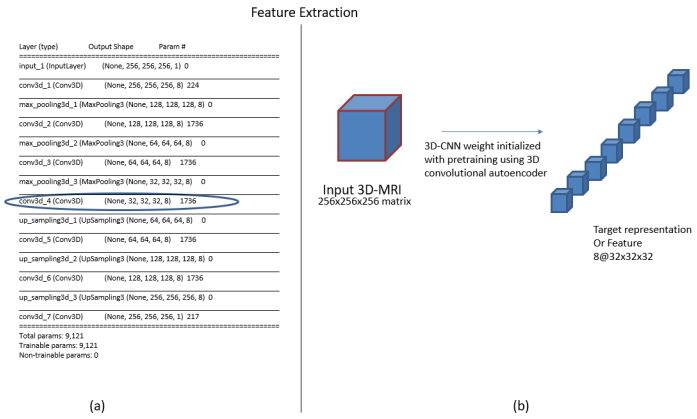


Figure 6.5: Extraction of features in unsupervised manner using 3D stacked convolutional auto encoder. a) The Keras summary of the neural network architecture indicating 9121 trainable parameters are used. b) Illustration of the the transformation of the input volumetric MRI of the brain.

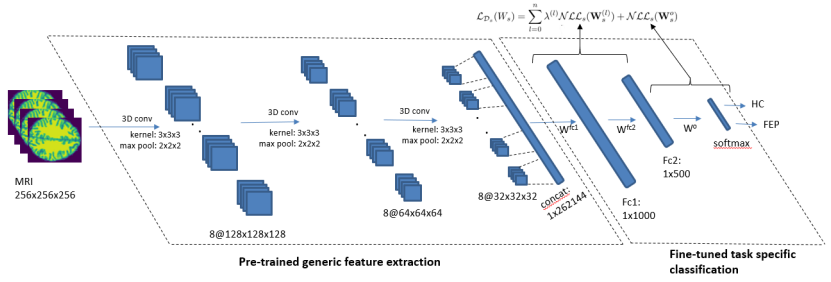


Figure 6.6: The final pretrained 3D CNN architecture for individualized predictive classification of FEP vs. HC.

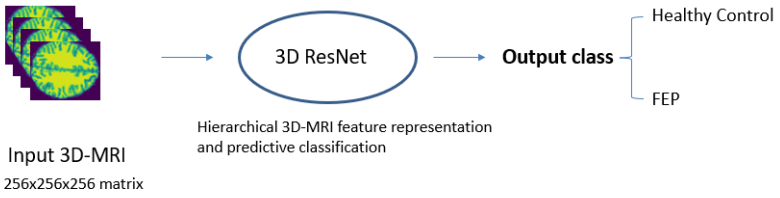


Figure 6.7: 3D CNN (ResNet) for individualized predictive classification of FEP. Proposed architecture 3. Hierarchical 3D-MRI feature representation using skip connections for achieving better learning representations.

6.3.6 REGION OF INTEREST (ROI) BASED ANALYSIS

In here, focusing the aim more on finding which brain region is affected by FEP, parcellated brain regions are considered for classification instead of using the whole brain image regions. The analysis scheme is illustrated in fig 6.8. The parcellated brain re-

gions are represented using 3D volumetric image, 2D image by projecting all the slices into a 2D plane side by side and 1D feature vector by slice wise stretching.

The experimented is carried on using only the data from Verona center. The results are compared with SVM. In addition to the 3D resnet, a 2D version with transfer learning is also tested. We trained the 2D resnet pretrained with imagenet natural images [106]. It is done in such a way so as to minimize the degree of freedom on the learnable kernels (or filter weights).

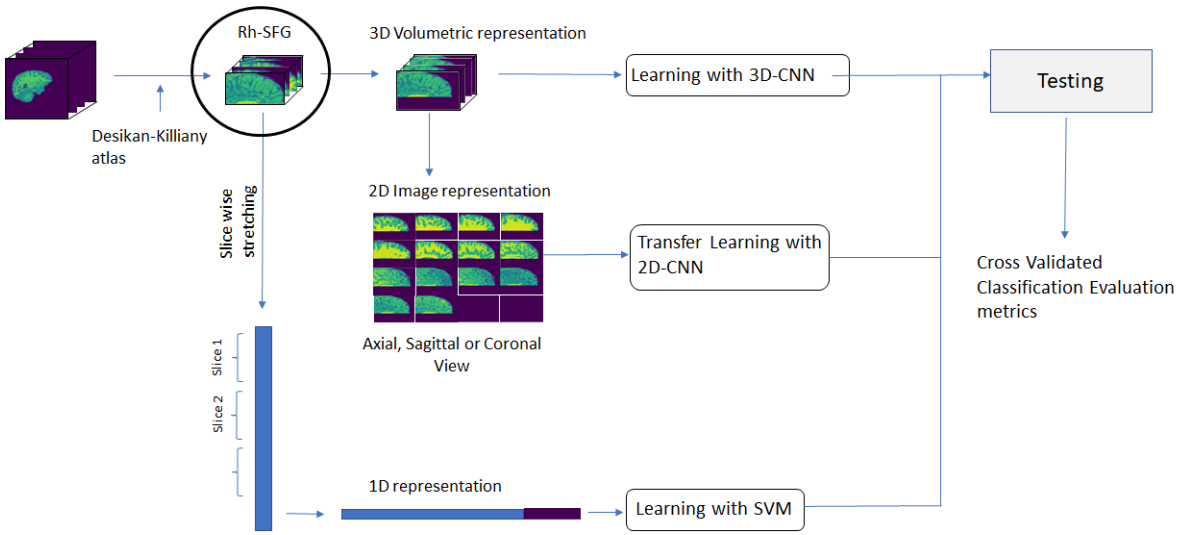


Figure 6.8: Experimental set up for ROI based neuroimage classification (i.e., HC vs FEP). The figure shows the analysis of right hemisphere superior frontal gyrus (Rh_SFG).

6.3.7 RESULTS AND DISCUSSIONS

When the whole brain images is considered the classification results of the first three NN architectures are shown in table 6.1.

	ACC	SE	SP
Simple 3D CNN	47.7 (chance level)	chance level	chance level
Pre-traind 3D CNN	chance level	chance level	chance level
3D Resnet	56.9 %	56.9 %	56.9 %

Table 6.1: Classification result of the three tested methods on the whole brain image as input.

From the first three experiments, it seems accurate classification of first episode psychosis using 3 tesla skull striped whole volumetric structural image is difficult, when differentiating healthy groups and FEP groups. The highest performance (i.e., 56% accuracy) is achieved when 3D resnet used. Both the simple and pre-trained architecture resulted in 47% accuracy. But it is still needed to do the experiment using different learning settings and to test with bigger dataset to confirm this conclusion.

Regarding the unsupervised feature extraction using 3D-CAE, as shown in figure 6.5 the input skull stripped brain image will be transformed into eight cubes, each with a

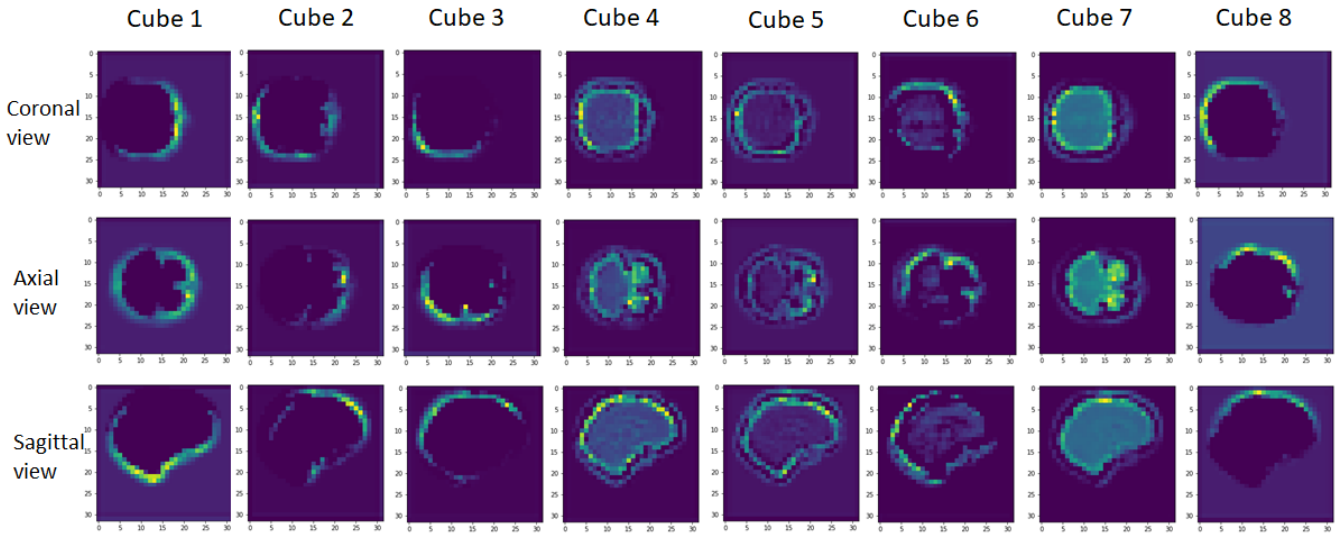


Figure 6.9: Qualitative result: Brain features extracted in unsupervised manner from a FEP patient. The images shown are the 16th slice of each of the 8, $32 \times 32 \times 32$ dimensional cubes (i.e., the transformed brain representations).

dimension of $32 \times 32 \times 32$ In figure 6.9, it is shown the automatically extracted feature in unsupervised fashion using 3D CAE that is used for pre-training of latent features that are useful in differentiating healthy class from FEP class. In the figure it is shown the sagital, coronal and axial view of the 16th slice. Most of the layers of the network learned cortical areas of the gray matter. It is coherent with the literature as it has been shown there is cortical thickness alteration on FEP patients [20].

The result of the ROI based classification is presented in table 6.2.

Brain regions	SVM (Acc)	2D res-net with transfer learning (Acc/Se/Sp/Ck)	3D res-net (Acc/Se/Sp/Ck)
rh_Amygdala	0.47	-	0.48 ± 0.06/ 0.75 ± 0.43/ 0.25 ± 0.43/ 0.0 ± 0.0
rh_Superior frontal gyrus	0.52	-	0.51 ± 0.07/ 0.75 ± 0.43/ 0.25 ± 0.43/ 0.0 ± 0.0
lh_Pars orbitalis	0.59	0.50 ± 0.08/ 0.65 ± 0.48/ 0.36 ± 0.47/ 0.01 ± 0.05	-

Table 6.2: Classification results using parcelatted brain regions. rh = right hemisphere, lh = left hemisphere.

7

CONCLUSION AND FUTURE WORKS

> ... said I took a break and I loved the break,
> now I'm off the brake, looking to make a break
> at a greater stake and the "flow" sweet like
> grater cake ...
>
> Protoje

In this thesis, a machine learning (ML) based perspective and analysis of structural neuroimages in the context of neuropsychiatry is presented. Furthermore, novel brain structure feature that measures cortical thickness distribution called cortical skewness is proposed. ML based computer aided classification pipeline for diagnosis of neuropsychiatric disorders is presented. The experiments are aimed on the analysis and understanding of first episode psychosis, bipolar disorder and schizophrenia.

Overall the results indicate that machine learning techniques have the potential to inform clinical practice and research, as they can make predictions about brain scan data from individual subjects. Furthermore, ML methods may reflect an objective biomarker of psychiatric disorder, based on abnormalities of brain structure. The key benefits of utilizing these methods including deep learning as compared to traditional, only statistics based analytical techniques, is that, from data modeling perspective, traditional statistical analysis, most of the times, deals with more assumptions of whole population data distribution and group analysis instead of individual subject analysis. When we see it from the perspective of the medical problem assessed in this thesis, the advantages of ML is that first, it handles individual prediction of the data which is a naturally desired characteristics when thinking of integrating these systems into daily use clinical practices. Where as traditional analytical techniques mostly deal with group analysis of subjects, which makes them limited in the practical daily uses of clinical settings. Secondly, ML methods naturally handles multiple variables in the analysis simultaneously, (i.e., clinicians can diagnose patients using multiple clinical scores and

biological data of subjects). In brief summary, besides the research purpose, when ML based computational methods are integrated in softwares that clinicians use for the diagnosis of neuropsychiatric disorders, they have more benefits and could provide more translation about the mental diseases using computational perspectives.

Using the newly proposed diagnostic metric 'cortical skewness', it is shown that the left hemisphere inferior parietal is one of the most altered brain region in bipolar disorder patients, when compared with healthy controls. Also according to the experiments on chapter 4, the following brain regions are associated with bipolar disorder: left and right hemisphere Banks of the Superior Temporal Sulcus, left and right hemisphere-Cuneus, left hemisphere-Inferior Parietal, right hemisphere-Lateral Occipital, left hemisphere-MiddleTemporal, left hemisphere-Pericalcarine, left hemisphere-Supramarginal, left hemisphere-Orbito Lateral, left hemisphere-Superior Frontal Gyrus, right hemisphere-Inferior Temporal, left hemisphere-Triangularis, left hemisphere-Frontal Pole, right hemisphere- Enthorinal, right hemisphere- Temporal Lobe, right hemisphere-Post Central, right hemisphere-Superior Parietal and left hemishpere-Opercularis.

As one of the main contributions of this thesis, a structural neuroimaging based classification pipeline for computer aided diagnosis of psychiatric disorders is proposed. The learning scheme follows a semi-supervised learning approach (that is less exploited in medical image analysis domain) to address a contextual similarity based classification of structural neuroimages. The main challenges in here are two; first, there is limited dataset in this domain and second the difference between neuroimages among different classes is very subtle. To address this issue we employed a graph transduction algorithm that learns both from labeled and unlabeled data and operates based on contextual similarity. Furthermore, with the aim of achieving better performance, the original framework of the graph transduction algorithm is extended by appending a pre-training procedure using the feature extracted from the brain images. The idea is that the graph based semi-supervised algorithm operates on data modeled as graph nodes and their similarity denoted with graph edges, so the framework heavily depends on the distance between the features, so we exploited a distance metric learning strategy to properly learn a distance metric on the data at hand. This pre-training resulted in the transformation of the graph causing displacement of the nodes, then the label propagation would happen on the new graph yielding better classification performance. The advantage of the semi-supervised graph based contextual similarity learning algorithm is that it has a deep mathematical foundation rooted with the notions of game theory that is developed with the aim of modeling the essentials of decision making in interactive situations. When we see it from the perspective of the problem in our hand, classification can be seen as a decision making process. And this pipeline is tested on the classification of schizophrenia using preselected brain region thickness. Further more it is used in the automatic biomarker identification of bipolar disorder and first episode psychosis. At last this pipeline is also used in the multi modal data classification (i.e, using structural neuroimages and psychiatric report scores) to address classification of SCZ, BD and FEP including the challenging differential diagnosis of BD vs. SCZ.

At last, it were exploited recent state of the art machine learning algorithms, based on convolutional neural networks. The motivation for that was the embedded feature extraction capability that avoid the engineering of human insight based brain features.

Finally as a future work, the proposed methods will be formed as a unified framework using ensemble techniques and further exploration of brain features, such as cortex-entropy, kurtosis etc. to be used as a diagnostic metrics and for characterization of psychiatric disorders.

[1] K. L. Possin, "Visual spatial cognition in neurodegenerative disease," *Neurocase*, vol. 16, no. 6, pp. 466–487, 2010. *(Cited on page 1)*

[2] P. Oswald, D. Souery, S. Kasper, Y. Lecrubier, S. Montgomery, S. Wyckaert, J. Zohar, and J. Mendlewicz, "Current issues in bipolar disorder: a critical review," *European Neuropsychopharmacology*, vol. 17, no. 11, pp. 687–695, 2007. *(Cited on page 2)*

[3] A. C. Altamura, B. Dell'Osso, H. A. Berlin, M. Buoli, R. Bassetti, and E. Mundo, "Duration of untreated illness and suicide in bipolar disorder: a naturalistic study," *European archives of psychiatry and clinical neuroscience*, vol. 260, no. 5, pp. 385–391, 2010. *(Cited on page 2)*

[4] A. C. Altamura, M. Buoli, A. Caldiroli, L. Caron, C. C. Melter, C. Dobrea, M. Cigliobianco, and F. Z. Quarantini, "Misdiagnosis, duration of untreated illness (dui) and outcome in bipolar patients with psychotic symptoms: a naturalistic study," *Journal of affective disorders*, vol. 182, pp. 70–75, 2015. *(Cited on page 2)*

[5] D. Bhugra, "The global prevalence of schizophrenia," *PLoS medicine*, vol. 2, no. 5, p. e151, 2005. *(Cited on pages 2 and 49)*

[6] R. W. Heinrichs and K. K. Zakzanis, "Neurocognitive deficit in schizophrenia: a quantitative review of the evidence." *Neuropsychology*, vol. 12, no. 3, p. 426, 1998. *(Cited on page 2)*

[7] K. J. Aitchison, R. M. Murray, P. J. Power, and E. M. Tsapakis, *First episode psychosis*. CRC Press, 1999. *(Cited on page 2)*

[8] P. R. Szeszko, K. L. Narr, O. R. Phillips, J. McCormack, S. Sevy, H. Gunduz-Bruce, J. M. Kane, R. M. Bilder, and D. G. Robinson, "Magnetic resonance imaging predictors of treatment response in first-episode schizophrenia," *Schizophrenia bulletin*, vol. 38, no. 3, pp. 569–578, 2010. *(Cited on page 2)*

[9] B. Olabi, I. Ellison-Wright, A. M. McIntosh, S. J. Wood, E. Bullmore, and S. M. Lawrie, "Are there progressive brain changes in schizophrenia? a meta-analysis of structural magnetic resonance imaging studies," *Biological psychiatry*, vol. 70, no. 1, pp. 88–96, 2011. *(Cited on page 2)*

[10] M. Bodnar, A. M. Achim, A. K. Malla, R. Joober, A. Benoit, and M. Lepage, "Functional magnetic resonance imaging correlates of memory encoding in relation to achieving remission in first-episode schizophrenia," *The British Journal of Psychiatry*, vol. 200, no. 4, pp. 300–307, 2012. *(Cited on page 2)*

[11] B. Mwangi, K. P. Ebmeier, K. Matthews, and J. Douglas Steele, "Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder," *Brain*, vol. 135, no. 5, pp. 1508–1521, 2012. *(Cited on page 2)*

[12] D. E. Linden, "The challenges and promise of neuroimaging in psychiatry," *Neuron*, vol. 73, no. 1, pp. 8–22, 2012. *(Cited on page 2)*

[13] G. Orru, W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 4, pp. 1140–1152, 2012. *(Cited on pages 2, 3, and 7)*

[14] B. Johnston, B. Mwangi, K. Matthews, D. Coghill, and J. Steele, "Predictive classification of individual magnetic resonance imaging scans from children and adolescents," *European child & adolescent psychiatry*, vol. 22, no. 12, pp. 733–744, 2013. *(Cited on pages 2 and 3)*

[15] B. Mwangi, T. S. Tian, and J. C. Soares, "A review of feature reduction techniques in neuroimaging," *Neuroinformatics*, vol. 12, no. 2, pp. 229–244, 2014. *(Cited on pages 2, 3, and 4)*

[16] B. Mwangi, K. Matthews, and J. D. Steele, "Prediction of illness severity in patients with major depression using structural mr brain scans," *Journal of Magnetic Resonance Imaging*, vol. 35, no. 1, pp. 64–71, 2012. *(Cited on page 2)*

[17] P. T. Reiss, "Cross-validation and hypothesis testing in neuroimaging: An irenic comment on the exchange between friston and lindquist et al." *NeuroImage*, vol. 116, pp. 248–254, 2015. *(Cited on page 2)*

[18] N. Agarwal, J. D. Port, M. Bazzocchi, and P. F. Renshaw, "Update on the use of mr for assessment and diagnosis of psychiatric diseases," *Radiology*, vol. 255, no. 1, pp. 23–41, 2010. *(Cited on page 3)*

[19] Z. Lao, D. Shen, Z. Xue, B. Karacali, S. M. Resnick, and C. Davatzikos, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," *Neuroimage*, vol. 21, no. 1, pp. 46–57, 2004. *(Cited on pages 3 and 7)*

[20] L. Squarcina, U. Castellani, M. Bellani, C. Perlini, A. Lasalvia, N. Dusi, C. Bonetto, D. Cristofalo, S. Tosato, G. Rambaldelli *et al.*, "Classification of first-episode psychosis in a large cohort of patients using support vector machine and multiple kernel learning techniques," *NeuroImage*, vol. 145, pp. 238–245, 2017. *(Cited on pages 3, 7, 63, and 70)*

[21] L. Squarcina, C. Perlini, D. Peruzzo, U. Castellani, V. Marinelli, M. Bellani, G. Rambaldelli, A. Lasalvia, S. Tosato, K. De Santi *et al.*, "The use of dynamic susceptibility contrast (dsc) mri to automatically classify patients with first episode

psychosis," *Schizophrenia research*, vol. 165, no. 1, pp. 38–44, 2015. *(Cited on pages 3, 14, and 36)*

[22] D. Peruzzo, U. Castellani, C. Perlini, M. Bellani, V. Marinelli, G. Rambaldelli, A. Lasalvia, S. Tosato, K. De Santi, V. Murino *et al.*, "Classification of first-episode psychosis: a multi-modal multi-feature approach integrating structural and diffusion imaging," *Journal of neural transmission*, vol. 122, no. 6, pp. 897–905, 2015. *(Cited on pages 3, 6, and 7)*

[23] S. Frangou, D. Dima, and J. Jogia, "Towards person-centered neuroimaging markers for resilience and vulnerability in bipolar disorder," *Neuroimage*, vol. 145, pp. 230–237, 2017. *(Cited on page 3)*

[24] E. Veronese, U. Castellani, D. Peruzzo, M. Bellani, and P. Brambilla, "Machine learning approaches: from theory to application in schizophrenia," *Computational and mathematical methods in medicine*, vol. 2013, 2013. *(Cited on pages 4, 5, 6, and 7)*

[25] R. C. Craddock, P. E. Holtzheimer III, X. P. Hu, and H. S. Mayberg, "Disease state prediction from resting state functional connectivity," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1619–1628, 2009. *(Cited on page 5)*

[26] V. A. Kovalev, M. Petrou, and J. Suckling, "Detection of structural differences between the brains of schizophrenic patients and controls," *Psychiatry Research: Neuroimaging*, vol. 124, no. 3, pp. 177–189, 2003. *(Cited on page 5)*

[27] C.-Y. Wee, P.-T. Yap, W. Li, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang, and D. Shen, "Enriched white matter connectivity networks for accurate identification of mci patients," *Neuroimage*, vol. 54, no. 3, pp. 1812–1822, 2011. *(Cited on page 5)*

[28] D. Gothelf, F. Hoeft, T. Ueno, L. Sugiura, A. D. Lee, P. Thompson, and A. L. Reiss, "Developmental changes in multivariate neuroanatomical patterns that predict risk for psychosis in 22q11. 2 deletion syndrome," *Journal of psychiatric research*, vol. 45, no. 3, pp. 322–331, 2011. *(Cited on page 5)*

[29] S. G. Costafreda, C. H. Fu, M. Picchioni, T. Toulopoulou, C. McDonald, E. Kravariti, M. Walshe, D. Prata, R. M. Murray, and P. K. McGuire, "Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder," *BMC psychiatry*, vol. 11, no. 1, p. 18, 2011. *(Cited on page 5)*

[30] V. Rocha-Rego, J. Jogia, A. Marquand, J. Mourao-Miranda, A. Simmons, and S. Frangou, "Examination of the predictive value of structural magnetic resonance scans in bipolar disorder: a pattern classification approach," *Psychological medicine*, vol. 44, no. 3, pp. 519–532, 2014. *(Cited on pages 6 and 7)*

[31] G. Fung, Y. Deng, Q. Zhao, Z. Li, M. Qu, K. Li, Y.-w. Zeng, Z. Jin, Y.-t. Ma, X. Yu *et al.*, "Distinguishing bipolar and major depressive disorders by brain structural

morphometry: a pilot study," *BMC psychiatry*, vol. 15, no. 1, p. 298, 2015. *(Cited on page 6)*

[32] H. Chen, C. DeLong, M. Bame, I. Rajapakse, T. Herron, M. McInnis, and K. O'shea, "Transcripts involved in calcium signaling and telencephalic neuronal fate are altered in induced pluripotent stem cells from bipolar disorder patients," *Translational psychiatry*, vol. 4, no. 3, p. e375, 2014. *(Cited on page 6)*

[33] H. G. Schnack, M. Nieuwenhuis, N. E. van Haren, L. Abramovic, T. W. Scheewe, R. M. Brouwer, H. E. H. Pol, and R. S. Kahn, "Can structural mri aid in clinical classification? a machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects," *Neuroimage*, vol. 84, pp. 299–306, 2014. *(Cited on pages 6 and 7)*

[34] M. H. Serpa, Y. Ou, M. S. Schaufelberger, J. Doshi, L. K. Ferreira, R. Machado-Vieira, P. R. Menezes, M. Scazufca, C. Davatzikos, G. F. Busatto *et al.*, "Neuroanatomical classification in a population-based sample of psychotic major depression and bipolar i disorder with 1 year of diagnostic stability," *BioMed research international*, vol. 2014, 2014. *(Cited on page 6)*

[35] J. Savitz, S. L. Rauch, and W. Drevets, "Clinical application of brain imaging for the diagnosis of mood disorders: the current state of play," *Molecular psychiatry*, vol. 18, no. 5, p. 528, 2013. *(Cited on page 6)*

[36] L. Squarcina, C. Perlini, M. Bellani, A. Lasalvia, M. Ruggeri, P. Brambilla, and U. Castellani, "Learning with heterogeneous data for longitudinal studies," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 535–542. *(Cited on page 6)*

[37] H. G. Schnack, N. E. Van Haren, M. Nieuwenhuis, H. E. Hulshoff Pol, W. Cahn, and R. S. Kahn, "Accelerated brain aging in schizophrenia: a longitudinal pattern recognition study," *American Journal of Psychiatry*, vol. 173, no. 6, pp. 607–616, 2016. *(Cited on page 6)*

[38] M.-J. Wu, B. Mwangi, I. E. Bauer, I. C. Passos, M. Sanches, G. B. Zunta-Soares, T. D. Meyer, K. M. Hasan, and J. C. Soares, "Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning," *Neuroimage*, vol. 145, pp. 254–264, 2017. *(Cited on page 6)*

[39] G. Gerig, M. Styner, M. E. Shenton, and J. A. Lieberman, "Shape versus size: Improved understanding of the morphology of brain structures," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2001, pp. 24–32. *(Cited on pages 6 and 7)*

[40] U. Castellani, E. Rossato, V. Murino, M. Bellani, G. Rambaldelli, C. Perlini, L. Tomelleri, M. Tansella, and P. Brambilla, "Classification of schizophrenia using feature-based morphometry," *Journal of Neural Transmission*, vol. 119, no. 3, pp. 395–404, 2012. *(Cited on pages 6 and 33)*

[41] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *Neuroimage*, vol. 56, no. 2, pp. 387–399, 2011. *(Cited on page 7)*

[42] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013. *(Cited on page 7)*

[43] B. Mwangi, M.-J. Wu, B. Cao, I. C. Passos, L. Lavagnino, Z. Keser, G. B. Zunta-Soares, K. M. Hasan, F. Kapczinski, and J. C. Soares, "Individualized prediction and clinical staging of bipolar disorders using neuroanatomical biomarkers," *Biological psychiatry: cognitive neuroscience and neuroimaging*, vol. 1, no. 2, pp. 186–194, 2016. *(Cited on pages 7 and 49)*

[44] Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos, "Compare: classification of morphological patterns using adaptive regional elements," *IEEE transactions on medical imaging*, vol. 26, no. 1, pp. 93–105, 2007. *(Cited on page 7)*

[45] N. Koutsouleris, E. M. Meisenzahl, C. Davatzikos, R. Bottlender, T. Frodl, J. Scheuerecker, G. Schmitt, T. Zetzsche, P. Decker, M. Reiser *et al.*, "Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition," *Archives of general psychiatry*, vol. 66, no. 7, pp. 700–712, 2009. *(Cited on page 7)*

[46] U. Yoon, J.-M. Lee, K. Im, Y.-W. Shin, B. H. Cho, I. Y. Kim, J. S. Kwon, and S. I. Kim, "Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia," *Neuroimage*, vol. 34, no. 4, pp. 1405–1415, 2007. *(Cited on page 7)*

[47] L. Palaniyappan and P. F. Liddle, "Aberrant cortical gyrification in schizophrenia: a surface-based morphometry study," *Journal of psychiatry & neuroscience: JPN*, vol. 37, no. 6, p. 399, 2012. *(Cited on page 7)*

[48] M. E. Shenton, C. C. Dickey, M. Frumin, and R. W. McCarley, "A review of mri findings in schizophrenia," *Schizophrenia research*, vol. 49, no. 1-2, pp. 1–52, 2001. *(Cited on page 7)*

[49] D. Arnone, J. Cavanagh, D. Gerber, S. Lawrie, K. Ebmeier, and A. McIntosh, "Magnetic resonance imaging studies in bipolar disorder and schizophrenia: meta-analysis," *The British Journal of Psychiatry*, vol. 195, no. 3, pp. 194–201, 2009. *(Cited on page 7)*

[50] S. Selvaraj, D. Arnone, D. Job, A. Stanfield, T. F. Farrow, A. C. Nugent, H. Scherk, O. Gruber, X. Chen, P. S. Sachdev *et al.*, "Grey matter differences in bipolar disorder: a meta-analysis of voxel-based morphometry studies," *Bipolar disorders*, vol. 14, no. 2, pp. 135–145, 2012. *(Cited on page 7)*

[51] E. Bora, A. Fornito, M. Yücel, and C. Pantelis, "Voxelwise meta-analysis of gray matter abnormalities in bipolar disorder," *Biological psychiatry*, vol. 67, no. 11, pp. 1097–1105, 2010. *(Cited on page 7)*

[52] K. Yu, C. Cheung, M. Leung, Q. Li, S. Chua, and G. McAlonan, "Are bipolar disorder and schizophrenia neuroanatomically distinct? an anatomical likelihood meta-analysis," *Frontiers in human neuroscience*, vol. 4, p. 189, 2010. *(Cited on page 7)*

[53] D. Anderson, B. A. Ardekani, K. E. Burdick, D. G. Robinson, M. John, A. K. Malhotra, and P. R. Szeszko, "Overlapping and distinct gray and white matter abnormalities in schizophrenia and bipolar i disorder," *Bipolar disorders*, vol. 15, no. 6, pp. 680–693, 2013. *(Cited on page 7)*

[54] K. M. Cecil, M. P. DelBello, R. Morey, and S. M. Strakowski, "Frontal lobe differences in bipolar disorder as determined by proton mr spectroscopy," *Bipolar disorders*, vol. 4, no. 6, pp. 357–365, 2002. *(Cited on page 7)*

[55] C. Yüksel, J. McCarthy, A. Shinn, D. L. Pfaff, J. T. Baker, S. Heckers, P. Renshaw, and D. Öngür, "Gray matter volume in schizophrenia and bipolar disorder with psychotic features," *Schizophrenia research*, vol. 138, no. 2-3, pp. 177–182, 2012. *(Cited on page 7)*

[56] P. Brambilla, M. A. Nicoletti, R. B. Sassi, A. G. Mallinger, E. Frank, D. J. Kupfer, M. S. Keshavan, and J. C. Soares, "Magnetic resonance imaging study of corpus callosum abnormalities in patients with bipolar disorder," *Biological psychiatry*, vol. 54, no. 11, pp. 1294–1297, 2003. *(Cited on page 7)*

[57] M. Bellani, P. Yeh, M. Tansella, M. Balestrieri, J. Soares, and P. Brambilla, "Dti studies of corpus callosum in bipolar disorder." *Biochemical Society transactions*, vol. 37, no. Pt 5, p. 1096, 2009. *(Cited on page 7)*

[58] J. Li, E. K. Edmiston, K. Chen, Y. Tang, X. Ouyang, Y. Jiang, G. Fan, L. Ren, J. Liu, Y. Zhou *et al.*, "A comparative diffusion tensor imaging study of corpus callosum subregion integrity in bipolar disorder and schizophrenia," *Psychiatry Research: Neuroimaging*, vol. 221, no. 1, pp. 58–62, 2014. *(Cited on page 7)*

[59] F. Benedetti, P.-H. Yeh, M. Bellani, D. Radaelli, M. A. Nicoletti, S. Poletti, A. Falini, S. Dallaspezia, C. Colombo, G. Scotti *et al.*, "Disruption of white matter integrity in bipolar depression as a possible structural marker of illness," *Biological psychiatry*, vol. 69, no. 4, pp. 309–317, 2011. *(Cited on page 7)*

[60] L. C. Hanford, G. B. Hall, L. Minuzzi, and R. B. Sassi, "Gray matter volumes in symptomatic and asymptomatic offspring of parents diagnosed with bipolar disorder," *European child & adolescent psychiatry*, vol. 25, no. 9, pp. 959–967, 2016. *(Cited on page 7)*

[61] J. Townsend and L. L. Altshuler, "Emotion processing and regulation in bipolar disorder: a review," *Bipolar disorders*, vol. 14, no. 4, pp. 326–339, 2012. *(Cited on page 7)*

[62] M. Wessa, C. Perlini, and P. Brambilla, "Neuropsychological underpinnings of the dynamics of bipolar disorder," *Epidemiology and psychiatric sciences*, vol. 24, no. 6, pp. 479–483, 2015. *(Cited on page 7)*

[63] A. Ferro, C. Bonivento, G. Delvecchio, M. Bellani, C. Perlini, N. Dusi, V. Marinelli, M. Ruggeri, A. C. Altamura, B. Crespo-Facorro *et al.*, "Longitudinal investigation of the parietal lobe anatomy in bipolar disorder and its association with general functioning," *Psychiatry Research: Neuroimaging*, vol. 267, pp. 22–31, 2017. *(Cited on page 7)*

[64] N. Just, "Improving tumour heterogeneity mri assessment with histograms," *British journal of cancer*, vol. 111, no. 12, p. 2205, 2014. *(Cited on page 11)*

[65] J. J. Kulynych, L. F. Luevano, D. W. Jones, and D. R. Weinberger, "Cortical abnormality in schizophrenia: an in vivo application of the gyrification index," *Biological Psychiatry*, vol. 41, no. 10, pp. 995–999, 1997. *(Cited on page 13)*

[66] M. Schaer, M. B. Cuadra, L. Tamarit, F. Lazeyras, S. Eliez, and J.-P. Thiran, "A surface-based approach to quantify local cortical gyrification," *IEEE transactions on medical imaging*, vol. 27, no. 2, pp. 161–170, 2008. *(Cited on page 13)*

[67] A. Cachia, M.-L. Paillère-Martinot, A. Galinowski, D. Januel, R. de Beaurepaire, F. Bellivier, E. Artiges, J. Andoh, D. Bartrés-Faz, E. Duchesnay *et al.*, "Cortical folding abnormalities in schizophrenia patients with resistant auditory hallucinations," *Neuroimage*, vol. 39, no. 3, pp. 927–935, 2008. *(Cited on page 13)*

[68] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, pp. 837–845, 1988. *(Cited on page 16)*

[69] V. Vapnik, *Statistical learning theory. 1998*. Wiley, New York, 1998, vol. 3. *(Cited on pages 19 and 34)*

[70] X. Zhu, J. Lafferty, and R. Rosenfeld, "Semi-supervised learning with graphs," Ph.D. dissertation, Carnegie Mellon University, language technologies institute, school of computer science Pittsburgh, PA, 2005. *(Cited on page 20)*

[71] A. Erdem and M. Pelillo, "Graph transduction as a noncooperative game," *Neural Computation*, vol. 24, no. 3, pp. 700–723, 2012. *(Cited on pages 20, 29, 30, 37, and 50)*

[72] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Advances in neural information processing systems*, 2004, pp. 321–328. *(Cited on page 20)*

[73] J. M. Smith and G. R. Price, "The logic of animal conflict," *Nature*, vol. 246, no. 5427, p. 15, 1973. *(Cited on page 22)*

[74] J. W. Weibull, *Evolutionary game theory*. MIT press, 1997. *(Cited on pages 22 and 37)*

[75] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, no. Feb, pp. 207–244, 2009. *(Cited on pages 23, 24, 28, 30, 32, 37, and 50)*

[76] L. G. Quintas, "A note on polymatrix games," *International Journal of Game Theory*, vol. 18, no. 3, pp. 261–272, 1989. *(Cited on page 27)*

[77] J. T. Howson Jr, "Equilibria of polymatrix games," *Management Science*, vol. 18, no. 5-part-1, pp. 312–318, 1972. *(Cited on page 27)*

[78] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The complexity of computing a nash equilibrium," *SIAM Journal on Computing*, vol. 39, no. 1, pp. 195–259, 2009. *(Cited on page 27)*

[79] C. Daskalakis, "On the complexity of approximating a nash equilibrium," *ACM Transactions on Algorithms (TALG)*, vol. 9, no. 3, p. 23, 2013. *(Cited on page 27)*

[80] J. Nash, "Non-cooperative games," *Annals of mathematics*, pp. 286–295, 1951. *(Cited on page 27)*

[81] E. Z. Mequanint, S. Rota Bulò, and M. Pelillo, "Dominant-set clustering using multiple affinity matrices," in *SIMBAD*, 2015, pp. 186–198. *(Cited on page 29)*

[82] E. Z. Mequanint, Y. T. Tesfaye, H. Idrees, A. Prati, M. Pelillo, and M. Shah, "Large-scale image geo-localization using dominant sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. *(Cited on page 29)*

[83] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering." in *NIPS*, 2004, pp. 1601–1608. *(Cited on pages 29 and 37)*

[84] N. Kriegeskorte, W. K. Simmons, P. S. Bellgowan, and C. I. Baker, "Circular analysis in systems neuroscience: the dangers of double dipping," *Nat Neurosci*, vol. 12, no. 5, pp. 535–540, 2009. *(Cited on page 30)*

[85] B. Mwangi, D. Spiker, G. B. Zunta-Soares, and J. C. Soares, "Prediction of pediatric bipolar disorder using neuroanatomical signatures of the amygdala," *Bipolar Disord*, vol. 16, no. 7, pp. 713–721, Nov 2014. *(Cited on page 30)*

[86] B. Mwangi, M. J. Wu, I. E. Bauer, H. Modi, C. P. Zeni, G. B. Zunta-Soares, K. M. Hasan, and J. C. Soares, "Predictive classification of pediatric bipolar disorder using atlas-based diffusion weighted imaging and support vector machines," *Psychiatry Res*, vol. 234, no. 2, pp. 265–271, Nov 2015. *(Cited on page 30)*

[87] I. Valli, A. F. Marquand, A. Mechelli, M. Raffin, P. Allen, M. L. Seal, and P. McGuire, "Identifying Individuals at High Risk of Psychosis: Predictive Utility of Support Vector Machine using Structural and Functional MRI Data," *Front Psychiatry*, vol. 7, p. 52, 2016. *(Cited on page 30)*

[88] Y. Xiao, S. Lui, W. Deng, L. Yao, W. Zhang, S. Li, M. Wu, T. Xie, Y. He, X. Huang *et al.*, "Altered cortical thickness related to clinical severity but not the untreated disease duration in schizophrenia," *Schizophr Bull*, vol. 41, no. 1, pp. 201–210, 2015. *(Cited on page 30)*

[89] M. E. Shenton, R. Kikinis, F. A. Jolesz, S. D. Pollak, M. LeMay, C. G. Wible, H. Hokama, J. Martin, D. Metcalf, M. Coleman *et al.*, "Abnormalities of the left temporal lobe and thought disorder in schizophrenia: a quantitative magnetic resonance imaging study," *N Engl J Med*, vol. 327, no. 9, pp. 604–612, 1992. *(Cited on page 30)*

[90] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2015. *(Cited on page 31)*

[91] S. Ebert, M. Fritz, and B. Schiele, "Pick your neighborhood–improving labels and neighborhood structure for label propagation," in *ICPR*, 2011, pp. 152–162. *(Cited on page 32)*

[92] B. Rashid, M. R. Arbabshirani, E. Damaraju, M. S. Cetin, R. Miller, G. D. Pearlson, and V. D. Calhoun, "Classification of schizophrenia and bipolar patients using static and dynamic resting-state fmri brain connectivity," *Neuroimage*, vol. 134, pp. 645–657, 2016. *(Cited on pages 33 and 50)*

[93] L. Squarcina, U. Castellani, M. Bellani, C. Perlini, A. Lasalvia, N. Dusi, C. Bonetto, D. Cristofalo, S. Tosato, G. Rambaldelli *et al.*, "Classification of first-episode psychosis in a large cohort of patients using support vector machine and multiple kernel learning techniques," *Neuroimage*, 2015. *(Cited on page 33)*

[94] S. Vascon, M. Frasca, R. Tripodi, G. Valentini, and M. Pelillo, "Protein function prediction as a graph-transduction game," *Pattern Recognition Letters*, 2018. *(Cited on page 34)*

[95] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman *et al.*, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006. *(Cited on pages 35, 36, and 62)*

[96] C. R. Lake, "Emil kraepelin (1856–1926) established the kraepelinian dichotomy and schizophrenia but then reneged," in *Schizophrenia Is a Misdiagnosis*. Springer, 2012, pp. 63–91. *(Cited on page 49)*

[97] E. Maggioni, M. Bellani, A. Altamura, and P. Brambilla, "Neuroanatomical voxel-based profile of schizophrenia and bipolar disorder," *Epidemiology and psychiatric sciences*, vol. 25, no. 4, pp. 312–316, 2016. *(Cited on page 49)*

[98] E. Maggioni, B. Crespo-Facorro, I. Nenadic, F. Benedetti, C. Gaser, H. Sauer, R. Roiz-Santiañez, S. Poletti, V. Marinelli, M. Bellani *et al.*, "Common and distinct structural features of schizophrenia and bipolar disorder: The european network on psychosis, affective disorders and cognitive trajectory (enpact) study," *PloS one*, vol. 12, no. 11, p. e0188000, 2017. *(Cited on page 49)*

[99] T. Insel, B. Cuthbert, M. Garvey, R. Heinssen, D. S. Pine, K. Quinn, C. Sanislow, and P. Wang, "Research domain criteria (rdoc): toward a new classification framework for research on mental disorders," 2010. *(Cited on page 49)*

[100] P. Brambilla, C. Perlini, M. Bellani, L. Tomelleri, A. Ferro, S. Cerruti, V. Marinelli, G. Rambaldelli, T. Christodoulou, J. Jogia *et al.*, "Increased salience of gains versus decreased associative learning differentiate bipolar disorder from schizophrenia during incentive decision making," *Psychological medicine*, vol. 43, no. 3, pp. 571–580, 2013. *(Cited on pages 50 and 54)*

[101] A. Bechara, A. R. Damasio, H. Damasio, and S. W. Anderson, "Insensitivity to future consequences following damage to human prefrontal cortex," *Cognition*, vol. 50, no. 1-3, pp. 7–15, 1994. *(Cited on page 50)*

[102] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, vol. 398. *(Cited on page 50)*

[103] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological physics and technology*, vol. 10, no. 3, pp. 257–273, 2017. *(Cited on page 61)*

[104] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015. *(Cited on pages 61 and 63)*

[105] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256. *(Cited on page 62)*

[106] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105. *(Cited on pages 62 and 69)*

[107] H.-I. Suk, S.-W. Lee, D. Shen, A. D. N. Initiative *et al.*, "Deep ensemble learning of sparse regression models for brain disease diagnosis," *Medical image analysis*, vol. 37, pp. 101–113, 2017. *(Cited on page 62)*

[108] N. J. Kabani, "3d anatomical atlas of the human brain," *Neuroimage*, vol. 7, pp. P–0717, 1998. *(Cited on page 62)*

[109] W. H. Pinaya, A. Gadelha, O. M. Doyle, C. Noto, A. Zugman, Q. Cordeiro, A. P. Jackowski, R. A. Bressan, and J. R. Sato, "Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia," *Scientific reports*, vol. 6, p. 38897, 2016. *(Cited on page 62)*

[110] E. Hosseini-Asl, G. Gimel'farb, and A. El-Baz, "Alzheimer's disease diagnostics by a deeply supervised adaptable 3d convolutional network," *arXiv preprint arXiv:1607.00556*, 2016. *(Cited on page 62)*

[111] A. Payan and G. Montana, "Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks," *arXiv preprint arXiv:1502.02506*, 2015. *(Cited on page 63)*

[112] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3d brain mri classification," in *Biomedical Imaging*

(ISBI 2017), 2017 IEEE 14th International Symposium on. IEEE, 2017, pp. 835–838. *(Cited on page 63)*

[113] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, and Z. J. Wang, "3d cnn based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural mri," *IEEE Access*, vol. 5, pp. 23 626–23 636, 2017. *(Cited on page 63)*

[114] G. Kang, K. Liu, B. Hou, and N. Zhang, "3d multi-view convolutional neural networks for lung nodule classification," *PloS one*, vol. 12, no. 11, p. e0188290, 2017. *(Cited on page 63)*

[115] S. Hoo-Chang, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, p. 1285, 2016. *(Cited on page 63)*

[116] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016. *(Cited on page 63)*

[117] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016. *(Cited on page 63)*

A

PUBLICATIONS

> ... Dreaming is good. But now your dream must be profitable. For example, you cannot find investors who will put funds in your project without them seeing the potential for profitability ...
>
> Olivier Madiba

- Learning from Enhanced Contextual Similarity in Brain Imaging Data for Classification of Schizophrenia, Dagnew T.M., Squarcina L., Rivolta M.W., Brambilla P., Sassi R., In: International Conference on Image Analysis and Processing (ICIAP), 2017. Lecture Notes in Computer Science, vol 10484. Springer, Cham.
- Automated Cortical Thickness and Skewness Feature Selection in Bipolar Disorder using a Semi-Supervised Learning Method: Letizia Squarcina, Tewodros M. Dagnew, Massimo W. Rivolta, Marcella Bellani, Roberto Sassi, Paolo Brambilla. Journal of Affective Disorders (Submitted, Jan. 2019).