Machine Learning Application In Neuroscience For Pre-Surgical Brain Tumor Removal Procedure

Lukman Enegi Ismaila

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

université angers

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ D'ANGERS
COMUE UNIVERSITÉ BRETAGNE LOIRE

ÉCOLE DOCTORALE N° 641
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Signal, Image, Vision*

Par

«Lukman Enegi ISMAILA (92011510)»

«Machine learning application in neuroscience for brain tumor resection procedure»

Thèse présentée et soutenue à « Campus du végétal - Angers», le « 03/07/2023 »
«LARIS - Laboratoire Angevin de Recherche en Ingénierie des Systèmes»

Rapporteurs avant soutenance :

Pr. Hasan DEMIREL, Eastern Mediterranean University, Cyprus
Dr. Stephanie BRICQ, Université de Bourgogne, France

Composition du Jury :

Examinateur :	Pr. Ilyess ZEMMOURA, Université de Tours, France
Dir. de thèse :	Pr. David ROUSSEAU, Université d'Angers, France.
Co-encadrants :	Dr. Pejman RASTI, CERADE, Université d'Angers, France.
	Dr. Jean-Michel LEMÉE, Département de Neurochirurgie, CHU d'Angers, Angers, France.

Invité(s) :

Dr. Kolawole BABALOLA, European Bioinformatics Institute (EMBL-EBI) United Kingdom.

ACKNOWLEDGEMENT

Praise be to *Allah* for the gift of life, wisdom, tranquility and abundant blessing in my life.

I would like to extend my deepest appreciation to Pr. David Rousseau, my supervisor. Under his tutelage, I developed invaluable skills in critical analysis, and scientific judgement. His unwavering patience, support, and enthusiasm were instrumental in my growth and development as a researcher, and it was an honor to have the opportunity to learn under his watch. My sincere gratitude to Dr. Pejman Rasti whose extensive support and valuable guidance made this scientific journey an absolute pleasure. He has relentlessly helped me through his professional experience and invaluable advice, I always appreciate the calm relationship. I also express my special appreciation to Dr. Jean-Michel Lemée for the immense support and advises. He practically prepared all the medical dataset used in this PhD, and through his professional support, he continues to ensure that this work is clinically sound.

I would like to express my heartfelt gratitude to the members of my committee, Pr. Ilyess Zemmoura, Pr. Hasan Demirel, and Dr. Stephanie Bricq, for graciously agreeing to evaluate my work, and for their valuable presence during my PhD defense. I also appreciate Dr. Kolawole Babalola who agrees to attend my PhD defense as the invited guest.

I am deeply grateful to my beautiful wife, Mrs. Fauwzziyyah Umar, for her unwavering kindness and steadfast support throughout my PhD journey. Her presence was a source of comfort and encouragement, particularly during challenging times. Being around you and Maheer brings immeasurable joy and happiness to my life that I will forever cherish.

I dedicate this PhD to my parents, Hajiya Memunat Sumaila and Alhaji Sumaila Salami for the endless favour, support and prayers. I would like to extend my special appreciation to my dear siblings, Arc. Yusuf, RN. Hajara, and Mr. Musa, for their valuable guidance and encouragement during this PhD. Their insightful advice and unwavering support were instrumental in helping me overcome obstacles and make significant progress towards the successful completion of this journey.

I am grateful to all other members of my family and friends for their helpful support and encouragement throughout my academic journey. Their love and belief in me have been a constant source of motivation and inspiration. I also appreciate the quality time in discussion and exchanging ideas with all colleagues in INRAe, Polytech Angers, especially all members of the ImHorPhen team.

Finally, I thank the Nigerian government who generously funded this doctoral study through the Petroleum Technology Development Fund (PTDF) under the management of Campus France international studies program.

ABBREVIATION

AAL - Anatomical Automatic Labeling
ACR – American College of Radiology
AD – Alzheimer's Disease
ADNI – Alzheimer's Disease Neuroimaging Initiative
AMIGO - Advanced Multimodality Image Guided Operating
BOLD – Blood Oxygenation Level Dependent
CPT – Current Procedural Terminology
CT - Computed tomography
CNN - Convolutional Neural Network
DBS – Deep Brain Stimulation
DAN - Dorsal Attention Network
DNN - Deep Neural Network
DMN – Default Mode Network
EPI - Echo planar imaging
EEG – Electroencephalography
FGS - Fluorescence Guided Surgery
FCD – Functional Connectivity Density
GNN - Graph Neural Network
ICA – Independent Component Analysis
ICN - Intrinsic Connectivity Network
ICs – Independent Components
iMRI - Intraoperative MRI
IoU - Intersection over union
LANG - Language Network
LDA - Linear Discriminant Analysis
lFPCN - Left Fronto-parietal Control Network
MEG – Magnetoencephalography
MNI - Montreal Neurological Institute
MRI – Magnetic Resonance Imaging

MLP - Multilayer Perceptron
NMR – Nuclear Magnetic Resonance
PET – Positron Emission Tomography
PCA - Principal component analysis
rFPCN - Right Fronto-parietal Control Network
RF – Radio Frequency
rs-fMRI - Resting-State Functional Magnetic Resonance Imaging
RSN - Resting-State Network
SAL - Salience Network
SAR – Specific Absorption Rate
SBCA - Seed Based Correlation Analysis
SICA - Spatial Independent Component Analysis
SPM - Statistical Parametric Mapping
SMF – Static Magnetic Field
SPECT – Single-Photon Emission Computed Tomography
T-SNE - t-distributed stochastic neighbor embedding
TNC - Total Number of Components
VAN - Ventral Attention Network

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Research context

Brain tumors represent a significant public health issue all around the world, with approximately 6,000 new cases diagnosed each year in France [1]. Despite advances in surgical and medical treatments, the prognosis for patients with brain tumors remains poor, with a 5-year survival rate of only 35% [2, 3]. Therefore, there is a pressing need to develop innovative approaches to improve the tumor resection procedure with regards to providing improved accuracy in identifying the tumor and other regions of the brain to avoid interference on the healthy tissue as this can lead to post-surgical and recovery complications amongst others.

In this PhD, we have narrowed our attention to one critical stage of brain tumor resection procedure as illustrated in the thesis focus (see Figure 1.1). The medical management of newly diagnosed brain tumors comprised of the following several steps; from the radiological diagnosis, to the surgical resection and the adjuvant treatment [4]. The brain tumor resection stage requires the neurosurgeon correct understanding of various brain functional regions in order to minimize the post-surgical neurological impairment. Indeed, this aspect needs extreme care and high level of precision because, any interference into unconcerned brain regions can lead to bleeding, brain swelling, infection, brain damage or even death.

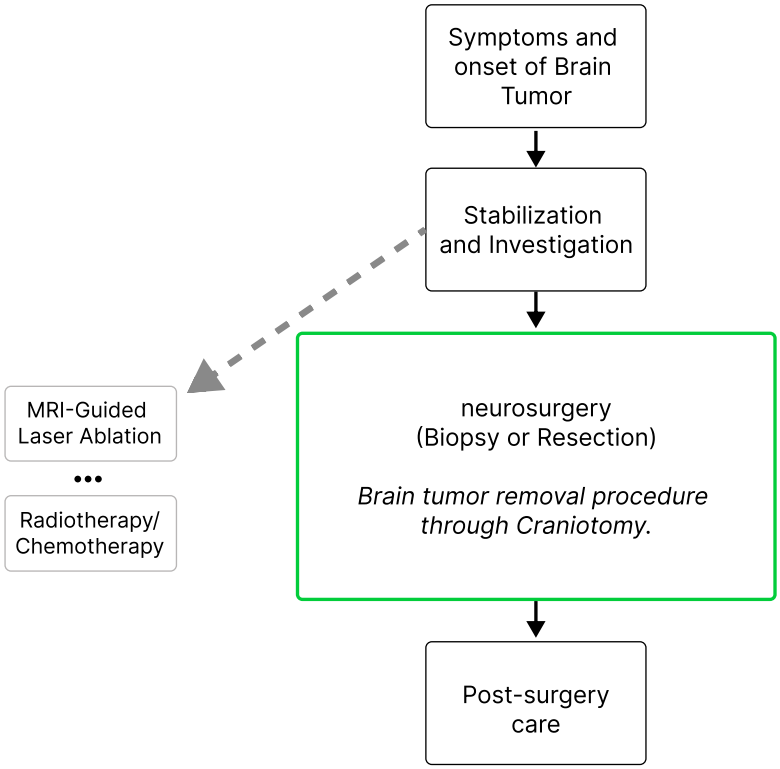


Figure 1.1 – Thesis focus: neurosurgical planning stage.

Medical surgery is the usual treatment for most brain tumors. During the process of brain tumor removal, a neurosurgeon creates an opening in the skull which is generally referred to as craniotomy as illustrated in figure 1.2. Whenever possible, the surgeon attempts to remove all the tumor while minimizing the risk of postoperative deficit, by sparing important functional brain areas. If the tumor cannot be completely removed without damaging vital brain tissue, doctors then aim to remove as much of the tumor as possible. This brain tumor surgical procedure cannot be done without adequate understanding of the various brain regions, since it helps to avoid non-tumor regions of the brain. Generally, clinicians use magnetic resonance imaging (MRI) machines to produce detailed images of the functional regions of the human brain, preoperatively to identify the surrounding functional areas. It allows the realization of a presurgical planning with the choice of the surgical access to the lesion and the appropriate resection goal tailored to each patient to minimize postoperative deficit. The MRI machine is a non-invasive imaging technology used to investigate anatomy and function of the body for both healthy subjects and unhealthy patients without the use of damaging ionizing radiation. This non-invasive neuro-imaging option has gotten a lot of atten-

tion for monitoring, and early diagnosis of neural disorders which allow useful efforts to mitigate the progress. Using the blood oxygen level dependent (BOLD) imaging technique, functional MRI (fMRI) indirectly measures the brain's neural activity by detecting blood flow changes. This is possible because, a local increase of cerebral blood flow, which is detectable with MRI through the BOLD signal, aligns with the activation of functional brain areas.

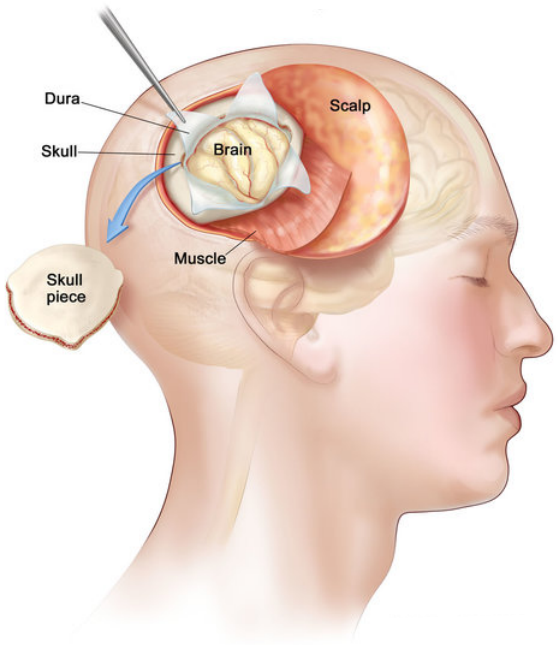


Figure 1.2 – Craniotomy approach of brain tumor removal [5].

The traditional task-based fMRI and resting-state fMRI are two of the main techniques used in neuro-imaging. While task-based method analyzes the variation of BOLD signals according to an activation paradigm to identify brain areas specifically involved in the activation paradigm (e.g. language network), which requires patients' cooperation and cognitive involvement. On the other hand, resting-state fMRI (rs-fMRI) allows identification of functional networks without explicit requirement of subjects to perform any cognitive task by analyzing the synchronicity of spontaneous low-frequency ($0.1Hz$) BOLD signal oscillation between brain areas. In addition to the wide interest among researchers on the advantages of resting-state fMRI over task-based fMRI approach [6, 7], task-based fMRI is associated with limitations such as high test-retest reliability [8], greater coverage of brain regions [9], reduced cognitive load [10],

more sensitive to clinical populations [11, 12] and no reliance on cognitive cooperation of the patient [13]. Resting-state fMRI approach is very interesting because, task-based fMRI technique may not be suitable for patients who are unable to cooperate or realize the task of the block paradigm due to special conditions such as unconscious patients, patients in pain, infants and so on. Although resting-state fMRI appears more appealing, it is not routinely available because of the necessity to have an expert reviewer who can manually identify each functional network. This revision stage is a laborious and time consuming process because, it involves the sorting and observation of all generated activation maps.

The application of machine learning in the field of neuroscience has the potential to revolutionize the way brain tumors are managed, by providing improved accuracy in diagnosis and better-informed treatment decisions. With the rapidly growing advancement and use of machine learning and computer vision to address critical problems in medical imaging research, there is an opportunity to leverage these advantages to support neurosurgeons by proposing state-of-the-art machine learning algorithms, that can be clinically validated and ensure better accuracy in brain tumor resection procedure [14].

1.2 Thesis statement and research questions

In order to improve predictive power and precision current medical imaging models, the neurosurgical literature is increasingly focusing on replacing traditional statistical models with more complex Machine Learning (ML) models [15, 16]. With a homogeneous distribution rate of machine learning adoption in clinical practice, it is clear that neurosurgeons have become open to adopt advanced machine learning techniques in different kinds of situation for example, in the survey carried our by V.E Staartijes *et al.* about 60.2% use ML to predict outcome, 51.5% for neural complications and 50.5% to interpret and quantify medical imaging [17]. Machine learning techniques have been used in various neurosurgery stages to detect cerebrospinal fluid leaks [18], functional brain network analysis for early dementia detection [19], or predict postoperative satisfaction [20]. Despite this interesting trend and availability of recent publications which reviews and proposes wide range of ML techniques in neurosurgery, a data-driven approach based on machine learning algorithms for full automation of the identification of functional brain networks has so far been hampered by lack of suf-

ficient unhealthy data. In this PhD work, we aim to exploit some advanced machine learning and computer vision techniques to propose and demonstrate functional brain network identification for brain tumor removal procedure using resting-state fMRI image data acquired from the department of neurosurgery of the University Hospital of Angers.

There have been several advances towards an improved technology for brain tumor removal. One of the most advanced technologies is fluorescence-guided surgery (FGS), which utilizes a fluorescent marker to highlight tumor tissue and distinguish it from surrounding healthy tissue. This technique enables neurosurgeons to achieve more complete resections and has been shown to improve patient outcomes [21]. Another promising technology is intraoperative MRI (iMRI), which allows for real-time imaging during surgery and helps neurosurgeons to accurately target and remove tumor tissue [22]. Advanced multimodality image-guided operating (AMIGO) [23] is another technology which allows neurosurgery operation to be guided for the removal of brain tumors. It involves the integration of multiple imaging modalities, such as magnetic resonance imaging (MRI), computed tomography (CT), and functional MRI (fMRI), to provide real-time images of the patient's brain during surgery. While all these technologies presents interesting alternative to high precision surgery for brain tumor removal, rs-fMRI has some advantages over these technologies because, it can identify all functional networks in a single 15-minute resting MRI sequence, without requiring the implementation of an MRI sequence and an experimental paradigm for each function to be tested. This allows for a more comprehensive evaluation of functional brain areas than FGS, which only highlights tumor tissue, and iMRI, which may not capture the full extent of the tumor due to its limited field of view. Additionally, rs-fMRI can detect functioning brain regions in patients unable to perform activation tasks, making it a valuable tool for preoperative planning. In the case of AMIGO, there is known risk of technical malfunctions, which can compromise the accuracy of the images and increase the risk of surgical errors. Additionally, the cost of AMIGO technology is already a significant barrier to its widespread adoption, as it requires specialized equipment and training for its use. AMIGO technology is currently only available in a limited number of hospitals, and its use may be restricted to high-volume centers with specialized expertise in brain tumor surgery. Lastly, the accuracy of the images produced by AMIGO depends on the skill and experience of the operator, who must be able to interpret and integrate multiple sources of information in real-time.

As a connection point, our focus in this research is to address the current limitation in the use of resting-state fMRI as a clinical routine. This could also address some critical neuroimaging limitations for general purpose usage [24].

Research Question 1 (RQ_1): In effort to standardize the resting-state approach of fMRI for functional brain network recognition, we identify that, an initial step is to explore an automatic function brain network recognition to avoid the manual review process, since it is one of the major limitations that affects routine adoption. We are interested in investigating the use of machine learning algorithms to recognize and provide better understanding of different brain areas to neurosurgeons while sparing healthy brain tissues as well as reduce the likelihood of postoperative deterioration. To ensure our deep learning model performs well with unhealthy data which may present spatial irregularities like displacement or even destruction of brain functional areas due to tumor invasion or local mass effect. It is necessary to train a machine learning model with unique dataset of comparative class however, unhealthy fMRI data for functional brain network recognition is limited owing to the fortunate rarity of brain tumors, as well as the substantial inter-individual variation in tumor formation or location. This motivates the extension of RQ_1 to allow us investigate ways of exploiting the observed similarity between healthy and unhealthy data. This approach could potentially, allow transferability of features learned from healthy data to boost recognition accuracy of fMRI functional brain networks in unhealthy data.

How can we efficiently automate functional brain network recognition in unhealthy patent by exploiting the information gain from healthy data ?

Research Question 2 (RQ_2): The manual review of functional brain networks by clinicians involves the process of data annotation as routinely required in rs-fMRI data preprocessing stage. This data annotation process is tedious and time consuming. It is also important to emphasize that this annotation process has no clinical benefit in the brain tumor resection process. At this stage, we aim to eliminate the process of healthy data annotation. This is important because, large amount of healthy data can be acquired for model learning, while annotation can be avoided since healthy volunteers can be easily enrolled for data acquisition in the non-invasive imaging modality.

Can we avoid the annotation of healthy subject data since it has no clinical relevance ?

Research Question 3 (RQ_3): In our demonstration of feature transferability from healthy data to unhealthy data, we observed interesting improvement in model accuracy which signifies that, indeed there is useful similarity between the two classes of data. However, we also observed sharp drop in the accuracy of model trained and validated with unhealthy data when compared with model trained and validated with healthy data, which underscores the existence of difference in data class. These two conflicting observations amount to data discrepancy which informed our intuition to better understand the indirect relationship that exist between healthy and unhealthy data. This investigation is interesting because, visual observation from all data class, shows that there is similarity between related functional brain network activation signals despite individual variability. Although, we expect the influence of brain tumor on functional brain network activation maps as a source of local difference in the case of unhealthy data, the overall relation requires further evaluation for better understanding.

What is the source of observed data discrepancy between healthy and unhealthy fMRI images ?

1.3 Research contributions

In this PhD, we proposed end-to-end deep learning models for the automatic identification of functional brain networks by analysing resting-state fMRI images. We expect that, this will open-up resting-state approach for fMRI brain region identification and further improve the precision of functional brain region identification in order to avoid interference and preserve the patient's neurological functions during brain tumor resection. We demonstrate that, it is possible to implement data-driven machine learning approach like deep learning, despite the well known limitation of small amount of unhealthy fMRI data. In this case, we proposed and end-to-end deep learning model for functional brain network identification. We also proposed to tackle unhealthy data limitation while benefiting from observed similarity between healthy and unhealthy data using transfer learning approach.

Despite the recorded progress of boosting our model accuracy in unhealthy patients using transfer learning approach, an observed limitation is the annotation of large dataset in this supervise learning technique. To address this problem, manual annotation of healthy data as required in the transfer learning approach from healthy to unhealthy fMRI data, needs to be avoided. We propose a self-supervision technique, which prevents the annotation of this healthy data for which there is no clinical interest. In this approach, we practically ensured that no healthy data annotation is necessary, to allow potentially large data collection since the process is non-invasive and healthy volunteers can easily be enrolled for data collection.

In effort to address the local and global relationship between healthy and unhealthy data, we demonstrated several statistical techniques to reveal local differences in fMRI images with respect to tumor influence on functional brain network activation. We also showed that latent space visualization can be a useful tool to understand the low dimensional distances between different classes of data in this difficult problem. Our final result shows that, setting the same brain network activation volume threshold for both healthy and unhealthy data resulted in the consistent ambiguous relationship between healthy and unhealthy data, since those of unhealthy data are also influences by brain tumor and other network activation map overlap-scenario.

The output of this PhD, demonstrates interesting insight in the use of deep learning techniques to automatically recognize critical regions of the brain and avoid disruption during brain surgery for tumor removal. Additionally, The exploitation of the similarity between healthy and unhealthy data is initiated by illustrating the use of transfer learning and self supervision technique to manage limited unhealthy data and avoid healthy data labeling respectively.

1.4 Significance of the contributions

Several studies have pointed out major limitations in task-based fMRI by highlighting that the performance of repetitive tasks may result in artifacts in the analysis of BOLD signals [25], and resting-state fMRI is considered a more reliable mapping technique in pre-operative surgical planning compared to task-based fMRI [26, 27], because of its ability to identify multiple networks at the same time, which saves scanning time when information from multiple networks is required [28].

The results presented in this report, opens a significant possibility in understanding functional brain regions for brain tumor removal procedure. A recent survey by M. Khosla *et al.* on machine learning in resting-state fMRI analysis argues that, there is need open-up some limitations of resting-state fMRI analysis using the optimized machine learning and medical imaging approach since resting-state is easier to standardize across sites compared to task-based protocols since it does not rely on external stimuli [29]. The results of this study, like the identification of default mode network (DMN) in resting-state connectivity provides interesting implications in neurological and psychiatric disorders, including autism, schizophrenia and alzheimer's [30, 31, 32]. Lastly, It is clear that clinical application of rs-fMRI is still limited, while many potential clinical adoption are currently being investigated to allow inclusion of this technique in presurgical planning for patients with brain tumor and epilepsy [28].

By addressing the main limitations of using resting-state fMRI method of functional brain network recognition for neurosurgical procedure, this work promotes accelerated standardization potential for this technique.

1.5 Structure of the document

The rest of this document is organized in three chapters (chapter 2, chapter 3 and chapter 4) before a concluding chapter 5. In these three chapters, we present our main methodological contributions:

In chapter 2, we present our work on machine learning algorithm for automatic identification of functional brain networks using shallow machine learning techniques and deep learning approach with further investigation of feature transferability [33, 34].

In chapter 3, we discuss our investigation on the use of self-supervision learning technique to address healthy data annotation which is time consuming and has no clinical benefit [35].

In chapter 4, we provide details of our study to better understand the local and global relationship between healthy and unhealthy fMRI data.

To allow better understanding of strategies discussed in this document, we present all relevant explanation of the terminology of machine learning and computer vision used in this thesis report in Annex A. In Annex B, we provide details of the clini-

cal procedure for fMRI functional brain image data acquisition and pre-processing for machine learning experiments to simplify understanding and distinguish resting-state fMRI data acquisition from our methodological contributions shown in chapters 2, 3 & 4. Lastly, in Annex C, we present a pilot study on the reduction of the complexity of the models via graph encoding to perform graph representation learning using graph neural network (GNN) [36].

CHAPTER 2

AUTOMATIC fMRI NETWORK RECOGNITION WITH SHALLOW AND DEEP LEARNING TECHNIQUES

In this chapter, we propose an end-to-end deep learning algorithm to classify functional brain networks in resting-state functional magnetic resonance imaging (rs-fMRI) data [33]. Considering the available small size data of unhealthy functional magnetic resonance imaging (fMRI) data, we demonstrated the use of transfer learning technique to improve our deep learning model which also highlights the existence of a similarity between healthy and unhealthy dataset [34]. Therefore this chapter is based on the above cited publications.

2.1 Introduction

One of the most researched applications of machine learning in healthcare is medical imaging [37]. While effort remains consistent in developing and improving algorithms, data availability is crucial for deploying efficient machine learning solutions [38]. The recent covid-19 pandemic has demonstrated, for instance, how the availability of a large annotated dataset could significantly boost the power of machine learning [39]. However, in most clinical practices, such an initiative to share a large dataset is still discouraged.

The machine learning community has developed several workarounds approaches to compensate for the lack of data. This compensation can be obtained using algorithms that learn faster, like in few-shot learning approaches [40]. The lack of data can also be compensated by automatically generating fake data, which are realistic enough to boost the training of algorithms. This includes the generation of synthetic data via

simulators [41], generative models [42] or via data augmentation [43]. Another approach known as transfer learning, uses pre-trained models on similar datasets [44]. In this chapter, we focus on this transfer learning approach.

While largely used in computer vision, transfer learning is still actively investigated especially in the medical domain [45, 46]. One of the most common transfer learning approach in computer vision is using pre-trained models from 2D color outdoor natural images. However, for specific application domains, such as medical imaging, this approach is neither optimal nor possible, due to the difference in data structure between medical images and 2D color natural images (3D images instead of 2D images, size of images, bit depth of images, ...) [47]. Also, the efficiency of transfer learning has been shown to be optimal when images share the similar content [48]. Lastly, transfer learning helps when the data used for the pre-training is less expensive when compared to images from the target domain. This analysis brings us to the simple and yet innovative idea of investigating the value of a pre-trained model on healthy subjects' data when transferred to unhealthy patients. We propose testing and illustrating this transfer learning idea in a specific clinical use case.

The rationale for the selection of this use case are:

a) We need a use case for which very few public datasets are indeed currently available

b) We need a use case for which the impact of the disease on the image is not too large to minimize the shift between healthy and unhealthy data

c) We need a use case for which imaging modality is non-invasive so that acquisition of images on healthy control is relatively easy.

This brought us to the selection of the clinical use case presented in the next section to test our original idea of transfer learning from healthy subjects' data to unhealthy patient data.

2.1.1 Clinical use case

Functional MRI (fMRI) is a method that eases the understanding of brain activation by analyzing the blood-oxygen-level-dependent (BOLD) signals, allowing the identification and localization of functional brain areas. The development of this technique

promotes a better understanding of the functional anatomy of the human brain and a more accurate characterization of inter-individual topographical variability of functional brain areas such as language areas [6]. Thus, some fMRI techniques are progressively included as a procedure in several pathologies for surgical planning [49, 50, 51, 52].

The standard fMRI approach is a task-based block paradigm contrasting brain activation at rest and when performing a specific task. However, despite its usefulness, this technique presents several drawbacks, and inconsistencies [53]; the patient's cooperation is needed, and it is unsuitable for young children and patients who are unable to perform the task. In addition, the study of several functional networks is time-consuming, and it requires the acquisition of each network with subsequent development of a specific activation task paradigm [54].

An alternative for the task-based characterization of functional networks is the resting-state fMRI (rs-fMRI), which studies the synchronization of low-frequency oscillation between brain areas at rest [55, 56]. It is possible and practical to identify from these signals the so-called Intrinsic Connectivity Networks (ICNs), which reflect the neuro-anatomical substrate that corresponds to brain functional networks [57, 58]. However, rs-fMRI for functional network identification is not yet part of the pre-operative routine, because of the high level of expertise needed for ICNs identification. Indeed, each of the ICNs needs to be visually reviewed by an expert to identify an individual functional network of interest [59]. To broaden the use of this technique in the pre-surgical planning for various surgical procedures, the initial stage consists of effective automation of fMRI brain network identification in patients' data.

In the literature, automated machine learning algorithms have been the subject of several studies to identify disease patterns in rs-fMRI data, especially in epilepsy [60, 61], as well as traumatic brain injuries [62], addiction [63], cognitive impairment [64], and psychiatric disorders like depression and schizophrenia [63, 61]. There have been relatively few attempts [65, 66, 57, 67] to automatically identify functional networks on rs-fMRI data using machine learning. Lu *et al.* [65] developed an instance-based automated method for identifying language networks in brain tumor subjects using independent component analysis (ICA)-based mapping on rs-fMRI. By contrast, we are data-driven and do not limit ourselves to only language networks. In fact, our study considers seven functional networks. Additionally, each of the previous studies has its defined scopes, data variants, and functional networks used for automated

identification in rs-fMRI for pre-surgical planning. In [66], the authors proposed a task-free paradigm for acquiring fMRI data, which was less demanding for patients and easy to administer. Further investigation was carried out on right-handed healthy control subjects. A semi-automated language component identification procedure was proposed and tested on healthy patients [66]. In this chapter, we consider unhealthy patients in addition to healthy subject data. In the study by [57], a model was trained to identify the main functional networks in a small number of healthy volunteers for different functional networks. The performance of the simple feed-forward network proposed in [57] is ultimately dependent on handcrafted features extracted from fMRI images. The above concerns motivated our proposition to design a specific end-to-end deep learning [68, 69] knowledge transfer method to identify and automate the detection of functional networks in rs-fMRI of unhealthy patients. This approach has the advantage of being applicable to patients in need of brain surgery due to brain tumors or other reasons.

While some efforts are being made to provide more and more public datasets of medical images of large interest, there are currently few available public datasets of resting-state fMRI of healthy or unhealthy individuals [70, 71]. However, these datasets have been produced with slightly different protocols than ours. The differences include the type of disease, number of participants, and MRI sequence for some areas. Therefore, these differences would prevent the transfer learning approach on our dataset. Other related datasets include the database of [72]. It is made of 227 healthy individuals aged 18 to 74 to investigate the impact of adult age on functional brain connectivity; the database of [73] includes 993 patients and 1, 421 healthy individuals to classify psychiatric disorders. We investigate patients with brain tumors. Therefore, these datasets would also not allow a direct transfer learning approach from healthy to unhealthy on our data. Therefore, the situation of clinical interest considered in this study is perfectly suited to test the possibility of transferring knowledge from healthy to unhealthy patients.

As an innovative elements, (i) we automatically identify functional networks on rs-fMRI data for the first time with an end-to-end deep learning method as opposed to handcrafted features that were previously proposed in the closest literature for this problem [57]. (ii) We demonstrate the value of transfer learning from a model of healthy control subjects to unhealthy patients with a brain tumor.

2.2 Materials and methods

In this study, preprocessed resting-state fMRI signals of functional brain network activation of 81 healthy control subjects and 55 unhealthy patients to acquire features which correspond to 7 biological networks of the brain, which are the Language Network (LANG), Salience Network (SAL), Ventral Attention Network (VAN), Default Mode Network (DMN), Left Fronto-parietal Control Network (lFPCN), Right Fronto-parietal Control Network (rFPCN), Dorsal Attention Network (DAN). The seven selected brain features represent the main intrinsic connectivity network (ICN) identified and described in resting-state fMRI literature. These particular networks were selected for the DMN to serve as a control for the others because of the inter-individual variability that makes them difficult to identify using detection software or by non-expert reviewers. The detailed process of fMRI data acquisition and preprocessing is provided in Annex B. All acquired images where processed in 'tmaps or gray level image and thresholded' format from healthy subject data as shown in Figure 2.1.

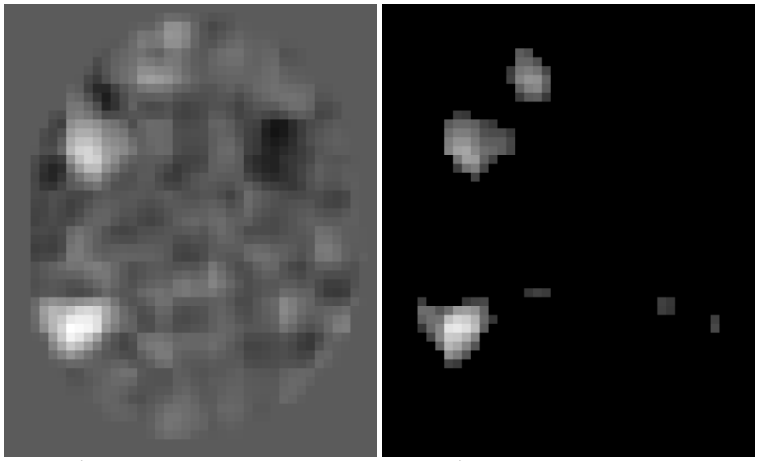


Figure 2.1 – Slice view of LANG network in a) gray level and b) threshold format.

2.3 Identification of functional brain networks through machine learning algorithms

The analysis of fMRI data involves the identification of functional brain networks that reflect the intrinsic organization of the human brain and its dynamic interactions

with the environment. However, this task is challenging due to the high dimensionality and complexity of fMRI data, as well as the limited prior knowledge about the underlying biological processes. To address these challenges, machine learning algorithms have emerged as a powerful tool for the analysis of fMRI data. These algorithms can be used to identify functional brain networks by capturing the intrinsic patterns of brain activity, as well as their correlations with external variables such as behavioral, demographic, and clinical factors. Over the past decade, a growing number of studies have applied machine learning algorithms to fMRI data, including independent component analysis (ICA) [74], graph-based methods [75], and deep learning [76]. In this approach, we explored both shallow and deep learning algorithms to better understand the applicability and evaluate its performance on our fMRI data task.

2.3.1 Shallow learning classification algorithms

These algorithms are characterized by their simple architecture, which consists of only a single layer of processing unit, and are well suited for tasks where the data has a low dimensionality or is relatively straightforward to model. There are numerous algorithms that have been developed for various purposes such as image classification, speech recognition, and natural language processing. Shallow learning, which is also known as shallow neural network or single-layer network, is among the most commonly used categories of machine learning algorithms. Our initial experiments focused on using some well known shallow learning algorithms for the classification of functional brain networks. These shallow learning algorithms are briefly discussed as follow:

Random forest: An approach for ensemble learning called "random forest" can be applied to both classification and regression problems. It is a form of decision tree method that mixes several decision trees rather than depending just on one decision tree to produce a prediction. The fundamental principle of random forest is to construct numerous decision trees on bootstrapped data samples and then utilize their combined forecasts to arrive at a final prediction. Problems involving regression and classification can both be solved using random forests. As applied to our fMRI brain network classification experiment, we used the entropy parameter to determine how nodes branch in a decision tree

$$Entropy = \sum_{i=1}^{c} -p_i * log_2(pi), \tag{2.1}$$

where p_i is the frequency of label i at a node and c is the number of unique labels.

Naïve bayesian classifier: The naïve bayes classifier is a simple probabilistic machine learning algorithm based on Bayes' theorem. it presumes that all aspects are independent hence, it is referred to as "Naïve". In a naïve bayes classifier, the goal is to predict the class of a given data point based on the values of its features. Given a set of classes C = c1, c2, ..., ck, the probability of a data point x belonging to a class c_i can be calculated using bayes' theorem

$$P(c_i/x) = \frac{P(x/c_i) * P(c_i)}{P(x)}, \tag{2.2}$$

where P(c_i/x) is the posterior probability of x belonging to class c_i, P(x/c_i) is the likelihood of observing x given that it belongs to class c_i, P(c_i) is the prior probability of class c_i, and P(x) is the evidence, which is calculated as the sum of the likelihoods of observing x for all classes.

Depending on the naïve bayes algorithm being utilized, the naïve bayes classifier can be trained using either bayesian inference or maximum likelihood estimation. Gaussian naïve bayes, multinomial naïve bayes, and bernoulli naïve bayes are the three primary varieties of naïve bayes algorithms. For example, gaussian naïve bayes are best suited to continuous data, while multinomial or bernoulli naïve bayes are better suited to discrete data. In our experiment, we explored the bernoulli naïve bayes for fMRI image classification task.

K-Nearest neighbors: This algorithm is a non-parametric, instance-based, supervised learning algorithm used for classification and regression tasks. The basic idea behind K-Nearest neighbors (KNN) is to find the K-nearest data points to a given test data point and make predictions based on the majority class or average value of the K-nearest neighbors

$$d(q,p) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}. \tag{2.3}$$

where $d(q,p)$ is Euclidean distance's between data points p and q.

Mathematically, given a training set of N points in a D-dimensional feature space, the KNN algorithm computes the distance between the test data point and each of the N training points, using a distance metric such as Euclidean distance, Manhattan distance, or Cosine similarity. The K-nearest neighbors are then selected and the prediction is made based on either the majority class of the K-nearest neighbors (in classification tasks) or the average value of the K-nearest neighbors (in regression tasks). Most commonly with euclidean distance as implemented in this experiment, KNN uses the Equation 2.3 to iterative compute the data point values related to fMRI functional network maps.

Support vector machine: Support vector machine (SVM) is another supervised learning algorithm used for classification and regression tasks. It is based on the concept of finding the maximum margin hyperplane that separates the data into two classes. The maximum margin hyperplane is the one that has the largest margin, or distance, from the closest data points, called support vectors. We describe the optimization problem that is tackled by SVMs as

$$min_{w,b,\zeta} \frac{1}{2} w^\top w + C \sum_{i=1}^{n} \zeta, \tag{2.4}$$

$$y_i(w^\top \phi(x_i) + b) \geqslant 1 - \zeta_i, \tag{2.5}$$

where ζ_i is the distance to the correct margin with $\zeta_i \geqslant 0$, i=1,...,n; C is the regularization parameter; $w^\top w = \|w^2\|$ is the normal vector; $\phi(x_i)$ denotes the transformed input space vector; b is a bias parameter; lastly, y_i denotes the i-th target value.

Although, SVM can also be extended to non-linear classification problems for mapping the data into a high-dimensional feature space using a kernel function, our interest point in this experiment is on linear kernel for our multi-class fMRI images classification. In this case, the objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies different data points.

Classification tree: A classification tree is a tree-based technique that iteratively divides the data into progressively smaller subsets depending on the characteristics that produce the greatest information gain or impurity reduction at each split. The end

result is a decision tree, where each leaf node represents a class prediction and each interior node represents a feature test. The path from the root to a leaf node based on the feature values is used to forecast a new data point. In our fMRI image classification, we explore this technique since its simpler and easier to interpret, although random forest models can be more complex as it combines multiple decision trees to improve classification performance.

Feed forward neural network: A feed-forward neural network is a type of artificial neural network that consists of multiple layers of interconnected nodes, or artificial neurons. The name 'feed-forward' refers to the fact that data flows through the network from input to output, without looping back.

Each neuron in a feed-forward neural network takes inputs from the previous layer, performs a weighted sum of these inputs, and applies an activation function to produce an output, which becomes the input for the next layer. The activation function is typically a non-linear function, such as a sigmoid or a rectified linear unit (ReLU), that introduces non-linearity into the model and allows the network to learn complex non-linear relationships between the inputs and outputs. Mathematically, a feedforward neural network is represented in Equation 2.7 while the cost function is given in Equation 2.6. Let x be the input vector, W and b is the weight and bias matrices, respectively, and f is the activation function. The cost function of the network is given by

$$C(W, b) = \frac{1}{2n} \sum_{x} \|y(x) - a\|^2, \tag{2.6}$$

where w is weights gathered in the network; b is biases n is number of inputs for training; $\|v\|$ is vector $v's$ normal length; x is input and a is output vectors. The output y of a single neuron in layer l is given by:

$$y = f(W_l x + b_l), \tag{2.7}$$

To identify the most suited family of machine learning algorithm for functional network classification, we implemented six machine learning algorithms mentioned above, which includes: random forest, feed forward neural networks, naïve bayesian classifier, K-nearest neighbors, support vector machine, and classification tree. In our experiment, the random forest classifier consists of a combination of tree classifiers, 100 in our experiment. Each classifier is generated using a random vector sampled

independently from the input vector. Each tree casts a unit vote for the most popular class to classify an input vector. Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability that a given sample belongs to a particular class. Naïve bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. The k-nearest neighbors classifier [77] stores the complete training data. New examples are classified by choosing the majority class among the k-closest examples in the training data. We used the Euclidean distance to measure the tile distance between examples for our particular problem. Support vector machine is another powerful method for building a classifier. It aims to create a decision boundary between two classes that enables the prediction of labels from one or more feature vectors. This decision boundary, known as the hyperplane, is orientated so that it is as far as possible from the closest data points from each of the classes. Decision trees [78] recursively split the feature space based on tests that evaluate one feature variable against a threshold value. We used the information gain criteria for choosing the best test and top-down pruning with a value of 0.95 to reduce over-fitting.

2.3.2 Deep learning with CNN

Following the implementation of shallow machine learning models we consider the implementation of an end-to-end deep learning algorithm. Since deep learning allows modelling to learn from complex data structure and extract features from various signals like speech and images by using neural networks with numerous hidden layers (hence the name "deep"). Unlike shallow machine learning algorithms which require manual feature engineering, deep learning algorithms can extract and learn rich as well as meaningful representations of the data from the raw inputs, such as activation signals in our fMRI image etc. This makes deep learning particularly well-suited for tasks where the data is high-dimensional and complex, and where manual feature engineering is challenging and non-feasible or needs to be avoided. Our focus at this stage is to demonstrate the possibility of an end-to-end deep learning methods in our benchmark test. This is interesting because, it avoids the use of handcrafted features of our medical image data which allows greater flexibility of our model. We selected the predominant approach in computer vision, namely deep convolutional neural networks [79]. The baseline approach resorts to standard supervised training of the prediction

model (the neural network) on the target training data. No additional data sources were used. In particular, given a training set comprised of K pairs of images f_i and labels $\hat{y}_i$, we train the parameters θ of the network r using stochastic gradient descent to minimize empirical risk.

$$\theta^* = \arg\min_{\theta} \sum_{i=1}^{K} \mathcal{L}(\hat{y}_i, r(f_i, \theta)), \tag{2.8}$$

$\mathcal{L}$ denotes the loss function, which is cross-entropy in our case. The minimization is carried out using the adam optimizer [80] with a learning rate of 0.001. The architecture of networks $r(\cdot, \cdot)$, shown in Figure 2.2, has been optimized on a cross-sample set and is given as follows: three convolutional layers with filters of size 3×3 and respective numbers of filters; 64, 128 and 256, each followed by ReLU activations and 2×2 max pooling; a fully connected layer with 256 units, ReLU activation and dropout (0.5) and a fully connected output layer for 7 classes and a softmax activation. The hyper-parameters of the optimized CNN were based on a grid-search operating on the depth of the neural network. Other dimensions could be further investigated such as width as done in Efficient Net [81]. Here, we do not seek an absolute best performance but rather focus on the possible relative gain of performance brought by transfer learning from healthy control to unhealthy patients. In addition to the optimized CNN of Figure 2.2, we also included comparison with standard CNN architectures like VGG16 [82], ResNet [83] and DenseNet [84].

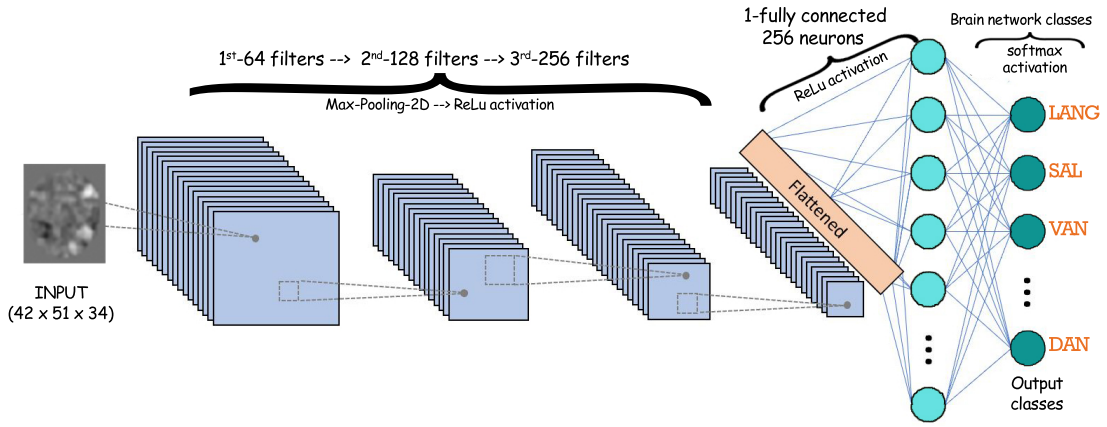


Figure 2.2 – Proposed end-to-end convolutional neural network(CNN) architecture for fMRI network classification.

The tested shallow and deep supervised learning classification algorithms were implemented based on fMRI data from 81 healthy subjects. The training dataset included 78 individual cartography with each of the seven main functional brain networks, corresponding to the 7 identified networks among the 55 ICs generated for each of the 78 healthy control subjects in the training group. In order to reduce the dimensionality and minimize over-fitting in shallow learning algorithms, we extracted the coordinates of the network activation peak of each cluster in order to minimize the number of variables considered for training before feeding the data into algorithms. Each algorithm was trained ten times with a cross-validation strategy to ensure robustness and confidence. Algorithms were then tested using the fMRI data from the four other healthy subjects. We used each of these algorithms for each patient to identify the seven identified networks among generated 55 ICs from the main functional networks. The identified networks were further compared to the reference networks by our two expert reviewers for validation. We identified the most suited algorithms for identifying seven main functional networks (DMN, lFPCN, LANG, rFPCN, SAL, DAN, and VAN). Finally, we tested the different parameters of the model to optimize the results. The best method was selected based on the highest classification performances. We consider further investigation on the possibility to increase our datasize to allow model training with more generalize data.

2.4 Data augmentation

In effort to strengthen the generality of our data, we explored the option of simulating some possible clinical noise as discussed in Table 2.1. This was a relevant approach because, it allows the opportunity to evaluate the influence of clinical noise in our model. Furthermore, this considerably increased the amount of our experiment data following systematic tune of augmentation parameters which provided the 9 variants in each case as shown in Table 2.2 which was observed to have significantly boost our model performance.

Some of the data augmentation options for simulating the clinical noise includes:

— Elastic transformation: Elastic image transformation is a technique used to change or distort a picture in a non-rigid way. While warping or deforming the pixels to conform with a new shape or position, elastic image transformation aims to preserve the semantic content of the image.

Table 2.1 – Clinical noise simulation from image data augmentation.

Name	Parameters	Medical simulation
Elastic transformation	Alpha, Sigma	Neuroplasticity
Horizontal flip	xyz-Axis	Symmetry of the brain
Pepper noise	probability	Common noise artifact in MRI images
Sharpen	alpha, brightness	Simulation of susceptibility artifacts
Scale	Zoom	Average human brain size

— Horizontal flip: This option provides the flip of both rows and columns horizontally for a given image. This allows the observation to understanding the level of horizontal symmetry of the brain.
— Pepper noise: Pepper noise is a type of image noise that appears as black pixels. Generally, this is referred to as "salt and pepper noise" when its random black and white pixels. In this case, It allows us to simulate the MRI noise in our images by imitating possible errors in the image acquisition process, such as a faulty sensor or transmission errors, or by image processing operations that introduce errors into the image.
— Sharpen: In effort to manipulate the sharp appearance of our input image, we applied a filter to the image, which enhances the high-frequency content of the image and increases the contrast between adjacent pixels.
— Scale: Image scale augmentation in this case is achieved through a variety of techniques, including resizing the images, cropping the images, or zooming in or out on the images.

The implementation of the above data augmentation options followed the procedure described in [85]. We observed non-compatibility for 3-dimension peppered noise augmentation, therefore, we implemented algorithm 2, which to allows 3-dimensional peppered noise augmentation option for our dataset. All other image augmentation options followed the described procedure in algorithm 1 and the visual output of the 6 initial data augmentation options are provided in Figure 2.3, while further efforts to use generated lesion mask in the strategic application of augmentation to simulate a more realistic effort of tumor presence is show in Figure 2.4.

Further data augmentation to simulate the impact of brain tumor in fMRI brain

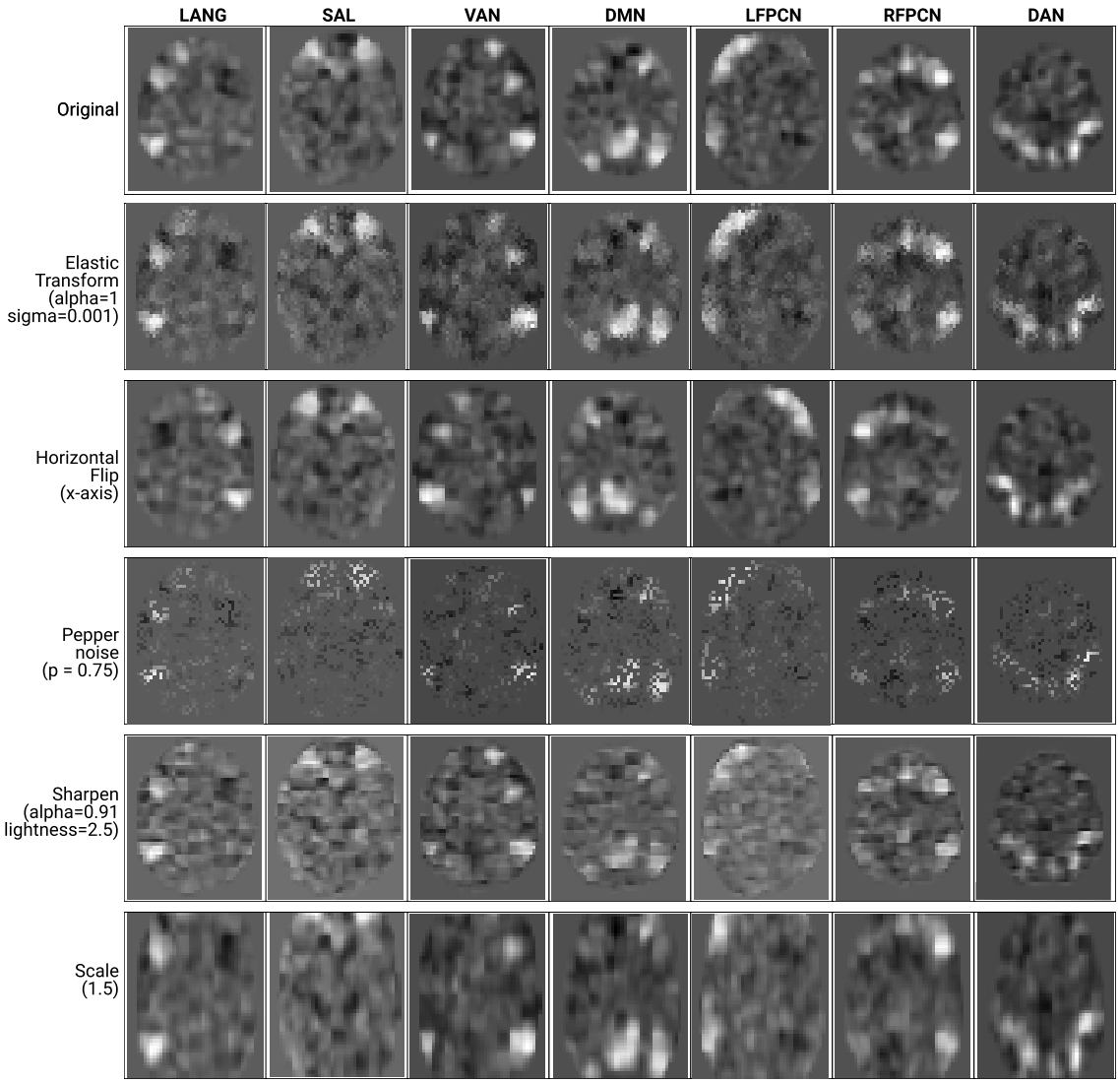


Figure 2.3 – Visualization of fMRI data augmentation using clinical noise simulation. Functional networks are represented column-wise. 1^{st} row represent original fMRI data while while 2^{nd} through 6^{th} represent the different augmentation options.

activation networks, was achieved in two ways. First, we computed a spatial stretch on the healthy fMRI network images similar to the effect of a brain tumor on the area within and about $3-5px$ range around the region of the lesion mask (see Figure 2.6). A classical filter known as pinch-explode was used for this purpose (see Figure 2.5). Secondly, we introduced a randomly generated 3D lesion mask. The lesion masks were chosen with a radius of $0-10px$ across the 10^{th} to 32^{nd} channels of our image data with dimension $42px \times 51px \times 34channels$ comparable to real tumor masks as shown

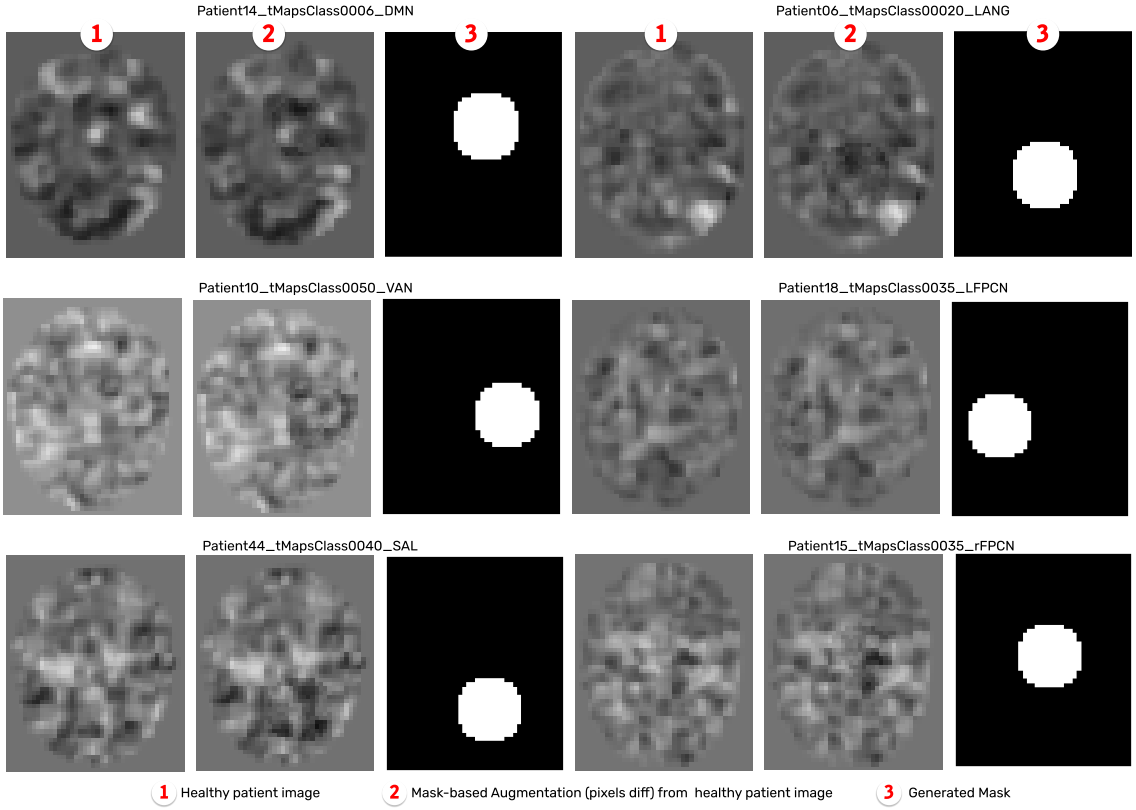


Figure 2.4 – Data augmentation via generated mask for pixel difference compensation.

in Figure 2.7. With such a signal void, we turned the image voxels of the brain tumor region using our masks into zero values, i.e., no signal, to mimic an observe drop in mean pixel intensity inside the tumor. In both data augmentation ways, the input images were healthy patients. The transformation were chosen (stretch and signal-void) to simulate the expected impact of the tumor on the fMRI signal. In this spirit, data augmentation is another form of transfer learning from healthy to unhealthy patients to be compared with the other transfer learning approaches of the previous sections.

2.5 Transfer learning strategy

In this machine learning strategy, a model that has been trained for one task is repurposed for a different, related task. The idea is that, some aspects of the problem, such as low-level features, are shared across tasks, and therefore can be leveraged to improve the performance of a model on the target task.

Transfer learning is particularly useful in this scenarios since the amount of un-

Algorithm 1: fMRI 3D image augmentation

Data: 3D image
Result: augmented 3D image
image data object ← load nibabel library;
width, height, channel ← image dimension;
$A \leftarrow$ augmentation parameter;
$B_{i,j} \leftarrow$ image pixel;
$\hat{\mathbb{X}} \leftarrow$ augmented image pixel;
for *each i in channel* **do**
  for *each j in height* **do**
    for *each k in width* **do**
      if $A_{i,j} \geqslant 0$ **then**
        apply $B_{i,j} \leftarrow A$; `// apply desired image augmentation`
        $\hat{\mathbb{X}}_{i,j} \leftarrow B_{i,j}$; `// return augmented image`
      end
      else
        $\hat{\mathbb{X}}_{i,j} \leftarrow B_{i,j}$
      end
    end
  end
end
return $\leftarrow \hat{\mathbb{X}}$; `// return augmented 3D image.`

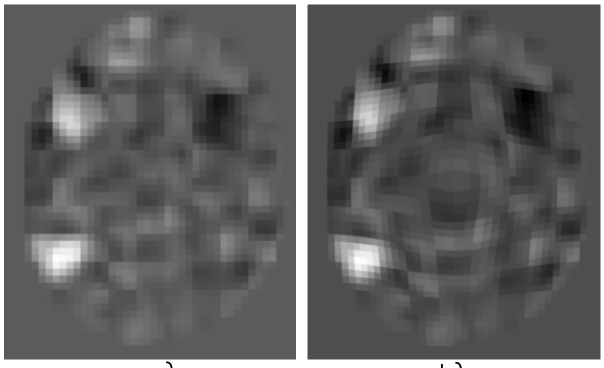


Figure 2.5 – LANG Network a) with Pinch-explode augmentation b).

healthy rs-fMRI data available for the target task is limited, but there is a larger amount of healthy rs-fMRI data available for the related functional brain network classification task. In these cases, the model can be pre-trained on the source healthy data, and then fine-tuned on the target task data. The pre-training step allows the model to learn

Algorithm 2: Pepper noise augmentation for 3D images

Data: 3D image
Result: 3D peppered noise image
image data object ← load nibabel library;
width, height, channel ← image dimension;
$\mathbb{M}$ ← new generated 3D image;
A ← augmentation parameter;
$B_{i,j}$ ← image pixel;
$\hat{\mathbb{X}}_{i,j}$ ← augmented image pixel;
for *each i in height* **do**
  for *each j in width* **do**
    let r ← random number ; // random application of noise
    if $r < x/2$ **then**
      $\mathbb{M}_{i,j}$ ← pepper noise ; // apply black pixels (pepper noise)
    end
    else
      $\mathbb{M}_{i,j} \leftarrow X_{i,j}$
    end
  end
end
return ← $\mathbb{M}$; // return image with pepper noise.

Table 2.2 – fMRI Image augmentation options for clinical noise simulation and variants.

Type Code	Description	Augmented variants
O	Original image	-
A1	Elastic transform	9
A2	Flip	9
A3	Pepper noise	9
A4	Sharpen	9
A5	Scale	9

useful representations of the data, which can then be fine-tuned to the target task.

The best model from the previous section was then investigated in its capability to transfer to unhealthy patients. We explored three main transfer learning techniques: brute transfer, mix transfer, and weight transfer. These techniques allow our test on unhealthy data to be better identified by some knowledge from healthy data and even simulated (augmented) data. In the brute transfer, a model was entirely trained on

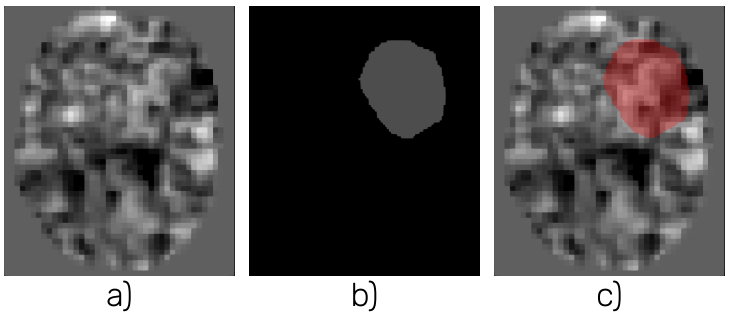


Figure 2.6 – LANG Network a), Lesion mask b) and mask overlay on functional network activation c).

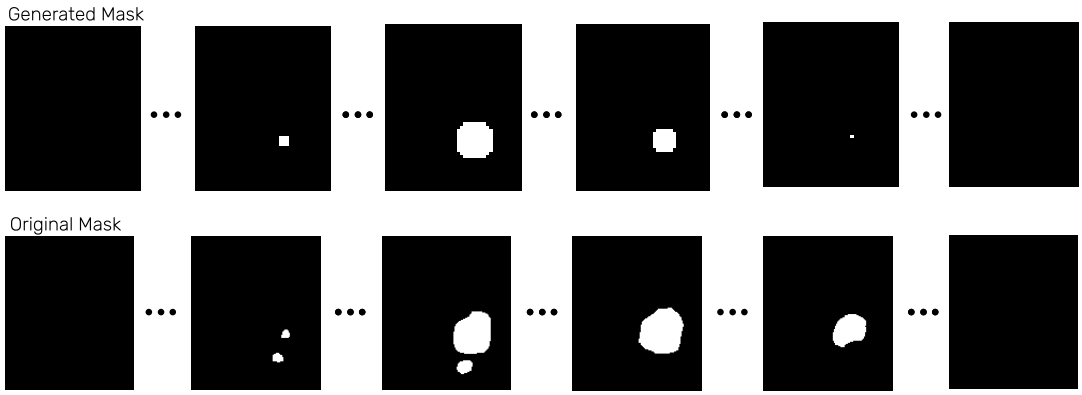


Figure 2.7 – Cross view between synthetic and original mask over 34 fMRI image channels.

data from healthy control and directly evaluated with unhealthy data, while in the mix transfer, the training database contained some unhealthy data. For the weight transfer method, our saved model weights from healthy data were loaded for further training and fine-tuning with unhealthy patient data. We tested the model with unseen unhealthy data (patients with tumors). We trained all transfer learning models at a learning rate of $1e-5$ with $500-1000$ epochs. To minimize over-fitting, we used an early stopping method based on the validation error increase. A grid-search algorithm chose optimal hyperparameters for the CNN model based on maximized precision of the training data: the stopping points for network training were ten validation failures followed by a model checkpoint.

2.6 Results and discussion

We implemented both shallow and deep learning algorithms described above, and all observed results were recorded for further evaluation. In this section, we compared the performance of several machine learning techniques to find the best baseline method which was used in the next phase for our transfer learning experiments. Finally, we compare our result with the closely related literature.

2.6.1 Performance comparisons

The comparison of the different algorithms in Table 2.3 identified the proposed CNN model as the most efficient approach for identifying the functional networks of interest on both healthy subjects and unhealthy patients data. In addition to the comparison presented in Table 2.3, we extended our effort to implement other well-known CNN architectures like VGG16, ResNet, and DenseNet on our dataset. However, the performance of these models was recorded in the range of 50% to 55% on healthy data and, therefore, perceived to be unreliable. The observed difficulty was in the dimension of the original images and the total number of images in our dataset. The typical image size for well-known CNN architectures for computer vision (like VGG16, ResNet, and DenseNet) is considered to be at $224 pixels \times 224 pixels$ as they are mainly designed to work on the ImageNet database [86]. Our original images are in multi-channel format and therefore have a size of $42 pixels \times 51 pixels \times 34$ (width, height, channel). In order to adjust the image size, a bi-cubic interpolation has been used to up-sample image size by a factor of 4. This up-sampling reduced the quality of images and caused a significant drop in the performance of the models. On the other hand, the number of training images is much lower than the number of image samples in well-known CNN architecture, which led the model to over-fit and reduced the model performance.

Following our deep learning model implementation, we recorded the classification accuracy as shown in Table 2.4. This results provides better insights on the possibility of strengthening the robustness of our model by allowing the use of data variants. With the accuracy of 0.89 ± 0.01 in tMaps image version of our healthy subject data, it is clear that more data samples could improve our result. We can also observe that some of our augmentation strategies did not increase our result in any way which indicates that further observation is in fact required to understand and quantify the amount of

Table 2.3 – fMRI brain network classification results of various supervised machine learning techniques with Healthy subjects data.

	Classification Techniques	**Healthy Data**
1	Proposed CNN	0.86 ± 0.01
2	Random Forest	0.82 ± 0.01
3	Feed forward NN	0.84 ± 0.02
4	Naïve Bayesian classifier	0.45 ± 0.02
5	K-Nearest neighbors	0.83 ± 0.02
6	Support vector machine	0.83 ± 0.01
7	Classification tree	0.64 ± 0.06

influence each augmentation option has in model improvement.

2.6.2 Transfer learning

We selected the best method identified in Table 2.3 for healthy data and conducted the transfer learning approaches on this method with data from unhealthy patients. The results in Table 2.5 show the recorded accuracy values for several experiments on the proposed CNN model. Each defines the data used for training and testing with their respective data sizes. It has to be mentioned that the trained model never sees the testing data, neither during the training process, nor the hyper-parameters tuning process.

Table 2.4 – fMRI brain network classification results with CNN using healthy data augmentation.

	Data Organisation		**tMaps images**		**Threshold**	
		Size	**Loss**	**Accuracy**	**Loss**	**Accuracy**
1	Original	567	0.67 ± 0.03	0.86 ± 0.01	0.70 ± 0.03	0.86 ± 0.01
2	Original + A1	5530	0.65 ± 0.06	0.88 ± 0.01	0.72 ± 0.06	0.88 ± 0.02
3	Original + A2	5530	0.76 ± 0.08	0.81 ± 0.02	0.69 ± 0.04	0.84 ± 0.02
4	Original + A3	5530	1.22 ± 0.14	0.85 ± 0.03	0.86 ± 0.09	0.87 ± 0.01
5	Original + A4	5530	0.51 ± 0.05	0.87 ± 0.02	0.48 ± 0.02	0.87 ± 0.01
6	Original + A5	5530	0.68 ± 0.05	0.86 ± 0.02	0.67 ± 0.04	0.88 ± 0.01
7	Original +A1+A3+A4+A5	5530	0.81 ± 0.06	0.89 ± 0.01	0.86 ± 0.07	0.88 ± 0.01

Several baseline experiments were conducted to assess the other added value of transfer learning approaches. First, we trained on healthy control data and tested on healthy control. This experiment provided an upper bound of performance with the

highest accuracy of 86%. This high score is possibly also due to the expected higher homogeneity of healthy control. The same experiment was carried out while training unhealthy and testing unhealthy patients. A drop of about 10% of accuracy was observed, which builds a second baseline with fewer patients. The investigated transfer learning approaches were expected to provide performances between these two bounds. We considered four transfer learning strategies for this experiment (i) brute transfer (training on healthy and testing on unhealthy data), (ii) mixed transfer (adding some unhealthy data to healthy data to train the model), (iii) weight transfer (fine-tuning on unhealthy data), and (iv) transfer learning with data augmentation.

On the brute transfer strategy, as indicated in Table 2.5 row 3, we trained our model with 81 healthy control subjects and conducted testing on all 55 unhealthy patients. We recorded an average accuracy of 0.74 ± 0.01 for all test data size ranges. The brute transfer is therefore not bringing any improvement here. For the mix transfer strategy, Table 2.5 row 4, we trained our model with 81 healthy control subjects and unhealthy patients (45). At the same time, we performed our model test with ten unhealthy patients. An improvement in accuracy to 0.77 ± 0.01 on test data was observed by comparison with the brute transfer. The addition of data helps, even with a mixture of healthy and unhealthy patients by comparison with pure unhealthy patients experiment of row 1. However, we do not reach the upper bound performance of row 1 despite having more data than in this experiment. This performance demonstrates a discrepancy between healthy and unhealthy patients. Figure 2.8 shows the validation accuracy (from validation data) of the trained model on healthy data for various amount of added unhealthy patients $(10, 20, 30, 45)$. We recorded $\cong 1\%$ increase in validation accuracy for every ten unhealthy patient data added to training data(7 functional network images per patient). As the third transfer learning strategy, in Table 2.5 row 5, we transferred the weight and bias of a model fully trained on healthy data (model of row 1) to a model for training on unhealthy data. The model was retrained and fine-tuned on 45 unhealthy patients and tested on 10 remaining patients. Performance of 0.78 ± 0.01 is obtained on unhealthy test data. This result is the highest performance among all tested transfer learning strategies. The three transfer learning strategies were repeated in the presence of augmented data (see Table 2.5 rows 6 to 10). Augmented data was produced by data augmentation techniques (see section 2.4) from healthy data to simulate unhealthy data. The recorded performances in these experiments remained in the same range as other transfer learning approaches.

Table 2.5 – fMRI network classification results for healthy and unhealthy data (patients counted as whole represent 7 fMRI network images in each case).

	Training		Testing		
SN	**Data description**	**Patient data**	**Data description**	**Patient data**	**Accuracy**
1	Healthy	71	Healthy	10	0.86 ± 0.02
2	Unhealthy	45	Unhealthy	10	0.75 ± 0.01
3	Healthy	81	Unhealthy	55	0.74 ± 0.01
4	Healthy + Unhealthy	81 + 45	Unhealthy	10	0.77 ± 0.01
5	Fine-tuning on Unhealthy from Healthy	45	Unhealthy	10	0.78 ± 0.01
6	Augmentation (unhealthy simulation)	81	Unhealthy	10	0.75 ± 0.01
7	Healthy + Augmentation (unhealthy simulation)	81 + 81	Unhealthy	10	0.73 ± 0.01
8	Fine-Tuning on Unhealthy from healthy + Signal void	45	Unhealthy	10	0.76 ± 0.00
9	Healthy + Signal Void	81	Unhealthy	10	0.73 ± 0.01
10	Signal Void + Unhealthy	45	Unhealthy	10	0.74 ± 0.02

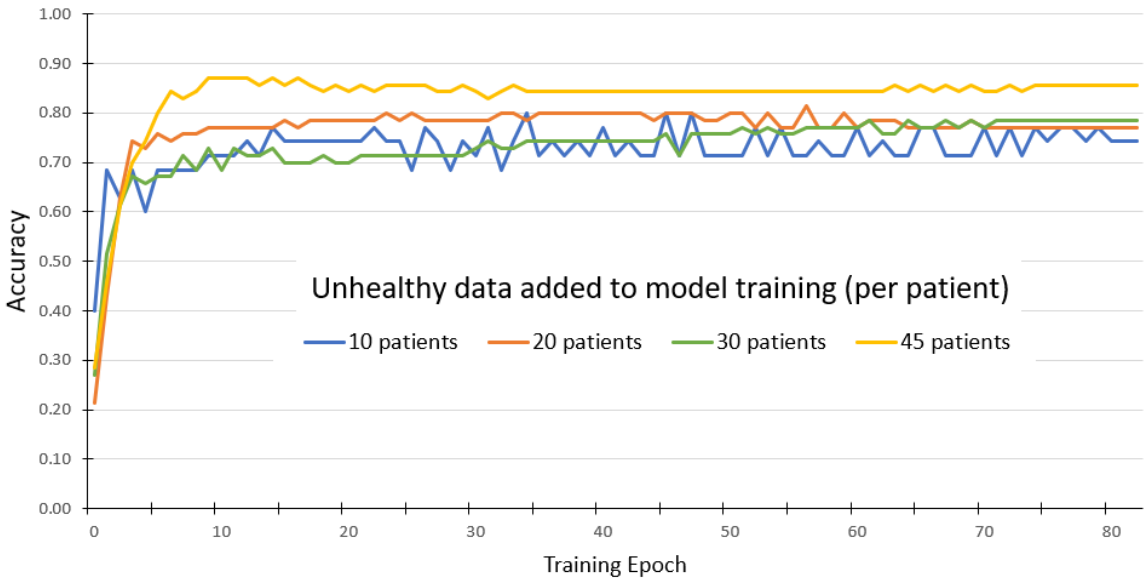


Figure 2.8 – Validation accuracy curve of unhealthy patients data added to model training with healthy subject data.

2.6.3 Comparison with prior works

As closely related work, Mitchell *et al.* [57] focus on identifying selected functional networks in 21 healthy volunteers by training a simple feed-forward neural network model. This approach was achieved using a Multilayer Perceptron (MLP), which usually follows the procedure of hand-crafted features extracted from data. Generally,

Multilayer Perceptrons (MLP) are fully connected neural networks which generate outputs based on inputs. Literature sometimes uses MLP interchangeably with Deep Neural Network (DNN); However, there is a sharp contrast because, MLP is a subset of DNN. In this case, there is a pre-selection of ICs of interest. Our ICs were generated using a bottom-up, data-driven approach using an independent component analysis. ICA has gained popularity as one of the two frequently selected analytical methods for rs-fMRI data, which requires no seed on any predefined region [87, 88]. In contrast, ICs generated in Mitchell *et al.* study used canonical seed regions of interest scattered across the brain. These two approaches may provide similar features for further analysis. However, hand-crafted feature extraction can limit the flexibility and potential of identifying certain functional brain areas, as demonstrated in our approach. In addition, the location of the seed regions could significantly impact the resulting pattern of a functional system like the Language network. Furthermore, sensitivity to systematic noise like head movement and physiological nuisance signals causes false identification of non-language areas (false positive) and false detection of putative language areas (false negative), which limits the clinical application of seed-based rs-fMRI in language mapping [65]. The comparison of our proposed CNN performed in the same conditions as Mitchell's method [57] is given in Table 2.6, and this further demonstrates the interest of our approach.

Table 2.6 – Comparison of our proposed end-to-end deep learning model with related approach.

	Proposed CNN	T. J. Mitchell et al.
Training: Healthy – Test: Healthy	0.86 ± 0.02	0.84 ± 0.01
Training: Unhealthy – Test: Unhealthy	0.75 ± 0.01	0.72 ± 0.01
Transfer Learning from Healthy to Unhealthy	0.74 ± 0.01	0.71 ± 0.01

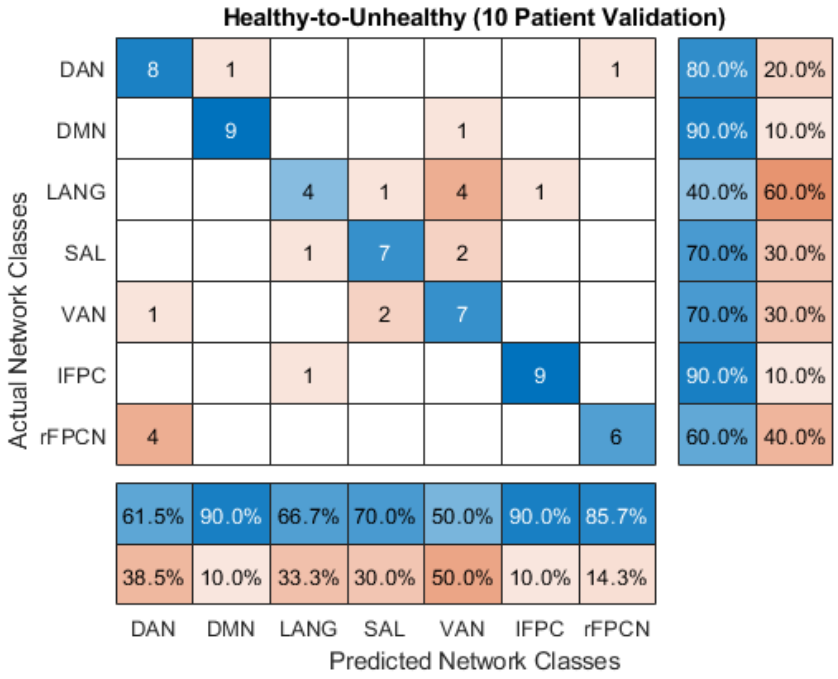
2.6.4 Error analysis

The results of this study indicate that, healthy control can help to boost the functional network identification for unhealthy patient data by adding the healthy data during the training process. In this section, we discuss the observed errors and further analyze the origin of the transferability between healthy to unhealthy data.

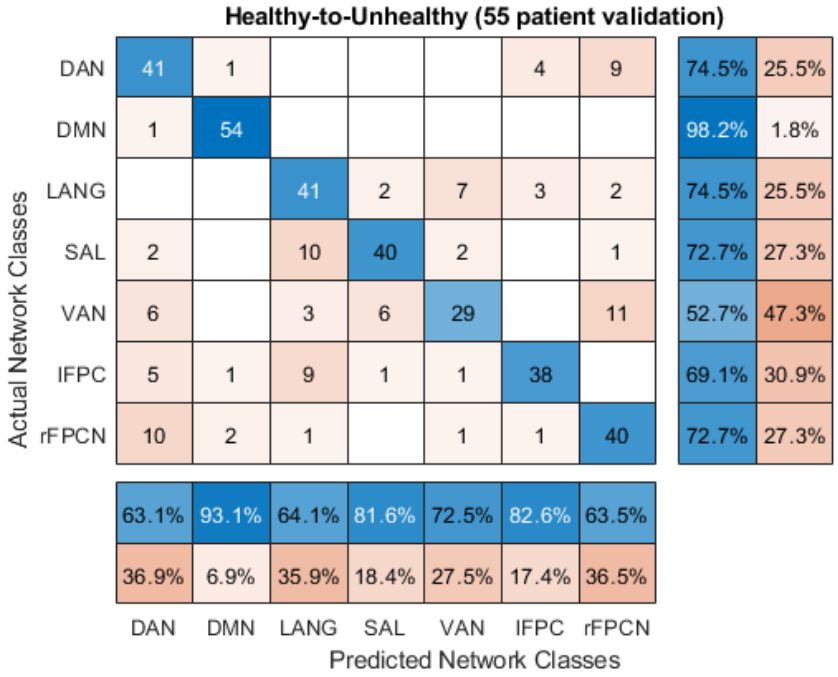
One may wonder "where did the classification errors in this experiment can come from ?". To reveal this, we generated the confusion matrix (see Figure 2.9) as well as

sensitivity (true positive rate) and specificity (true negative rate) of the classification individual functional brain networks to discover the most sensitive cases. Table 2.7 shows model evaluation of each individual networks for classification of healthy subjects, unhealthy patients and transfer learning respectively. The primary source of confusion between the different functional networks is the spatial overlap between the activated areas. We segmented the functional network identification into classification steps, identifying in each of them between the 55 ICs the best fitted ICs for all seven functional networks. We realized that the main sources of error came from the confusion between LANG and the VAN as well as DAN and rFPCN as shown in Figure 2.9. The difficulty in differentiating between DAN and rFPCN may be explained by the spatial overlapping between the two networks[89]. In contrast, the relationship between VAN and LANG networks is more complex than in other networks. The distinction between the language and ventral attentional networks in rs-fMRI may be difficult, as they present similar activations in the ventrolateral prefrontal cortex, inferior frontal cortex, and temporal gyrus in right-handed patients [90]. However, slight differences in the activation may allow for discrimination between these two networks in the inferior parietal lobule, in which the activation is more anterior, located in the temporoparietal junction and the supramarginal gyrus for the attentional network, and more posterior in the angular gyrus for the language network [90, 91, 6]. The ventral attentional network is also located in the non-dominant hemisphere, almost symmetrical to the language network in the dominant hemisphere, which may also explain the difficulties of discriminating between these two networks. Considering the lateralization of these two networks, the handedness assessment using the Edinburgh handedness inventory has been considered as a supplement to discriminate between ventral attentional and language networks [6]. However, while this information may be useful in right-handed patients where the left-hemisphere dominance exists in 96% of patients. Left-handed patients should be considered with caution since only 27% of left-handed patients have a dominant right hemisphere, and therefore, a left-lateralized ventral attentional network [92].

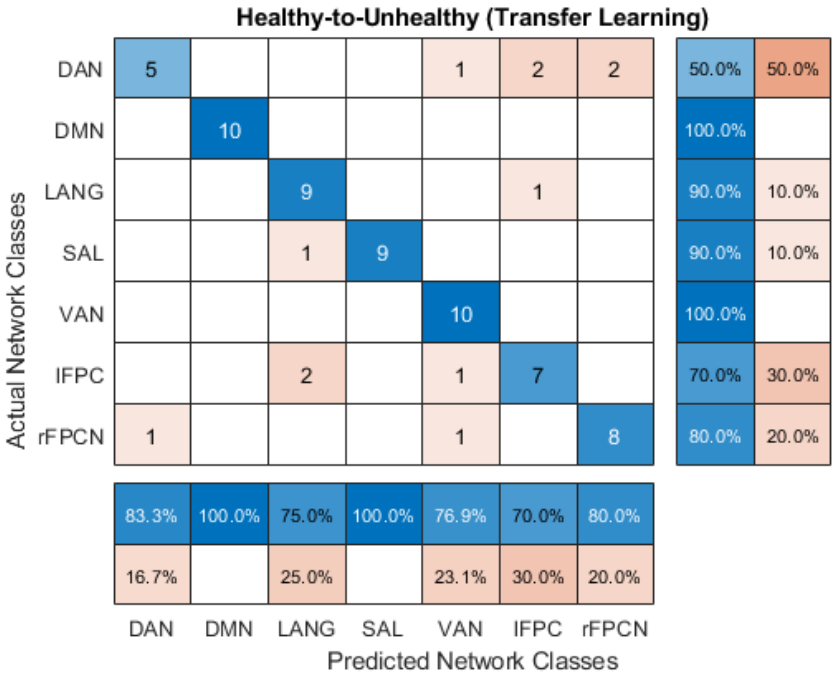
We investigated the possible overlapping surface of thresholded functional networks and the lesion mask in unhealthy patients to understand better the possibility of transfer from healthy to unhealthy data. The distribution of intersection over union (IoU) values of 3D binary images of all unhealthy patients data for correct and wrong classification. Most of the thresholded functional networks have little or no overlap



(a) Confusion matrix of brute transfer learning from healthy to unhealthy data.



(b) Confusion matrix of brute training on healthy and validation with all unhealthy data.



(c) Confusion matrix of weight transfer learning from healthy to unhealthy data.

Figure 2.9 – Confusion matrix of functional network prediction by proposed CNN model on all individual functional brain networks as classes LANG, SAL, VAN, DMN, lFPCN, rFPCN, and DAN.

Table 2.7 – Model performance evaluation for each functional networks of healthy subjects, unhealthy patients and transfer learning (Healthy to unhealthy) approach.

Networks	**Healthy Subjects**		**Unhealthy Patients**		**Healthy-to-Unhealthy**	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
DMN	1.00	1.00	0.97	1.00	0.98	0.90
LANG	0.98	0.70	0.97	0.80	0.97	0.60
LFPCN	1.00	1.00	0.95	0.70	0.98	0.90
RFPCN	0.98	0.90	0.95	0.80	0.98	0.70
VAN	0.95	0.90	0.95	0.80	0.88	0.80
DAN	1.00	0.90	1.00	0.60	0.92	0.80
SAL	0.98	1.00	0.98	0.97	0.95	0.8

with the lesion mask. The normalized versions of these histograms are provided in Figure 2.10.

In our statistical evaluation, the two distributions were observed to be highly-skewed values of 2.11 and 1.94 for IoU of correctly and wrongly classified images, respectively, indicating non-Gaussian distribution. These histograms show that the

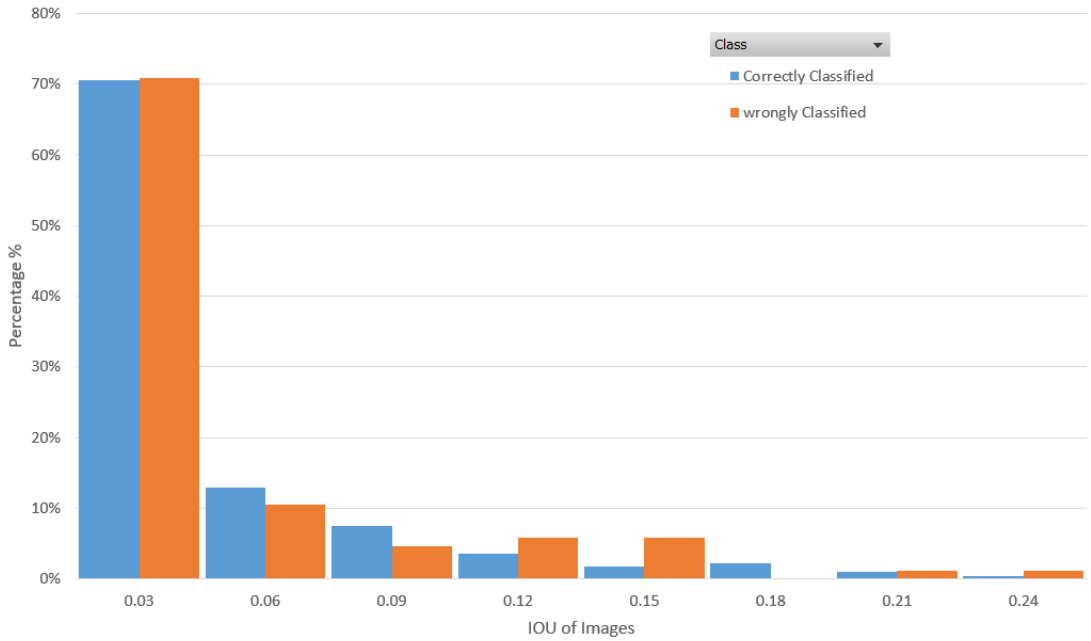


Figure 2.10 – Normalized histogram of IoU of functional network activation and lesion mask of unhealthy subject data.

category (correctly classified or wrongly classified) are estimated to be equal across the different IoU values as also confirmed by the p-value of 0.75 in the t-test carried out from the IoU distribution, which indicates non-significance (>0.05) in a difference between the two categories. To qualitatively illustrate this statistical fact, Figure 2.11 provides a scenario where images with or without overlap are correctly or wrongly classified. No direct effect of the tumor on the thresholded functional networks targeted is observed in our dataset. This observation can explain the possibility of transfer learning from healthy to unhealthy data. Nonetheless, we found a useful but not perfect transferability, and therefore, a discrepancy exist. This could be in the intrinsic shape of the functional network of unhealthy patients, which may be distorted when located in the vicinity of the tumor. The further investigation of this observed discrepancy to establish a reliable relationship between healthy and unhealthy data is provided in chapter 4.

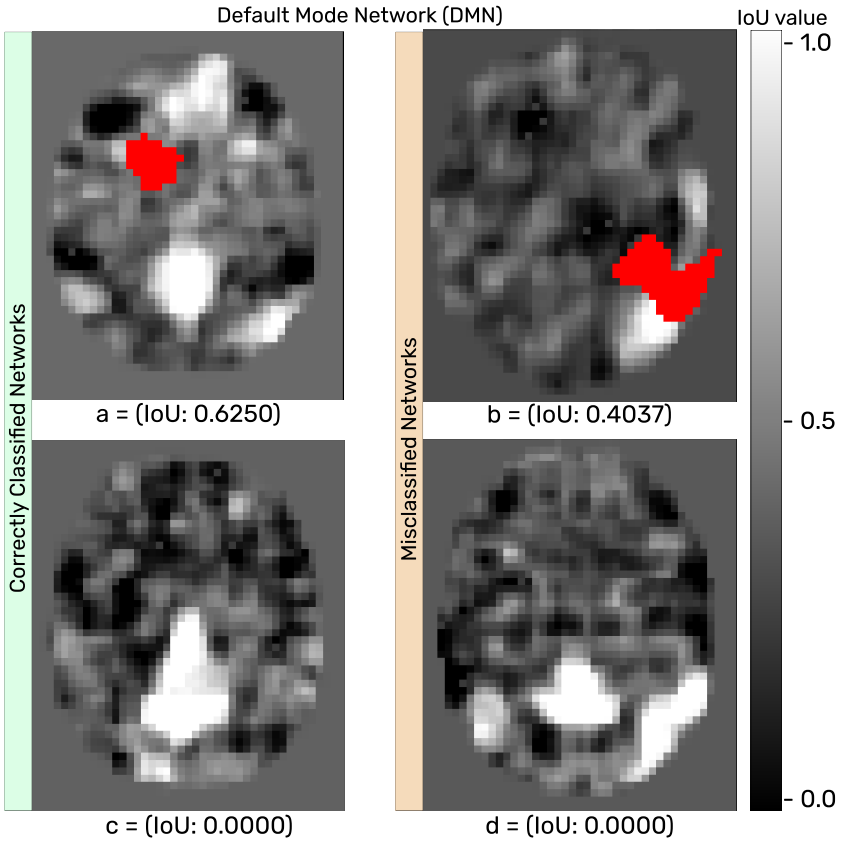


Figure 2.11 – Visualization of correctly and wrongly classified images (with and without Overlap)- a is correctly classified with overlap; b is wrongly classified with overlap; c is correctly classified without overlap; d is wrongly classified without overlap

2.7 Conclusion and perspective

This work demonstrates the interesting possibility of transfer learning from healthy control to unhealthy patient data. This was illustrated for the automatic identification of functional brain networks in rs-fMRI for patient with brain tumors. This result is important because, it opens an easy way to overcome the lack of data in machine learning for biomedical imaging. We demonstrated that healthy control data could boost the classification of functional brain networks in rs-fMRI for patient with brain tumors. This was obtained with an optimized classical CNN, which was shown to outperform standard CNN architectures and shallow learning methods, including the one previously tested in the literature on healthy subjects. The overall best performance obtained with unhealthy patients after transfer learning was 0.78%. The remaining errors where found to be indeed corresponding to difficult cases. The gain brought by the

transfer from healthy subjects was about 4%, which is a classical order of magnitude in transfer learning. These performances remain smaller than the best performance obtained only on healthy control subjects (0.86%). Brain tumors make the classification harder than in healthy subjects; nonetheless, the knowledge gained from healthy control subjects can help classify functional brain networks in rs-fMRI with unhealthy patients. It is, therefore, an interesting result since healthy control subjects can be easily enrolled for data acquisition through the non-invasive rs-fMRI studies.

The limiting factor in transferring knowledge from healthy to unhealthy may be the discrepancy between healthy control and unhealthy patients, which occur due to the influence of tumor on a region of the functional brain network. Several paths to compensate for this discrepancy could be investigated. Style transfer from healthy to unhealthy could be investigated to perform this compensation in the image domain. Also, one could consider domain adaptation in the neural network to operate this shift in the latent space rather than in the image. Lastly, one could also consider the preprocessing image approach to compensate in the image domain for the distortion (spatial deformation, bold signal attenuation, ...) brought by the tumors in the images.

In this chapter, we proposed an end-to-end deep learning method for functional brain network identification and demonstrated the possibility of transfer of knowledge from healthy to unhealthy patients. This initial approach aim to compensate for limited availability of unhealthy dataset by feature transfer from healthy data to improve model prediction on unhealthy data. In effort to achieve this, we acquired more healthy data which is relatively easier in this non-invasive fMRI approach however, annotation requirement is a bottleneck. This observation motivates us to further investigate the use of self-supervision learning to avoid healthy data annotation since it constitutes significant time loss for the clinician, and has no benefit in the tumor resection procedure. This methodology can be extended to address other biomedical imaging problems for the production of large cohort is essential to improve the deep learning accuracy. As a possible improvement, this work still rely on manual data annotation by a clinical expert which can be complex and time consuming process.

CHAPTER 3

SELF-SUPERVISED LEARNING WITH FMRI DATA

This chapter describes a follow-up study on our deep learning model for functional brain network identification to propose a methodology for learning from healthy data to improve functional brain network recognition accuracy in unhealthy data without the need of annotating them, using contrastive supervision learning approach. This method will allow the possibility to acquire more healthy data through the non-invasive medical imaging modality as well as exploit the similarity between healthy and unhealthy data. This chapter is based on a publication titled "Self-Supervised Learning for Functional Brain Networks identification in fMRI from Healthy to Unhealthy Patients".

3.1 Introduction

Although, the recorded progress in computer vision driven medical imaging has been well related to the possibility of using labeled data to train machine learning model (supervised machine learning approaches) [93]. However, this supervised learning technique poses some common limitations. One of these limitations is the usual lack of large annotated data sets which may be because it is hard to acquire large data of rare disease, or because the international community maintain limited distribution of public dataset, or because human expertise for the annotation of the dataset is limited.

There are several workarounds to compensate for the limited availability of dataset [94, 95]. These include few-shot learning, creation of artificial data, generative models, or data enhancement. Transfer learning, another common strategy, makes use of models that have already been trained on comparable dataset. As demonstrated in a recent work [34], which propose to use such transfer learning approach from healthy subjects

to unhealthy patients. This is very interesting approach for medical imaging modalities which are purely non-invasive and therefore for which it is rather easy to enroll healthy control. This proposal was illustrated for functional brain network identification task in resting-state functional magnetic resonance imaging (rs-fMRI) data. Initial observation suggest that the gain in classification performance recorded was because, the brain tumor of the unhealthy patients have very limited impact on the rs-fMRI signals hence, transferability of useful features was possible.

Transfer learning approach as illustrated in Figure 3.1, allows a mode to learn the features from healthy subject data and perform significantly better in unhealthy data prediction. One common limitation as observed in this approach with transfer learning from healthy patients to unhealthy patients [34], is the need of manual annotation of the healthy patients data. This annotation process is time consuming while it has to be performed on patients for which there is clearly no clinical interest. To avoid this unnecessary step while trying to take benefit from the similarity between healthy subjects and unhealthy patients by transfer learning, we propose to investigate the possibility of self-supervision and use healthy data without the need of annotating them.

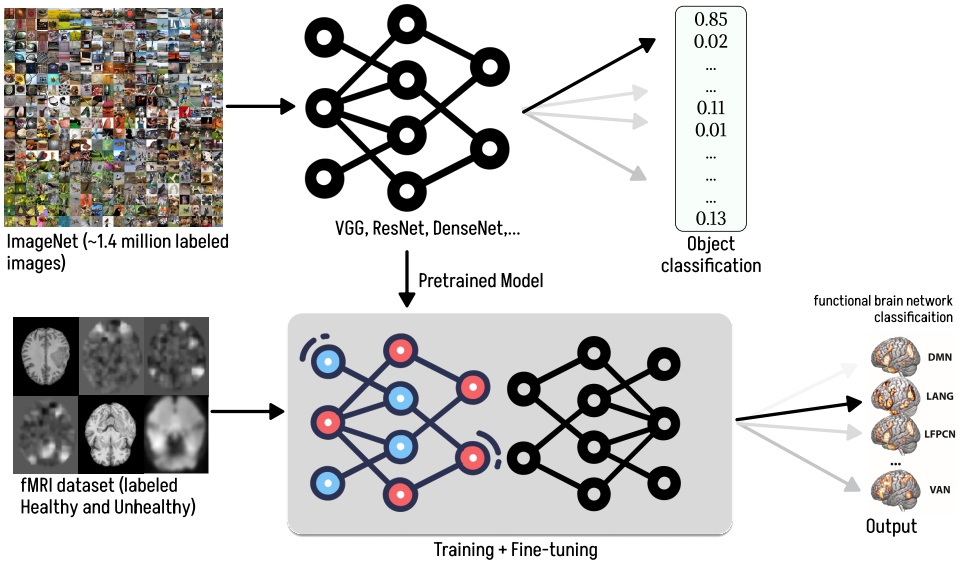


Figure 3.1 – Visual description of transfer learning strategy.

Self-supervision is an unsupervised learning approach, where machine learning models are trained with unlabeled data for a pre-defined task to allow the model learn useful information in the data and can be used for further prediction. Self-supervision

learning allows the understanding of underlying pattern in unlabeled sample data [96]. It can be regarded as an intermediate form between supervised and unsupervised learning. It is usually based on an artificial neural networks. The training of the network is performed in two stages. First, a pretext task is solved to earn pseudo-labels which contributes to initialize the network weights. Secondly, the target task is performed with supervised learning but with much fewer need of data annotation due to the initialisation from the weights trained on the pretext task.

In the following sections, we describe a methodology based on self-supervised machine learning for avoiding healthy data annotation. We demonstrate the potential of this approach using resting-state fMRI data for functional brain network classification.

3.1.1 Application of self-supervision in medical imaging

In medical imaging, self-supervision can be applied to various tasks such as image reconstruction [97], denoising [98], segmentation [99], and registration [100]. One of the benefits of self-supervision in medical imaging is that it can significantly reduce the dependence on annotated medical data, which is often subjective to availability of expert annotator. Self-supervision allows the models to learn underlying representations from the medical data by avoiding data labeling or uses a small database of annotated images [101]. This makes self-supervision an attractive solution for medical imaging problems, where obtaining annotations is often difficult or expensive.

Self-supervision technique is also used in medical imaging modalities, such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound images [101]. In MRI, self-supervision has been used to reconstruct images from undersampled data, reduce noise, and improve the accuracy of segmentation algorithms [97]. In CT, self-supervision has been used to improve the accuracy of image registration and denoising algorithms [102]. In ultrasound images, self-supervision has been used to improve the accuracy of image segmentation and registration algorithms [103].

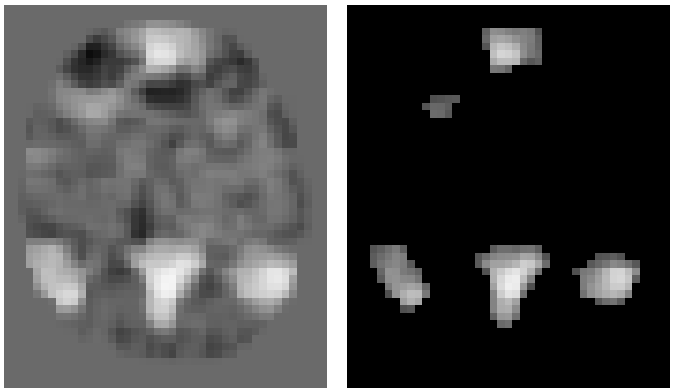
In functional magnetic resonance imaging (fMRI), self-supervision has been used to learn representations of functional connectivity patterns in the brain, without the need for explicit annotations [104]. These learned representations can be used for a variety of tasks, such as the classification of patients into different diagnostic groups, the prediction of clinical outcomes, and the identification of biomarkers for neurological and psychiatric disorders. Some interesting application of self-supervision includes; de-

noising fMRI data, self-supervision has been used to remove noise and artifacts from the data, while preserving the underlying functional connectivity patterns. In the reconstruction of functional connectivity patterns, self-supervision has been used to estimate the functional connectivity patterns from undersampled fMRI data, without the need for explicit annotations. In the classification of patients into different diagnostic groups, self-supervision has been used to learn representations of functional connectivity patterns in the brain, which can be used to discriminate between patients with different neurological and psychiatric disorders. Self-supervision is now applied in all fields of computer vision but recently, began to receive consideration for fMRI related data [105, 106]. In [105] a regression task to predict the fatigue from patient based on their fMRI patterns is targeted. In [106], the authors proposed a transformer framework which uses a two-phase training approach where the model is first trained to reconstruct 3D volume data using self-supervised training and then fine-tuned on specific tasks using ground truth labels, for various fMRI tasks such as age and gender prediction and schizophrenia recognition. In this study, we explore the strategy of self-supervision to avoid healthy data annotation in feature transfer process from healthy to unhealthy data as illustrated in the use case of [34] where functional brain networks have to be identified from rs-fMRI images.

Our focus to tackle this prevailing limitation in the use of resting-state fMRI for functional brain network identification is very crucial for scalability of our model. This is because, manual annotation of healthy data is not practical when dealing with large amounts of data. Furthermore, this process has no clinical relevance in brain tumor removal procedure. To specify the extend for which data annotation consumes the time of the clinician, it generally takes $\approx$10 minutes to annotate data of a healthy control subject or volunteer and $\approx$15 minutes to annotate unhealthy patient data as described by the clinical expert who conducted the rs-fMRI data preprocessing. It is understood that, unhealthy data annotation takes more time to complete because of the extra care needed to understand the regional displacements due to the presence of tumor in the brain. To further standardize the approach of resting-state fMRI for brain tumor removal procedure, we propose to contribute through self-supervision learning technique, to avoid the tedious and time consuming healthy data annotation which is routinely done by clinicians manually. As an initial step, we start by discussing details of our database and how we approach this investigation.

3.2 Materials and methods

At this stage, we adopt the described procedure of resting-state fMRI signals acquisition and preprocessing as provided in section 2.2. Using the selected key networks of LANG, DMN, SAL, VAN, lFPCN, DAN, and rFPCN from the 55 produced ICs for each patient based on fMRI spatial distribution and activation peaks of these activations. We also process two versions of the annotated images: complete gray level images (connectivity map) and equivalent thresholded image copy. Figure 3.2 illustrates a DMN network image sample from unhealthy group show both gray level and thresholded version. At the cluster level, individual spatial components were thresholded at z = 2, corresponding to the 5% most active voxels in each intrinsic connection network. This approach is consistent with the literature and allows for the identification of the anatomical location of activated brain regions despite background clinical noise [107]. Following the data acquisition stage, we perform random visual examination on the acquired data to better understand the similarity relationship.



(a) Gray level image (b) Thresholded image

Figure 3.2 – Visualization of healthy fMRI image data variants with example from Default Mode Network (DMN).

Based on visual observation of the acquired dataset of rs-fMRI images in Figure 3.3. One can suggest that, although the influence of brain tumor result to displacement and distortion of the functional brain activation volume, it is indeed non-visible in unhealthy patients data. This is interesting because, it justify the reason for the observed transferability as demonstrated using transfer learning technique in chapter 2 of this document.

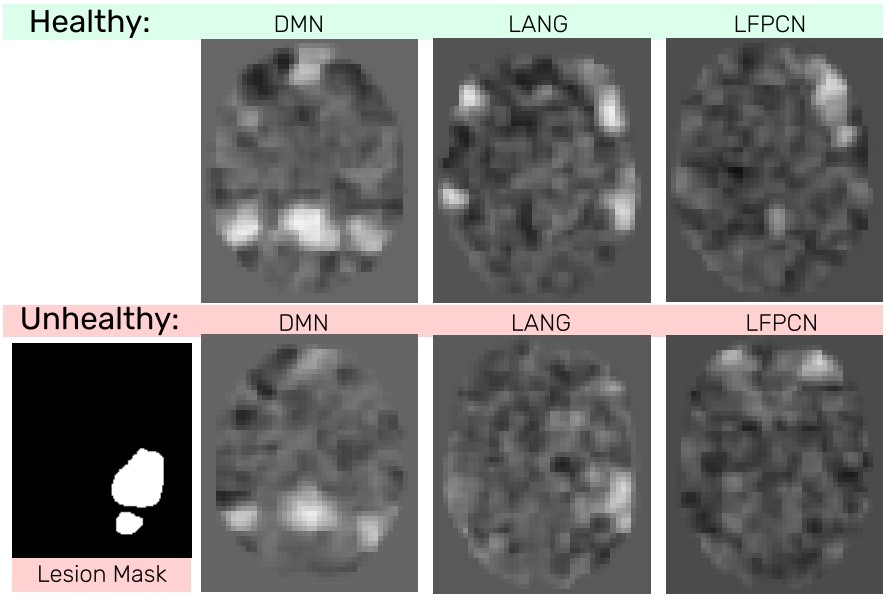


Figure 3.3 – Visual observation of sample image slices from selected functional networks in unhealthy and unhealthy patient data.

3.3 Self-Supervision learning for fMRI image classification

In recent years, a number of self-supervised learning algorithms have been proposed for image recognition and classification tasks [108]. The literature on self-supervised learning for fMRI image classification is growing rapidly, and a number of recent studies have reported promising results. For example, some studies have demonstrated that self-supervised learning algorithms can outperform traditional supervised learning algorithms in terms of accuracy and generalization performance. Other studies have shown that self-supervised learning algorithms can be effective in learning complex representations of fMRI images [109], which can be used for a variety of tasks, including classification, segmentation, and regression. While all these techniques represent a clear demonstration of learning good representation from comparative data, it inspires our adaptation to propose self-supervision in the usecase of rs-fMRI as illustrated in Figure 3.4. First, we start by reviewing some common techniques to provide better understanding and ensure that the most suitable option is implemented.

A number of interesting self-supervision approaches has been proposed in recent years, and two of the most prominent ones are contrastive and non-contrastive self-

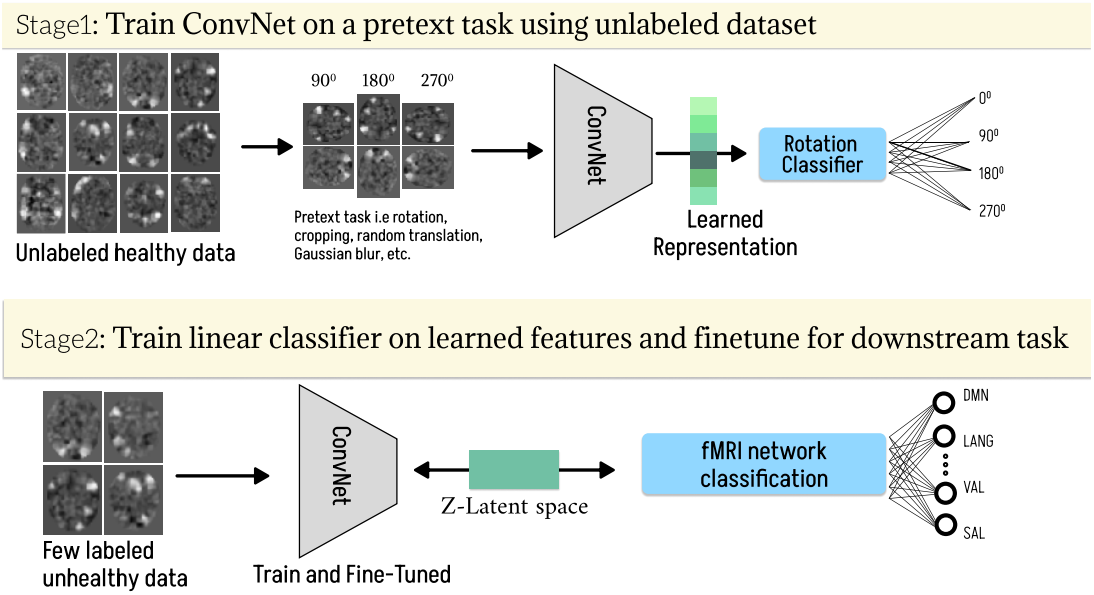


Figure 3.4 – Learning good representation by context prediction from image augmentation.

supervision. Non-contrastive self-supervised learning involves training the model to predict some form of auxiliary information, such as the location of an object in an image or the next frame in a video. This approach is regarded as counter-intuitive because, it only uses positive sample pairs to train the representation (and only the distance between them is minimized), it may appear that the representation will collapse into a constant solution, where all inputs map to the same output. The loss function is expected to reach zero which represent smallest possible value, with a collapsed representation. This technique shows the ability to learn good representation regardless of the lack of negative examples [110]. This approach demonstrates that, training a non-contrastive self-supervised learning framework leads to a useful local minimum but not to the global trivial minimum which is what informed our intuition towards investigative contrastive learning approach for our high dimensional fMRI data.

3.3.1 Contrastive Self-supervision learning

Contrastive self-supervision is a technique in which the model is trained on positive and negative pairs of samples, with the aim of learning to differentiate between similar and dissimilar images. This approach learns representations by minimizing the distance between two views of the same data point and maximising views from dif-

ferent data points. Essentially, it reduces the distance between positive and negative data to a minimum and increases the distance between negative and positive data to a maximum. In this work, we centered our focus on the approach of contrastive self-supervision learning and we describe the general intuition of two of the key algorithms in contrastive learning below:

a) Momentum contrast — The rationale for this approach is the intuition that in order to learn an effective representation, we require a large dictionary with a wide variety of negative examples while maintaining the dictionary key encoder's maximum consistency. Using the dictionary as a queue rather than as a static memory bank or a mini-batch is the fundamental component of this strategy. By divorcing the dictionary size from the mini-batch size, this produces a dynamic lexicon that offers a wealth of negative instances and can be expanded as necessary [111].

b) SimCLR — The key here is to employ significantly higher batch sizes (8192, to obtain a rich set of negative examples), greater data augmentation (cropping, color distortion, and gaussian blur), non-linear processing of the embeddings prior to similarity matching, a larger model, and longer training times. This research provides empirical evidence that experimenting with these are some of the obvious things to do in order to increase performance [112].

3.3.2 Self-Supervision experiments

We adopted SimCLR [113] for our implementation, a technique based on contrastive learning, to efficiently learn visual representations from unlabeled images. Through a contrastive loss in a hidden representation of neural networks, SimCLR learns representations by maximizing agreement [114] between many augmented views of the same data sample as illustrated in Figure 3.5.

In this method, two augmented views are generated from an image for contrastive prediction. A typical possibility in the generated augmentation such as cropping, can be global and local or adjacent views as illustrated in Figure 3.6. In this case, Solid rectangle represent images and dashed rectangles are generated augmentation views. By randomly cropping images, we sample contrastive predication tasks that include global to local view ($A_i \rightarrow A_j$) or adjacent view ($B_i \rightarrow B_j$) prediction as proposed by [113].

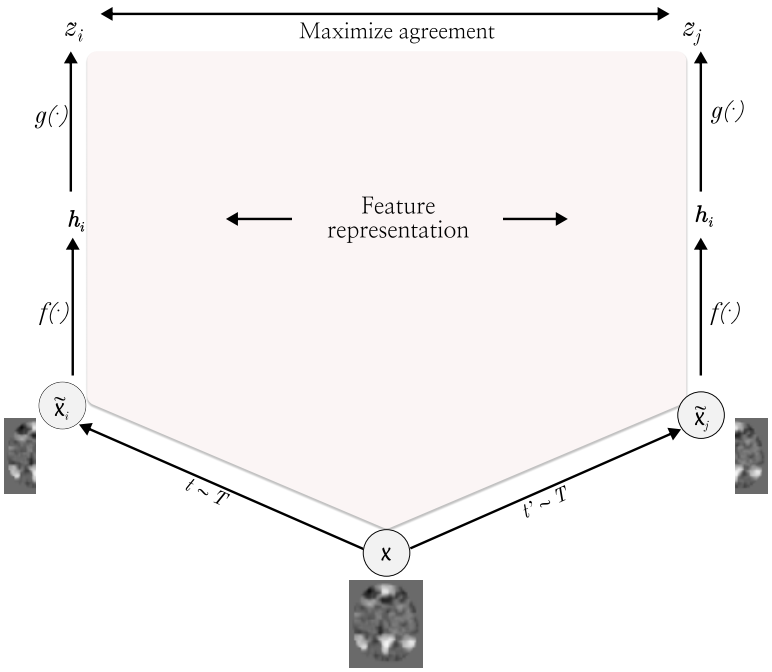
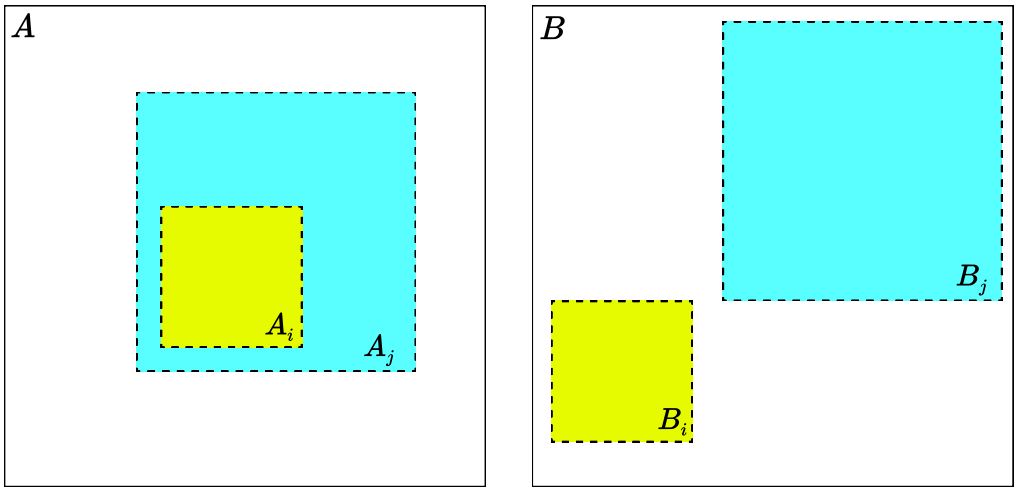


Figure 3.5 – Illustration of contrastive training with input fMRI image x to obtain two correlated views $\hat{x}_i$ and $\hat{x}_j$.



a) Global and local views. a) Adjacent views.

Figure 3.6 – Image augmentation views for contrastive prediction.

Generally, the encoder network in divided into two parts: a based encoder $f(\cdot)$ and, a projection head $g(\cdot)$. The base network works just like the deep convolutional neural network and its responsible for extracting a representation features from the augmented data samples. The projection head $g(\cdot)$ maps the feature representation h

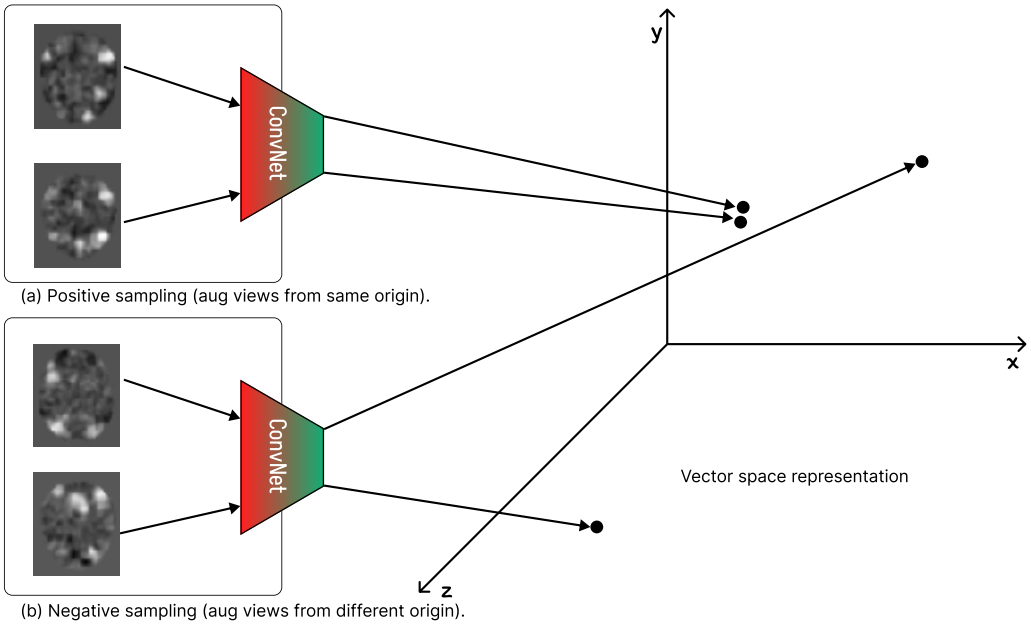


Figure 3.7 – Vector space representation of positive and negative similarity views.

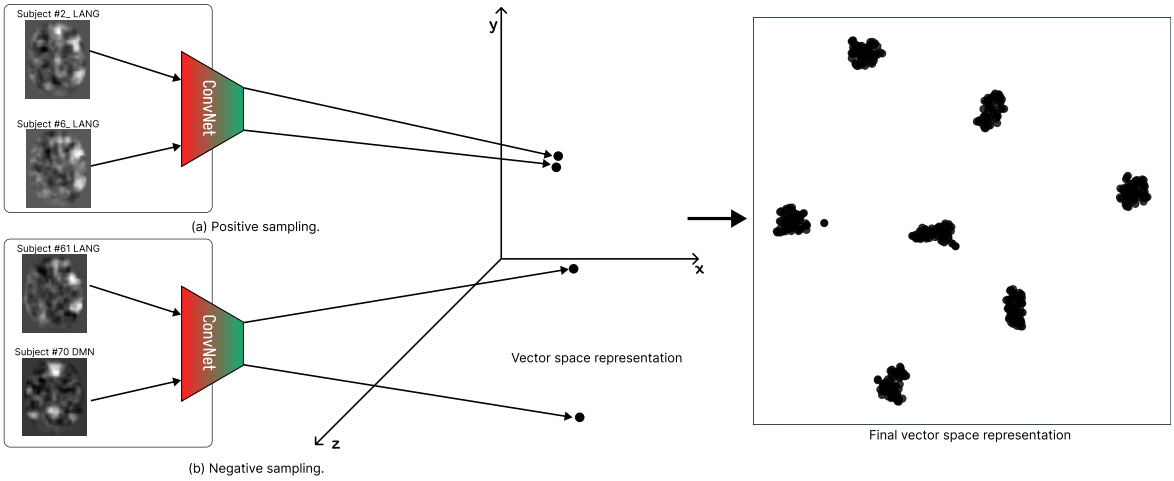


Figure 3.8 – Vector space representation of positive and negative at the end of pretext task stage.

into a space where we apply the contrastive loss like aggregate similarities between vectors (see Figure 3.7). The vector space representation due the repeated maximization and minimisation of agreements at the pretext task stage is shown in Figure 3.8 while the vector space representation at the end of downstream task stage is shown in Figure 3.9. The projection head $g(\cdot)$ is normally discarded after the contrastive training is completed, while $f(\cdot)$ is used as a pretrained feature extractor. The representations z

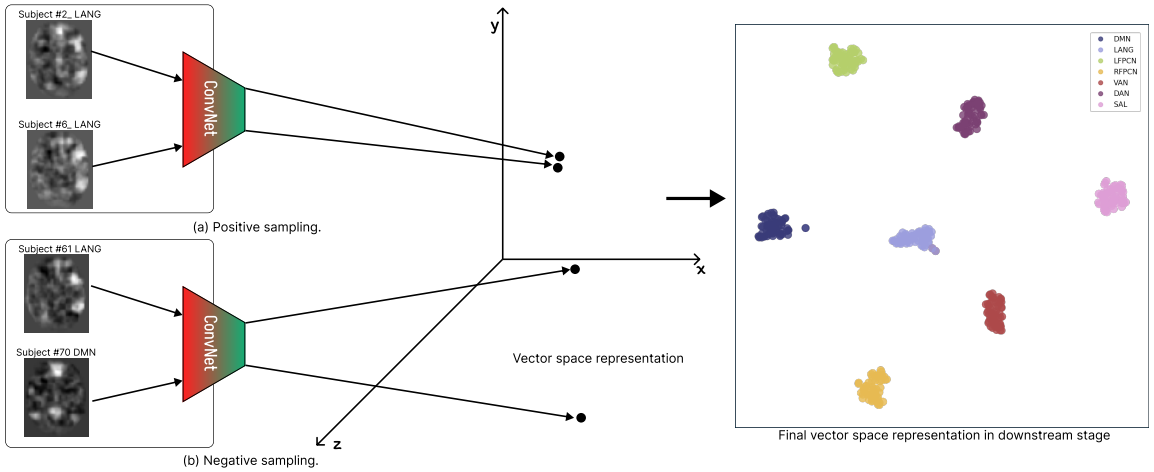


Figure 3.9 – Vector space representation of positive and negative at the end of downstream task stage.

obtained from the projection head $g(\cdot)$ have been shown to perform worse than those of the base network $f(\cdot)$ when finetuning the network for a new task. This is due to the fact that the representation z are trained to become invariant to many features like the local embeddings that can be important for downstream tasks. Therefore $f(\cdot)$ is only required at the contrastive stage.

Given a mini-batch of images that were chosen at random, each image x_i is augmented twice using random rotation, gaussian blur, and random crop resulting in two views of the same example both x_{2k-1} and x_{2k}. To construct representations h_{2k-1} and h_{2k}, the two images are encoded using an encoder network $f(\cdot)$ (ResNet). After that, the representations are altered again using a non-linear transformation network $g(\cdot)$, yielding z_{2k-1} and z_{2k} that are used for the contrastive loss. The contrastive loss between two positive cases i, j (augmented from the same image) is presented using a mini-batch of encoded samples as follows;

$$l_{i,j}^{NT-Xnet} = -log\frac{exp(Cosim(z_i, z_j)/\tau))}{\sum_{k=1}^{2N} \mathbf{1}_{[k\neq i]} exp(Cosim(z_i, z_j)/\tau)}, \tag{3.1}$$

$$sim(z_i, z_j) = \frac{z_i^\top \cdot z_j}{\|z_i\| \cdot \|z_j\|}, \tag{3.2}$$

where $\mathbf{1}_{[k\neq i]} \in \{0, 1\}$ is an indicator function evaluating to 1 iff $[k \neq i]$, $Cosim(\cdot, \cdot)$ is cosine similarity between two vectors as show in equation 3.2, and τ is a temperature scalar.

We trained the model at a learning rate of $1e-5$ with $100,000$ epochs. To reduce model over-fitting, we adopt an early stopping method which uses the value of increase in validation error to make decision. Furthermore, we used grid-search algorithm to select optimal hyper-parameters for the SimCLR model related to increased precision of training data. The halting point of the training model was after $10,000$ validation failures and then a model checkpoint.

In SimCLR cosine similarity, the maximum cosine similarity obtainable is 1, while -1 is the minimum obtainable. By implementation, we observe that the features of two different views of images converge to a cosine similarity around zero since the minimum, -1, actually required z_i and z_j to be in the direct opposite location in all feature dimensions, and this limits its flexibility. The iterative training process with the cosine similarity functions results in attraction of related views and repulsion of views from different image augmentation as illustration in Figure 3.10 and Figure 3.11 respectively.

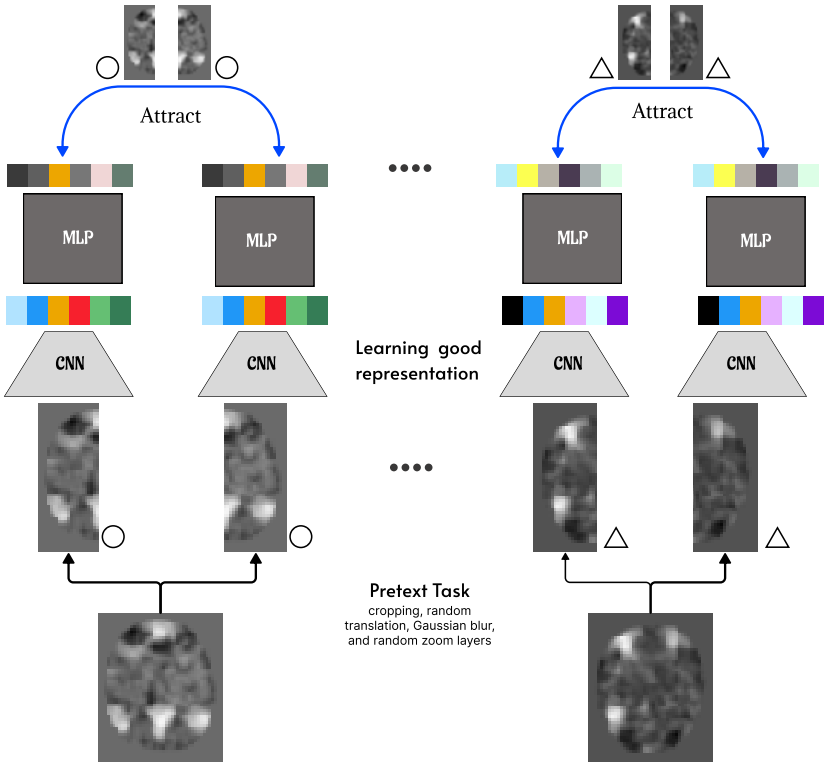


Figure 3.10 – Contrastive similarity maximization in related augmented features.

For contrastive learning, we employed image augmentations including cropping,

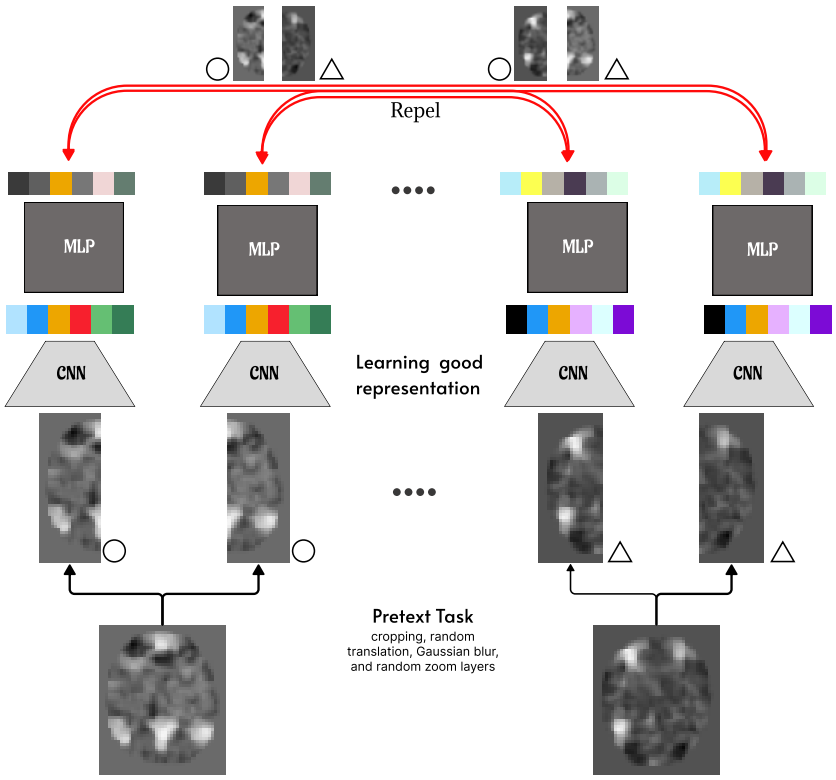


Figure 3.11 – Contrastive similarity minimization in unrelated augmented features.

which pushes the model to encode various portions of the same image, as well as random translation, Gaussian blur, and random zoom layers. We concurrently loaded a large batch of unlabeled data from healthy subject images and a smaller batch of annotated samples from unhealthy subject images during training. We also used random horizontal flips as the second image augmentation method. To prevent overfitting on the few labeled samples, stronger augmentations, such cropping, are used for contrastive learning together with weaker ones like horizontal flips at the pretext task stage.

The encoder model was pretrained on unannotated images with a defined contrastive loss. The encoder's top is equipped with a nonlinear projection head, which enhances the quality of encoder representations. We employed the NT-Xent loss (Normalized Temperature-scaled Cross Entropy), which has the following meaning: Each image in the batch is treated as if it were its own class. Then, for each "class", we have two instances (a pair of augmented views). The representation of each perspective is compared to the representation of every possible pair (for both augmented versions).

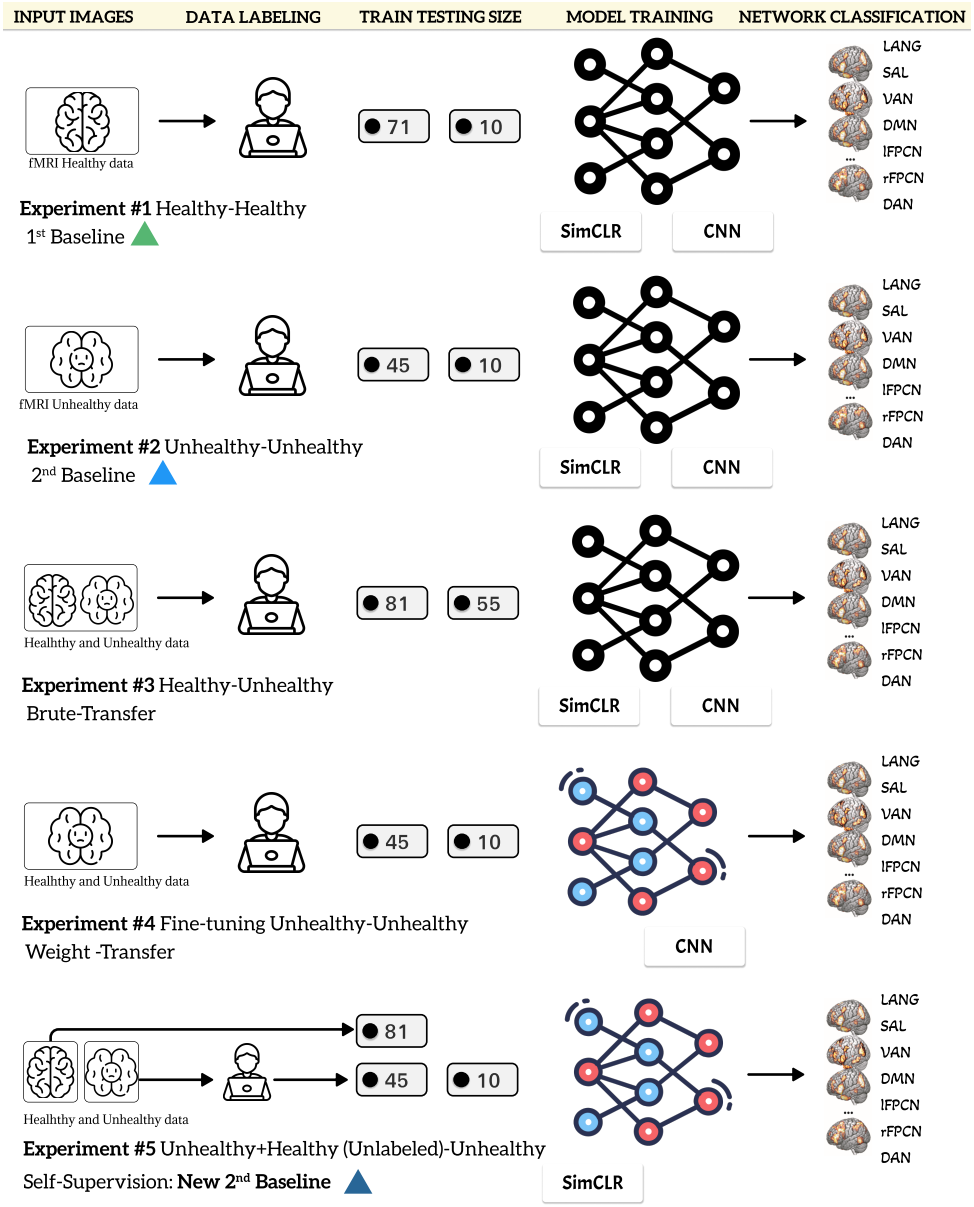


Figure 3.12 – Illustration of our experiments with supervised and self-supervised approach.

As logits, we employ the temperature-scaled cosine similarity of comparing representations. Finally, as the "classification" loss, we employ categorical cross-entropy. In order to monitor the pretraining performance, we used two metrics of contrastive accuracy [113] and linear probing accuracy. We fine-tuned the encoder on the annotated subjects, by adding a single, fully connected classification layer with a random initial-

ization on top.

3.4 Results and discussion

In this section, we provide details of outcomes from our experiments by using data collection procedure and training techniques explained in section 3.2 and illustrated in Figure 3.12. The values in table Table 3.1 display the experimental accuracy numbers that were recorded from different experiments organized from the adopted self-supervised model as well as a comparison with the proposed supervised learning model in [34]. Data sizes that are utilized for testing and training were specified in each case. It is important to note that neither during training nor during hyper-parameter adjustment does the trained model ever view testing data.

Table 3.1 – Supervised and Self-supervised fMRI brain network classification results with healthy and unhealthy data (7 fMRI network activation image corresponds to single patient in all cases).

	No. of training subjects	**No. of testing subjects**	**SimCLR**	**CNN** [34]
Healthy to Healthy	71 (labeled healthy)	10	81.74%	86%
Unhealthy to Unhealthy	45 (labeled unhealthy)	10	73.48%	75%
Healthy to Unhealthy	81 (labeled healthy)	55	69.21%	74%
Unhealthy to Unhealthy With unlabeled healthy	**81(unlabeled healthy) + 45 (labeled unhealthy)**	**10**	**76.39%**	—
Fine-tune on Unhealthy data from Healthy data	45 (labeled unhealthy)	10	—	78%

We performed data randomization at several points in the model training pipeline to provide a more consistent and reliable output, and we make sure the model has never seen test data before. Although the use of cross-validation techniques could be an alternative, we were unable to consider this option in order to keep our model simple and avoid further training complexity, which would have increased the computing resources needed for our contrastive learning model.

Initially, we trained and evaluated our model using data from healthy control subjects. This approach gave an absolute limit of performance with the highest accuracy of 81%, which is almost 5% less than the CNN model proposed in chapter 2. The accuracy evaluation in this case is very encouraging, owing to the known spatial consistency of healthy image data. In this experiment, the CNN model proposed in chapter 2 reached the best performance on the CNN model compared with SimCLR. However, it has to

be mentioned that a CNN method has an additional cost from data annotation. In contrast, the self-supervised method can perform similarly with few annotated data.

In similar experiment, where training and testing of our model was organized with solely unhealthy patients, a reduction of about 8% compared to our previous result was recorded, which created a second baseline with fewer data. The same behavior was observed in the chapter 2 between the classification of unhealthy patients and healthy subjects, where a performance drop was around 11%. Although, the performance on this baseline is less than the first experiment, the results are more interesting as this performance is achieved with fewer annotated data, therefore very applicable in clinical purposes.

On the brute-transfer strategy, (learning from healthy subjects data without fine-tuning on unhealthy data), as shown in Table 3.1 row 3, we trained both our self-supervised and supervised model with 81 annotated healthy control subjects and conducted testing on 55 unhealthy patients data. This time, we recorded an average accuracy of 69% for different ranges of test data sizes. It can be agreed that the brute-transfer learning does not introduce any accuracy enhancement in this case, similar to what was observed in chapter 2. Although, this observation highlights a significant difference between healthy and unhealthy patient data which quantifies its impact on transferability.

The fourth row of Table 3.1 shows the performance of a new experiment, where the SimCLR model is trained on a portion of unhealthy patients (45) and all unlabelled healthy data, which is fed to the model during the training among augmented images (pretext task). This experiment shows the most important result as its performance is more than the CNN model on unhealthy patients (2^{nd} row) and the brute-transfer learning (3^{rd} row) with about 3% and 7% respectively. The advantage of the SimCLR model in this experiment compared with other models in chapter 2 is the use of non-labeled and few labeled data to train a model, while for CNN and transfer learning models, a large amount of label data is required.

The last row of the Table 3.1, indicates the best performance of the transfer learning model in [34] while the CNN model has been trained once on all annotated healthy data. Then the model weights have been transferred and fined-tuned on unhealthy data. Although this model has the maximum accuracy among other experiments, the cost of the training model is too high as we need to use all 81 annotated healthy subjects and 45 annotated unhealthy subjects during the training. This cost can reduce

the method's applicability for clinical purposes as we always lack annotated database in this domain, while the self-supervised method can gain similar performance with fewer data.

3.5 Conclusion and perspective

Self-supervised learning is a rapidly growing area of research in fMRI image analysis, and has shown great promise for tasks such as image classification. While there are still a number of challenges and limitations that need to be addressed, the future of self-supervised learning for fMRI image classification is promising, and there is a growing body of evidence to support its progressive development and refinement for standardization of resting-state fMRI image processing.

In this work, we proposed ways to tackle the dependence on large annotated dataset and demonstrated the use of self-supervision to directly avoid healthy data annotation. Firstly, few shot learning remains a promising alternative to reduce the dependence on large annotated healthy data. This is possible, since only a few label samples are required to train a model to understand the underline feature patterns in our data. Secondly, synthetic data generation via generative model, data augmentation or enhancement is another promising method since in this case, no data annotation is required. Lastly, we applied contrastive self-supervision method to illustrate the use of unannotated healthy data to learn useful features in unhealthy data prediction. The technique demonstrated is interesting because, it accelerates learning in medical imaging, where healthy patients can be easily enrolled. In contrast to traditional fine tuning process of transfer learning as required in supervised learning approach, it does not require annotation of healthy subject data, while few labeled unhealthy patients data can be used. This strategy therefore, opens the possibility to greatly improve our model by benefiting from the demonstrated transferability of healthy data features for unhealthy data prediction. Large size of healthy control subject data (volunteer subjects) can be easily acquired by enrolling more subjects into this non-invasive medical imaging modality without data annotation requirement since there is no clinical interest for labeling healthy data.

In the experiments discussed in this chapter, we provide an application of self-supervision technique in the identification of functional biological networks, and fortunately, the same method can be used for any non-invasive medical imaging task for

which healthy controlled subject can be acquired easily and there is similarity between healthy and unhealthy data.

CHAPTER 4

fMRI IMAGE DATA DISCRIMINATION IN HEALTHY AND UNHEALTHY SUBJECTS

In this chapter, we explored various techniques to demonstrate local and global relationship that exist in our fMRI image database. This approach has become necessary because of the observed discrepancy identified in our data, since we have benefited through feature transferability which indicates similarity, while at the same time, model validation with unhealthy data shows a significant drop.

4.1 Introduction

In previous chapters (see chapter 2), we demonstrated the transferability of functional brain network recognition by allowing our model to learn from features of healthy subject fMRI images to improve prediction in unhealthy data. In this study, we aim to further develop our investigations on the origin of the similarity and dissimilarity between the healthy and unhealthy fMRI data. Since we performed classification of the functional activation networks with convolutional neural networks the most direct approach to start with is to consider the perspective of the healthy and unhealthy data discrimination with CNN.

4.2 Deep learning approach

A systematic accuracy drop between healthy and unhealthy fMRI data across the 7 functional brain networks was observed, as shown in Figure 4.1, from performance analysis with the best CNN model from chapter 2. The best performance accuracy when training on healthy data and testing on healthy data was 0.86 while, on average, training on unhealthy data and testing on unhealthy data provides an accuracy of

0.75. This implies that unhealthy data are more difficult to classify when compared to healthy data. We extend our follow up on this observation to binary discrimination of healthy and unhealthy data to evaluate the extent of this difficulty.

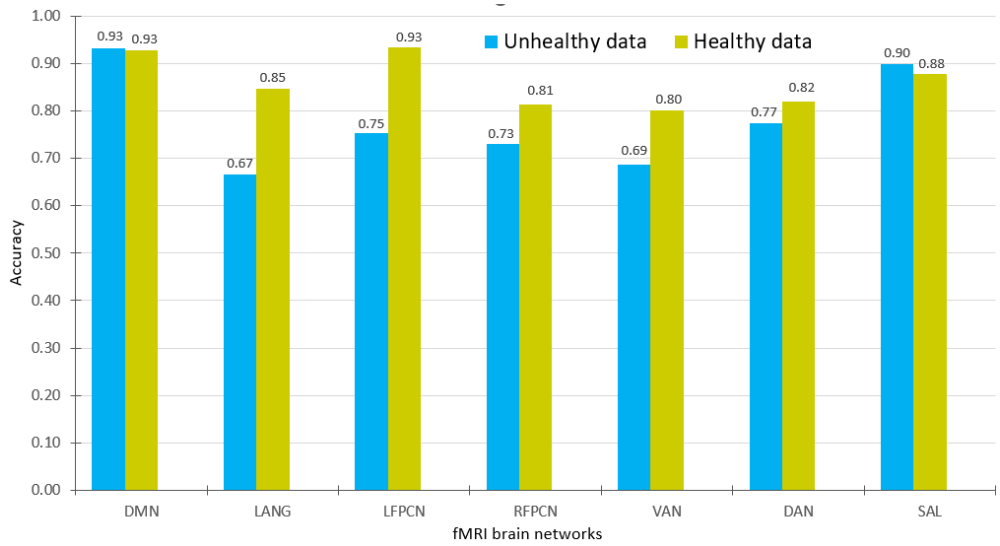


Figure 4.1 – CNN classification result of individual functional brain network in fMRI healthy and unhealthy data.

We implemented a PyTorch based CNN model for binary classification between healthy and unhealthy fMRI images. This was achieved by assigning the label "1" for all healthy data across all functional network while assigning the label "0" to all unhealthy data across 7 functional networks. With our output layer of 2, we indicated the early stopping of 20 steps, and train our model with learning rate starting from 1e-7 to 1e-3. It is important to note that, the CNN model explained here is different from the proposed model in chapter 2, which is a multi-class classifier model 7 functional brain network classification. In the evaluation phase of this experiment, we recorded binary classification results of healthy and unhealthy data as shown in row 4 of Table 4.1. As a basis of meaningful comparison, as well as to underscore the prevailing data discrepancy, we included results of 7 functional brain network classification from our previous experiments as show in rows 1, 2 & 3 of Table 4.1 respectively. This clearly shows that despite the usefulness of feature transfer demonstrate in previous experiment we were unable to fully compensate for the accuracy drop due to data discrepancy. Our current observations with the binary classification shows strong evidence of data distinguishability between healthy and unhealthy data. Furthermore, we provided the confusion

matrix in Figure 4.2 of the model evaluation to understand cases of misclassification errors.

Table 4.1 – CNN binary discrimination of healthy and unhealthy data with other fMRI network classification approach with similar data.

CNN Classification Model Description	Train/Validation/Test data size	Classification Accuracy
Healthy fMRI brain networks	427/70/70	0.86 ± 0.02 [33]
Unhealthy fMRI brain networks	245/70/70	0.75 ± 0.01 [33]
Transfer learning	567 - 245/70/70	0.78 ± 0.01 [34]
Healthy and Unhealthy discrimination	574/98/89	0.95 ± 0.02

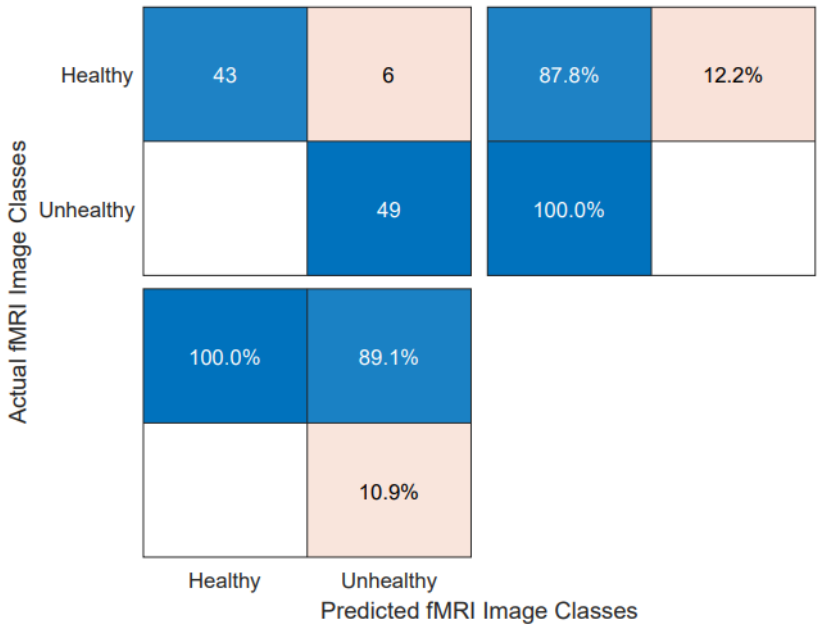


Figure 4.2 – Confusion matrix of CNN binary classification of healthy and unhealthy fMRI data with row and column summary as percentage of true-positive and true-negative respectively.

To understand how this end-to-end deep learning model was able to discriminate between the two classes, we observed the latent space representation of the network. We provide a t-distributed stochastic neighbor embedding (T-SNE) of a combined representation of healthy and unhealthy data with respect to the 7 functional brain networks in the same latent space in Figure 4.3 to understand the relationship between

the features of healthy and unhealthy data in 7 functional brain networks. The visualization of healthy data representation in the same latent space with unhealthy data allows us to measure the local displacement between different functional brain networks. The corresponding visualization obtained from binary classification improve our understanding of the local similarly among related networks while the difference between the two data sources remains a close distance in the groups.

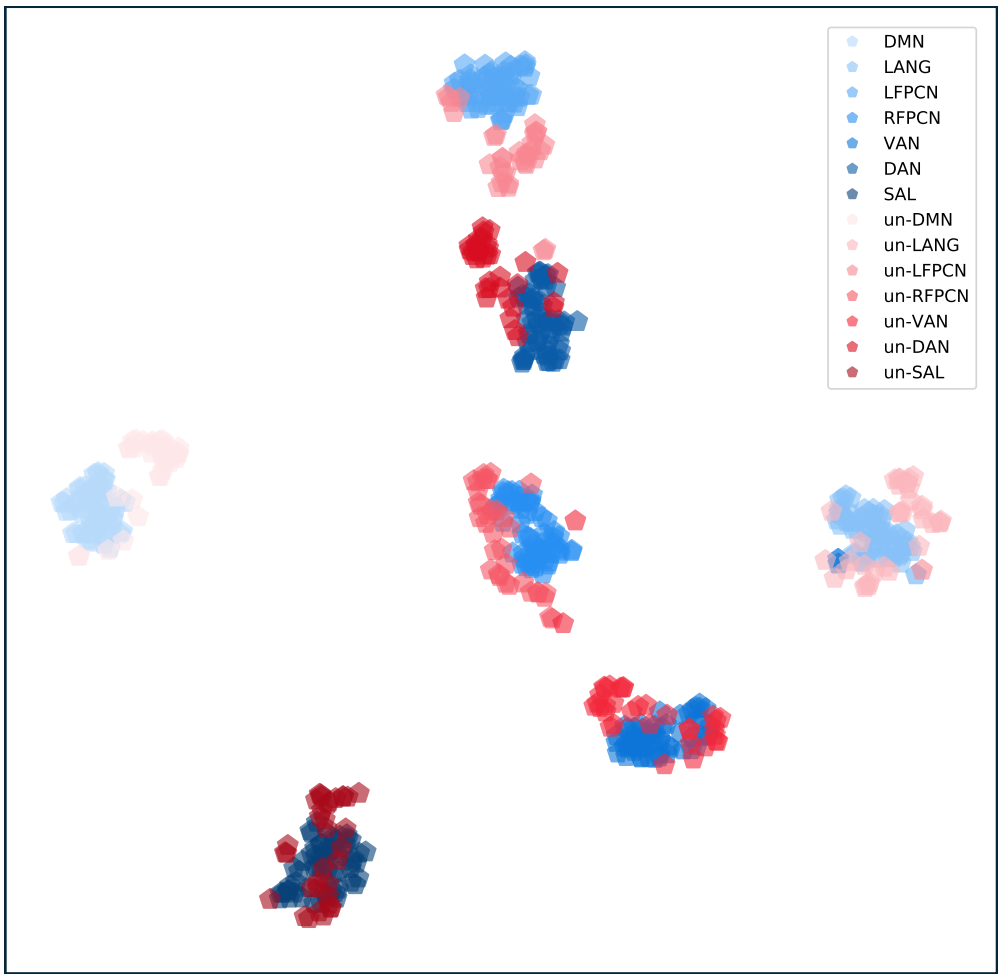


Figure 4.3 – Latent space visualization of fMRI healthy and Unhealthy data with T-SNE, across 7 functional networks in the two groups.

We also explored the latent space visualization in respect to our binary image classification as shown in Figure 4.4 which represent our data in 2 dimensional space using principal component analysis (PCA) and Figure 4.5 to allow effective representation of non-linear relationships in T-SNE. This is interesting because, we aim to understand the representation of these two visualization options since, PCA aim to maintain the

global layout of our fMRI data while T-SNE focuses on preserving the local structures of data points to allows more realistic visualization. Although, we can see clear distinction in the healthy and unhealthy data representation shown in both cases, the T-SNE option provides a better view of the similarity in healthy data compared to the PCA option.

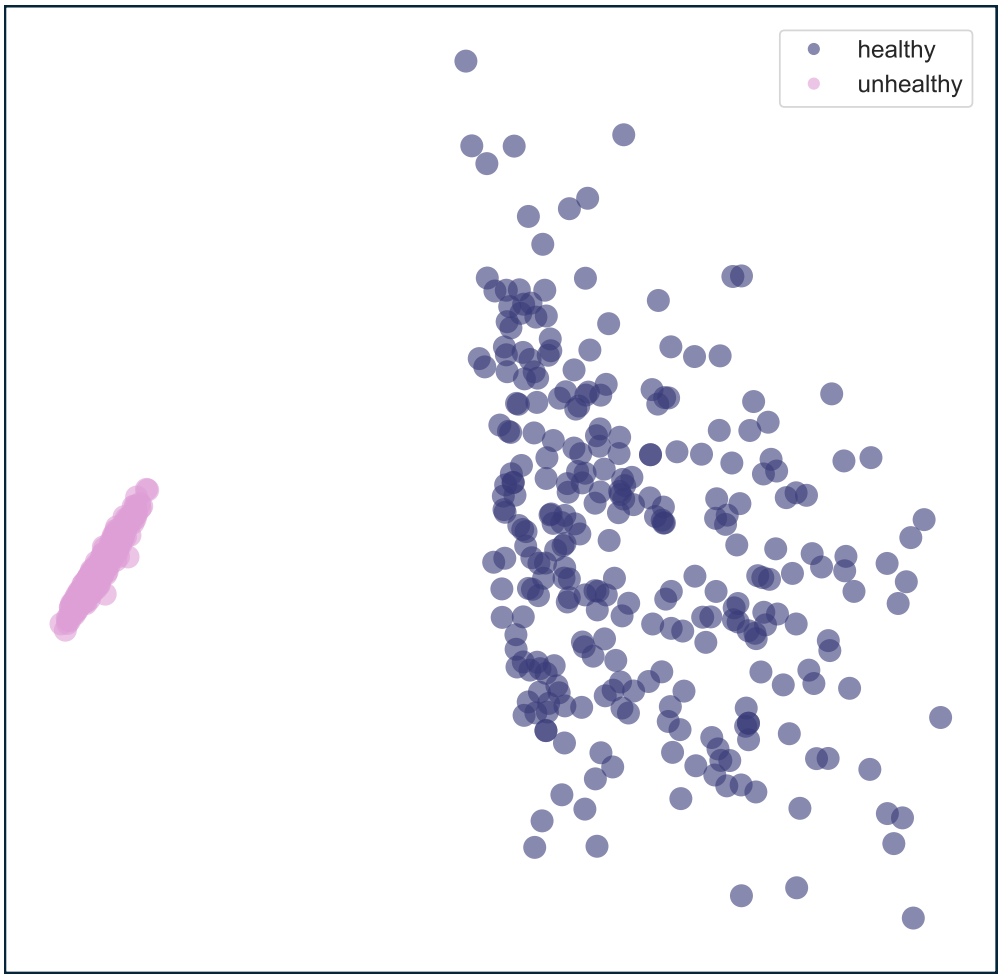


Figure 4.4 – PCA visualization of healthy and unhealthy fMRI data.

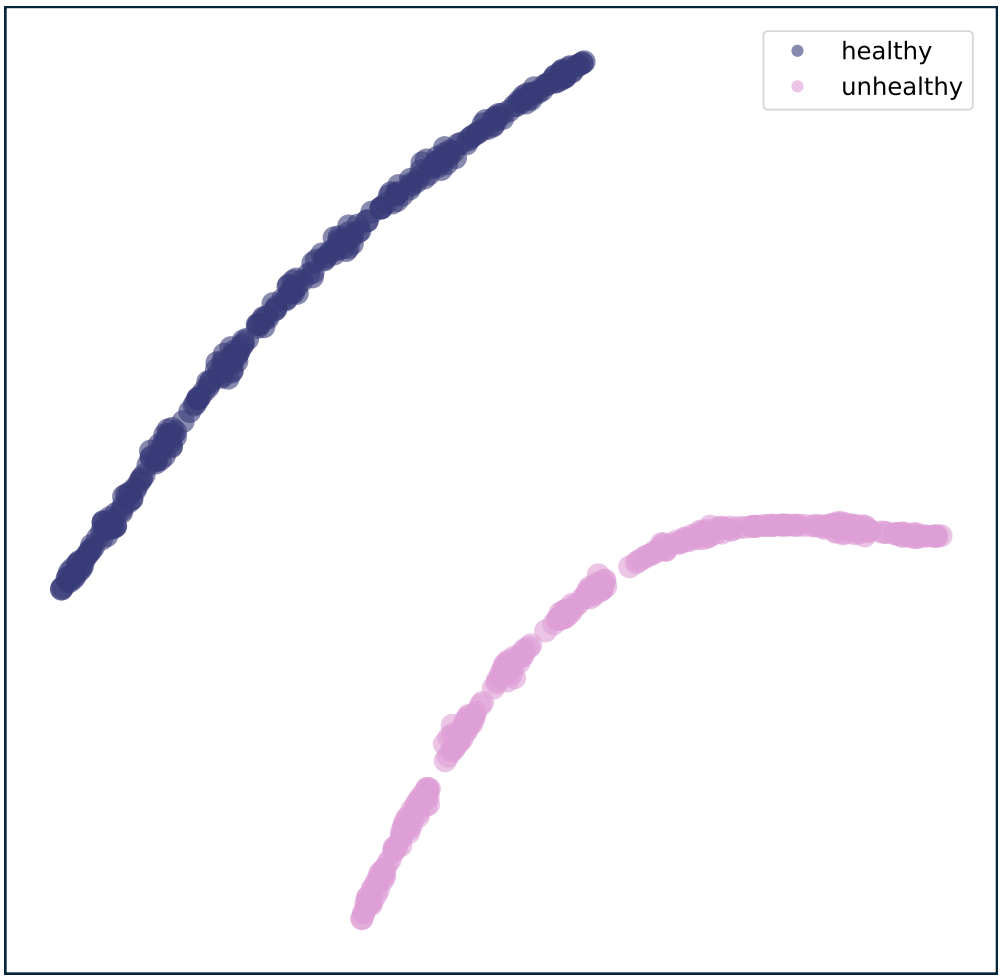


Figure 4.5 – T-SNE visualization of healthy and unhealthy fMRI data.

We extend our effort to the use of GRADient-weighted Class Activation Mapping (Grad-CAM), a more versatile version of CAM, which can provide visual explanations for any arbitrary convolutional neural network (CNN) [115]. This process allows us to produces a coarse localization map that highlights the important regions in our fMRI brain network images for healthy and unhealthy data discrimination. Our implementation computes the gradient of the logits in each class with respect to the activation maps of the final convolutional layer, and then averages the gradients across each feature map to generate an importance score which help to improve interpretability, and easier understanding of the local features used by the network for classification by our CNN model.

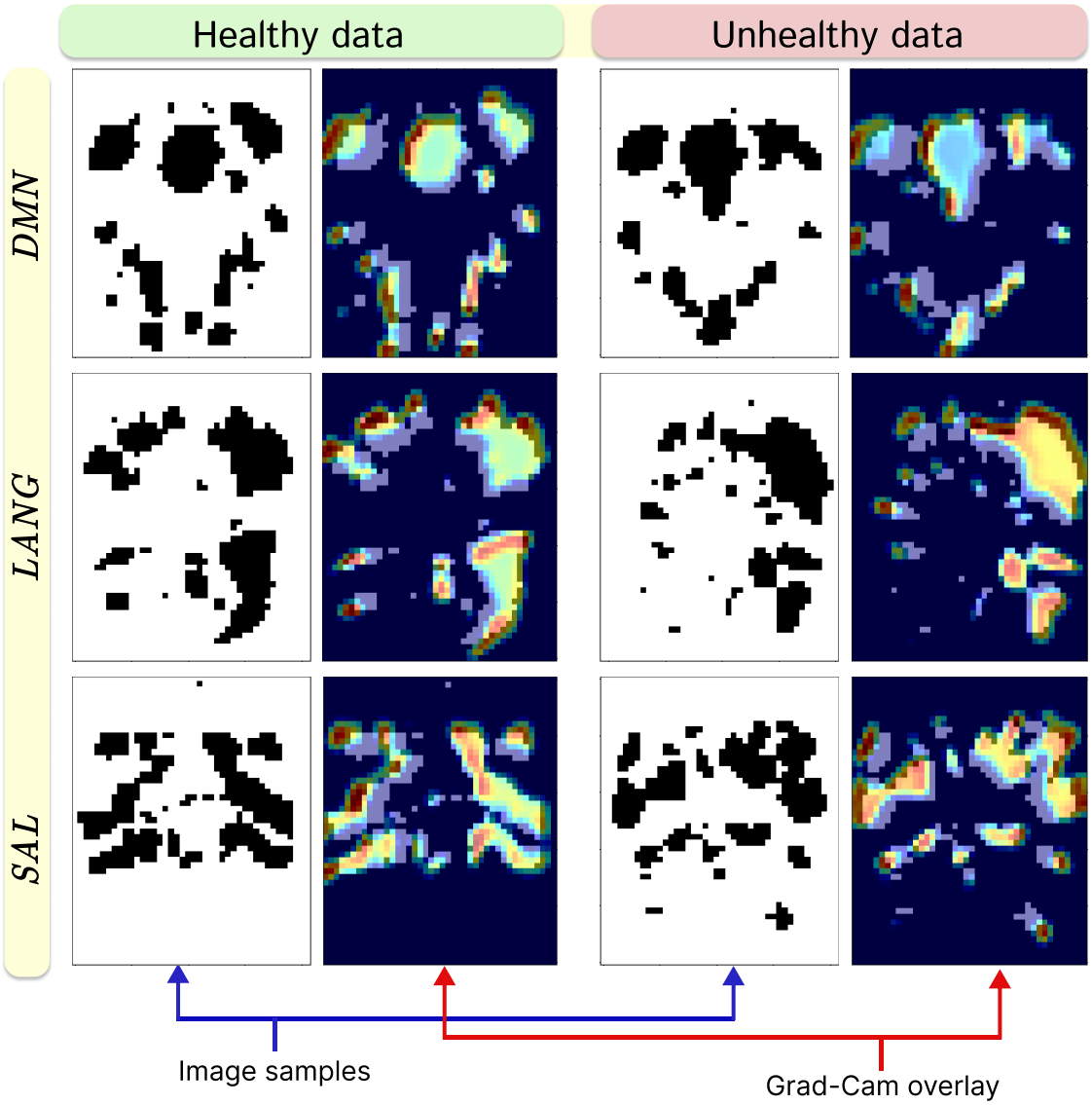


Figure 4.6 – Grad-CAM visualization of healthy and unhealthy data across 3 sample functional brain networks.

Our deep learning, latent space and Grad-CAM visualization provided great insight to evaluate the level of the identified discrepancy. Clearly, there is a systematic shift between healthy and unhealthy data as demonstrated by the strong discriminability of our CNN model. This explains why a model trained on healthy data does not performs well on unhealthy data validation, which in turn explains why the model performance on unhealthy data is less compared to healthy data. Nevertheless, when looking at the functional network in the latent space, the discrepancy between healthy and unhealthy data is systematically less obvious than the discrepancy between func-

tional networks. This observation explains the transferability of the learned representation in our CNN mode from healthy to unhealthy data, since the latent space has the same "macro" structure. In the case of Grad-CAM visualization, it is apparent that feature maps used for discrimination in both healthy and unhealthy data are consistently global across healthy and unhealthy data as well as in different functional brain network as shown in Figure 4.6. In effort to reveal the local relationship between healthy and unhealthy data, we extend this investigation to statistical analysis of local features in our fMRI data.

4.3 Statistical analysis approach

4.3.1 Pixel intensity difference between region of lesion and region of non-lesion in unhealthy patient data

We performed statistical evaluation of the difference between pixel intensities of lesion and non-lesion region of unhealthy patient image in order to better understand the cause of accuracy drop as seen in our deep learning classification model. This was done by comparing the pixel intensities of tumor regions which is referred to as "lesion" via the provided masks and the remaining region of the network image, referred to as "non-lesion"

$$x = \mu_L - \mu_{nL}, \tag{4.1}$$

where we denote x as the difference of average gray level between lesion (tumor) area and non-lesion area as depicted in Figure 4.7

We provide an illustration of lesion and non-lesion region separation in a mathematical expression (see Equation 4.1), and visual observation in Figure 4.7 as applied to our unhealthy fMRI data. While this operation helps to observe the mean pixel intensity differences across all images (see Figure 4.8), which was in fact dynamic within the 0.0 - 0.2 pixel range. We also conducted statistical analysis on the two variables in order to determine the significance of the associated differences between the two collections.

To evaluate the significance of pixel intensity difference between region of lesion and region of non-lesion in unhealthy patient data, we computed the chi-square p-value X^2 using

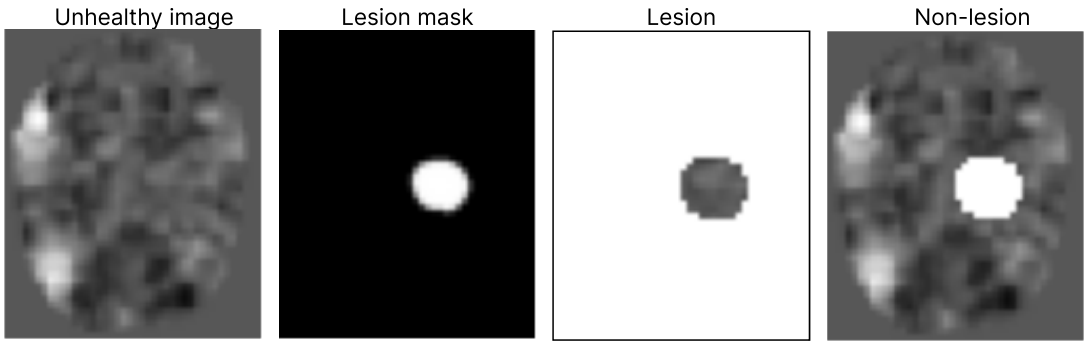


Figure 4.7 – Visual presentation of lesion and non-lesion region subtraction process from unhealthy fMRI images.

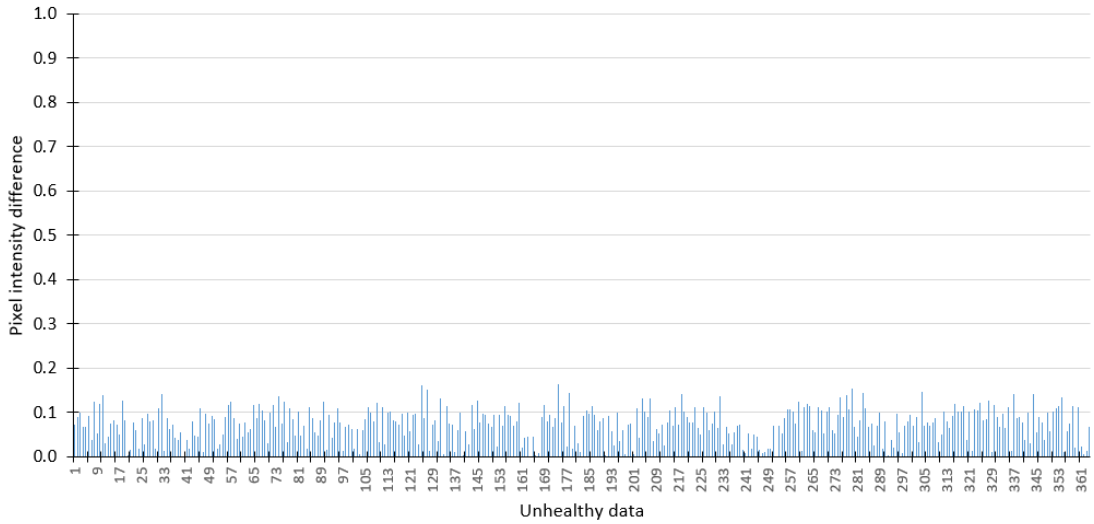


Figure 4.8 – Histogram of pixel difference between lesion and non-lesion region shown on the actual scale of pixel intensity range $(1 - 10)$.

$$X^2 = \sum \frac{(O_i - E_i)}{E_i}, \tag{4.2}$$

O_i = observed value (pixel intensities) and E_i = expected value. We obtained a p-value of 0.3242 (>0.05), as a test of significance between the pixel intensities of each image and the expected value of 1.0 (normalized image pixel range is 0-1). This observation suggests that the difference in pixel intensity of those obtained from lesion region, against those from non-lesion region is statistically insignificant. Since this approach compares small region of brain tumor (lesion) against the remaining region of the brain, we believe that more accurate relationship between the data can be uncovered by proposing a more realistic comparison.

4.3.2 Pixel intensity evaluation of lesion area and surrounding region of unhealthy patient data

At this stage, we investigate the differences between the lesion region and its surrounding area, as opposed to the differences between the lesion location and the remainder of the brain region. This was achieved by comparing pixels intensities of tumor region with pixel values from $5 - 10px$ surrounding region of the tumor. This approach is more subtle because, in clinical perspective, the actual effect of brain tumor, which includes distortion or total displacement of brain region may be observed within this indicated surrounding area of the brain. A surrounding mask of the lesion is generated via morpho-mathematical 3D dilation process as shown in Figure 4.9. This defines new set of pixels intensity distribution of lesion and surrounding region of lesion. Then, we revisit Equation 4.1 to analyze the difference of both lesion and surrounding region of lesion.

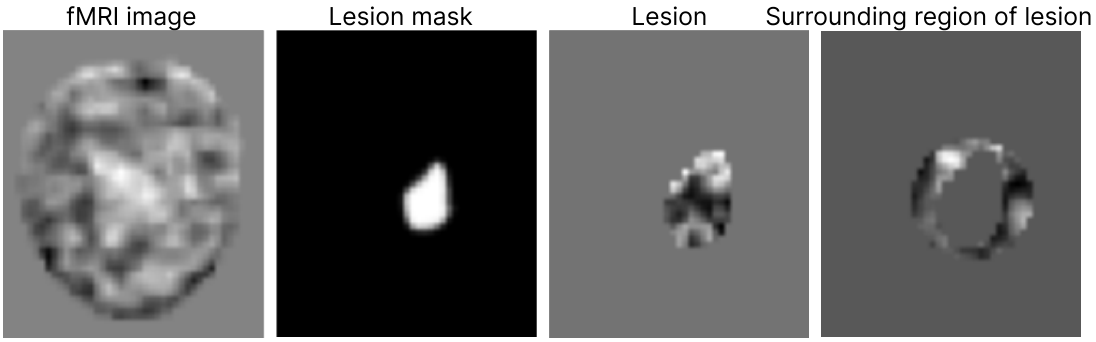


Figure 4.9 – Functional brain network image of lesion and non-lesion(surrounding) area of unhealthy data.

The chart in Figure 4.10 is the normalized histogram of all images showing the pixel distribution for both lesion and surrounding non-lesion region of the fMRI data. It can be seen that, the frequency across different pixel groups are progressively similar and to further justify this, we used

$$Skewness = \frac{\sum_i^N (X_i - \hat{X})^3}{(N-1) * \sigma^3}, \tag{4.3}$$

where $X_i = i^{th}$ is a random variable, X = Mean of the pixel intensity distribution, N = Number of image features in the distribution and σ = Standard distribution. We evaluated the two variables and recorded a highly skewed value of 1.3094 and 1.6903 for

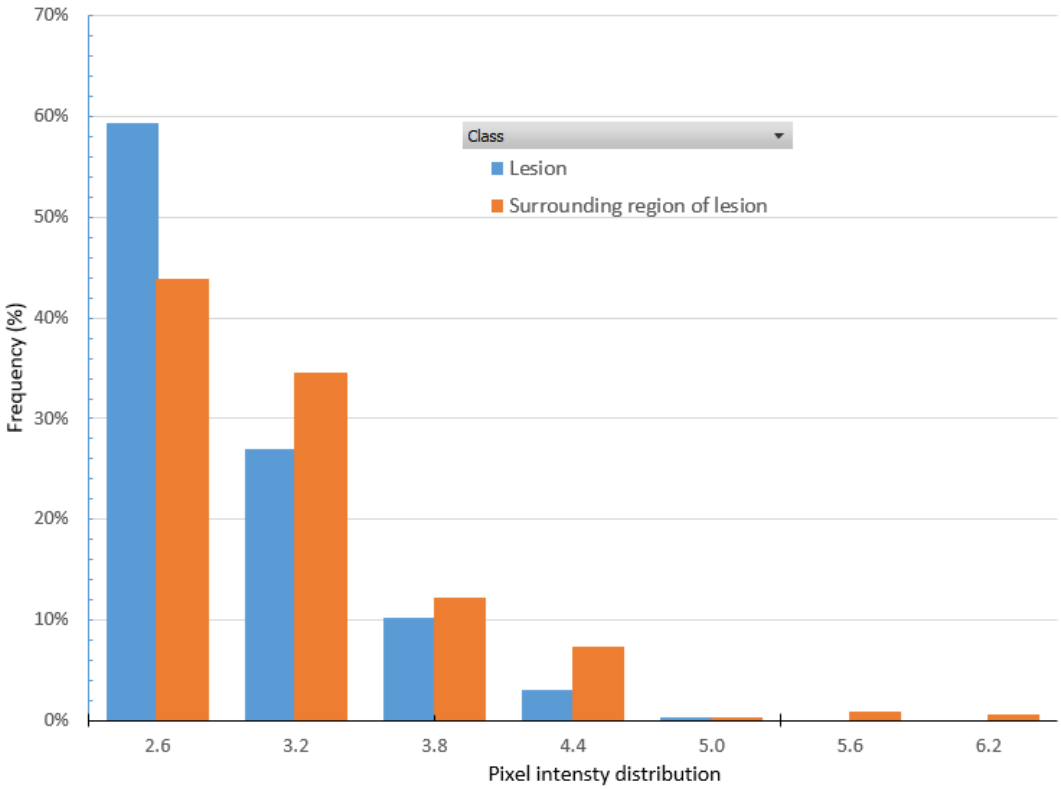


Figure 4.10 – Pixel distribution of fMRI images with respect to lesion and surrounding region of lesion.

lesion and lesion surrounding respectively, which also established that the two variables fall in the same distribution. Furthermore, using the expression in Equation 4.2, we provided the result of chi-square p-value test of 0.1300 (>0.05) which shows that the difference between the pixel distribution of lesion and lesion surrounding (see Equation 4.1) is statistically insignificant. While the above listed approaches failed to find the origin of the discrepancy between healthy and unhealthy region of our data, we now get interested in understanding the possible influence of brain tumor overlap with functional network activation maps.

4.3.3 Intersection over union (IoU) of rs-fMRI network activation images of unhealthy patients and Lesion mask

In order to understand the effect of the overlap between lesion mask and functional brain activation network, we conducted a statistical examination on the values

of IoU and the prediction of each test image, whether correctly classified or misclassified by our CNN model. We achieved this by finding the area of intersection between functional brain network activation and the brain tumor against the area of the union between the functional brain network activation and brain tumor as expressed in

$$IoU = \frac{a \cap b}{a \cup b}, \tag{4.4}$$

and visually illustrated in Figure 4.11.

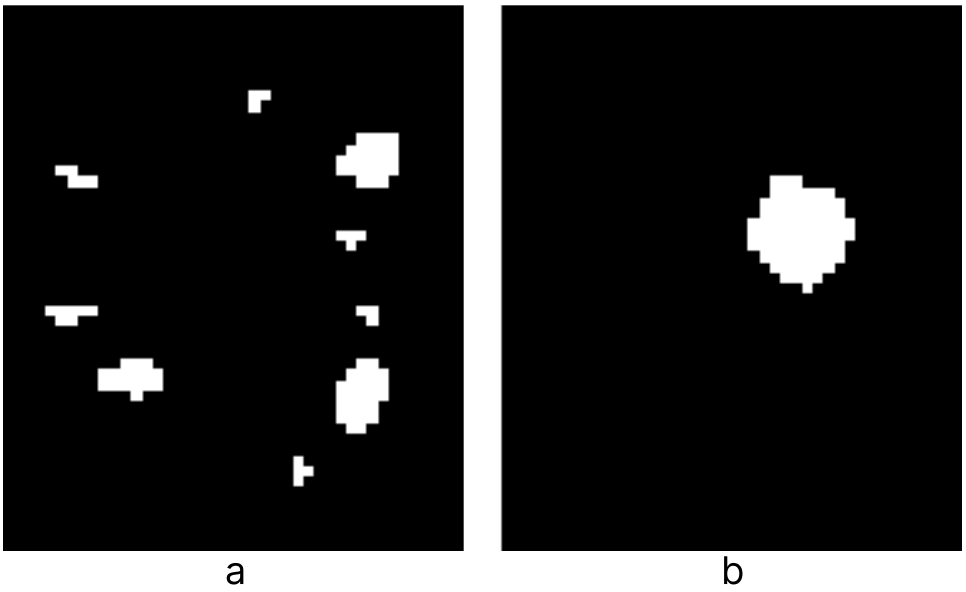


Figure 4.11 – Binarized image sample of **a)** network activation; **b)** lesion mask.

To understand the influence of tumor overlap with network activation and ability of our model to correctly classify functional brain network, we observed the skeweness value of both data at 2.1124 and 1.9409 for IoU of true positive (correctly classified images) and true negative (wrongly classified images) respectively, this indicates that, although the distribution is highly positively skewed, the two variables belong to the same statistical distribution. This fact is also supported by the visualization provided in Figure 4.12, which shows that there is no pattern between the value of IoU and model classifiability. Our investigation at this stage, only reveals the effect of overlap on functional network classifiability. Therefore, we extend our investigation to analyze the local relationship in activation region with lesion, and region without lesion overlap.

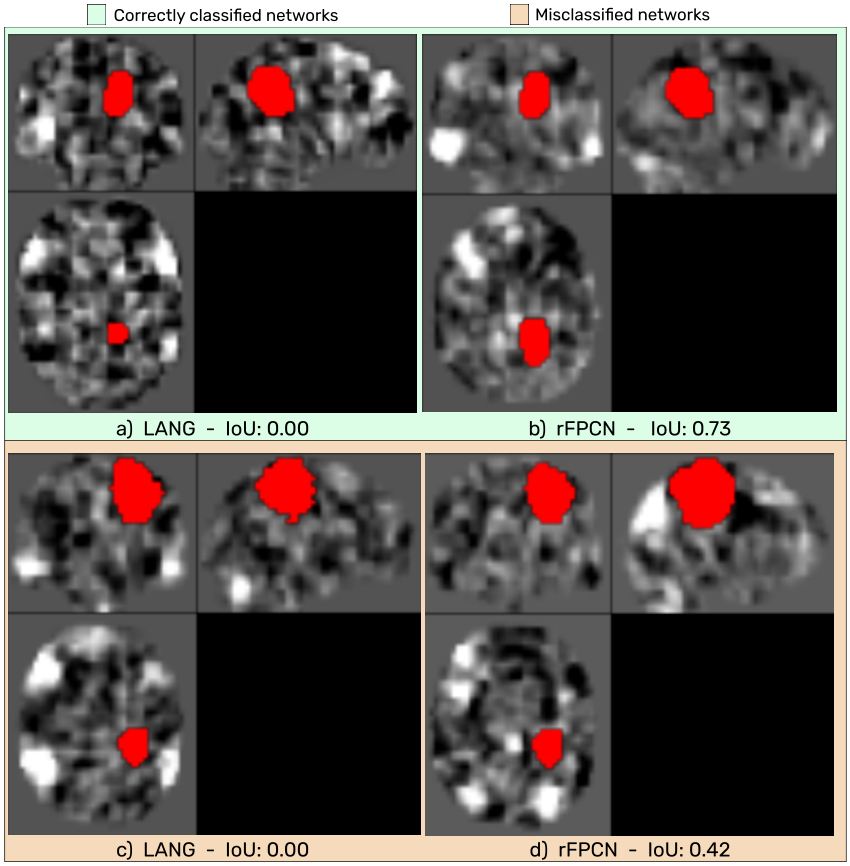


Figure 4.12 – Visualization of correctly and wrongly classified LANG and rFPCN networks with and without lesion overlap - a) correctly classified without overlap; b) correctly classified with overlap; c) wrongly classified without overlap; d) wrongly classified with overlap

.

4.3.4 Pixel intensity difference between region of network activation without lesion and region of network activation with lesion overlap

We performed pixel intensity analysis between region of network activation without lesion and region of activation with lesion in thresholded images. This observation could reveal whether the difference between healthy and unhealthy data emanated from the influence of tumor overlap with network activation or not (see figure 4.13). To evaluate the differences in the pixel intensity between these two regions, we ex-

pressed the non-negative difference $\mathbb{X}$ as in

$$\mathbb{X} = |\mu(\alpha) - \mu(\beta)| \tag{4.5}$$

where $\mathbb{X}$ is the pixel intensity difference between region of overlap α and non-overlap β with activation network.

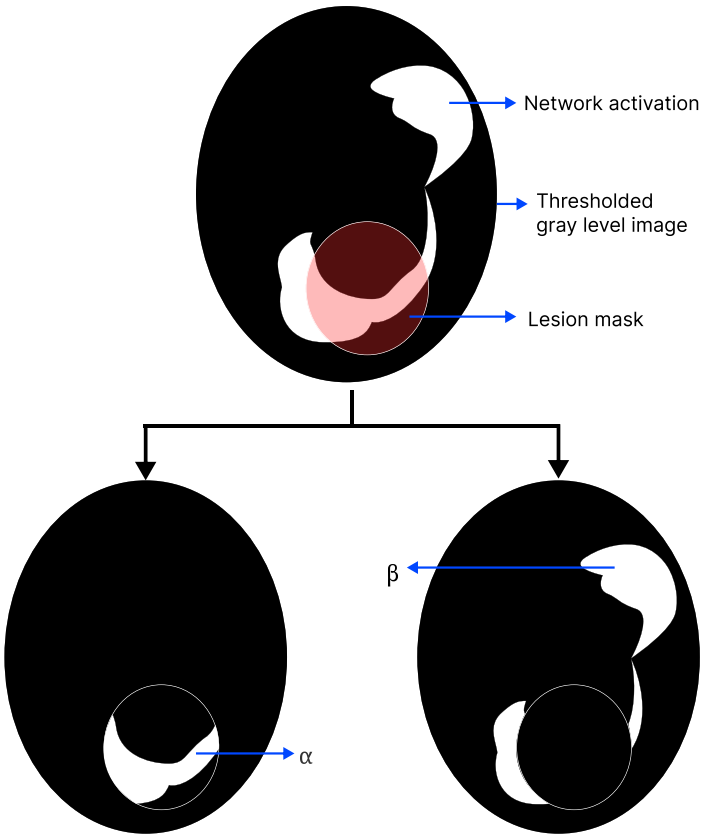


Figure 4.13 – Description of activation network separation between region of lesion overlap α and non-overlap β in a single brain network image.

In the statistical test conducted, result shows a chi-square p-value of 0.4858 (>0.05) from our pixel difference distribution whose intensity ranges from 0-10 across all region of activation as shown in Figure 4.14. This values therefore, demonstrate that the difference in pixel intensity in region of network activation with lesion overlap and region of network activation without lesion overlap is statistically insignificant.

Following the investigation on local relations between healthy and unhealthy data, we extend our findings to understand whether separate components in fMRI brain network activation map of healthy and unhealthy data observe any systematic spatial

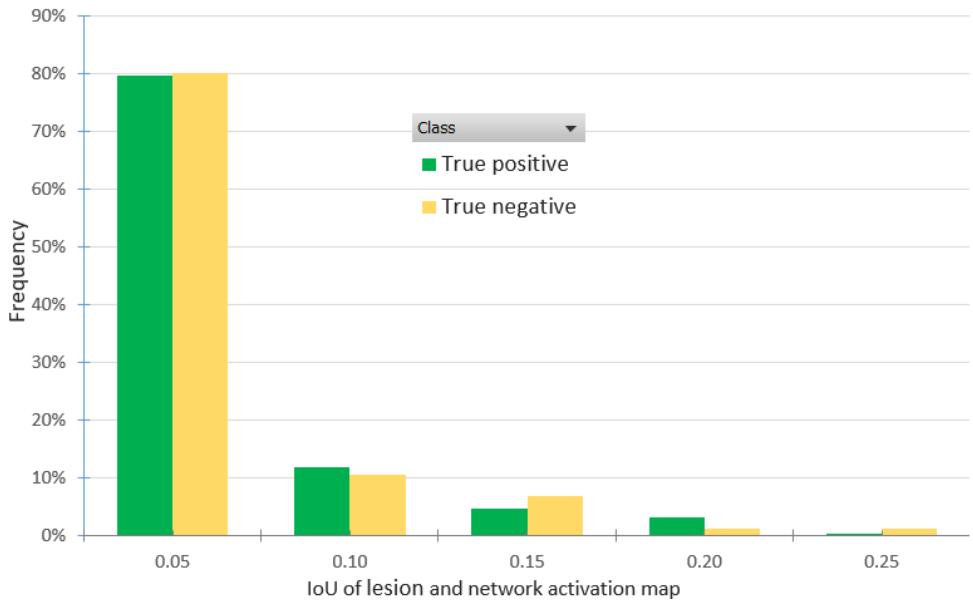


Figure 4.14 – Normalized IoU of images (network activation and lesion mask) in true positive (correctly classified) and true negative (wrongly classified) cases in functional brain network classification.

pattern.

4.3.5 fMRI brain network activation components

We computed a basic feature concerning the number of component of the activation map in presence or absence of lesion as shown in Figure 4.15.

To ensure fair assessment between healthy and unhealthy data, we sampled through equal amount of data from the two groups. The result shown in Figure 4.16, suggests that the number of connected components across different functional brain networks were not significant and may not have contributed to the similarity or difference between healthy and unhealthy data.

4.3.6 fMRI brain network activation volume

Using the binary version of our image data, we computed the volume of functional brain network activation pixels for each corresponding network in healthy and unhealthy data. This was necessary to understand how the volume of pixels present in each network map, possibly influence model recognizable features, and consequently

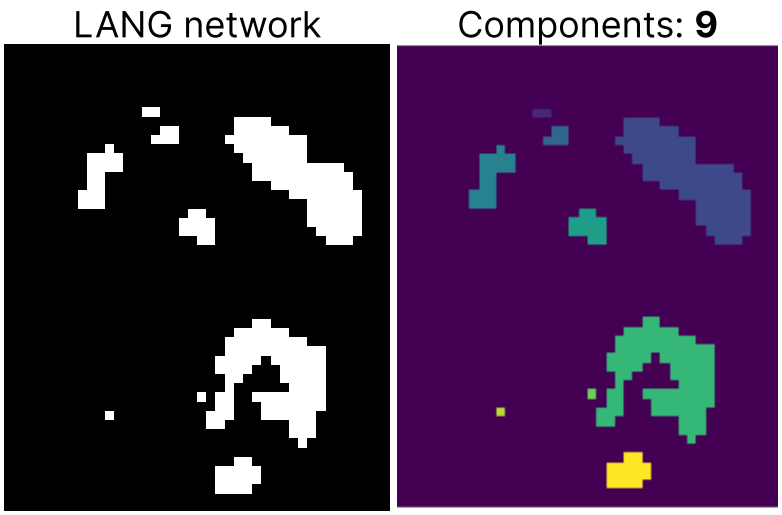


Figure 4.15 – Counting disconnected components in fMRI images of network activation map

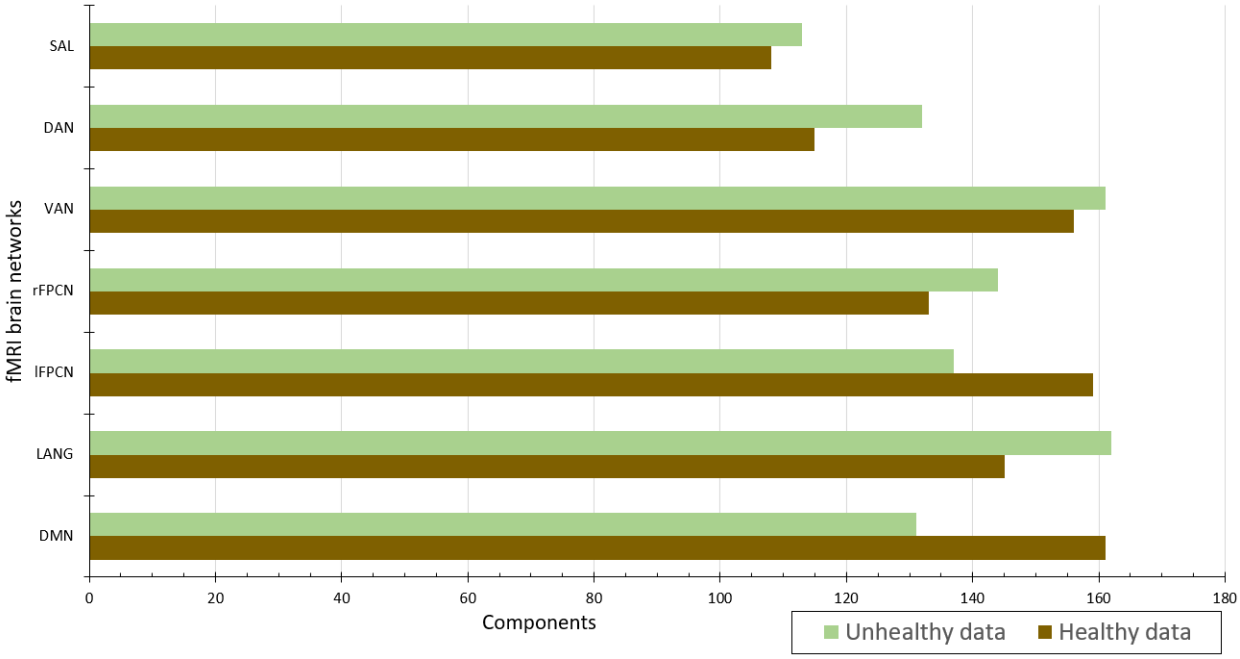


Figure 4.16 – Network map components size distribution across functional brain networks for healthy and unhealthy data.

fMRI network classifiability. To achieve this, the overall pixel volume representation of each functional network activation map which indicate activated pixels were collected (see Figure 4.11b). After computing the pixel volumes, we analyzed result for each corresponding network map for healthy and unhealthy data.

The observation of the 7 distinct functional brain network volume for the 55 subjects from both healthy and unhealthy data as shown in Figure 4.17 attracts our attention. This was because, all other discriminating factors we explored so far depends on the presence of the network feature representation while the brain network activation volume reveal the level of abundance of these features. We also extend our statistical test here, to understand whether the differences between these two volume distribution across 7 functional networks (see Figure 4.18).

We recorded a p-value of $0.00076 (< 0.05)$, which in fact shows that the difference in this volume distribution is statistically significant. Furthermore, a closer look into this volume distribution, reveals the big picture, which suggest that functional brain network activation pixel representation in healthy subject data is consistently more abundant than those observed in unhealthy data. This means that, more of the activation volumes as observed in healthy subjects data, supports better detection accuracy as obvious in our deep leaning models. On the other hand, fewer pixels activation value is responsible for the drop in accuracy and this can easily be attributed to the presence of tumor which may have overlapped network activation map (see Figure 4.18).

4.4 Discussion and conclusion

In this chapter, we investigated the origin of the feature transferability and data discrepancy as observed in our healthy and unhealthy fMRI dataset used for experiments in the previous chapters. While a convolutional approach demonstrated clear discriminability between the healthy and unhealthy data, we struggled to find simple features that could be used to measure this data discriminability. This further emphasize that, a similarity exist between them hence, transferability was possible.

To further measure the level of local or regional similarities or differences in our data, we explored a number of statistical experiments to establish whether the observed relationship in our data is significant or insignificant, as shown in Table 4.2. As a summary of our statistical tests, it is evident that all our tests to determine the significance of the possible differences in image features were negative, except the volume of activated pixels in healthy and unhealthy data, which reliably shows that, these images are indeed hard to distinguish at sub-regional level. The fact that a global feature such as the volume of activated voxels is significantly discriminative between healthy and unhealthy is coherent with the clinical knowledge that the presence of brain tumor

Figure 4.17 – fMRI network activation pixel volume distribution by different functional networks.

is expected to decrease the BOLD signal. Observation on the impact of the volume is also consistent with the observation made on the Grad-CAM which shows feature highlight in all network activation maps, i.e. as a global feature.

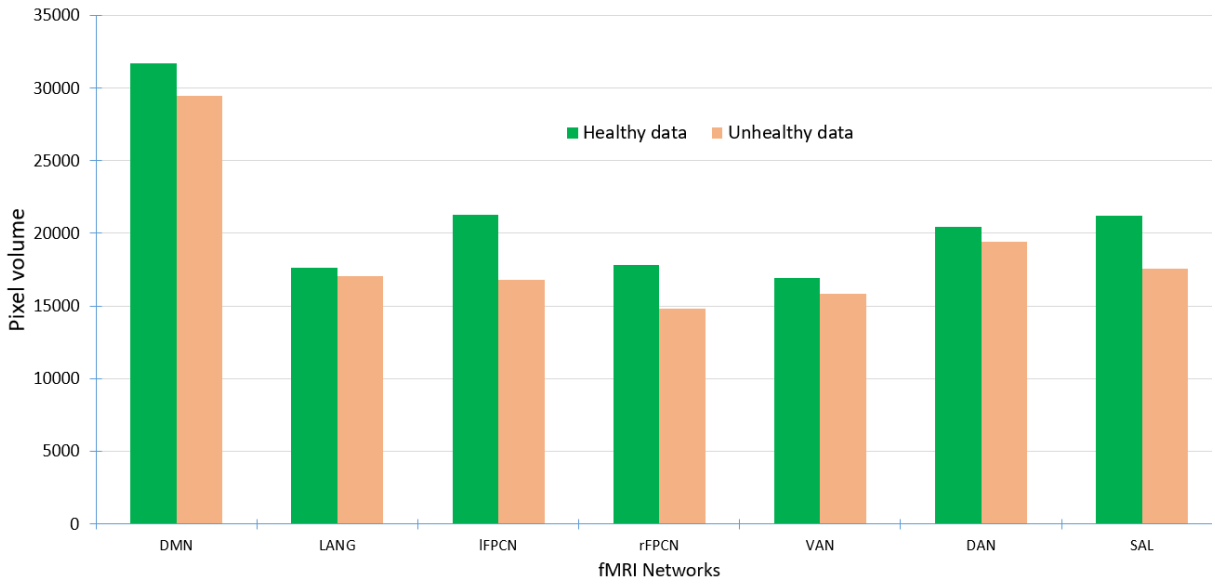


Figure 4.18 – Distribution of pixels volume in fMRI network activation map for healthy and unhealthy data.

Table 4.2 – Summary of statistical tests to discriminate between healthy and unhealthy fMRI data.

Statistical test description	**Results**
Lesion and non-lesion region	Statistically insignificant
Lesion and surrounding region of lesion	Statistically insignificant
Intersection over union of RSN network activation and brain tumor	Statistically similar
RSN activation with and without lesion overlap	Statistically insignificant
Activated pixels volume distribution in healthy and unhealthy data	Statistically significant

In summary, we investigated the identified fMRI data discrepancy between healthy and unhealthy images to provide a reliable relationship that explains why we were able to demonstrate transferability of healthy features to better recognize unhealthy data, and at the same time, why we observe a decrease in classification accuracy from our deep learning models.

CHAPTER 5

CONCLUSION AND PERSPECTIVES

5.1 Methodological contributions in medical imaging modalities

In this thesis, we contributed to neuroscience using computer vision and machine learning techniques to propose methodology for automatic functional brain network recognition by learning from provided healthy and unhealthy resting-state functional MRI (rs-fMRI) data. We have organize our contributions in three parts.

In a first contribution (chapter 2), we proposed an end-to-end deep learning algorithm for functional brain network classification. This model was evaluated with fMRI images from healthy control subjects and unhealthy (tumor) patients data. The main novelty of this work is that, we demonstrated feature transferability from healthy control data to improve model prediction accuracy in unhealthy patients data using transfer learning approach. This transfer learning is promising since healthy patient are easy to enroll in clinical studies. However, the limitation of such an approach is that healthy patient have to be annotated which represent a loss of time for clinicians. In a second contribution (chapter 3), we focused on addressing the problem of healthy data annotation in the process of predicting unhealthy data using contrastive self-supervision technique. Learning on healthy data to predict unhealthy data is rather counter intuitive and can be seen as paradoxical. In order to better understand what makes this transfer possible but also what makes healthy distinct from unhealthy data in resting state fMRI, we proposed in the third contribution (chapter 4), to investigate the prevailing discrepancy between healthy and unhealthy data in different strategic ways including latent space representation of convolutional network, local contrast, global volume of activated signals in functional networks, and number of disconnected network components.

5.2 Perspectives

The proposed end-to-end deep learning algorithm for functional brain network classification represents a significant step forward in the field of neuroimaging. By demonstrating the transferability of features learned from healthy control data to improve model prediction accuracy on unhealthy patient data using transfer learning approach, the model's accuracy can be improved significantly. Moreover, this study showed that learning on healthy data can scale the model prediction capability, since unhealthy data is generally limited. From a clinical point of view, future research could extend this work to identify the various types of brain tumors. At another end, from a signal processing point of view, we can process the BOLD signals for investigating voxel-level analysis in raw temporal activation signals as depicted in Figure 5.1.

One could also investigate the translation of our approach to other clinical applications. Specifically, the proposed self-supervision learning approach which allows pretraining on healthy data without the need for annotation is by nature generic. This method has the potential to reduce the laborious and time-consuming data annotation required for transfer learning, which is a significant advantage. Further research could investigate the potential of this method in other medical imaging applications. This could apply in all imaging modalities where the disease itself is not directly visible yet the contrasts of these imaging modalities constitute a necessity to visualize anatomical or functional areas of interest.

Another direction we envisioned to follow up our work, is the use of alternative encoding of data in order to reduce the computational cost of the proposed models. The results obtained with CNN on functional neural network are rather convincing. However, they are obtained with huge amount of parameters and one can wonder if more frugal end-to-end learning approach could deliver similar performances. Since we are targeting biological networks, one can think of an encoding of the images in the form of graphs connecting the main activated areas in the functional network. We started such investigations and present some preliminary results in Annex C related to this direction. Our initial results to reduce model parameters using dimension reduction and consequently, graph representation is very encouraging however, a better graph encoding approach could be explored to mitigate the observed accuracy loss in our graph neural network (GNN) model.

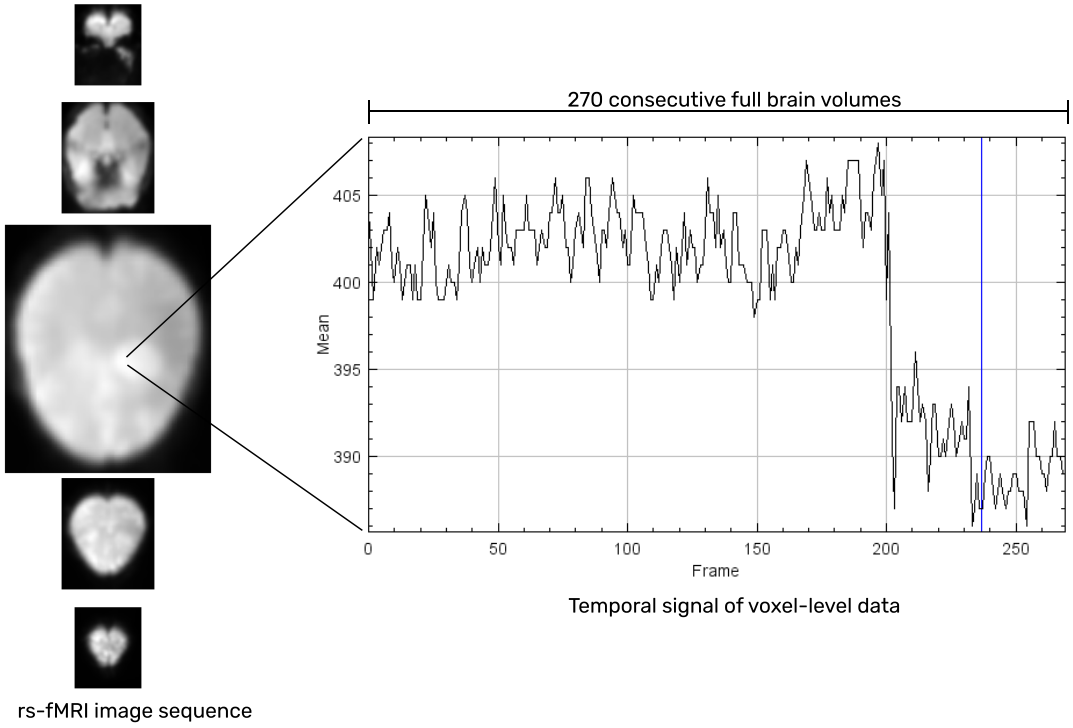


Figure 5.1 – Voxel-level analysis for more efficient data discrimination between healthy and unhealthy data.

5.3 Publications

Journal Articles

✍ Lukman E. Ismaila, Pejman Rasti, Florian Bernard , Mathieu Labriffe, Philippe Menei, Aram Ter Minassian, David Rousseau, and Jean-Michel Lemée: "Transfer Learning from Healthy to Unhealthy Patients for the Automated Classification of Functional Brain Networks in fMRI" MDPI Applied Sciences doi: `https://doi.org/10.3390/app12146925` (July 2022).

International conferences

✍ Lukman E. Ismaila, Pejman Rasti, Jean-Michel Lemee and David Rousseau: "Self-Supervised Learning for Functional Brain Networks Identification in fMRI from Healthy to Unhealthy Patients" 16th International Conference on Signal-Image Technology & Internet-Based Systems SITIS'22 (Oct 2022).

National conferences

- ✍ Lukman E. Ismaila, Pejman Rasti, Jean-Michel Lemee and David Rousseau: "Deep Learning Pour La Classification Automatique de Réseaux Cérébraux Fonctionnels Par Irmf de Repos" Journal of Neuroradiology doi: `https://doi.org/10.1016/j.neurad.2022.01.017` (Mar. 2022).
- ✍ Lukman E. Ismaila, Pejman Rasti, Jean-Michel Lemee and David Rousseau: "Toward more frugal models for functional cerebral networks automatic recognition with resting state fMRI" GRETSI Symposium Signal and Image Processing, 2023 (Accepted).

CHAPTER 6

ANNEX A: DEFINITION OF TERMINOLOGY

6.1 Machine learning

Machine learning is a subfield of artificial intelligence (AI), that involves the use of statistical algorithms to enable computer systems learn and improve from experience. Machine learning models are designed to automatically learn and improve from data, without being explicitly programmed for every task. By using patterns and statistical inference, these models can be trained to make predictions or decisions based on previously unseen data. The data are transcoded by feature vectors. The feature space is the reference frame where data are represented. Often this feature space is in $\mathbb{R}^n$ where $n > 3$. Dimension reduction techniques [116] such as principal component analysis (PCA) [117], t-distributed stochastic neighbor embedding (T-SNE) [118] and linear discriminant analysis (LDA) [119], are used to project the feature space from $\mathbb{R}^n$ to $\mathbb{R}^2$ or $\mathbb{R}^3$ as applicable in the case of our high dimensional fMRI images, to visualize the structure of data in the reduced space. There are several ways of learning the structure of the data in the feature space including supervised learning, unsupervised learning, self-supervised learning and reinforcement learning. Throughout this study, we investigated and addressed problems using supervised and variant of unsupervised machine learning (self-supervision). Supervised machine learning refers to the algorithms where you have input variables x and an output variable Y called labels, and you use an algorithm to learn the mapping function from the input to the label. The goal is to approximate the mapping function so well that you can predict the correct Y for a new input data x. On the other hand, unsupervised machine learning seeks to identify structure among unlabeled data x.

6.2 Deep learning

Deep learning is a type of machine learning that involves training artificial neural networks to learn from data. In this case, neural networks consist of layers of interconnected nodes that process and transform input data to produce an output. Deep learning algorithms use multiple (hidden) layers of these nodes to learn complex representations of data (see Figure 6.1). By learning these representations, deep learning models can perform tasks such as image classification, speech recognition, natural language processing, and more.

Deep learning has become increasingly popular in recent years due to its ability to handle large, complex datasets and its state-of-the-art performance on various tasks. Some examples of deep learning applications include highly sensitive intelligent or predictive systems, virtual assistants, medical imaging and computer aided diagnosis, and image and speech recognition.

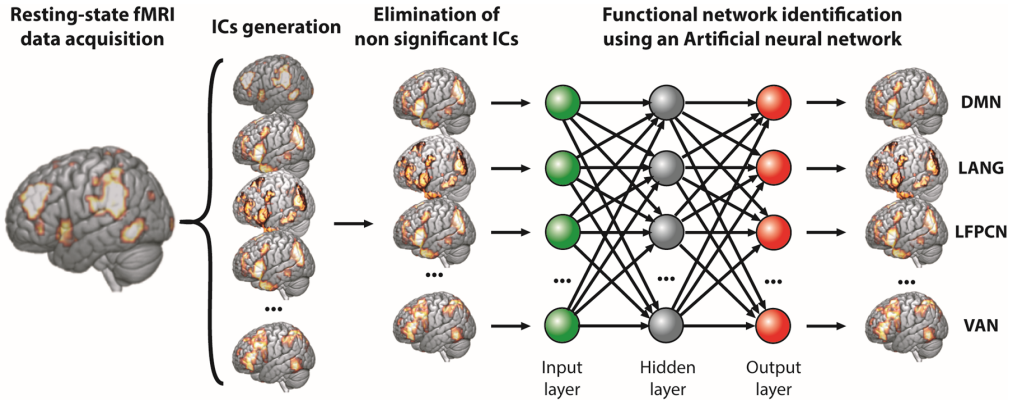


Figure 6.1 – Deep convolutional network model for rs-fMI network classification.

Although traditional machine learning algorithms are resilient, they still require human intervention to establish the features, and there is no guarantee that the characteristics chosen are the most appropriate for solving a problem. In practice, traditional machine learning algorithms are frequently simple and have few parameters to optimize during the training phase. As a result, they attain their full performance peak (see Figure 6.2) irrestpective of the additional amount of training data size.

Deep learning, on the other hand, which is a subset of a family of machine learning methods based on artificial neural networks, improves with the addition of data. The complexity of deep learning algorithms, which have millions of parameters to change and hence require big datasets to be resilient, explains this. Deep learning architec-

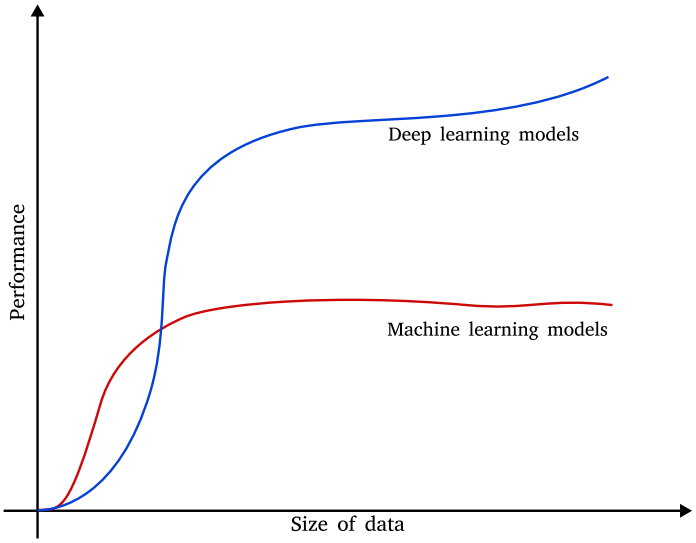


Figure 6.2 – Machine learning and Deep learning performance comparison in respect to data.

tures are made up of end-to-end raw data transformations that are chained together from top to bottom. Unlike traditional machine learning, deep learning architectures select features from raw data. Today, supervised machine learning methods, including deep learning, are the most robust in computer vision applications. They do, however, necessitate image annotation.

6.3 Transfer learning

Transfer learning is a machine learning technique where a model trained on one task is reused or adapted as a starting point for another related task. In transfer learning, the knowledge learned from one task, which typically involves a large dataset, is transferred to a different but related task where the data set is relatively smaller. This approach helps in situations where there may not be enough labeled data available for the new task, or training from scratch would be computationally expensive or time-consuming.

For example, a model trained on a large dataset for object recognition can be used as a starting point for another task, such as image classification or object detection (see Figure 6.3). The pre-trained model can be fine-tuned on the new task using the smaller dataset, or the model's weights can be frozen, and additional layers can be added to the

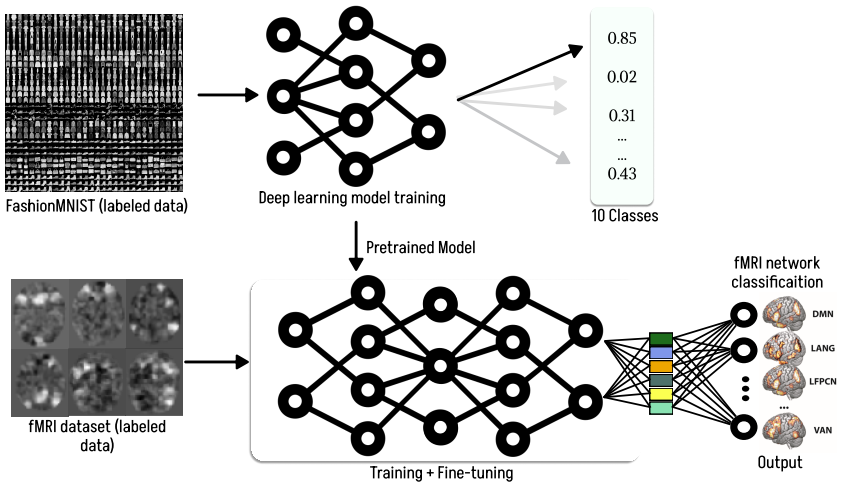


Figure 6.3 – Illustration of transfer learning strategy.

model to learn the new task's specific features. Transfer learning has been successfully applied to many applications, including computer vision, natural language processing, and speech recognition.

6.4 Self-supervision learning

Self-supervised learning is a technique in machine learning that enables a model to learn from the data itself, without requiring explicit supervision. The goal of self-supervised learning is to leverage the structure and patterns present in the data to generate supervisory signals that can be used to train the model (see Figure 6.4). Self-supervised learning has been successfully applied in various domains, including natural language processing, computer vision, and speech recognition.

One of the most popular approaches for self-supervised learning in computer vision is contrastive learning, where the model is trained to distinguish between pairs of similar and dissimilar examples. For instance, in the case of image data, the model is trained to distinguish between pairs of images that are taken from the same scene or object (positive examples) and those that are not (negative examples). This approach has been shown to be effective in pre-training image representations that can be fine-tuned for downstream tasks such as image classification and object detection.

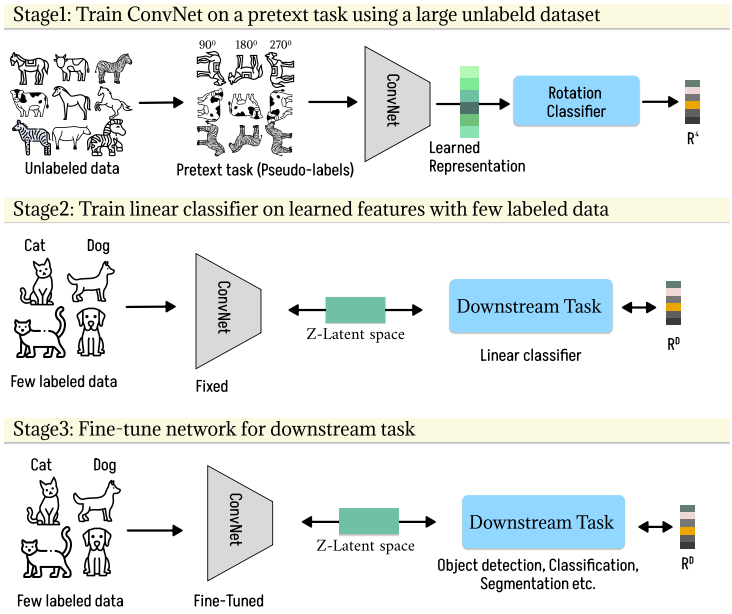


Figure 6.4 – Self-supervision learning on comparative dataset using context prediction.

6.5 Medical imaging

Medical imaging refers to the use of various technologies and techniques to create visual representations of the interior of the human body for clinical analysis and medical intervention. Medical imaging techniques include X-ray, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, and nuclear medicine imaging. These techniques can provide information about the structure, function, and physiological processes of the body, allowing medical professionals to diagnose and monitor diseases and injuries, plan and guide treatments, and evaluate the effectiveness of interventions (see Figure 6.5). Medical imaging has revolutionized the practice of medicine, enabling non-invasive and precise diagnosis, reducing the need for exploratory surgery and improving patient outcomes.

6.6 Image classification

Image classification is a fundamental task in computer vision that involves predicting a label or category for an input image. The goal of image classification is to build a model that can accurately assign a label to an image based on its visual features. In recent years, deep learning has emerged as a powerful technique for image classification,

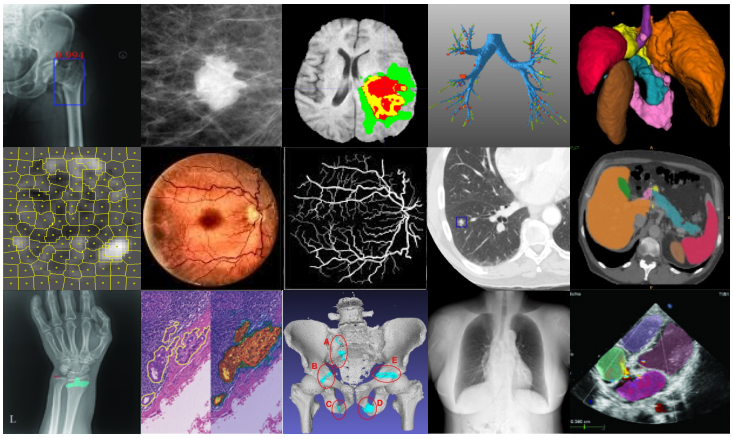


Figure 6.5 – Application of computer vision in medical imaging.

with convolutional neural networks (CNNs) being the most popular deep learning approach for this task. CNNs consist of multiple layers of learnable filters that extract increasingly complex and abstract features from an input image.

Image classification (see Figure 6.1) refers to a predictive modeling problem where a class label is predicted for a given functional brain network image. A model will use the training dataset (fMRI images and labels) and will calculate how to best map examples of input data to specific class labels. Some popular examples of classification algorithms are: k-Nearest Neighbors [120], Decision Trees [121], Support Vector Machine [122], Naive Bayes [123] and convolutional neural network (CNN) [124].

6.7 Graph representation learning

Graph representation learning is a popular research area in machine learning that aims to learn low-dimensional vector representations of nodes or subgraphs in a graph, capturing the structural and relational information of the graph. Node embedding is one of the popular approaches for graph representation learning that learns a low-dimensional representation of each node that can capture its structural and relational context in the graph. Several methods for node embedding include matrix factorization-based methods, random walk-based methods, and neural network-based methods such as Graph Convolutional Networks (GCNs).

One of the challenges in graph representation learning is to learn node embeddings that capture both local and global information of the graph. Recent research has pro-

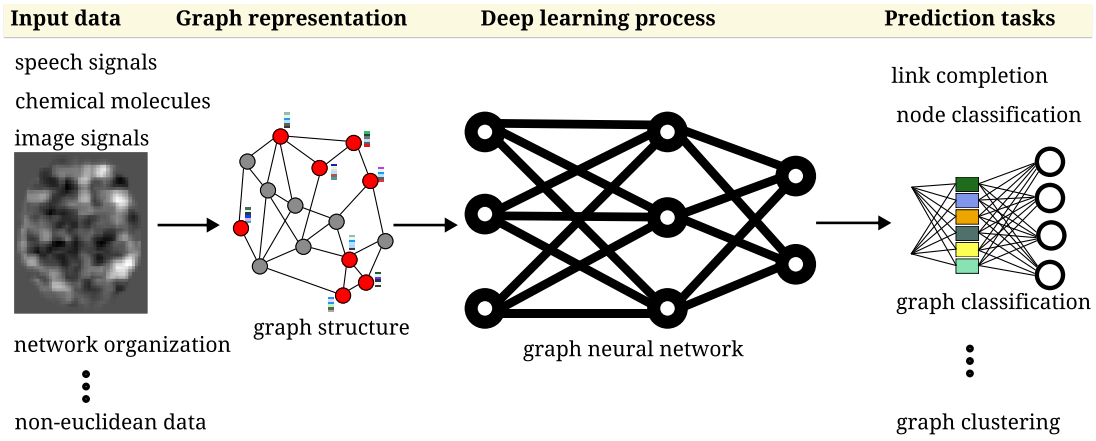


Figure 6.6 – Critical stages of graph representation learning method.

posed several methods to address this challenge, such as the Graph Attention Network (GAT) [125] and the GraphSAGE algorithm [126]. GAT uses attention mechanisms to learn node embeddings that capture the importance of different nodes and edges in the graph. GraphSAGE algorithm uses a sampling strategy to generate subgraphs around each node, then applies a multi-layer perceptron to learn node embeddings that capture both the local and global context of the subgraph. These methods have shown promising results in various applications, such as node classification, link prediction, and graph classification (see Figure 6.6).

CHAPTER 7

ANNEX B: fMRI FUNCTIONAL BRAIN IMAGE DATA ACQUISITION

7.1 Introduction

In this section, we provide background details of functional magnetic resonance imaging (fMRI) data and discuss the various steps followed to acquired the data and preprocess our resting state fMRI (rs-fMRI) images for machine learning application.

Certain brain functions, such as motor skills, vision, and memory, have constant inter-individual functional localization, but others, such as language, social cognition, and attention, exhibit substantial inter-individual anatomical heterogeneity as proposed by Vigneau *et al.* in there work [91]. In order to preserve these functional areas after tumor resection, it is required to design and implement in routine surgical practice techniques that allow clinicians to identify these areas and preserve them.

Neuro-monitoring and awake surgery allows intra-operative identification and preservation of cortical functional areas and bundles of white fibers, it also represent the standard of surgical therapy for brain lesions located in the eloquent zone. Preoperative identification of eloquent brain areas is necessary for surgical risk assessment, surgical planning, and to guide cortical mapping during surgery in order to preserve the neurological status of the patient and optimize the quality of surgical resection. magnetic resonance imaging (MRI) machine, is one of the preferred methods for detecting functioning brain regions and white fiber bundles (see Figure 7.1a).

The Blood oxygen level dependent (BOLD) imaging effect is a local increase in cerebral blood flow on the surface of activated cortical areas that is observable in MRI by measuring the deoxyhemoglobin/oxyhemoglobin ratio. Once quantified, this BOLD impact can be studied using various approaches to identify functioning brain areas and utilized to plan eloquent brain region during neurosurgery [127].

Functional activation MRI is one of the modalities for identifying different func-

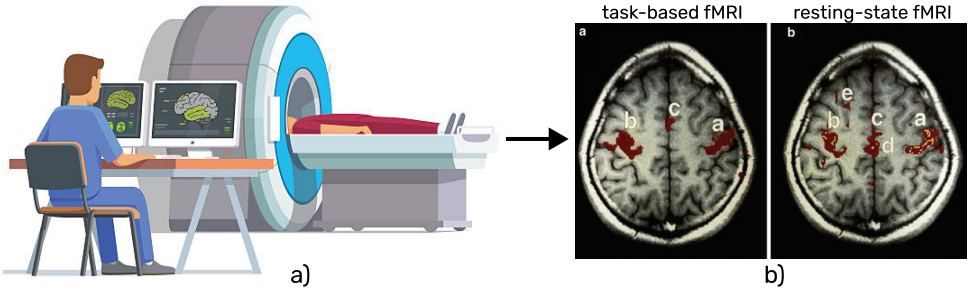


Figure 7.1 – Task-based and resting-state fMRI image acquisition with MRI machine (Fig. 7.3b [13]).

tional regions of the brain. This identification is based on a comparison of activation between periods of performing a task that is particular to the brain function we aim to identify and succeeding rest periods. The block paradigm (see Figure 7.2) is achieved by alternating between stages of stimulus and repose, and by comparing the variation of the BOLD signal with the anticipated theoretical hemodynamic response obtained by multiplying the canonical cerebral hemodynamic response by the block paradigm as proposed by [128] and illustrated in Figure 7.3, it is possible to statistically identify the brain regions that were specifically active during the task that was performed.

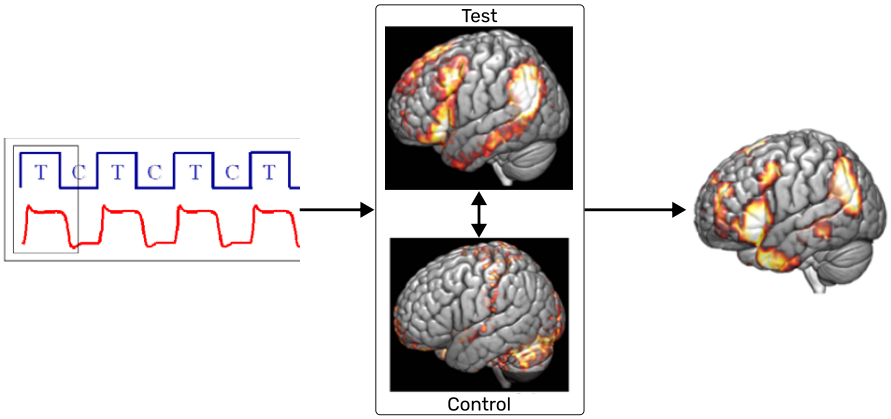


Figure 7.2 – MRI in block paradigm.

This fMRI method is the most widely used in everyday practice and is easy to setup. Unfortunately, there are a number of drawbacks to this method. It takes time to set up and have the patient complete a paradigm that is unique to each brain function whose functional brain areas we are trying to identify. This paradigm also places a great deal of emphasis on the patient's capacity to complete the requested task, which can

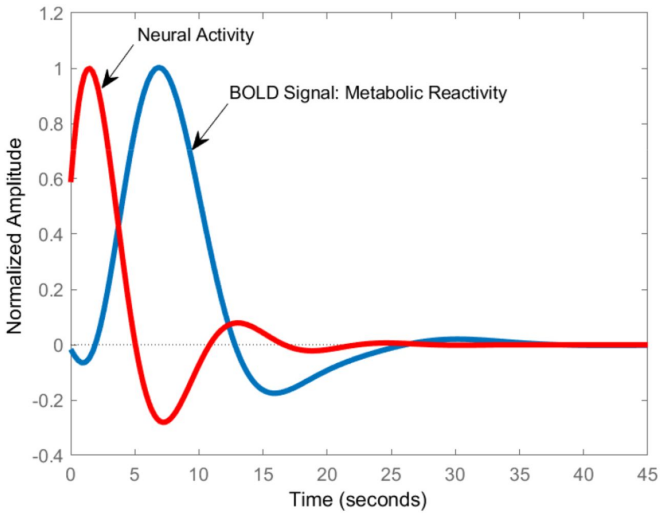


Figure 7.3 – Correspondence of neural function and metabolic reactivity in BOLD fMRI [128].

be challenging for patients who are young, old, claustrophobic, sedated patients, or moderately impaired before surgery.

The analysis of the synchronicity of spontaneous BOLD signal oscillation between brain areas during resting state fMRI (rs-fMRI) enables the identification of functional networks without performing any explicit tasks, in contrast to task-based fMRI, which studies the BOLD signal increase in brain areas during a language task to identify brain language areas.

By measuring the spontaneous fluctuations of the BOLD signal of various brain regions during a so-called "resting" MRI session, where the patient is only instructed to look at a fixation cross and let go of his internal thoughts, it is also feasible to identify functioning brain areas [27]. Without requiring the patient to perform tasks during the MRI, it is possible to identify intrinsic connectivity networks by performing an independent component analysis starting from the supposition that brain areas with synchronous variation of their BOLD signal in low frequencies (<0.1Hz) are a part of the same network.

Many studies [128, 129] have demonstrated an association between functional brain regions identified by activation MRI and cortical stimulation and intrinsic connection networks especially for language identification as shown in Figure 7.4. This analysis, which was done in separate components, creates brain maps of the regions with synchronous BOLD signal activity. These maps must then be manually inspected to

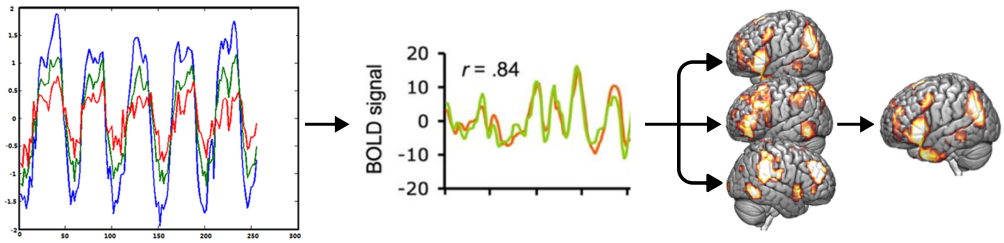


Figure 7.4 – Resting-state fMRI language network identification with Independent component analysis.

identify the networks of interest and assist the preservation of brain processes.

7.2 Clinical background and data acquisition stages

The data collected in the acquisition stage of our experiment is from two groups of participants, which are: healthy control subjects or volunteers and patients with tumors. While healthy data was acquired from the volunteer group, unhealthy data were obtained from patients with brain tumors with a specific lesion region as indicated by the provided binary lesion mask. A further description of the unhealthy population is provided in [6]. This data acquisition experiment, is a single-center, prospective, open-label trial, in compliance with regulation and ethical guidelines for clinical research, approved by the local ethics committee (Comité de protection des personnes Ouest II, decision reference CPP 2012-25). Eighty-one healthy volunteers (36 females and 45 males) aged from 23 to 38 years old were included and signed written informed consent. Fifty-five adult patients with a brain lesion treated in the Department of Neurosurgery of the university hospital of Angers (CHU Angers) underwent a preoperative fMRI language mapping with both rs-fMRI and task fMRI as well as a perioperative cortical mapping of eloquent brain language areas in awake condition. All subjects gave their written, informed consent before enrolling in this study.

Summary of Preprocessing stages:

a) Data acquisition: Resting-state fMRI scans are acquired using magnetic resonance imaging (MRI) techniques, capturing the spontaneous neural activity of the brain in the absence of specific tasks.

b) Preprocessing: The acquired fMRI data undergo several preprocessing steps to correct for various artifacts and noise. These steps may include slice-timing correction

to align slices in time, motion correction to compensate for head motion during the scan, and spatial normalization to align the data to a standard brain template.

c) Brain extraction: The next step involves extracting the brain region from the fMRI data while removing non-brain tissue, such as skull and scalp, to focus only on the brain activity.

d) Spatial smoothing: Smoothing is applied to the fMRI data to reduce noise and enhance the signal-to-noise ratio. This is typically achieved by convolving the data with a spatially defined kernel.

e) Temporal filtering: Low-frequency drifts and high-frequency noise are removed through temporal filtering. Common approaches involve high-pass filtering to remove slow drifts and low-pass filtering to eliminate high-frequency noise.

f) Independent component analysis (ICA): ICA is applied to the preprocessed fMRI data to identify independent components or spatially distinct brain networks. ICA separates the fMRI signal into spatially independent components, each representing a distinct pattern of neural activity.

g) Identification of functional brain networks: After performing ICA, the independent components corresponding to functional brain networks are identified based on their spatial patterns and known neuroanatomical information. These networks represent coherent patterns of activity across different brain regions.

h) Activation map image volume generation: Activation maps are generated by quantifying the strength of connectivity or functional activity within each identified brain network. This involves calculating measures such as correlation coefficients or z-scores to represent the level of activity or connectivity within each network.

A process diagram show the complete flow of preprocessing steps in resting-state fMRI from acquisition to independent component identification and functional brain network activation map image volume generation is illustrated in Figure 7.5.

7.2.1 Data preprocessing

In this experiment, all datasets were acquired on a 3.0 Tesla MR Scanner (Magnetom® Skyra Medical Systems™). During image acquisition, patients laid supine with the head immobilized by foam pads and straps, with earphones, and kept in darkness. Patients watched a black screen with a red fixation cross in the center through a prism.

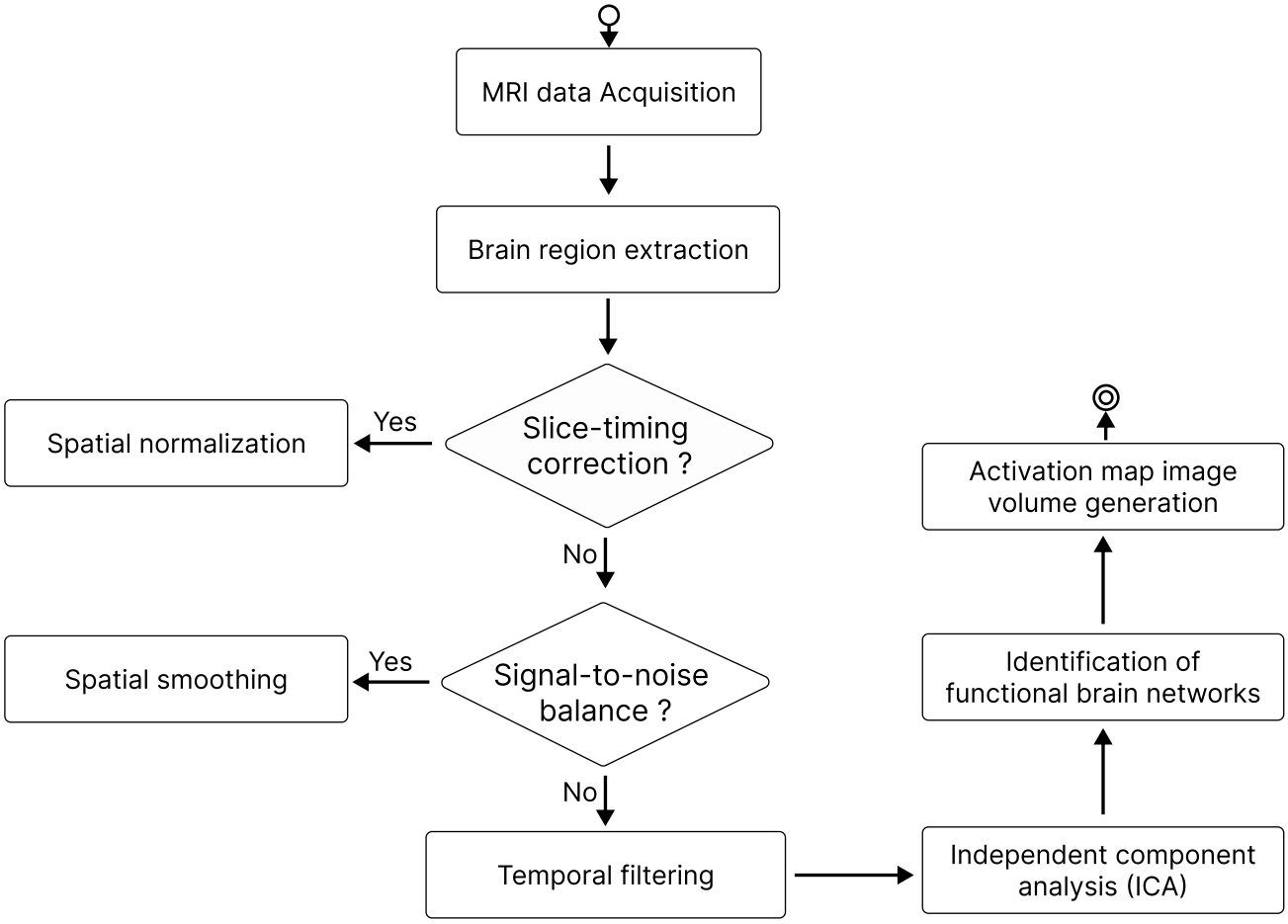


Figure 7.5 – Summarized flow of fMRI data collection and preprocessing stages.

The first three acquisition volumes in each functional series were discarded, to allow the longitudinal magnetization to stabilize. Preprocessing was carried out using SPM8 (Wellcome Department of Imaging Neuroscience, University College, London, UK, `http://www.fil.ion.ucl.ac.uk/spm`) running under MATLAB (The MathWorks). Each patient's native space images were corrected for time delays between slices. Then, all images were realigned to the first volume of the first session and unwrapped to correct head movement and susceptibility distortions. The three-dimensional dataset was segmented in native space, using the VBM 8.0 toolbox for statistical parametric mapping (SPM®) and co-registered to the mean functional image using gray matter segmentation as a reference image. The coregistered gray matter segmentation was then used to spatially normalize data into a standard template provided by the Montreal Neurological Institute (MNI template) [130]. Finally, the images were spa-

tially smoothed with a 6-mm kernel of full width at half-maximum. + Echo planar imaging (EPI) sequence was used for each fMRI with the following parameters TR = 2,280 ms, TE = 30 ms, flip angle = 90°, 42 axial interleaved slice of 4 mm slice thickness, in-plane matrix = 64 × 64 with a field of view = 168 × 187 mm, yielding a voxel size of 3 × 3 × 4 mm^3, covering the whole brain including the cerebellum. During task fMRI, we acquired 270 functional volumes per session over two sessions, and for rsfMRI, we acquired 270 functional volumes over one session. A T1-weighted anatomical three-dimensional dataset was also obtained, covering the whole brain to coregister and normalize EPI images, with the following parameters: 192 contiguous sagittal slices, in-plane matrix 256 × 256, yielding a voxel size of 1 × 1 × 1 mm^3.

For task fMRI analysis, the two conditions were the two successive epochs of a trial: TL and SG. A generalized linear model approach was used with regressors corresponding to each of the two conditions SG and TL convolved with a model of canonical hemodynamic response incorporated in the SPM8 package. Each individual time series of the preprocessed datasets was then analyzed by voxel-wise multiple regression. Low-frequency noise was removed by 128-s cutoff high-pass filtering. No global signal normalization was applied.

For rs-fMRI data analysis, a spatial independent component analysis (SICA) approach was used, employing a customized version of the Infomax algorithm running under MATLAB, for the identification of large-scale networks [131]. Fifty-five spatial independent components (ICs) were computed on preprocessed images of each individual run. Individual spatial components were thresholded at z = 2 which produces two variants of images the gray scale (tMaps) and thresholded images.

For all healthy and unhealthy data, we have extracted 55 features independent component analysis (ICA) with a specific interest in 7 brain features. One of the main difficulties with independent component analysis in resting-state fMRI is the determination of the total number of components (TNC) to be used, which may lead to suboptimal decompositions with the merging of multiple networks in case of low TNC, or the fragmentation of a functional network into multiple components in case of high TNC [132, 133]. Our choice to analyze 55 ICs among all patients was based on previous works and appeared to be a good compromise to identify functional brain networks. [134, 58]

Matlab function to merge tmaps images During this PhD, we extend our effort to support the preprocessing stage of resting-state fMRI data, by creating a matlab mod-

ule to assist the neurosurgeon to generate a 4-dimensional version (4D.nii) of the high dynamic range (.hdr) fMRI images faster. We provide a simple flowchart in Figure 7.6 to describe how our algorithm works. This function was provided through a intuitive and non-technical user interface as show in Figure 7.7, and it is available as a standalone application for windows, Linux Ubuntu and Mac operating system. Early feedback from clinician who used this matlab function, emphasize how much it saves time in the prepossessing stage.

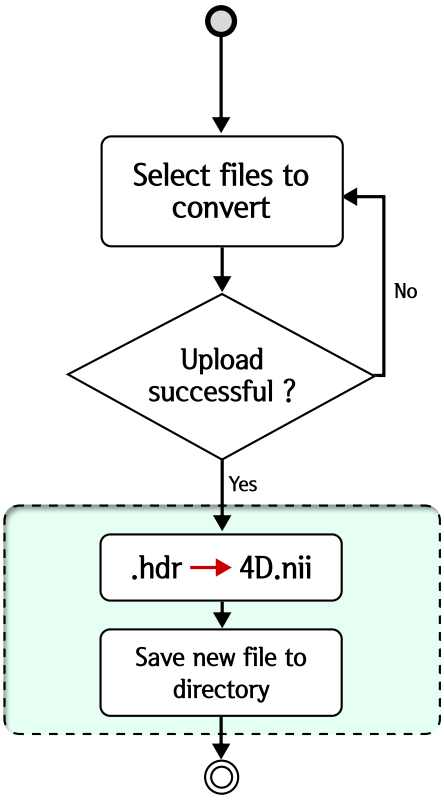


Figure 7.6 – Flowchart of Matlab module for merging 3D fMRI images.

fMRI image dataset description

The extracted 7 brain features correspond to 7 biological networks of the brain, which are the Language Network (LANG), Salience Network (SAL), Ventral Attention Network (VAN), Default Mode Network (DMN), Left Fronto-parietal Control (lFPCN), Right Fronto-parietal Control Network (rFPCN), Dorsal Attention Network (DAN). The seven brain features that were chosen to reflect the key ICN found and discussed in the literature about resting-state fMRI [135, 133]. Because of the inter-individual heterogeneity that makes it difficult to discover using detection tools or by non-expert

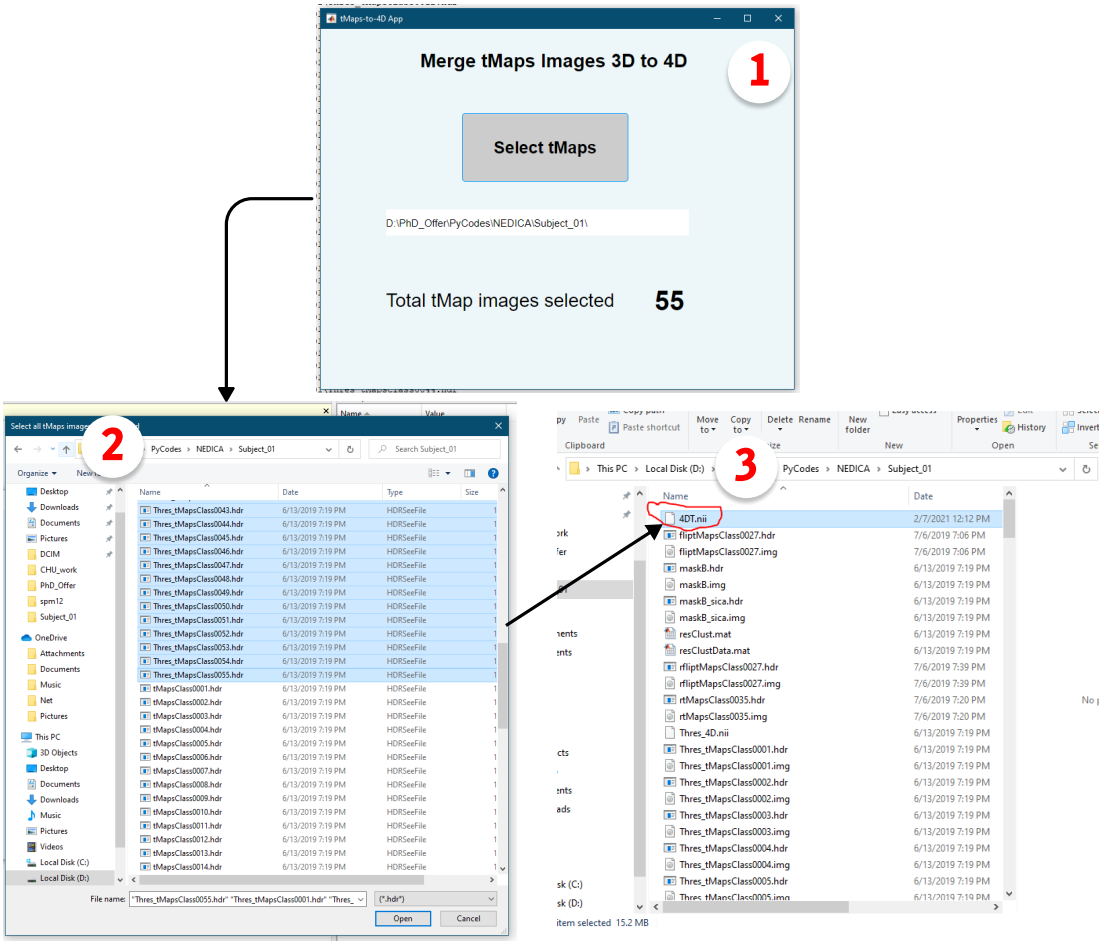
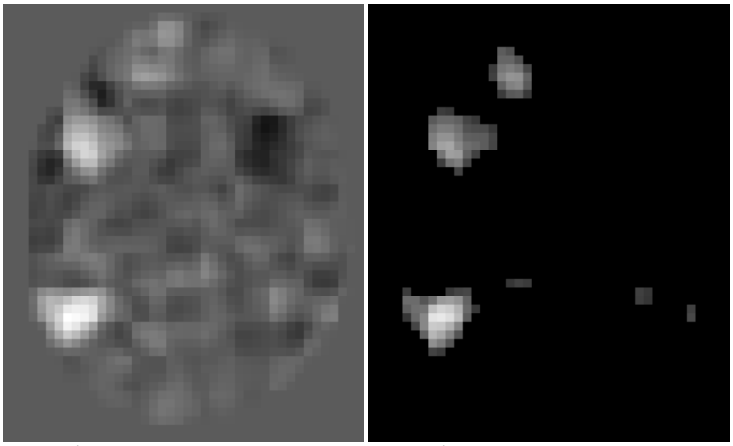


Figure 7.7 – Process illustration of 4D image generator function.

reviewers, these specific networks were chosen for the DMN to serve as a control for the others. This was done in order to ensure the accuracy of the results. These connectivity networks match to recognized functional networks that are critical for maintaining cognitive processes and have been incorporated into pre-surgical planning. The connectivity networks were also found to be consistent between rs-fMRI various fMRI data acquisition and analysis techniques [136]. Functional networks without anatomical variabilities, such as the motor, sensory, or visual cortex, were not considered for algorithm training and automated identification because of their consistent anatomical location.

Image labels for each healthy and unhealthy data file marked by domain experts were used to assign each image to its respective network class. In addition to the two variants of network images provided for both healthy and unhealthy data as shown in



a) gray level image b) threshold image

Figure 7.8 – Output rs-fMRI functional brain network activation images in 2 variants (a)-gray level and 2)-threshold) after preprocessing stage.

Figure 7.8, unhealthy data includes details of the brain tumor as described in Table 7.1 and shown in Figure 7.9.

Table 7.1 – Description of preprocessed fMRI data of unhealthy patients.

	Unhealthy Patient Image Data Description	
	Files provided	**Description**
1	Lesion.nii	This file is the binary mask for the brain tumour, each corresponds to a patients
2	Grey matter mask (mrwp1)	Is the mask for the grey matter (useful since the activation are all in the grey matter)
3	White matter mask (mrwp2)	Is the mask for the white matter (no activation inside the white matter, but may be a good way to estimate the brain deformations linked to the tumour and the peritumoural edema)
4	Cerebrospinal fluid mask (mrwp3)	The mask for the cerebrospinal fluid (like for the white matter, no activation inside, but may be useful to estimate brain deformations)
5	Whole brain-white grey matter (wms)	The whole brain (white and grey matter) in T1 anatomical MRI sequence, with the skin and skull clipped
6	Whole brain (wmrs)	This provides view of the whole brain cerebrospinal fluid, skull and skin included

7.3 Conclusion

In this Annex, we provide detailed description of the steps follow to acquire resting-state fMRI images of 7 selected functional brain network activations which are the main

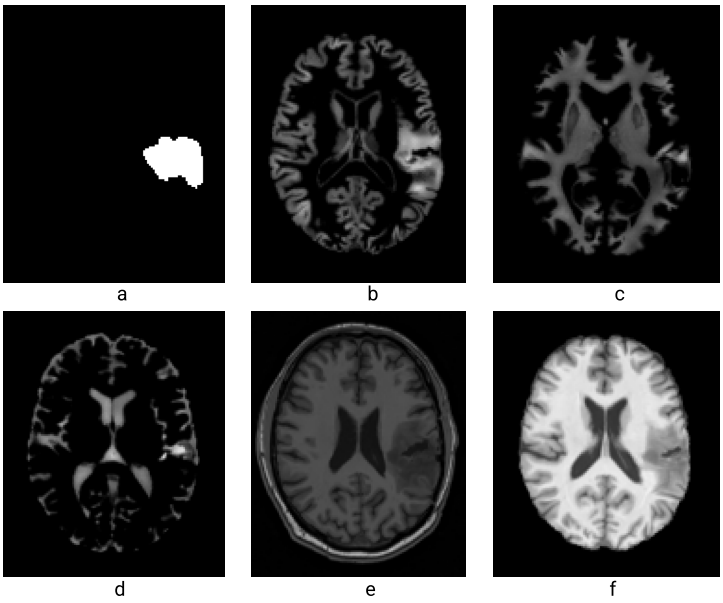


Figure 7.9 – Visual representation of the described data of unhealthy patients in Table 7.1 (a is the lesion mask, b is the grey matter mask; c is white matter mask; d is cerebrospinal fluid mask; e is whole brain, cerebrospinal fluid (skull and skin included); f is whole brain(white & grey matter).

data sources used in the experiments describe in other chapters of this documents.

The functional MRI approach described in this Annex allows the identification of functional connectivity networks in a single 15-minute resting MRI sequence, eliminating the need to implement an MRI sequence and an experimental paradigm for each function that needs to be evaluated. Additionally, this method enables the recognition of functional brain regions in individuals unable to complete activation tasks (children, elderly patients, confused and so on). The main drawback of resting MRI is that the maps produced by independent component analysis must be manually reviewed in order to find the maps of interest. This method requires a reviewer with extensive understanding of the anatomy of functional brain networks and the examination is meticulous, time-consuming, and requires a professional proofreader and the possibility of human error is to be expected because some functional networks' have very similar morphologies, such as the language network and the ventral attention network.

The preprocessing stage of our fMRI data, followed standard procedure according to the template provided by the Montreal Neurological Institute (MNI template) [130], for both healthy and unhealthy data. Our observation from data discrimination experiments from chapter 3, suggest that indeed, the observed discrepancy in this work

could be better managed by introducing recalibration based on the severity of brain tumor influes in the case of unhealthy data.

CHAPTER 8

ANNEX C: TOWARD MORE FRUGAL MODELS FOR FUNCTIONAL CEREBRAL NETWORKS AUTOMATIC RECOGNITION WITH RESTING STATE FMRI

In this Annex, we advance our investigation on one of the perspectives in the contributions of this PhD. In our proposed deep learning model in chapter 2, we anticipate that the addition of more data will increase model prediction accuracy, and to guarantee the feasibility of this objective, model parameters must be kept at minimal to ensure a simple and efficient model as well as avoid the possibility of overfitting. Unfortunately, this was not the case due to the fact that our input fMRI data are of high dimension, which has resulted in a large model parameters thus resulting in a complex model. An approach to simplify and efficiently represent the functional activation signals in our image data is required to ensure an optimized and scalable model.

In the following pages, we described our strategy to simplify our high dimensional data, and discussed our graph encoding strategy to demonstrate a more efficient representation with superpixels graphs from resting-state fMRI images of functional cerebral networks recognition.

Toward more frugal models for functional cerebral networks automatic recognition with resting state fMRI

Lukman E. ISMAILA[1], Pejman RASTI[1,2], Jean-Michel LEMÉE[3], David ROUSSEAU[1*]

[1]LARIS, UMR INRAe, IRHS Angers, Université d'Angers, France

[2]CERADE ESAIP, École d'Ingénieurs

[3]Service de Neurochirurgie CHU d'Angers, France

david.rousseau@univ-angers.fr

Résumé – Nous considérons une situation d'apprentissage machine où des modèles à base de réseaux de neurones convolutionnels classiques ont montré de bonnes performance. Nous investiguons différentes techniques d'encodage sous forme de supervoxel puis de graphes pour réduire la complexité du modèle tout en limitant la perte de performance. Cette approche est illustrée sur une tâche de reconnaissance de réseaux fonctionnels de repos pour des patients atteints de tumeurs cérébrales. Les graphes encodant des supervoxels préservent les caractéristiques d'activation des réseaux cérébraux fonctionnels à partir des images, réduisant les paramètres du modèle de 26 fois tout en maintenant les performances du modèle CNN.

Abstract – We refer to a machine learning situation where models based on classical convolutional neural networks have shown good performance. We are investigating different encoding techniques in the form of supervoxels, then graphs to reduce the complexity of the model while tracking the loss of performance. This approach is illustrated on a recognition task of resting-state functional networks for patients with brain tumors. Graphs encoding supervoxels preserve activation characteristics of functional brain networks from images, optimize model parameters by 26 times while maintaining CNN model performance.

1 Introduction

Convolutional neural networks (CNN) are powerful tools to perform computer vision tasks. CNN are however very demanding in terms of energy, data and annotation due to the large amount of parameters to be tuned during their training. These limitations are specially important in medical imaging where the constitution of large cohorts of unhealthy patients can be a bottleneck as frequently observed in cases of rare diseases like brain tumor. Recently, we have shown the possibility to circumvent this limitation by the use of transfer learning from self-supervised training on healthy data to unhealthy data [1]. We used small data in our experiments, and approach opens the possibility for scalability when a larger model is trained from additional data acquired.

This was obtained for the automatic recognition of functional cerebral networks via resting-state functional magnetic resonance imaging (rs-fMRI) [2] for patient with brain tumors. The CNN architecture proposed for the classification of functional brain network with 3D fMRI images by Ismaila *et al.* , was observed with high model training parameters despite the small data size [2] which constitutes a complex model and struggles with risks of overfitting.

In this work, we test possible ways to simplify deep learning models by reducing the overall parameter size. To this purpose, we propose to compare a basic CNN method with the approach depicted in Fig. 1. Based on a recent work by Gousia *et al.* , which highlighted the benefits of graph encoding in optimizing CNN model parameters especially in medical imaging [3]. We investigate various ways of encoding the rs-fMRI 3D volume data in more compacted fashions and systematically compare our observation with the performance obtained in [2]. This effort only represent an initial attempt towards more efficient encoding of our brain volume images, as well as opens the possibility for scalability when a larger model is trained from additional data acquired.

2 Database

fMRI brain network activation image data of 81 healthy subjects and 55 unhealthy patients were collected. Regular volunteers provide the healthy data, while patients with brain tumors where a binary mask indicate region of lesion in the brain constitute the unhealthy data. This analysis, was done in separate components which creates brain maps of the regions with synchronous blood oxygen level dependent (BOLD) signal activity. In the data acquisition stage, we extracted the intrinsic connectivity networks (ICNs) by using methods that combine the information of both the temporal and spatial dimensions, such as independent component analysis. The extracted signals represent the neuro-anatomical basis for the functional

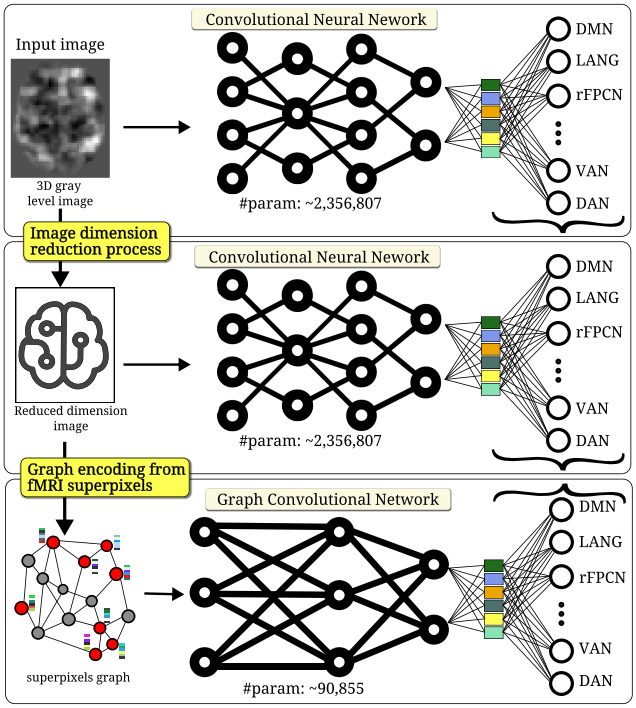


Figure 1: Visual abstract of our method. We consider as baseline performance either in terms of number of parameters and accuracy of CNN applied on raw rs-fMRI activation maps for functional cerebral network automatic recognition. We compare this performance with neural networks applied on compacted versions of the images.

networks in the brain [4]. The statistical parametric mapping (SPM) anatomy toolbox for Matlab was used to generate the 3D brain volume images, from the initial spatio-temporal fMRI signals. Among the 55 ICNs processed for each patients, 7 of these signals where recognized manually by experts to be biological networks of the brain such as Default Mode Network (DMN), Language Network (LANG), Right Fronto-parietal Control Network (rFPCN), Left Fronto-parietal Control (lFPCN), Salience Network (SAL), Dorsal Attention Network (DAN) and Ventral Attention Network (VAN). The annotated images were used in two versions: full images (connectivity map) and corresponding thresholded images.

3 Spatial dimension reduction

One may wonder if the entire 3D volume in gray levels is fully informative for automatic recognition of the functional cerebral networks. Several dimension reduction approaches can be envisioned. From the acquired brain volumes of resting-state fMRI images $42px \times 51px \times 34channels$, we normalized the pixel intensity range to 0-1 and computed several reduced version of these raw data as depicted in Fig. 2.

First, one can reduce the number of spatial dimension via a projection. We produced 2D gray level image by performing *Mean* operation on pixel intensity across the axial (A) plane as shown in Fig. 2 Secondly, to understand whether the intensity of the activation map holds discriminative information, we created 2D binary images by performing an *OR* operation in respect to sagittal, coronal and axial (SCA) plane respectively, which were further stacked together to provide SCA binary stack image. Also, we performed another *OR* operation across the axial plane to obtain a 3D binary volume image which overall, resulted in 4 variants of generated images as illustrated in Fig. 2. Lastly, we tested if the full resolution of voxels is necessary for the classification of the functional network, which are rather formed by large structures than fine details. To this purpose, segmentation of the gray level activation map was performed using SLIC algorithm [5, 6]. We processed the 2D segmented labels to obtain a superpixels image, while the 3D segmented labels provided the supervoxels image as shown in Fig. 4. Furthermore, we averaged (smoothened) the pixel intensities within each segment of our superpixels and supervoxels images. This step allows us to evaluate the integrity of the functional brain network features which was done by training a CNN model for 7 distinct functional brain network classification using the generated superpixels/supervoxels images.

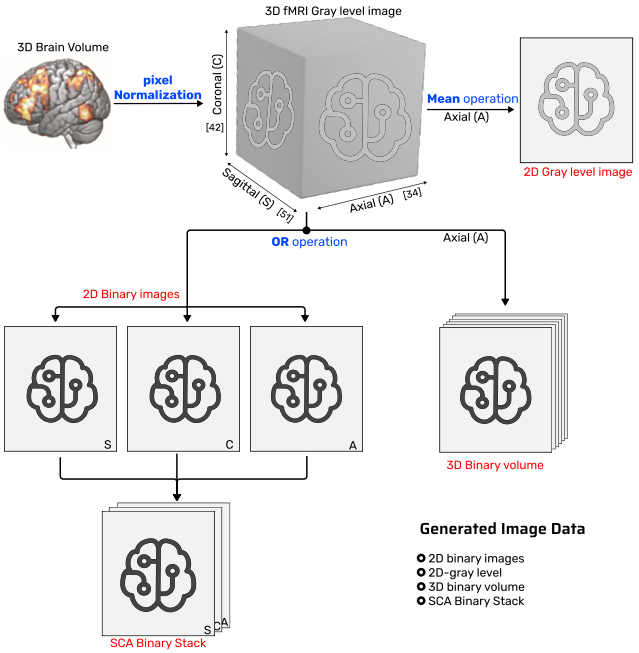


Figure 2: fMRI image dimension reduction process.

Table 1: CNN based fMRI brain network classification with unhealthy data.

Data	Train-Test	Accuracy	Parameters
3D gray level	315-70	0.75 ± 0.01	2,356,807
3D binary	315-70	0.66 ± 0.02	2,356,807
2D gray level	315-70	0.68 ± 0.01	2,337,799
2D binary	315-70	0.63 ± 0.01	2,337,799
SCA-binary stack	315-70	0.68 ± 0.02	2,011,271

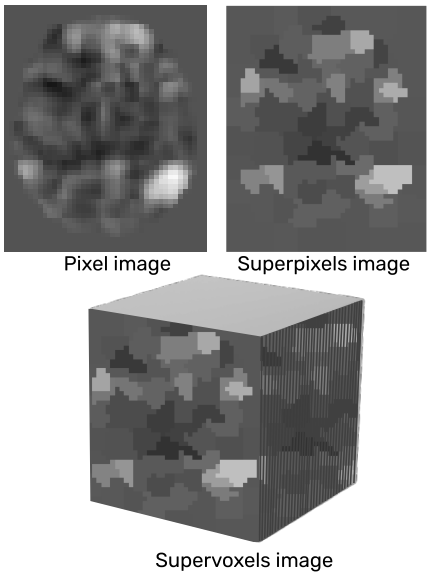


Figure 3: Pixel image segmentation into superpixels and supervoxels.

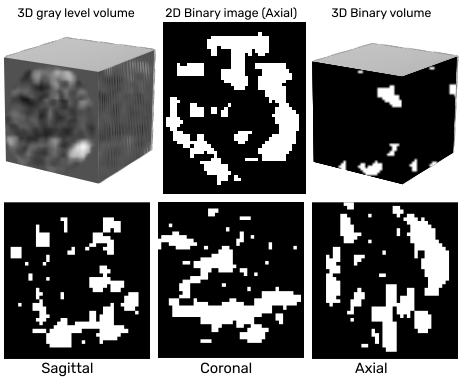


Figure 4: fMRI (LANG network) image spatial transformation.

When using the dimension reduction from 3D to 2D or from grey level to binary images, we observe performance drop as provided in Tab. 1. This suggests that, there is information in the gray level distribution and the 3D shape of the network which are not preserved via the simple spatial dimension reduction tested. By contrast, the values in Tab. 2 represent the functional brain network classification results with CNN model using pixels, superpixels and supervoxels data respectively. Interestingly the loss of performance is very limited when one reduces the gray levels to the average value of the pixels inside a supervoxel or even a superpixel image. Therefore, despite the spatial dimension reduction tested, the reduction of the number of parameters in the models is so far very limited or negligible. To produce this reduction of the model, we proposed to encode the most promising dimension reduction technique (supervoxels) in a compact way as described in the next section.

Table 2: CNN classification of functional brain networks using superpixels/supervoxels images generated in the segmentation stage of graph encoding process with unhealthy subjects as input data.

Data	Train-Test	Accuracy	Parameters
3D gray level	315-70	0.75 ± 0.01	2,356,807
Superpixels image	315-70	0.69 ± 0.02	2,356,807
Supervoxels image	315-70	0.73 ± 0.02	2,356,807

4 Graph encoding

To further benefit from the spatial dimension reduction of the previous section, we investigate the possibility to reduce the complexity of the associated neural networks models with limited reduction of performance on the functional cerebral network recognition. To this purpose, we consider to encode our supervoxelized images into graphs. Commonly in graphs, interacting nodes are connected by edges whose weights can be defined by either temporal connections or anatomical junctions, because, graphs are naturally good at relational organization between entities, which makes them great option for representing the 3D capture of voxelwise signals mapped to a specific region of the brain [7]. Therefore, a possibly efficient representation of these fMRI network activations in images can be tested using a graph relation network, which connects nodes of related regions via graph edges.

To obtain a graph representation of our supervoxels images, we connected the segmented neighboring regions through an edge, and denoted the center of each region as a graph node, segment-wise attributes were encoded as node spatial embeddings. This step was repeated until all neighboring nodes were traversed (see Fig.5). We implemented this approach using the region adjacency graph technique [8], which simply represents each region of the segment as graph nodes and the link between two touching regions as edge using the provided labels of the segmented regions [9]. From the extracted relative spatial coordinates of each superpixel of our image data via the cartesian function, we computed the node position as edge attribute $(pos[i] - pos[j])$ via k-NN graph transformation.

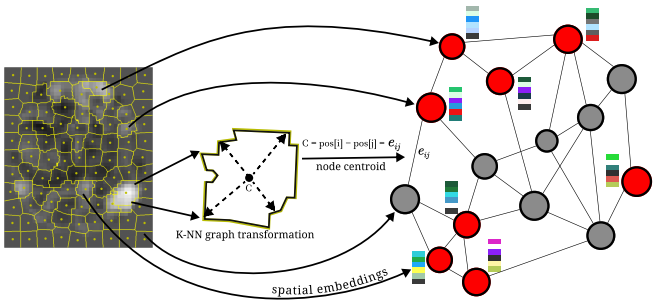


Figure 5: Graph encoding process from superpixels/supervoxels images.

The number of supervoxels was fixed empirically based on the typical size of the activation spots. The resulting graphs

from the encoding stage were observed to be structurally indistinguishable from the connectivity point of view. The contrastive information is expected to stand on the distribution of edge values, which differ from one structural network map to another.

We implemented our method using SplineCNN, a graph neural network which uses a novel type of spline-based convolutional layer for learning [10]. This state-of-the-art GNN is suitable for image-based graph classification task because, it allows the capture of local patterns using spatial relationship between graph nodes by performing global graph pooling. We trained our model parameters with 2 convolutional layers and 2 fully connected output layers with indication of 7 classes in the output layer and a softmax activation. Best results were obtained by training with 2-step learning rate values of $1e-3$ for epochs $0-200$ and $1e-5$ for epochs $200-500$ with early stopping.

For fair comparison with the best result obtained with CNN model in [11], we performed transfer learning during the training of the CNN and GNN models using 80% - 10% - 10% ratio for train-validation-test data slit respectively, as well as early stopper with patient set to 10 misses. The performance provided in Tab. 3 shows the recorded result from fMRI functional network classification using this transfer learning strategy. Brute transfer indicates the strategy of training directly on healthy data and testing on unhealthy data for both CNN and GNN models. In this cohort, results were compared with values from training and testing on unhealthy data using CNN and GNN model, which provided the 1^{st} baseline and 2^{nd} baseline values of 0.75 ± 0.01 and 0.64 ± 0.03 respectively, while 0.78 ± 0.01 and 0.70 ± 0.01 were recorded in the transfer learning approach with CNN and GNN respectively. As a consequence, we demonstrate the possibility to obtain a compression of a factor of 26 on the number of model parameters after supervoxeization and graph encoding with only a reduction of 8%.

Table 3: Transfer learning classification with CNN and GNN models.

Data	Train-Test	Accuracy	Parameters
CNN brute-transfer	healthy-unhealthy	0.75 ± 0.01	2,356,807
CNN fMRI healthy (pretrained)	unhealthy-unhealthy	0.78 ± 0.01	2,356,807
GNN brute-transfer	healthy-unhealthy	0.67 ± 0.02	90,855
GNN fMRI healthy (pretrained)	unhealthy-unhealthy	0.70 ± 0.01	90,855

5 Conclusion

In this study, we investigated ways to reduce the complexity of end-to-end machine learning models based on convolutional neural networks for the automatic recognition of functional cerebral networks via resting-state fMRI data. A compaction of the activation maps into superpixels or supervoxels shows limited impact on the classification performance. We emphasize the anticipated influence of our 3D multi-channel images in model parameters, which motivates exploration of a dimension reduction technique before introducing the graph encoding technique. Model evaluation based on spatial dimension reduction was done to investigate its minimal influence in reducing our model parameter. However, this stage was important towards more efficient data encoding (graph structure), which was later shown to have significantly reduced the model parameter. Our initial encoding effort produces a compression of a factor $26\times$ where associated reduction in performance was observed at only 8%.

The effort to reduce the complexity of the models was concentrated on the encoding approach of our fMRI data. It would naturally be interesting to couple such effort with investigation on the architecture of the models [12, 13].

References

[1] L. Ismaila, P. Rasti, J.-M. Lemée, and D. Rousseau, "Self-supervised learning for functional brain networks identification in fmri from healthy to unhealthy patients," in *16th International Conference on signal image technology & internet based systems-sitis 2022*. IEEE, 2022.

[2] L. Ismaila, J.-M. Lemée, D. Rousseau, and P. Rasti, "Deep learning pour la classification automatique de réseaux cérébraux fonctionnels par irmf de repos," *Journal of Neuroradiology*, vol. 49, no. 2, p. 119, 2022.

[3] G. Habib and S. Qureshi, "Optimization and acceleration of convolutional neural networks: A survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 7, pp. 4244–4268, 2022.

[4] J.-M. Lemée, D. H. Berro, F. Bernard, E. Chinier, L.-M. Leiber, P. Menei, and A. Ter Minassian, "Resting-state functional magnetic resonance imaging versus task-based activity for language mapping and correlation with perioperative cortical mapping," *Brain and Behavior*, vol. 9, no. 10, p. e01362, 2019.

[5] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[6] A. Bakkari and A. Fabijańska, "Features determination from super-voxels obtained with relative linear interactive clustering," in *IPC*, vol. 21, no. 3, 2016, pp. 69–80.

[7] D. Ahmedt-Aristizabal, M. A. Armin, S. Denman, C. Fookes, and L. Petersson, "Graph-based deep learning for medical diagnosis and analysis: past, present and future," *Sensors*, vol. 21, no. 14, p. 4758, 2021.

[8] A. Hagberg, P. Swart, and D. S Chult, "Exploring network structure, dynamics, and function using networkx," Los

Alamos National Lab.(LANL), Los Alamos, NM (United States), Tech. Rep., 2008.

[9] L. Wu, P. Cui, J. Pei, L. Zhao, and X. Guo, "Graph neural networks: foundation, frontiers and applications," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 4840–4841.

[10] M. Fey and J. E. Lenssen, "Fast graph representation learning with pytorch geometric," *arXiv preprint arXiv:1903.02428*, 2019.

[11] L. E. Ismaila, P. Rasti, F. Bernard, M. Labriffe, P. Menei, A. T. Minassian, D. Rousseau, and J.-M. Lemée, "Transfer learning from healthy to unhealthy patients for the automated classification of functional brain networks in fmri," *Applied Sciences*, vol. 12, no. 14, p. 6925, 2022.

[12] R. Sanchez-Iborra and A. F. Skarmeta, "Tinyml-enabled frugal smart objects: Challenges and opportunities," *IEEE Circuits and Systems Magazine*, vol. 20, no. 3, pp. 4–18, 2020.

[13] S. Guo and Q. Zhou, *Machine Learning on Commodity Tiny Devices: Theory and Practice*. CRC Press, 2022.

BIBLIOGRAPHY

[1] I Baldi et al., « Descriptive epidemiology of CNS tumors in France: results from the Gironde Registry for the period 2000–2007 », *in*: *Neuro-oncology* 13.*12* (2011), pp. 1370–1378.

[2] Valérie Rigau et al., « French brain tumor database: 5-year histological results on 25 756 cases », *in*: *Brain Pathology* 21.*6* (2011), pp. 633–644.

[3] Kimberly R Porter et al., « Conditional survival of all primary brain tumor patients by age, behavior, and histology », *in*: *Neuroepidemiology* 36.*4* (2011), pp. 230–239.

[4] Rimas V Lukas et al., « Newly diagnosed glioblastoma: a review on clinical management », *in*: *Oncology (Williston Park, NY)* 33.*3* (2019), p. 91.

[5] National Cancer Institute, *Craniotomy illustration*, en, Feb. 2011.

[6] Jean-Michel Lemée et al., « Resting-state functional magnetic resonance imaging versus task-based activity for language mapping and correlation with perioperative cortical mapping », *in*: *Brain and Behavior* 9.*10* (2019), e01362.

[7] Ki Yun Park et al., « Mapping language function with task-based vs. resting-state functional MRI », *in*: *PLoS One* 15.*7* (2020), e0236423.

[8] Zarrar Shehzad et al., « The resting brain: unconstrained yet reliable », *in*: *Cerebral cortex* 19.*10* (2009), pp. 2209–2229.

[9] Koene RA Van Dijk, Mert R Sabuncu, and Randy L Buckner, « The influence of head motion on intrinsic functional connectivity MRI », *in*: *Neuroimage* 59.*1* (2012), pp. 431–438.

[10] Rasmus M Birn et al., « The effect of scan length on the reliability of resting-state fMRI connectivity estimates », *in*: *Neuroimage* 83 (2013), pp. 550–558.

[11] Zhiqun Wang et al., « The baseline and longitudinal changes of PCC connectivity in mild cognitive impairment: a combined structure and resting-state fMRI study », *in*: *PloS one* 7.*5* (2012), e36838.

[12] Haihong Liu et al., « Decreased regional homogeneity in schizophrenia: a resting state functional magnetic resonance imaging study », *in*: *Neuroreport* 17.*1* (2006), pp. 19–22.

[13] Bharat Biswal et al., « Functional connectivity in the motor cortex of resting human brain using echo-planar MRI », *in*: *Magnetic resonance in medicine* 34.*4* (1995), pp. 537–541.

[14] Zeynettin Akkus et al., « Deep learning for brain MRI segmentation: state of the art and future directions », *in*: *Journal of digital imaging* 30 (2017), pp. 449–459.

[15] Joeky T Senders et al., « Machine learning and neurosurgical outcome prediction: a systematic review », *in*: *World neurosurgery* 109 (2018), pp. 476–486.

[16] Alessandro Siccoli et al., « Machine learning–based preoperative predictive analytics for lumbar spinal stenosis », *in*: *Neurosurgical Focus* 46.*5* (2019), E5.

[17] Victor E Staartjes et al., « Machine learning in neurosurgery: a global survey », *in*: *Acta neurochirurgica* 162 (2020), pp. 3081–3091.

[18] Victor E Staartjes et al., « Neural network–based identification of patients at high risk for intraoperative cerebrospinal fluid leaks in endoscopic pituitary surgery », *in*: *Journal of neurosurgery* 133.*2* (2019), pp. 329–335.

[19] Yangyang Zhang et al., « Modularity-guided functional brain network analysis for early-stage dementia identification », *in*: *Frontiers in Neuroscience* 15 (2021), p. 720909.

[20] Parisa Azimi et al., « Use of artificial neural networks to predict surgical satisfaction in patients with lumbar spinal canal stenosis », *in*: *Journal of Neurosurgery: Spine* 20.*3* (2014), pp. 300–305.

[21] Walter Stummer et al., « Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial », *in*: *The lancet oncology* 7.*5* (2006), pp. 392–401.

[22] Christopher Nimsky et al., « Intraoperative high-field magnetic resonance imaging in transsphenoidal surgery of hormonally inactivepituitary macroadenomas », *in*: *Neurosurgery* 59.*1* (2006), pp. 105–114.

[23] Daniel F Kacher et al., « The advanced multimodality image-guided operating (AMIGO) suite », *in*: *Intraoperative Imaging and Image-Guided Therapy* (2014), pp. 339–368.

[24] Clare MC Tempany et al., « Multimodal imaging for improved diagnosis and treatment of cancers », *in*: *Cancer* 121.*6* (2015), pp. 817–827.

[25] Shu Zhang et al., « Characterizing and differentiating task-based and resting state fMRI signals via two-stage sparse representations », *in*: *Brain imaging and behavior* 10 (2016), pp. 21–32.

[26] Giuseppe Roberto Giammalva et al., « Brain mapping-aided supratotal resection (SpTR) of brain tumors: the role of brain connectivity », *in*: *Frontiers in Oncology* 11 (2021), p. 645854.

[27] David M Cole, Stephen M Smith, and Christian F Beckmann, « Advances and pitfalls in the analysis and interpretation of resting-state FMRI data », *in*: *Frontiers in systems neuroscience* (2010), p. 8.

[28] Megan H Lee, Christopher D Smyser, and Joshua S Shimony, « Resting-state fMRI: a review of methods and clinical applications », *in*: *American Journal of neuroradiology* 34.*10* (2013), pp. 1866–1872.

[29] Meenakshi Khosla et al., « Machine learning in resting-state fMRI analysis », *in*: *Magnetic resonance imaging* 64 (2019), pp. 101–121.

[30] Minyoung Jung et al., « Default mode network in young male adults with autism spectrum disorder: relationship with autism spectrum traits », *in*: *Molecular autism* 5.*1* (2014), pp. 1–11.

[31] Dost Öngür et al., « Default mode network abnormalities in bipolar disorder and schizophrenia », *in*: *Psychiatry Research: Neuroimaging* 183.*1* (2010), pp. 59–68.

[32] Walter Koch et al., « Diagnostic power of default mode network resting state fMRI in the detection of Alzheimer's disease », *in*: *Neurobiology of aging* 33.*3* (2012), pp. 466–478.

[33] Lukman Ismaila et al., « Deep learning for the automatic classification of functional brain networks by rest mri », *in*: *Journal of Neuroradiology* 49.*2* (2022), p. 119.

[34] Lukman E Ismaila et al., « Transfer Learning from Healthy to Unhealthy Patients for the Automated Classification of Functional Brain Networks in fMRI », *in*: *Applied Sciences* 12.*14* (2022), p. 6925.

[35] Lukman Ismaila et al., « Self-Supervised Learning for Functional Brain Networks identification in fMRI from Healthy to Unhealthy Patients », *in*: *16th International Conference on Signal Image Technology & Internet Based Systems-SITIS 2022*, IEEE, 2022.

[36] Jean-Michel Lemee Lukman E. Ismaila Pejman Rasti and David Rousseau, « Toward more frugal models for functional cerebral networks automatic recognition with resting state fMRI », *in*: *XXIXème Colloque Francophone de Traitement du Signal et des Images, GRETSI'23*, 2023.

[37] Erik R Ranschaert, Sergey Morozov, and Paul R Algra, *Artificial intelligence in medical imaging: opportunities, applications and risks*, Springer, 2019.

[38] Marco Aiello et al., « The challenges of diagnostic imaging in the era of big data », *in*: *Journal of clinical medicine* 8.3 (2019), p. 316.

[39] Nandhini Subramanian et al., « A review of deep learning-based detection methods for COVID-19 », *in*: *Computers in Biology and Medicine* (2022), p. 105233.

[40] Jai Kotia et al., « Few shot learning for medical imaging », *in*: *Machine learning algorithms for industrial applications*, Springer, 2021, pp. 107–132.

[41] Tristan Glatard et al., « A virtual imaging platform for multi-modality medical image simulation », *in*: *IEEE transactions on medical imaging* 32.1 (2012), pp. 110–118.

[42] Xin Yi, Ekta Walia, and Paul Babyn, « Generative adversarial network in medical imaging: A review », *in*: *Medical image analysis* 58 (2019), p. 101552.

[43] Phillip Chlap et al., « A review of medical image data augmentation techniques for deep learning applications », *in*: *Journal of Medical Imaging and Radiation Oncology* 65.5 (2021), pp. 545–563.

[44] Hassaan Malik et al., « A comparison of transfer learning performance versus health experts in disease diagnosis from medical imaging », *in*: *IEEE Access* 8 (2020), pp. 139367–139386.

[45] Laith Alzubaidi et al., « Towards a better understanding of transfer learning for medical imaging: a case study », *in*: *Applied Sciences* 10.13 (2020), p. 4523.

[46] Christos Matsoukas et al., « What Makes Transfer Learning Work For Medical Images: Feature Reuse & Other Factors », *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9225–9234.

[47] Juan Miguel Valverde et al., « Transfer learning in magnetic resonance brain imaging: a systematic review », *in*: *Journal of imaging* 7.*4* (2021), p. 66.

[48] Jason Yosinski et al., « How transferable are features in deep neural networks? », *in*: *Advances in neural information processing systems* 27 (2014).

[49] Sebastian Ille and Sandro M Krieg, « Functional mapping for glioma surgery, part 1: Preoperative mapping tools », *in*: *Neurosurgery Clinics* 32.*1* (2021), pp. 65–74.

[50] Maeike Zijlmans, Willemiek Zweiphenning, and Nicole van Klink, « Changing concepts in presurgical assessment for epilepsy surgery », *in*: *Nature Reviews Neurology* 15.*10* (2019), pp. 594–606.

[51] Charles Okanda Nyatega et al., « Altered Dynamic Functional Connectivity of Cuneus in Schizophrenia Patients: A Resting-State fMRI Study », *in*: *Applied Sciences* 11.*23* (2021), p. 11392.

[52] Faria Zarin Subah et al., « A deep learning approach to predict autism spectrum disorder using multisite resting-state fMRI », *in*: *Applied Sciences* 11.*8* (2021), p. 3636.

[53] Dongyang Zhang et al., « Preoperative sensorimotor mapping in brain tumor patients using spontaneous fluctuations in neuronal activity imaged with functional magnetic resonance imaging: initial experience », *in*: *Operative Neurosurgery* 65.*suppl_6* (2009), pp. 226–236.

[54] Ali Mahdavi et al., « Functional MRI in clinical practice: Assessment of language and motor for pre-surgical planning », *in*: *The neuroradiology journal* 28.*5* (2015), pp. 468–473.

[55] Joshua S Shimony et al., « Resting-state spontaneous fluctuations in brain activity: a new paradigm for presurgical planning using fMRI », *in*: *Academic radiology* 16.*5* (2009), pp. 578–583.

[56] Michael G Hart, Stephen J Price, and John Suckling, « Functional connectivity networks for preoperative brain mapping in neurosurgery », *in*: *Journal of neurosurgery* 126.*6* (2016), pp. 1941–1950.

[57] Timothy J Mitchell et al., « A novel data-driven approach to preoperative mapping of functional cortex using resting-state functional magnetic resonance imaging », *in*: *Neurosurgery* 73.*6* (2013), pp. 969–983.

[58] Aram Ter Minassian et al., « The presupplementary area within the language network: a resting state functional magnetic resonance imaging functional connectivity analysis », *in*: *Brain connectivity* 4.*6* (2014), pp. 440–453.

[59] Yanmei Tie et al., « Defining language networks from resting-state fMRI for surgical planning—A feasibility study », *in*: *Human brain mapping* 35.*3* (2014), pp. 1018–1030.

[60] Sharon Chiang, Harvey S Levin, and Zulfi Haneef, « Computer-automated focus lateralization of temporal lobe epilepsy using fMRI », *in*: *Journal of Magnetic Resonance Imaging* 41.*6* (2015), pp. 1689–1694.

[61] Ling-Li Zeng et al., « Unsupervised classification of major depression using functional connectivity MRI », *in*: *Human brain mapping* 35.*4* (2014), pp. 1630–1641.

[62] Rahul Raj et al., « Machine learning-based dynamic mortality prediction after traumatic brain injury », *in*: *Scientific reports* 9.*1* (2019), pp. 1–13.

[63] Darya Chyzhyk, Alexandre Savio, and Manuel Graña, « Computer aided diagnosis of schizophrenia on resting state fMRI data by ensembles of ELM », *in*: *Neural Networks* 68 (2015), pp. 23–33.

[64] Dajiang Zhu et al., « Connectome-scale assessments of structural and functional connectivity in MCI », *in*: *Human brain mapping* 35.*7* (2014), pp. 2911–2923.

[65] Junfeng Lu et al., « An automated method for identifying an independent component analysis-based language-related resting-state network in brain tumor subjects for surgical planning », *in*: *Scientific reports* 7.*1* (2017), pp. 1–16.

[66] Yanmei Tie et al., « Defining language networks from resting-state fMRI for surgical planning—A feasibility study », *in*: *Human brain mapping* 35.*3* (2014), pp. 1018–1030.

[67] Victor Nozais et al., « Deep Learning-based Classification of Resting-state fMRI Independent-component Analysis », *in*: *Neuroinformatics* 19.*4* (2021), pp. 619–637.

[68] Geert Litjens et al., « A survey on deep learning in medical image analysis », *in*: *Medical image analysis* 42 (2017), pp. 60–88.

[69] Zexun Zhou et al., « 3D dense connectivity network with atrous convolutional feature pyramid for brain tumor segmentation in magnetic resonance imaging of human heads », *in*: *Computers in Biology and Medicine* 121 (2020), p. 103766.

[70] Hannelore Aerts et al., « Modeling brain dynamics after tumor resection using The Virtual Brain », *in*: *Neuroimage* 213 (2020), p. 116738.

[71] Yuan Tao and Brenda Rapp, « Investigating the network consequences of focal brain lesions through comparisons of real and simulated lesions », *in*: *Scientific reports* 11.*1* (2021), pp. 1–17.

[72] Xia Li et al., « Dataset of whole-brain resting-state fMRI of 227 young and elderly adults acquired at 3T », *in*: *Data in brief* 38 (2021), p. 107333.

[73] Saori C Tanaka et al., « A multi-site, multi-disorder resting-state magnetic resonance image database », *in*: *Scientific data* 8.*1* (2021), pp. 1–15.

[74] Vince D Calhoun et al., « A method for making group inferences from functional MRI data using independent component analysis », *in*: *Human brain mapping* 14.*3* (2001), pp. 140–151.

[75] Ed Bullmore and Olaf Sporns, « Complex brain networks: graph theoretical analysis of structural and functional systems », *in*: *Nature reviews neuroscience* 10.*3* (2009), pp. 186–198.

[76] Lars Jansson and Tobias Sandsröm, « Graph convolutional neural networks for brain connectivity analysis », *in*: (2020).

[77] Richard O Duda, Peter E Hart, and David G Stork, *Pattern classification and scene analysis*, vol. 3, Wiley New York, 1973.

[78] J. Ross Quinlan, « Induction of decision trees », *in*: *Machine learning* 1.*1* (1986), pp. 81–106.

[79] Ian Goodfellow et al., *Deep learning*, vol. 1, MIT press Cambridge, 2016.

[80] Diederik P Kingma and Jimmy Ba, « Adam: A method for stochastic optimization », *in*: *arXiv preprint arXiv:1412.6980* (2014).

[81] Aytaç Altan and Seçkin Karasu, « Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique », *in*: *Chaos, Solitons & Fractals* 140 (2020), p. 110071.

[82] Karen Simonyan and Andrew Zisserman, « Very deep convolutional networks for large-scale image recognition », *in*: *arXiv preprint arXiv:1409.1556* (2014).

[83] Kaiming He et al., « Delving deep into rectifiers: Surpassing human-level performance on imagenet classification », *in*: *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[84] Gao Huang et al., « Densely connected convolutional networks », *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[85] Alexander Jung, « Imgaug documentation », *in*: *Readthedocs. io, Jun* 25 (2019).

[86] O. Russakovsky et al., « ImageNet Large Scale Visual Recognition Challenge », *in*: *IJCV* 115.*3* (2015), pp. 211–252.

[87] Dongyang Zhang and Marcus E Raichle, « Disease and the brain's dark energy », *in*: *Nature Reviews Neurology* 6.*1* (2010), pp. 15–28.

[88] Michael D Fox and Marcus E Raichle, « Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging », *in*: *Nature reviews neuroscience* 8.*9* (2007), pp. 700–711.

[89] R Nathan Spreng et al., « Intrinsic architecture underlying the relations among the default, dorsal attention, and frontoparietal control networks of the human brain », *in*: *Journal of cognitive neuroscience* 25.*1* (2013), pp. 74–86.

[90] Maurizio Corbetta, Gaurav Patel, and Gordon L Shulman, « The reorienting system of the human brain: from environment to theory of mind », *in*: *Neuron* 58.*3* (2008), pp. 306–324.

[91] Mathieu Vigneau et al., « Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing », *in*: *Neuroimage* 30.*4* (2006), pp. 1414–1432.

[92] Stefan Knecht et al., « Handedness and hemispheric language dominance in healthy humans », *in*: *Brain* 123.*12* (2000), pp. 2512–2518.

[93] Ana Barragán-Montero et al., « Artificial intelligence and machine learning for medical imaging: A technology review », *in*: *Physica Medica* 83 (2021), pp. 242–256.

[94] Angela Zhang et al., « Shifting machine learning for healthcare from development to deployment and from models to data », *in*: *Nature Biomedical Engineering* (2022), pp. 1–16.

[95] Ying Liu et al., « Few-Shot Image Classification: Current Status and Research Trends », *in*: *Electronics* 11.*11* (2022), p. 1752.

[96] Ashish Jaiswal et al., « A survey on contrastive self-supervised learning », *in*: *Technologies* 9.*1* (2020), p. 2.

[97] Liang Chen et al., « Self-supervised learning for medical image analysis using image context restoration », *in*: *Medical image analysis* 58 (2019), p. 101539.

[98] András Kalapos and Bálint Gyires-Tóth, « Self-supervised Pretraining for 2D Medical Image Segmentation », *in*: *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, Springer, 2023, pp. 472–484.

[99] Mangal Prakash et al., « Leveraging self-supervised denoising for image segmentation », *in*: *2020 IEEE 17th international symposium on biomedical imaging (ISBI)*, IEEE, 2020, pp. 428–432.

[100] Fengze Liu et al., « SAME: Deformable image registration based on self-supervised anatomical embeddings », *in*: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*, Springer, 2021, pp. 87–97.

[101] Saeed Shurrab and Rehab Duwairi, « Self-supervised learning methods and applications in medical imaging analysis: A survey », *in*: *PeerJ Computer Science* 8 (2022), e1045.

[102] Junshen Xu and Elfar Adalsteinsson, « Deformed2self: Self-supervised denoising for dynamic medical imaging », *in*: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*, Springer, 2021, pp. 25–35.

[103] Jianbo Jiao et al., « Self-supervised contrastive video-speech representation learning for ultrasound », *in*: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23*, Springer, 2020, pp. 534–543.

[104] Itzik Malkiel et al., « Self-Supervised Transformers for fMRI representation », *in*: *International Conference on Medical Imaging with Deep Learning*, PMLR, 2022, pp. 895–913.

[105] Ashish Jaiswal et al., « Understanding Cognitive Fatigue from fMRI Scans with Self-supervised Learning », *in*: *arXiv preprint arXiv:2106.15009* (2021).

[106] Itzik Malkiel et al., « Self-Supervised Transformers for fMRI representation », *in*: *Medical Imaging with Deep Learning*, 2021.

[107] Brent R Logan, Maya P Geliazkova, and Daniel B Rowe, « An evaluation of spatial thresholding techniques in fMRI analysis », *in*: *Human brain mapping* 29.*12* (2008), pp. 1379–1389.

[108] Shekoofeh Azizi et al., « Big self-supervised models advance medical image classification », *in*: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3478–3488.

[109] Florin C Ghesu et al., « Self-supervised learning from 100 million medical images », *in*: *arXiv preprint arXiv:2201.01283* (2022).

[110] Yuandong Tian et al., « Understanding self-supervised learning with dual deep networks », *in*: *arXiv preprint arXiv:2010.00578* (2020).

[111] Kaiming He et al., « Momentum contrast for unsupervised visual representation learning », *in*: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729–9738.

[112] Ting Chen et al., « Big self-supervised models are strong semi-supervised learners », *in*: *Advances in neural information processing systems* 33 (2020), pp. 22243–22255.

[113] Ting Chen et al., « A simple framework for contrastive learning of visual representations », *in*: *International conference on machine learning*, PMLR, 2020, pp. 1597–1607.

[114] Suzanna Becker and Geoffrey E Hinton, « Self-organizing neural network that discovers surfaces in random-dot stereograms », *in*: *Nature* 355.*6356* (1992), pp. 161–163.

[115] Ramprasaath R Selvaraju et al., « Grad-cam: Visual explanations from deep networks via gradient-based localization », *in*: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[116] Salifu Nanga et al., « Review of dimension reduction methods », *in*: *Journal of Data Analysis and Information Processing* 9.*3* (2021), pp. 189–231.

[117] Zhongheng Zhang and Adela Castelló, « Principal components analysis in clinical studies », *in*: *Annals of translational medicine* 5.*17* (2017).

[118] Laurens Van der Maaten and Geoffrey Hinton, « Visualizing data using t-SNE. », *in*: *Journal of machine learning research* 9.*11* (2008).

[119] Quanquan Gu, Zhenhui Li, and Jiawei Han, « Linear discriminant dimensionality reduction », *in*: *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2011, Athens, Greece, September 5-9, 2011. Proceedings, Part I 11*, Springer, 2011, pp. 549–564.

[120] Oliver Kramer and Oliver Kramer, « K-nearest neighbors », *in*: *Dimensionality reduction with unsupervised nearest neighbors* (2013), pp. 13–23.

[121] Oded Z Maimon and Lior Rokach, *Data mining with decision trees: theory and applications*, vol. 81, World scientific, 2014.

[122] William S Noble, « What is a support vector machine? », *in*: *Nature biotechnology* 24.*12* (2006), pp. 1565–1567.

[123] Irina Rish et al., « An empirical study of the naive Bayes classifier », *in*: *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, 22, 2001, pp. 41–46.

[124] Neena Aloysius and M Geetha, « A review on deep convolutional neural networks », *in*: *2017 international conference on communication and signal processing (ICCSP)*, IEEE, 2017, pp. 0588–0592.

[125] Petar Velickovic et al., « Graph attention networks », *in*: *stat* 1050.*20* (2017), pp. 10–48550.

[126] Will Hamilton, Zhitao Ying, and Jure Leskovec, « Inductive representation learning on large graphs », *in*: *Advances in neural information processing systems* 30 (2017).

[127] Han Lv et al., « Resting-state functional MRI: everything that nonexperts have always wanted to know », *in*: *American Journal of Neuroradiology* 39.*8* (2018), pp. 1390–1399.

[128] Charles D Schaper, « Analytic Model of fMRI BOLD Signals for Separable Metrics of Neural and Metabolic Activity », *in*: *bioRxiv* (2019), p. 573006.

[129] Jarod L Roland et al., « Resting-state functional magnetic resonance imaging for surgical planning in pediatric patients: a preliminary experience », *in*: *Journal of Neurosurgery: Pediatrics* 20.*6* (2017), pp. 583–590.

[130] Vladimir S Fonov et al., « Unbiased nonlinear average age-appropriate brain templates from birth to adulthood », *in*: *NeuroImage* 47 (2009), S102.

[131] Guillaume Marrelec et al., « Partial correlation for functional brain interactivity investigation in functional MRI », *in*: *Neuroimage* 32.*1* (2006), pp. 228–237.

[132] Yi-Ou Li, Tülay Adalı, and Vince D Calhoun, « Estimating the number of independent components for functional magnetic resonance imaging data », *in*: *Human brain mapping* 28.*11* (2007), pp. 1251–1266.

[133] Haris I Sair et al., « Presurgical brain mapping of the language network in patients with brain tumors using resting-state f MRI: Comparison with task f MRI », *in*: *Human brain mapping* 37.*3* (2016), pp. 913–923.

[134] Fatemeh Geranmayeh et al., « Overlapping networks engaged during spoken language production and its cognitive control », *in*: *Journal of Neuroscience* 34.*26* (2014), pp. 8728–8740.

[135] Cristina Rosazza and Ludovico Minati, « Resting-state brain networks: literature review and clinical applications », *in*: *Neurological sciences* 32.*5* (2011), pp. 773–785.

[136] Megan H Lee et al., « Clustering of resting state networks », *in*: *PloS one* 7.*7* (2012), e40370.

Titre : Application de l'apprentissage automatique en neurosciences pour la procédure d'ablation pré-chirurgicale d'une tumeur cérébrale

Mot clés : apprentissage automatique, apprentissage profond, vision par ordinateur, imagerie médicale, apprentissage par transfert, IRMf au repos, réseaux cérébraux fonctionnels, autosupervision, classification des images, réseau neuronal graphique, augmentation des données.

Résumé :

La résection d'une tumeur cérébrale est une procédure médicale essentielle pratiquée par les neurochirurgiens. Les zones fonctionnelles du cerveau doivent être identifiées et préservées pendant l'opération afin de maintenir la fonction neurologique altérée par la tumeur. L'imagerie par résonance magnétique fonctionnelle (IRMf) est utilisée pour la planification chirurgicale afin d'identifier ces zones sur la base du signal dépendant de l'oxygène du sang (BOLD), qui augmente le flux sanguin dans les régions cérébrales activées. Des paradigmes basés sur des tâches, comme le tapotement des doigts ou la parole, sont traditionnellement utilisés pour cette identification, mais ils prennent du temps et nécessitent la coopération du patient. L'IRMf à l'état de repos (IRMf-R) est une méthode alternative qui analyse les oscillations spontanées du signal BOLD pour identifier les réseaux de connectivité indépendants. Cependant, la reconnaissance manuelle de ces réseaux pendant l'intervention chirurgicale est difficile et sujette à des erreurs. Pour y remédier, nous avons proposé d'utiliser des techniques d'apprentissage automatique et de vision par ordinateur pour reconnaître automatiquement sept réseaux cérébraux fonctionnels à partir de données d'IRMf.

La collecte de grandes quantités de données d'IRMf est difficile en raison des limites du partage des données. Nous avons exploré l'apprentissage par transfert en formant des modèles avec des données saines et en les appliquant à des données malsaines pour surmonter cette limitation. Cette approche tire parti de la similitude entre les réseaux d'activation fonctionnelle du cerveau chez les sujets sains et non sains. En outre, nous avons développé un modèle d'apprentissage par autosupervision qui utilise des ensembles de données saines non étiquetées pour pré-entraîner le modèle, éliminant ainsi la nécessité d'une annotation fastidieuse des données par les cliniciens.

Nous avons également étudié les différences entre les données saines et malsaines afin de comprendre leur relation et la manière dont elles affectent la transférabilité et la classification des réseaux cérébraux fonctionnels. En outre, nous avons proposé une méthode de réduction des dimensions et un encodage graphique des images IRMf en utilisant l'apprentissage de la représentation graphique pour éviter les paramètres d'apprentissage importants, les modèles complexes et les exigences informatiques. Cette approche permet d'apprendre efficacement les signaux utiles dans les images d'IRMf et d'obtenir des résultats comparables aux modèles de réseaux neuronaux convolutionnels (CNN) avec des paramètres de modèle réduits.

Title: Machine Learning Application In Neuroscience For Pre-Surgical Brain Tumor Removal Procedure

Keywords: machine learning, deep learning, computer vision, medical imaging, transfer learning, resting-state fMRI, functional brain networks, Self-supervision, image classification, graph neural network, data augmentation.

Abstract:

Brain tumor resection is a critical medical procedure performed by neurosurgeons. Functional brain areas need to be identified and preserved during the surgery to maintain neurological function impaired by the tumor. Functional magnetic resonance imaging (fMRI) is used for surgical planning to identify these areas based on the blood oxygen dependent (BOLD) signal, which increases blood flow in activated brain regions. Task-based paradigms like finger tapping or speech are traditionally used for this identification but are time-consuming and require patient cooperation. Resting state fMRI (rs-fMRI) is an alternative method that analyzes spontaneous BOLD signal oscillations to identify independent connectivity networks. However, manual recognition of these networks during surgery is challenging and prone to errors. To address this, we proposed using machine learning and computer vision techniques to automatically recognize seven functional brain networks from rs-fMRI data.

Collecting large amounts of fMRI data is difficult due to medical data sharing restrictions. We explored transfer learning by training models with healthy data and applying them to unhealthy data to overcome this limitation. This approach leverages the similarity between functional brain activation networks in healthy and unhealthy subjects. Additionally, we developed a self-supervision learning model that uses unlabeled healthy datasets to pretrain the model, eliminating the need for time-consuming data annotation by clinicians.

We further investigated the differences between healthy and unhealthy data to understand their relationship and how it affects transferability and the classification of functional brain networks. Additionally, we proposed a dimension reduction method and graph encoding of fMRI images using graph representation learning to avoid large training parameters, complex models, and computational demands. This approach effectively learns useful signals in fMRI images and achieves comparable results to convolutional neural network (CNN) models with reduced model parameters.