UNIVERSITAT POLITÈCNICA DE VALÈNCIA

_ TELECOM ESCUELA TÉCNICA VLC SUPERIOR DE INGENIERÍA DE TELECOMUNICACIÓN

Application of Multilevel Time-Frequency Decomposition to fMRI for Brain Connectivity Network Estimation: Diagnosis of Schizophrenia

PAN Jinglun

Tutor:

Dr. Addisson Salazar (UPV)

Cotutor:

Dr. Yan (Cindy) Sun (QMUL)

Prof. Luís Vergara Domínguez (UPV)

Bachelor's Thesis presented at Escuela Técnica Superior de Ingenieros de Telecomunicación of Universitat Politècnica de València,

Academic Year 2019-2020

Valencia, 5 de abril de 2020

Escuela Técnica Superior de Ingeniería de Telecomunicación
Universitat Politècnica de València
Edificio 4D. Camino de Vera, s/n, 46022 Valencia
Tel. +34 96 387 71 90, ext. 77190
www.etsit.upv.es

VLC/ CAMPUS VALENCIA, INTERNATIONAL CAMPUS OF EXCELLENCE

EMAS Gestión ambiental verificada REG.NO. ES-CV- 000030

Resumen

Con el desarrollo de la tecnología de imágenes médicas, especialmente la tecnología de imágenes de resonancia magnética funcional (fMRI), los investigadores han descubierto que la conectividad cerebral ha ocurrido como una característica crítica de la esquizofrenia. Este estudio tuvo como objetivo implementar la descomposición de wavelet para extraer la serie temporal de datos de fMRI para apoyar el diagnóstico de la enfermedad de esquizofrenia. Después de esto, este estudio determinó la correlación wavelet de 116 regiones cerebrales para estimar la conectividad de la red cerebral y utilizó algoritmos de máquina de vectores de soporte (SVM) y kNN (vecino k-más cercano) para clasificar los datos de fMRI de un sujeto para probar si el sujeto está sufriendo de la esquizofrenia. Este estudio combinó un enfoque de análisis wavelet con algoritmos avanzados de aprendizaje automático: clasificadores de máquina de vectores de soporte (SVM) y kNN (vecino k-más cercano) para investigar la red de conectividad cerebral con diez sujetos. La matriz de conectividad fMRI determinada por la correlación cruzada de wavelet se construyó mediante el etiquetado anatómico automático del Atlas (AAL), que constaba de 116 regiones. Los coeficientes máximos de correlación wavelet en bandas de frecuencia específicas se extrajeron como el componente esencial de la clasificación, que se aplicó a los clasificadores para identificar a las personas con enfermedades de esquizofrenia de los controles saludables. Para mejorar la precisión de los resultados de clasificación, diseñamos e implementamos diferentes tipos de núcleos para clasificadores SVM (es decir, núcleos lineales, gaussianos y polinómicos) y diferentes tipos de distancias para clasificadores KNN (es decir, distancia euclidiana estándar, de Hamming y gaussiana) en este proyecto. Los resultados experimentales demuestran que el KNN con distancia euclidiana y gaussiana estándar muestra un gran rendimiento de clasificación con un índice de precisión del 90%, lo que significa que el método podría usarse eficazmente en el diagnóstico auxiliar de la enfermedad de esquizofrenia.

Resum

Amb el desenvolupament de la tecnologia d'imatge mèdica, especialment la tecnologia de ressonància magnètica funcional (fMRI), els investigadors han descobert que la connectivitat cerebral ha transcorregut com una característica crítica de l'esquizofrènia. Aquest estudi tenia com a objectiu la implementació de la descomposició d'ona ondulada per extreure la sèrie temporal de dades de la RMF per donar suport al diagnòstic de la malaltia d'esquizofrènia. Després d'aquest fet, aquest estudi va determinar la correlació d'ona d'ones de 116 regions cerebrals per estimar la connectivitat de la xarxa cerebral i va utilitzar els algorismes de Support Vector Machine (SVM) i kNN (k-Near Neighbor) per classificar les dades fMRI d'un subjecte a provar si el subjecte pateix. de l'esquizofrènia. Aquest estudi va combinar un enfocament d'anàlisi d'ondeletes amb algoritmes avançats d'aprenentatge automàtic: Support Vector Machine (SVM) i kNN (k-Near Neighbor) classificadors per investigar la xarxa de connectivitat cerebral amb deu subjectes. La matriu de connectivitat fMRI determinada per la correlació de les ondulacions es va construir mitjançant un Etiquetatge Automàtic d'Atlas Anatòmic (AAL), format per 116 regions. Els coeficients màxims de correlació d'ona en bandes de freqüència específiques es van extreure com a component essencial de la classificació, que es va aplicar als classificadors per identificar individus amb malalties d'esquizofrènia a partir dels controls saludables. Per millorar la precisió dels resultats de classificació, vam dissenyar i implementar diferents tipus de nuclis per als classificadors SVM (és a dir, nuclis lineals, gaussians i polinòmics) i diferents tipus de distàncies per als classificadors KNN (és a dir, Distància Euclidiana Estàndard, Hamming i Gaussiana) en aquest projecte. Els resultats experimentals demostren que el KNN amb Standard Euclidiana i Gaussian Distance mostra un gran rendiment de classificació amb un índex de precisió del 90%, cosa

que significa que el mètode es podria utilitzar eficaçment en el diagnòstic auxiliar de la malaltia de l'esquizofrènia.

Abstract

With the development of medical imaging technology, especially functional magnetic resonance imaging(fMRI) technology, researchers have discovered that brain connectivity has transpired as a critical feature of Schizophrenia. This study aimed to implement wavelet decomposition to extract the time series of fMRI data for supporting the diagnosis of schizophrenia disease. After this, this study determined the wavelet correlation of 116 brain regions to estimate brain network connectivity and used Support Vector Machine (SVM) and kNN (k-Nearest Neighbour) algorithms to classify the fMRI data of a subject to test if the subject is suffering from the Schizophrenia. This study combined a wavelet analysis approach with advanced machine learning algorithms: Support Vector Machine (SVM) and kNN (k-Nearest Neighbour) classifiers to research the brain connectivity network with ten subjects. The fMRI connectivity matrix determined by wavelet cross-correlation was constructed by automated Anatomical Atlas Labelling (AAL), consisted of 116 regions. The maximum wavelet correlation coefficients in specific frequency bands were extracted as the essential component of classification, which was applied to the classifiers to identify individuals with Schizophrenia Diseases from the Healthy Controls. To improve the accuracy of classification results, we designed and implemented different types of kernels for SVM classifiers (i.e., Linear, Gaussian, and Polynomial Kernels) and different kinds of distances for KNN classifiers (i.e., Standard Euclidean, Hamming, and Gaussian Distance) in this project. The experimental results demonstrate that KNN with Standard Euclidean and Gaussian Distance shows a great performance of classification with an accuracy index of 90%, which means the method could be effectively used in the auxiliary diagnosis of Schizophrenia Disease.

Index

Abstract 4

Capítulo 1. Introduction 3

Capítulo 2. Background 4

2.1 Functional Magnetic Resonance Imaging and Brain Activity 4

2.2 Brain Network Connectivity 5

2.2.1 Brain Network based on Graph Theory 5

2.2.2 Define Brain Network Nodes 6

1.2.3 Define the Brain Network Connectivity 6

2.3 Classification 7

Capítulo 3. Design and Implementation 8

3.1 Dataset 8

3.2 Pre-Processing 8

3.2.1 Brain Skull Removal (Using FSL FEAT's Brain Extraction Technique) 8

3.2.2 Slice Time Correction (SPM) 9

3.2.3 Motion Correction (SPM) 10

3.2.4 Spatial Smoothing (SPM) Using a Gaussian Filter 11

3.3 Network Estimation 11

3.4 Classification 13

3.4.1 Supported Vector Machine (SVM) 13

3.4.2 K- Nearest Neighbour (kNN) Classifier 16

3.5 Evaluation 16

3.5.1 Confusion Matrix 16

3.5.2 Accuracy Index 17

Capítulo 4. Results and Discussion 18

4.1 Network Estimation Results 18

4.2 Classification Results 22

4.2.1 Classification Results Based on Support Vector Machine 23

4.2.1.1 Support Vector Machine with Linear Kernel 23

4.2.1.2 Support Vector Machine with Gaussian Kernel 23

4.2.1.3 Support Vector Machine with Polynomial Kernel 24

4.2.2 Classification Results Based on K-Nearest Neighbour 24

4.2.2.1 k-Nearest Neighbour with Standard Euclidean Distance 24

4.2.2.2 k-Nearest Neighbour with Hamming Distance 24

4.2.2.3 k-Nearest Neighbour with Gaussian Distance 25

4.3 Confusion Matrix 25

4.4 Classification Accuracy 27

Capítulo 5. Conclusion and Further Work 28

Bibliography 29

Capítulo 1. Introduction

In recent years, with the development of functional magnetic resonance imaging (fMRI) technology, researchers have gradually explored the brain from structural analysis to focus on functional connections in the brain. A human brain can be considered as a complex network with numerous structurally or functionally interconnected brain regions [1]. More and more experiments are also exploring some mental diseases such as Alzheimer's disease, depression, and schizophrenia, which are suspected to link to functional connections in the brain. Schizophrenia is a severe mental disease with extensive cognitive impairment. However, there is still no clear conclusion on its pathological mechanism and affected brain regions, and it is generally believed to be caused by the mal-integration of brain functions [2][3][4]. Machine recognition technology can be used as a clinical diagnostic indicator of brain diseases from the brain function connection matrix obtained from fMRI data [5]. Nevertheless, brain function connection matrices are generally high-dimensional data, and existing machine recognition technologies cannot be directly used for data processing. Therefore, some traditional linear dimensionality reduction methods, such as PCA (Principal Component Analysis) and ICA (Independent Component Analysis), are used in the data pre-processing[6]. However, these dimensionality reduction methods have various restrictions. Thus, before applying the above methods, it must be assumed that the data conform to specific statistical characteristics. In order to solve the above problems, **H.Shen et al.**, [7] introduced low-dimensional embedding to fMRI; **S.Lee et al.**, [8] constructed a Gaussian process classifier to study abnormal brain functional connections; **H.GA et al.**,[9] researched on multi-scale extraction method of event-related fMRI data. However, these methods are limited to regional analysis. The inability to analyse the differences of the brain network as a whole leads to the loss of some critical classification information. Therefore, this study focused on the overall brain functional network, which transformed the fMRI data of each subject into the time series of each brain region. The wavelet decomposition method was chosen to establish the brain connection network. Based on the connectivity matrix, machine learning classifiers (SVM and kNN) were trained to classify whether the subject suffers from schizophrenia. With accuracy indexes as evaluation for classifiers, the study concluded the choice of the classifier to support the diagnosis of schizophrenia. **Figure 1** shows the experiment process of this study.

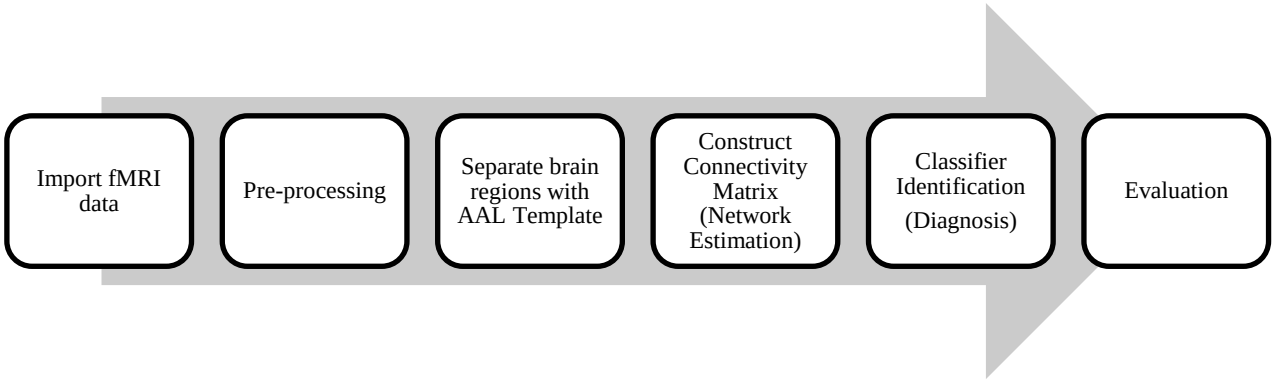


Figure 1 Overview of this study

Capítulo 2. Background

2.1 Functional Magnetic Resonance Imaging and Brain Activity

fMRI (functional Magnetic Resonance Imaging) is one of the most broadly used neuroimaging technologies which measures relative variations in deoxygenated haemoglobin for continuous brain activity. The most common method of fMRI is blood oxygen level-dependent (BOLD) imaging [10]. fMRI relies on the magnetization vector difference between oxyhaemoglobin and deoxygenated haemoglobin to generate fMRI signals. When a brain region becomes active, the haemoglobin will transport more blood flow with oxygen to the neurons. As a result, it could illustrate various magnetic properties during the process from oxygenation to deoxygenation. **Figure 2** shows the mechanism of the BOLD signal. It is a subordinate indicator of neural events based on blood flow regulated by local brain metabolism.

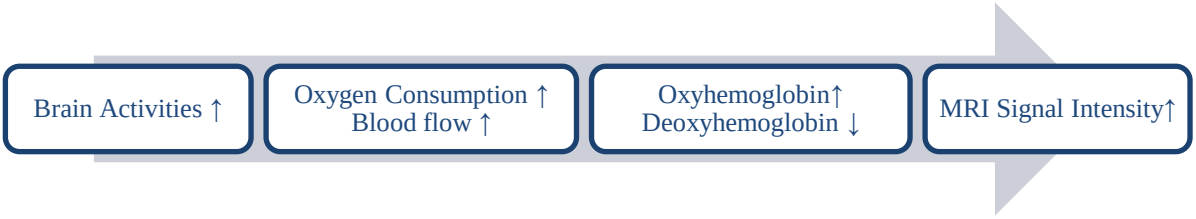


Figure 2 BOLD signal mechanism for fMRI.

The recent fMRI analyses have observed changes in BOLD amplitude by means of external stimulus, which determine the specialised function of a specific brain region (functional separation).

The precise amplitude of the BOLD signal cannot be compared across different subjects because fMRI is a contrast but not a quantitative imaging technique. Therefore, typically, the process aims to determine the relative differences in BOLD signal amplitude across two tasks of one subject. This has been designed in a block: a subject performs a certain task for about half a minute before fixed time for rest, and the cycle is repeated (**Figure 3**). At the same time, the corresponding BOLD signals are collected. Alternating between task and rest generates the images required for inferring brain activity. The voxel time courses are then collected for further analysis.

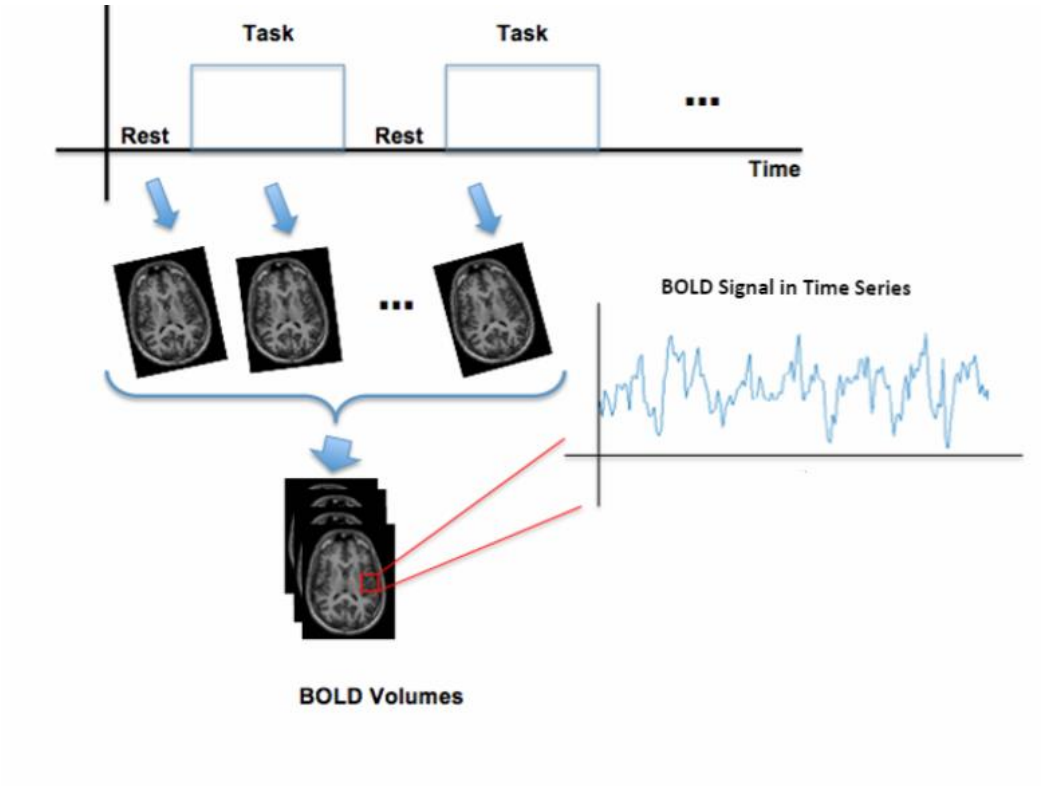


Figure 3 Example of fMRI design experiment.

The figure is adapted from [11]

2.2 Brain Network Connectivity

2.2.1 Brain Network based on Graph Theory

The human brain is one of the most complex systems in nature. It is estimated that there are about 1011 neuronal cells in the brain of an adult. These huge numbers of neuronal cells are connected through about 1015 synapses, forming a highly complex brain structure network. More and more evidence shows that this complex and huge network is the physiological basis of the brain for information processing and cognitive expression [12].Graph theory, as an important analysis tool in the field of complex network analysis, has been widely used in brain network study to detect the internal organization pattern of the brain. In graph theory analysis, a complex network can be abstracted into a graph *G*. Graph *G* is a set of nodes connected by edges [13]. **Figure 4** shows an example of a network. The black dots with numbers are the nodes and the blue line between the vertices is the edge. In the brain network, the brain area or voxel could be regarded as a node, and the functional connection or structural connection between the brain area or voxel could be viewed as an edge. In this way, the brain network can be abstracted into a graph to represent, and then the graph theory analysis method could be used to study the topological properties of brain networks.

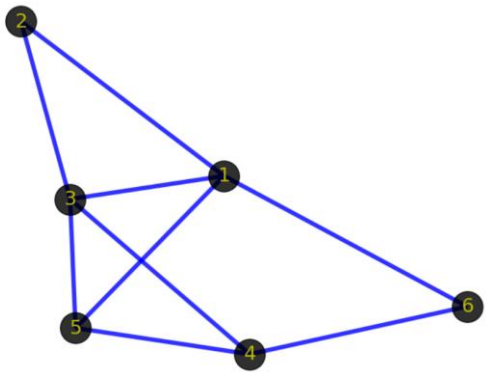


Figure 4 An example of Graph *G*

2.2.2 Define Brain Network Nodes

There are three commonly used methods to define nodes in brain network analysis:

(1) Each voxel is defined as a node. This method can analyse the brain network at the maximum resolution, but the amount of calculation is not conducive to further research.

(2) Each independent component as a node. According to the spatial independence between voxels, the whole brain is divided into several independent components using the ICA algorithm. This method does not require any prior knowledge and is a purely data-driven method. However, the independent components of the segmentation lack physiological interpretation, and it is challenging to apply the experimental results to clinical treatment.

(3) The physiological anatomy templates are used to segment the brain area. The method could take advantage of numerous templates. Also, the experimental method is simple and reproducible, which is used by most researchers now. Therefore, in this study, the AAL (Anatomical Automatic Labelling) template is used to divide the whole brain into 116 brain regions previously validated and reported by **Tzourio-Mazoyer et al.** [14]. Each brain region represents an independent node in the brain network, and the average voxel time series of each node is extracted to define the node time.

2.2.3 Define the Brain Network Connectivity

The edge in the brain network often consists of the anatomical connection, functional connection, or effective connection of the brain. Among the different types of connections, functional connection represents the cross-correlation degree of signal changes of the BOLD level in brain regions, and effective connection represents a causal connection in the brain region. Traditionally, the Pearson correlation is used for estimating the cross-correlation degree:

$$\rho_{x,y} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} \tag{2}$$

However, the traditional Pearson correlation cannot reflect the time series details. In order to overcome this problem, the wavelet correlation coefficient between nodes is selected as the edge weight of the functional connection. The specific algorithm for calculating wavelet correlation will be discussed in **Part 3.3**.

2.3 Classification

In general, there are four basic approaches to machine learning: supervised, unsupervised, semi-supervised, and reinforcement learning. Classification is a form of supervised learning.[15] At the first stage of training, the decision function (i.e., "classifier") learns from the values of features in a set of independent sample inputs and corresponding outputs. For a neuroimaging case, the "features" could be voxels' information and the "class label" could be the physical condition of subjects (**Figure 5**).

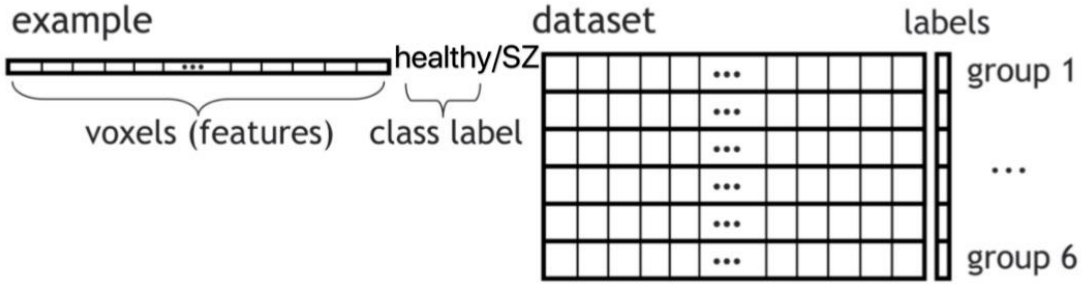


Figure 5 Example of the distribution of voxels and datasets

We will denote a sample with the row vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_v]$ and its class label as y. Then the classifier is used to generate category assignments (i.e. "labels") with a given precision. The learned classifier shows the relationship between the features and the class labels in the training dataset. We can use the formula "$\hat{y} = f(\boldsymbol{x})$." to signifies the relationship, where $\hat{y}$ the represents predicted labels.

Once the classifier has been trained based on the features, the relationship should be tested by different datasets. In other words, if the classifier truly "learned" the relationship between features and classes, it can automatically predict and assign the classification labels to the new classes of examples. When we compare the predicted labels with classifier with true labels, we will find an estimate of its performance. To be specific, we will represent the training and testing datasets with X_{train} and X_{test} and label matrices as the column vectors y_{train} and y_{test} respectively. Its accuracy could be measured to evaluate how well the classifier manifests (**Figure 6**).

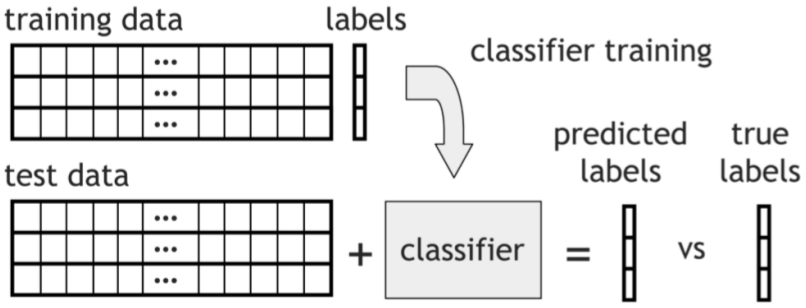


Figure 6 The sketch map of training and testing process

Capítulo 3. Design and Implementation

3.1 Dataset

The fMRI dataset is from OpenNEURO, a free and open platform for sharing MRI, MEG, EEG, iEEG, and ECoG data. The specific dataset used in this study is "Working memory in healthy and schizophrenic individuals" uploaded by Chris Gorgolewski on 2016-10-2. The participants for this dataset were recruited through the Conte Centre for the Neuroscience of Mental Disorders(CCNMD) at Washington University School of Medicine in St. Louis included: (1) individuals with DSM-IV Schizophrenia; (2) the non-psychotic siblings of individuals with schizophrenia; (3) healthy controls; and (4) the siblings of healthy controls. Siblings were full siblings, based on self-report. All participants gave written informed consent for participation and all participants had been included in a previous report on resting-state functional connectivity changes in schizophrenia[16]. For this study, ten subjects were chosen for implementation. **Table 1** shows the specific information of 10 subjects.

Participant_id	Condition	Gender	Age
sub-01	SCZ	MALE	28.961
sub-02	SCZ-SIB	MALE	26.8419
sub-03	SCZ-SIB	FEMALE	29.4648
sub-04	SCZ-SIB	MALE	25.8344
sub-05	SCZ	FEMALE	25.6454
sub-06	CON-SIB	FEMALE	24.5941
sub-07	CON	MALE	27.5838
sub-08	CON	FEMALE	18.768
sub-09	CON-SIB	FEMALE	21.2594
sub-10	CON-SIB	FEMALE	21.3005

Table 1 The subjects' basic information in this study

Note: SCZ represents the subjects with Schizophrenia, CON represents the subject is healthy. SCZ-SIB represents the siblings of subject affected Schizophrenia and CON-SIB represents the siblings of subject who are unaffected (i.e., healthy).

3.2 Pre-Processing

The goals of pre-processing are to remove uninteresting variability from the data, improving functional signal-to-noise ratio (SNR) as much as possible. So the data after pre-processing could be efficiently used for further statistical analysis.

3.2.1 Brain Skull Removal (Using FSL FEAT's Brain Extraction Technique)

Since fMRI studies are centred on brain tissue, our first step is to remove the skull and non-brain areas from the image. FEAT is a software tool for high quality model-based fMRI data analysis, with an easy-to-use graphical user interface (GUI). FEAT is part of FSL (FMRIB's Software Library)[17].

Firstly, the tool creates the intensity histogram to find the stable minimum and maximum intensity values and determines a coarse brain outline. At the same time, the centre of the head image was generated, as well as the approximate size of the head of the image. Next, the image goes through initialisation of tessellated surface, and gradually be warped by one vertex for one time. When the process goes on moving towards the brain's edge, it keeps the surface well-spaced and smooth all the time. If a properly clean solution is not acquired, then we should re-run with a higher smoothness constraint for the iterative process [18]. **Figure 7** illustrates the comparison of the brain image before and after the brain extraction process.

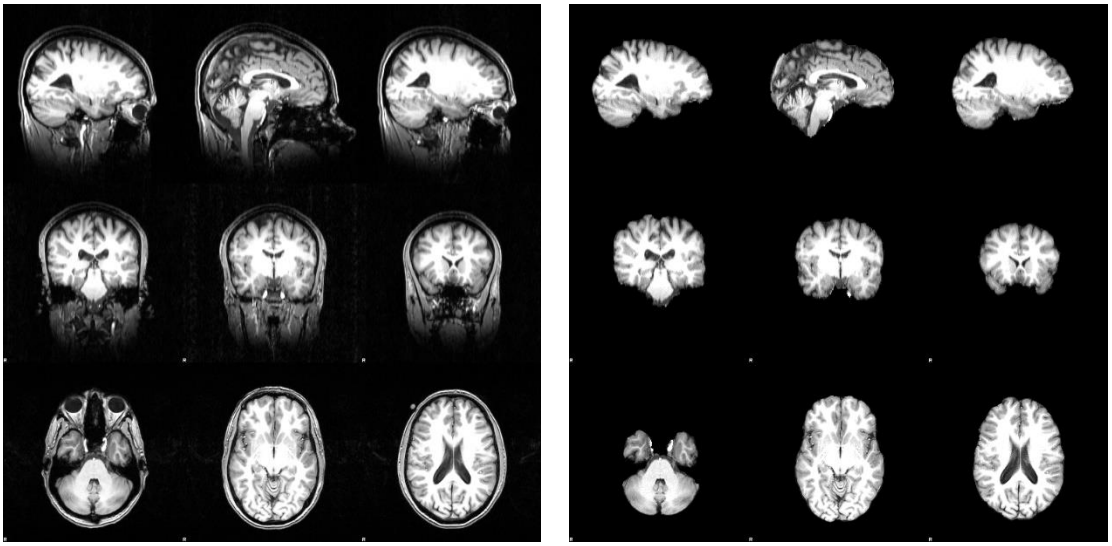


Figure 7 The comparison between the image before (left) and after (right) brain skull removal

3.2.2 Slice Time Correction (SPM)

Due to the order of slices that can be acquired in different ways, the BOLD signal in time series could be thus totally different. For example, the part of the **Figure 8 (A)** marked in red shows one brain region is consistently active after the boost of a stimulus. **Figure 8 (B)**. shows the interleaved fifteenth to seventeenth slices sequence in the brain region. The BOLD signals for these three slices very different because they are not obtained at the same time. **Figure 8 (C)** shows the correct and real BOLD signal of the slices. However, **Figure 8 (D)** illustrates the BOLD signal in time series could be not coincident for each repeat time because of the slices acquired at different times, which will interfere with subsequent analysis.

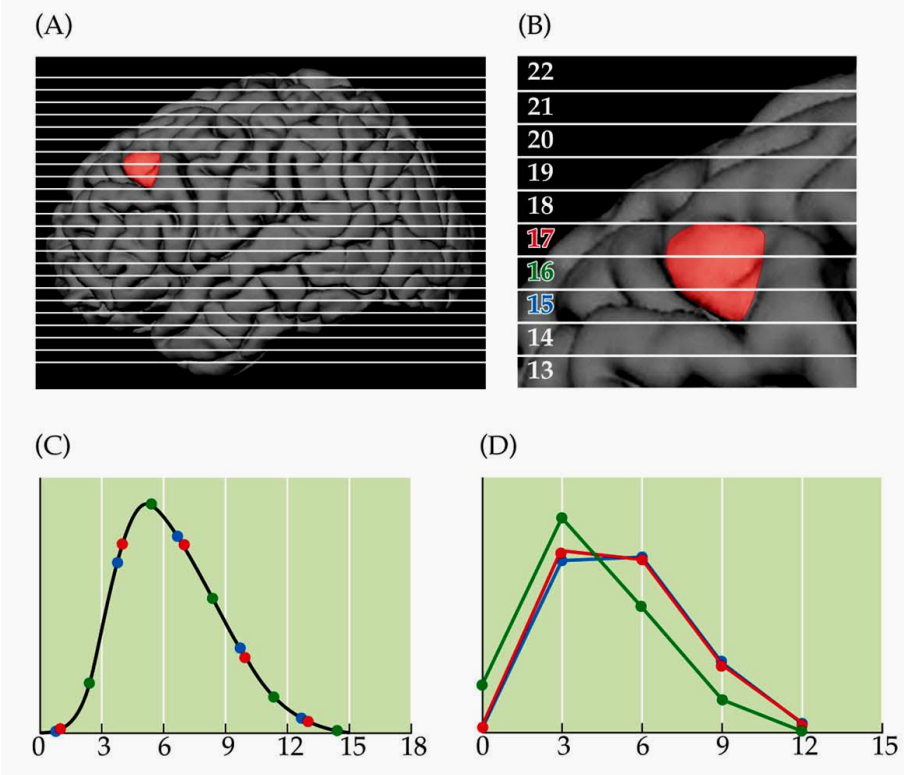


Figure 8 Effects of slice acquisition time upon the hemodynamic response

Slice Time Correction was achieved by adjusting specific parameters in the Statistical Parametric Mapping Software (SPM12, available at https://www.fil.ion.ucl.ac.uk/spm) to guarantee slices are coincident the same time. During the slice time correction process, all slices of one volume are interpolated with one specific slice as a reference which is considered as the foremost accurate slice without any interpolation. In contrast, other slices are generated by interpolation according to the reference [19].

3.2.3 Motion Correction (SPM)

Even small head movements can be a major problem: an increase in residual variance and fMRI data may get completely lost if sudden movements occur during a single volume. Therefore, motion correction process is to guarantee the identical brain of all fMRI images is at the same position. A function called Motion Correction in SPM software is used to solve the motion problem. In the first step, the motion parameter matrix is estimated between each image and the reference image. Then the parameters gained from each image are used to re-slice the image that best matches the reference image[20]. **Figure 9** shows an example of the motion correction process and the effect on the original image after motion correction.

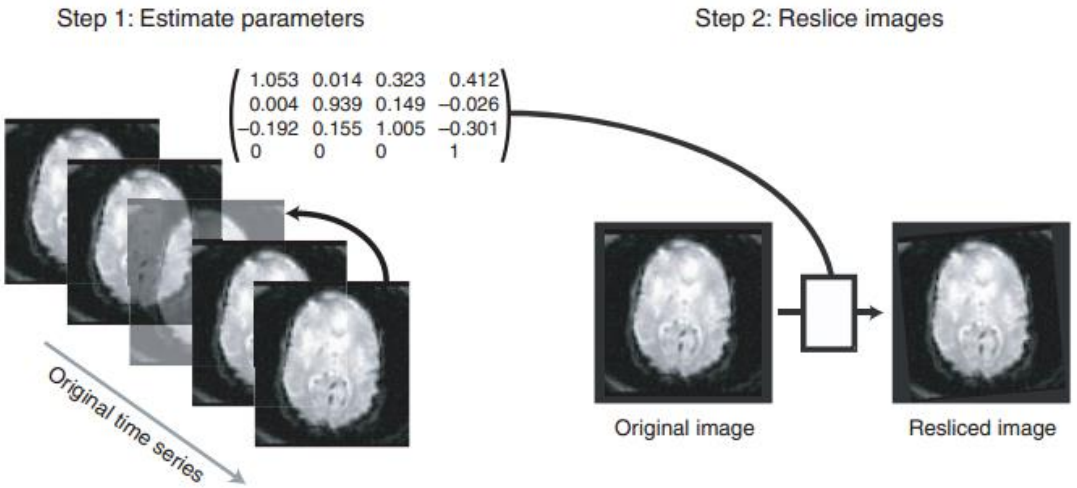


Figure 9 An example of Motion Correction process

3.2.4 Spatial Smoothing (SPM) Using a Gaussian Filter

Spatial smoothing in neuroimaging is to make the image seem smoother, which applies a blurring filter across the image to calculate the mean value of the part of the intensities from neighbouring voxels together. The effect is to blur the sharp edges and to increase SNR of fMRI images. In addition, spatial smoothing can offer assistance compensation for mistakes in the inter-subject alignment process. Therefore, adjusting the parameter of a Gaussian filter in the SPM software and making its diameter to be equal to 8 voxels had been done in this project for spatial smoothing. **Figure 10** shows the results before and after spatial smoothing process.

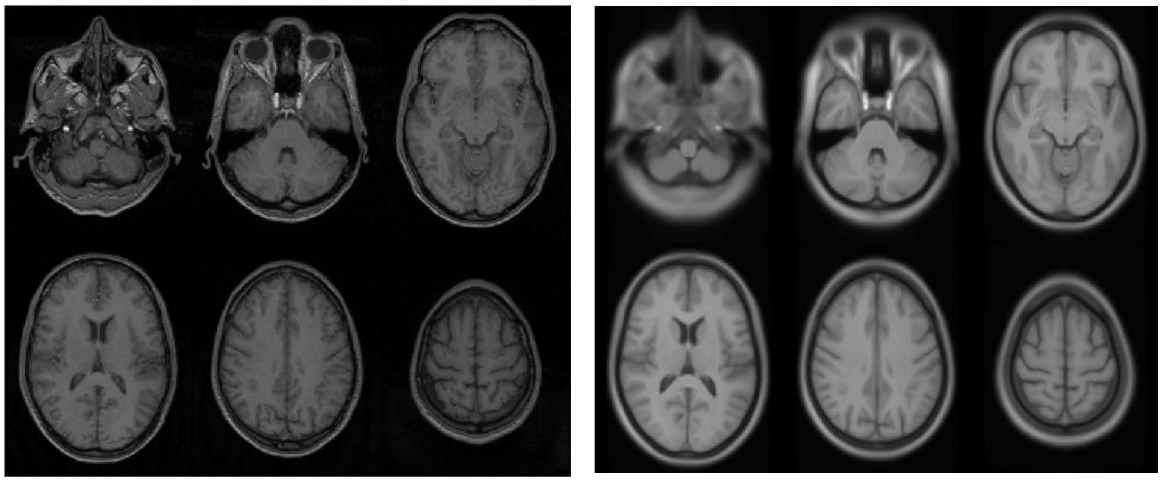


Figure 10 The brain image comparison before smoothing (left) and after smoothing (right)

3.3 Network Estimation

In **Part 2.2**, we have discussed some basics of brain network connectivity. Nodes and edges define the brain network, and the brain can be separated into 116 regions using the AAL template. Thus,

after pre-processing fMRI data with four steps above, the fMRI data could be expressed as a 4-D matrix with the size of 64*64*36*137. The first step for network estimation is to match each voxel with its corresponding brain region. Then the BOLD signal in time series for each region is generated to estimate every mean signal for each of the regions. **Figure 11** illustrates BOLD mean signal in time series of subject 1 as an example. Every coloured line represents each brain region, plotted by time on the horizontal axis and signal intensity on the vertical. At last, this study had implemented wavelet cross-correlation as the element of the brain network connectivity matrix.

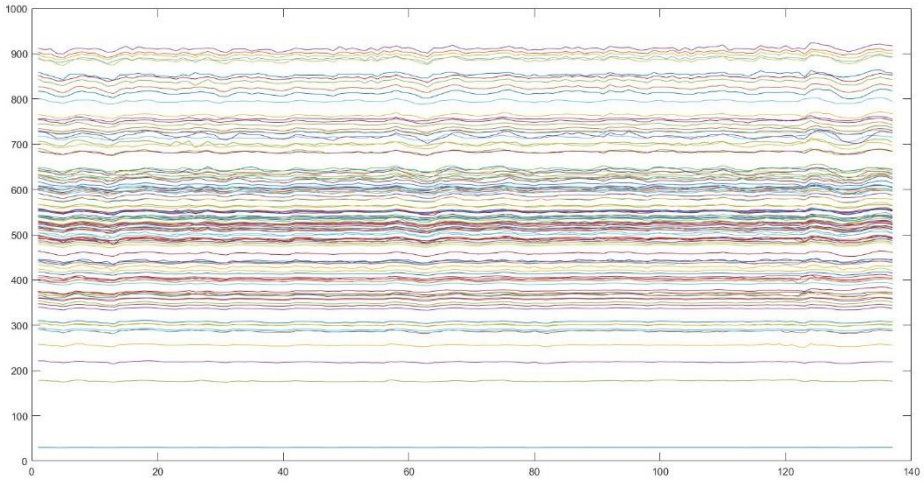


Figure 11 BOLD mean signal in time series for 116 regions of subject 1

The specific wavelet cross-correlation provide more information of time series in details than traditional Pearson correlation. Suppose the time series $\boldsymbol{X} = \{x_{i,1}, x_{i,2}, \dots, x_{i,N}\}$ is the average voxel time series for brain region i. According to the Maximal Overlap Discrete Wavelet Transform (MODWT), we can suppose $\{h_{j,l};\ l = 0,1,\dots,L_j - 1\}$ and $\{g_{j,l};\ l = 0,1,\dots,L_j - 1\}$ as the wavelet filter and scale filter respectively at the scale of j , where $L_j = (2j - 1) * (L - 1) + 1$ and L represents the length of initial filter. Assume the wavelet and scale coefficients are W_j and V_j respectively1 at the scale of j and satisfy the equations:

$$W_{j,t}^{(X)} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,1} X_{t-l \bmod V} \tag{3}$$

$$V_{j,t}(X) = \sum_{l=0}^{L_j-1} \tilde{g}_{j,1} X_{t-l \bmod N} \tag{4}$$

where $\tilde{h}_{j,1} = h_{j,1}/2^{j/2}$ and $\tilde{g}_{j,1} = g_{j,1}/2^{j/2}$.

Then the time series $\{\boldsymbol{X_t}\} = \{x_{i,1}, x_{i,2}, \dots x_{i,N}\}$ and $\{\boldsymbol{Y_t}\} = \{x_{j,1}, x_{j,2}, \dots x_{j,N}\}$ represents the average voxel time series for brain region i and j respectively, where N is the number of available time slices satisfying $N \geq L_j$, $t = 0,1,\dots,N-1$, which are gaussian processes with steady

increments. And the scale dependent covariance could be defined as:

$$\eta_{XY}(\lambda_j) = \frac{1}{N}\sum_{l=L_j-1}^{N-1} W_{j,1}^{X} W_{j,1}^{Y} \tag{5}$$

The unbiased estimate is

$$\gamma_{XY}(\lambda_j) = \frac{1}{2\lambda_j}\mathrm{Cov}\left\{W_{j,t}^{(X)} W_{j,t}^{(Y)}\right\} \tag{6}$$

where $\{\boldsymbol{W}_{\boldsymbol{j,l}}^{(\boldsymbol{X})}\}$ and $\{\boldsymbol{W}_{\boldsymbol{j,l}}^{(\boldsymbol{Y})}\}$ are the MODWT coefficients at the $\lambda_j = 2^{j-1}$ for $\{\boldsymbol{X_t}\}$ and $\{\boldsymbol{Y_t}\}$.

Then wavelet correlation coefficients of time series can be defined as:

$$\rho_{XY}(\lambda_j) = \frac{\gamma_{XY}(\lambda_j)}{v_X(\lambda_j)v_Y(\lambda_j)} \tag{7}$$

where $v_X^2(\lambda_j) = \mathrm{Var}(W_j)/2\lambda_j$ [21].

This study decomposed the average voxel time series as four levels. The cross-correlation coefficients could be high at some level(s) and be low at other level(s). That is to say, the signals could be correlated in some frequency band(s) and could be not correlated in other frequency band(s). Two signals are highly correlated if a high correlation coefficient is obtained in any of the frequency bands they were decomposed. Therefore, the maximum correlation was chosen to establish the brain network. The brain connectivity matrix then could be defined by cross-correlation coefficients computed between each pair of these regions as:

$$\boldsymbol{C} = \begin{bmatrix} \rho_{1,1} & \rho_{1,2} & \cdots & \rho_{1,116} \\ \rho_{2,1} & \rho_{2,2} & \cdots & \rho_{2,116} \\ \vdots & \vdots & & \vdots \\ \rho_{116,1} & \rho_{116,2} & \cdots & \rho_{116,116} \end{bmatrix} \tag{8}$$

The brain connectivity network matrix and distribution results will be shown in the **Part 4.1**.

3.4 Classification

After we built the brain network and obtained the network connectivity, we use some popular machine learning classifiers to distinguish if the subjects suffer from schizophrenia. We have defined 116 brain regions, i.e., 116 nodes as parameters to train the specific classifiers. In this study, Support Vector Machine (SVM) and K-Nearest Neighbour (kNN) were performed as classifiers to classify the fMRI data.

3.4.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) could be a data processing technique supported by applied mathematics learning theory, which belongs to belongs to supervised learning model with relevance

learning algorithm. The analytical data are used for classification and regression analysis. The mechanism of SVM is to find an optimal dividing hyperplane to separate datasets while making the blank area (i.e., margin) on both sides of the hyperplane as wide as possible for maximum separation of datasets.

Given a training sample set (xi, yi), where $i = 1,2,\dots,l, \boldsymbol{x} \in R^n, y \in \pm 1$, and hyperplane is written as $(\boldsymbol{w} \cdot \boldsymbol{x}) + \mathrm{b} = 0$. For the sake of classifying all samples accurately and correctly and guarantee the classification interval, the following constraints need to be met: $yi[(\boldsymbol{w} \cdot \boldsymbol{x}) + \mathrm{b}] \geq 1$, where $i = 1,2,\dots,l$. Therefore, we can estimate the classification interval as $2/||\boldsymbol{w}||$. Thus, the problem of generating a hyperplane is translated into a constraint to be calculated:

$$\min \Phi(w) = \frac{1}{2}\|w\|^2 = \frac{1}{2}(w' \cdot w) \tag{9}$$

To solve this constrained optimization equation, the Lagrange function should be introduced:

$$\mathrm{L}(w, b, \mathrm{a}) = \frac{1}{2}\|w\| - a(y((w \cdot x) + b) - 1) \tag{10}$$

where Lagrange multiplier $a_i > 0$. The solution of the constrained optimization problem is determined by the saddle point of the Lagrange function, meeting the condition of the partial derivatives of *w* and *b* at the saddle point to be zero. This quadratic programming problem could be converted into the dual problem:

$$\begin{aligned} &\max Q(a) = \textstyle\sum_{j=1}^{l} a_j - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_i a_j y_i y_j (x_i \cdot x_j) \\ &\text{s.t.} \textstyle\sum_{j=1}^{l} a_j y_j = 0 \quad j = 1, 2, \cdots, l, a_j \geqslant 0, j = 1, 2, \cdots, l \end{aligned} \tag{11}$$

Therefore, the optimal solution: $\boldsymbol{a}^* = (a_1^*, a_2^*, \dots, a_l^*)^T$ *can be used to* calculate the superior weight vector $\boldsymbol{w}^*$ and optimal bias b^* as:

$$w^* = \sum_{j=1}^{l} a_j^* y_j x_j \tag{12}$$

$$b^* = y_i - \sum_{j=1}^{l} y_j a_j^* (x_j \cdot x_i) \tag{13}$$

Where $j \in \{j \mid a_j^* > 0\}$. Therefore, optimal classification hyperplane $(\boldsymbol{w} \cdot \boldsymbol{x}) + \mathrm{b} = 0$ and optimal classification function is obtained:

$$f(\boldsymbol{x}) = \operatorname{sgn}\{(\boldsymbol{w}^* \cdot \boldsymbol{x}) + b^*\} = \operatorname{sgn}\left\{\left(\sum_{j=1}^{l} a_j^* y_j (x_j \cdot x_i)\right) + b^*\right\}, \boldsymbol{x} \in R^n \quad (15)$$

For the case of linear inseparability, the main idea of SVM is to map the input vector to a high-dimensional eigenvector space and construct an optimal classification surface in the feature space.

Let apply Φ to ***x***, which transforms ***x*** from input space $\mathbf{R^n}$ to characteristic space **H** as follows:

$$\boldsymbol{x} \to \Phi(\boldsymbol{x}) = (\Phi_1(\boldsymbol{x}), \Phi_2(\boldsymbol{x}), \cdots, \Phi_l(\boldsymbol{x}))^{\mathrm{T}} \quad (16)$$

By replacing the input vector x with the characteristic vector $\Phi(\boldsymbol{x})$, the optimal classification function can be obtained as [22]:

$$f(\boldsymbol{x}) = \operatorname{sgn}(\boldsymbol{w} \cdot \Phi(\boldsymbol{x}) + b) = \operatorname{sgn}\left(\sum_{i=1}^{l} \mathrm{a}_i y_i \Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}) + b\right) \quad (17)$$

In the dual problem above, the target function and the decision function only contain the inner product operation between training samples. However, the calculation of the inner product in a high-dimensional space also has a problem of a large amount of calculation. So the next step is to find a function from origin space $K(x_i, x_j) = <\Phi(x_i), \Phi(x_j)>$. In this way, we can simplify the calculation of the inner product after the mapping process. This simplification is called the kernel technique and the function K is called the kernel function. In this project, I designed and implemented three kinds of kernel: Linear, Gaussian and Polynomial.

(1) Linear Kernel: $K(x_i, x_j) = <x_i, x_j>$

The linear kernel function is the most common and lowest-level kernel function. SVMs with linear kernel have been used in many applications and are still in use today because of the increased robustness and computational speed with respect to other kernels. When the samples are separable in the low-dimensional space, the linear kernel function can be used to classify the samples without converting to the high-dimensional space.

(2) Polynomial Kernel: $K(x_i, x_j) = (<x_i, x_j> +1)^d$

The polynomial kernel function belongs to the global kernel function, its locality is poor, and the sample points that are far away can also affect the classifier. The parameter ***d*** represents the dimensionality of the kernel function. The larger ***d***, the higher the dimensionality of the mapping function. At this time, it is easier to classify the samples, but the computational complexity also increases. Although the complex classifier can have a good classification effect and the training sample achieves a high recognition rate, the classification performance is weak for new samples, that is, the phenomenon of "overfitting" has occurred.

(3) Gaussian Kernel: $K(x_i, x_j) = exp\left(-\gamma \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2\right)$

γ is the scope of the kernel function. The Gaussian kernel has a good classification effect on relatively close sample points, and the local performance is quite good. However, as γ increases, its generalization ability weakens so the global performance is poor.

3.4.2 K- Nearest Neighbour (kNN) Classifier

kNN is also a supervised learning algorithm at first proposed by Cover and Hart and has been broadly utilized in different areas like pattern recognition and data processing. The classification idea is to calculate the distance between the data points ***x*** to be classified, and all the data points in the existing data set ***X***. Then Take the first *K* points with the smallest distance, and divide this data point into the category with the highest number of occurrences.

The kNN algorithm could be described as the following steps:
(1) Pre-set a training dataset ***X***.
(2) Set the initial value of *K*. There is no uniform method for determining the *K* value (the *K* value selected according to the specific problem may be quite different). The general approach determine *K* value is to assign a rough initial value, and then continuously debug according to the experimental results, and finally reach the optimal. In this study, *K* was chosen to be 3 after debugging.

(3) Select three samples closest to the sample data to be tested from the training data. It is supposed that the sample point x in dataset ***X*** exists in n-dimensional space R^n, and the "nearest neighbors" between samples are measured by different types of distances $d(x_i, x_j)$. Suppose the *i*-th sample $x_i = \left(x_1^i, x_2^i, \ldots, x_n^i\right) \in R^n$, where x_l^i represents the *l*-th feature attribute value of the *i*-th sample. [23] Then three kinds of distance in this project could be defined as:

i. Standard Euclidean Distance: $d(x_i, x_j) = \sqrt{\sum_{l=1}^{n} \frac{(x_i - x_j)^2}{s_i}}$, where s_i is the standard deviation of x_i
ii. Hamming Distance: Given two vectors $x_i, x_j \in F^n$, we define the hamming distance between x_i and x_j, $d(x_i, x_j)$, to be the number of places where x_i and x_j differ.
iii. Gaussian Distance: In this study, we choose $\gamma = \frac{1}{2}$ so the Gaussian Distance could be defined as : $d(x_i, x_j) = 1 - \exp\left(-\frac{1}{2}\|x_i - x_j\|^2\right)$

(4) Given a sample to be classified $\mathbf{x}_q$, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ represents *K* samples nearest to $\mathbf{x}_q$. Let the discrete objective function (classification problem) be $f: R^n \rightarrow v_i$,where v_i is the label of *i*-th class and label set could be defined as $V = \{v_1, v_2, \ldots, v_s\}$. Then we can express the prediction function as:

$$\tilde{f}(\mathbf{x}_q) = \arg\max_{v \in V} \sum_{i=1}^{n} \delta(v, f(\mathbf{x}_i)) \tag{18}$$

where $\tilde{f}(\mathbf{x}_q)$ is the estimated $f(\mathbf{x}_q)$ and $\delta(v, f(\mathbf{x}_q))$ is Dirac delta function.
(5) $\tilde{f}(\mathbf{x}_q)$ is the class of the sample $\mathbf{x}_q$ to be classified.

3.5 Evaluation

3.5.1 Confusion Matrix

The confusion matrix is a situation analysis table that summarises classification result in the method of matrix. Take the binary classification problem as an example: there are two types of records in the

dataset: positive and negative categories. The classifier may make two types of judgments based on the categories in the dataset: positive or negative judgment.

The confusion matrix is a 2 × 2 situation analysis table displaying the quantity of the four types of records: positive records with correct judgments (True Positives, TP), positive records with incorrect judgments (False Negatives, FN), negative records with correct judgments (True Negative, TN) and negative records with false judgments(False Positives, TN). The structure of the confusion matrix could be defined as [24]:

$$\begin{bmatrix} number\ of\ TP & number\ of\ FP \\ number\ of\ FN & number\ of\ TN \end{bmatrix} \tag{20}$$

3.5.2 Accuracy Index

Based on the above confusion matrix, the following information can be derived:

(1) Total number of records in the data set = TP + FP + FN + TN

(2) Number of positive records in the data set = TP + FN

(3) Number of negative records in the data set = FP + TN

(4) Number of records for which the classification model made a positive judgment = TP + FP

(5) Number of negative judgments made by the classification model = FN + TN

(6) Number of records that the classification model correctly classified = TP + TN

(7) Number of records misclassified by the classification model = FP + FN

The classification accuracy index can be calculated by:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{21}$$

In order to make more precise comparisons among the classifiers with the same accuracy index, the accuracy with confidence weight could be determined as:

$$Weighted\ Accuracy = \frac{\sum TP\ subject * Confidence + \sum TN\ subject * Confidence}{TP+FN+FP+TN} \tag{22}$$

Capítulo 4. Results and Discussion

4.1 Network Estimation Results

As discussed in **Part 5.3**, the brain network estimation can be defined as a matrix with a size of 116*116. **Figure 12** - **Figure 21** illustrate the visualised estimated matrix of all ten subjects. **Figure 22** - **Figure 31** shows the wavelet correlation distribution of subject 1 to 10, plotted by wavelet correlation coefficient on the horizontal axis and the number of regions correlated in specific range on the vertical. Note that it has been optimised to make the distribution of cross-correlation coefficients more distributed in the range [0, 1].

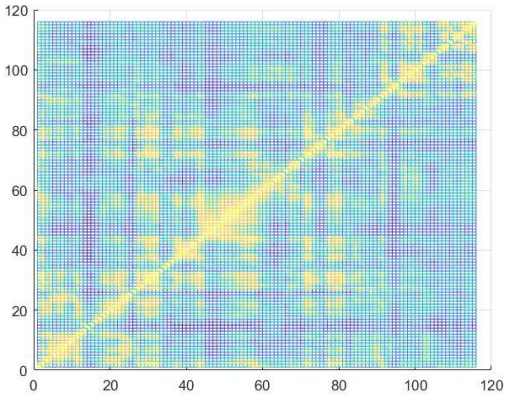


Figure 12 The visualised matrix of Subject 1

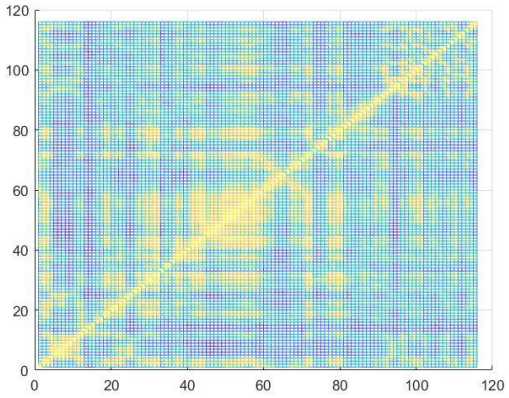


Figure 13 The visualised matrix of Subject 2

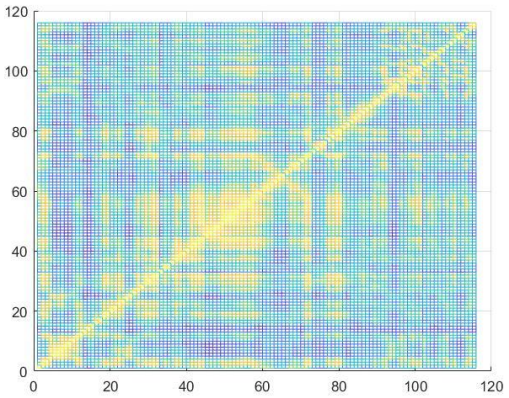


Figure 14 The visualised matrix of Subject 3

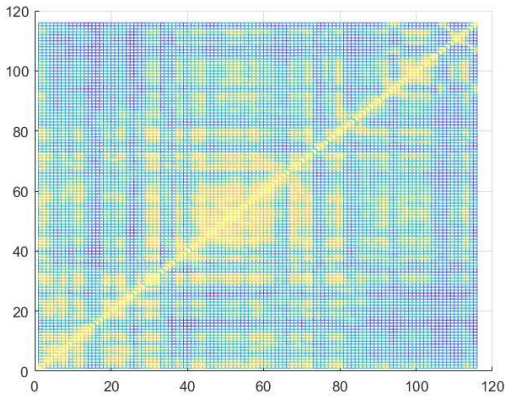


Figure 15 The visualised matrix of Subject 4

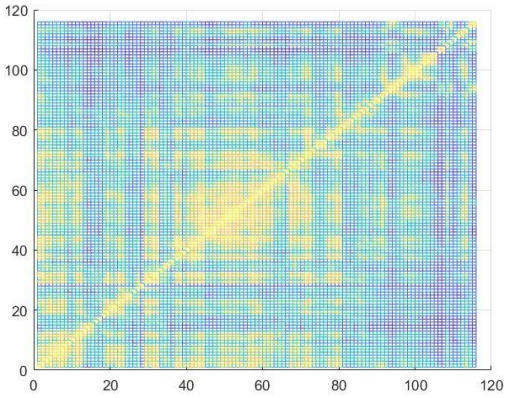


Figure 16 The visualised matrix of Subject 5

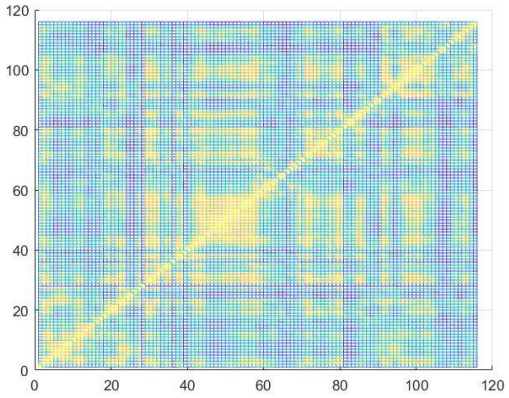


Figure 17 The visualised matrix of Subject 6

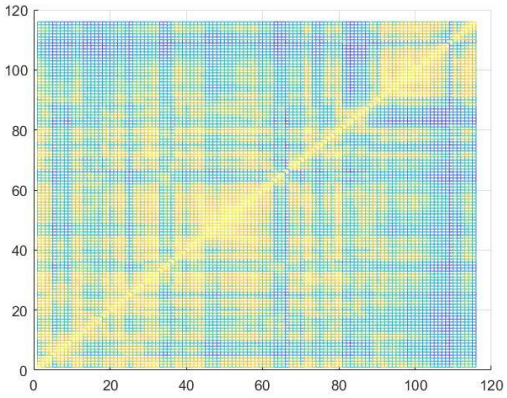


Figure 18 The visualised matrix of Subject 7

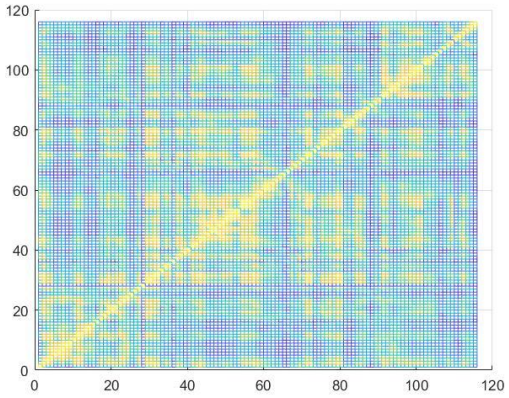


Figure 19 The visualised matrix of Subject 8

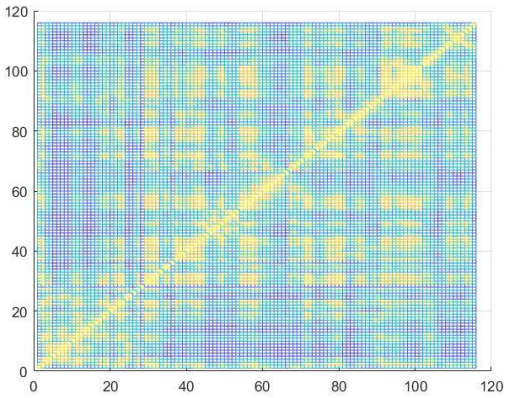


Figure 20 The visualised matrix of Subject 9

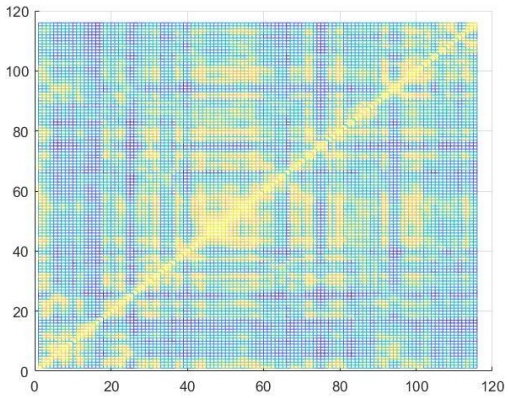


Figure 21 The visualised matrix of Subject 10

These figures show the correlation between two regions for corresponding subjects: yellow parts represent the regions are high-correlated, and blue parts represent they are low-correlated. This correlation information is used for subsequent classification.

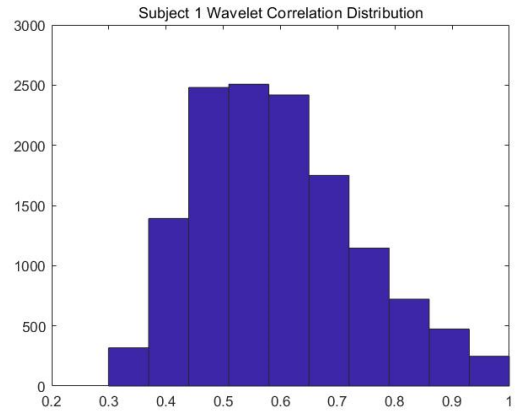


Figure 22 Subject 1 wavelet correlation distribution

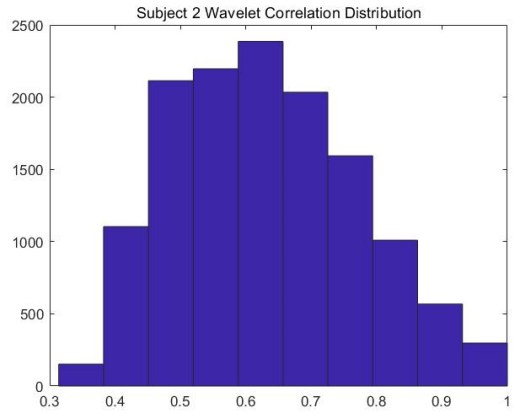


Figure 23 Subject 2 wavelet correlation distribution

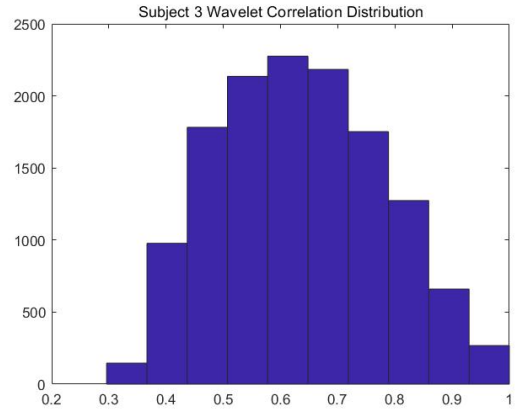


Figure 24 Subject 3 wavelet correlation distribution

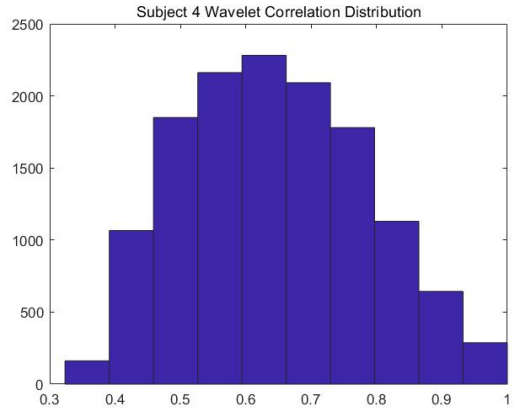


Figure 25 Subject 4 wavelet correlation distribution

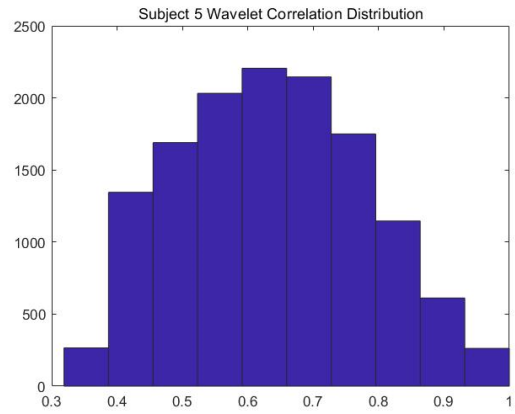


Figure 26 Subject 6 wavelet correlation distribution

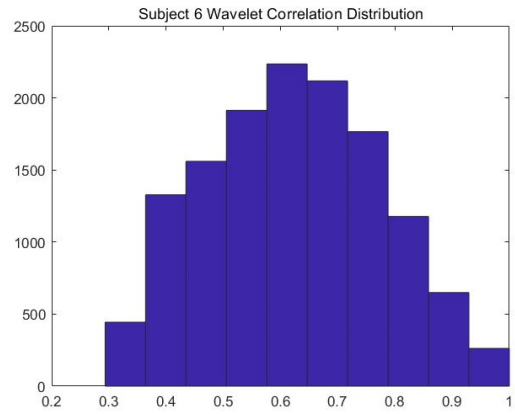


Figure 27 Subject 7 wavelet correlation distribution

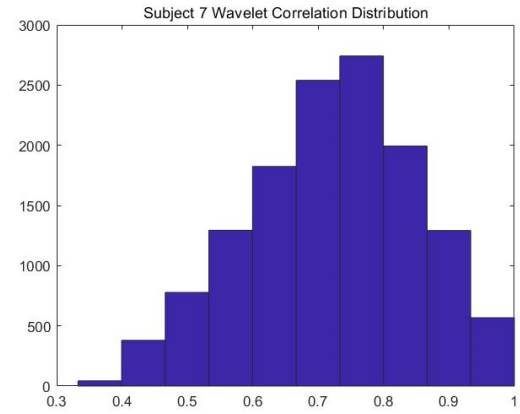


Figure 28 Subject 7 wavelet correlation distribution

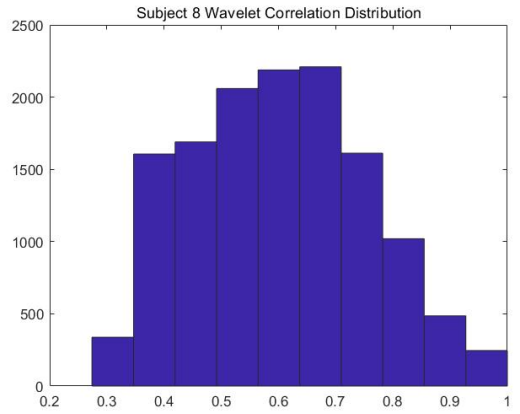


Figure 29 Subject 8 wavelet correlation distribution

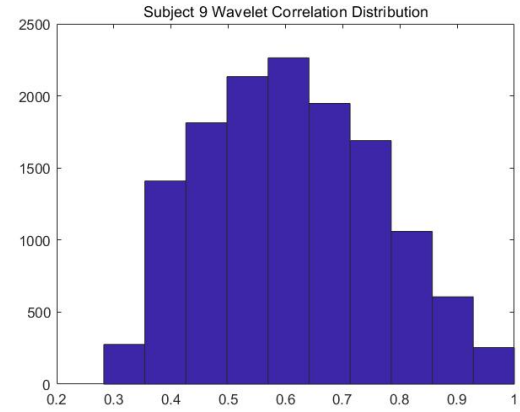


Figure 30 Subject 9 wavelet correlation distribution

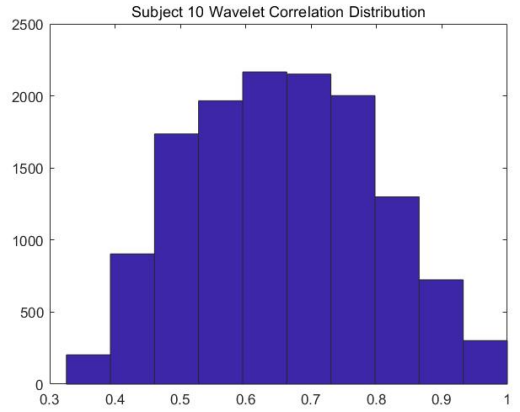


Figure 31 Subject 10 wavelet correlation distribution

4.2 Classification Results

Classification is based on the network estimation results. The leave-one-out (LOO) method was used to train and test the classifier. Suppose the data set contains N samples $(x_1, x_2, \ldots, x_n)$, and divide this sample into two parts: the first part with $N-1$ samples is used to train the classifier, and the other part with 1 sample is used to test. Iterating N times from N_1 to N_n in this way, all samples have undergone testing and training. The results of the classifiers can be evaluated as **Table 2** shows:

	Real Condition	Predicted Results
True Positive (TP)	Schizophrenia	Schizophrenia
False Positive (FP)	Healthy	Schizophrenia
True Negative (TN)	Healthy	Healthy
False Negative (FN)	Schizophrenia	Healthy

Table 2 Classifier evaluation table

In this study, the classification for each subject is based on 116 brain regions correlation coefficients. So if over 50% of regions' classification results (i.e., more than 58 regions) belong to a specific condition, the classification result for this subject is determined as a predicted condition. Additionally,

the confidence degree could be defined as:

$$Confidence = \frac{Number\ of\ regions\ of\ predicted\ condition}{Total\ number\ of\ regions\ to\ be\ classified} * 100\% \quad (19)$$

4.2.1 Classification Results Based on Support Vector Machine

4.2.1.1 Support Vector Machine with Linear Kernel

Participant_id	Predicted Results (Confidence)	Real Condition	Evaluation
sub-01	SCZ (81.9%)	SCZ	True Positive
sub-02	SCZ (97.4%)	SCZ-SIB	True Positive
sub-03	SCZ (97.4%)	SCZ-SIB	True Positive
sub-04	SCZ (51.7%)	SCZ-SIB	True Positive
sub-05	SCZ (88.0%)	SCZ	True Positive
sub-06	Healthy (98.2%)	CON	True Negative
sub-07	Healthy (68.1%)	CON	True Negative
sub-08	Healthy (98.3%)	CON-SIB	True Negative
sub-09	SCZ (75.0%)	CON-SIB	False Positive
sub-10	SCZ (99.1%)	CON-SIB	False Positive

Table 3 SVM with Linear Kernel classification results

4.2.1.2 Support Vector Machine with Gaussian Kernel

Participant_id	Predicted Results (Confidence)	Real Condition	Evaluation
sub-01	SCZ (83.6%)	SCZ	True Positive
sub-02	SCZ (100%)	SCZ-SIB	True Positive
sub-03	SCZ (100%)	SCZ-SIB	True Positive
sub-04	SCZ (53.4%)	SCZ-SIB	True Positive
sub-05	SCZ (95.7%)	SCZ	True Positive
sub-06	Healthy (100%)	CON	True Negative
sub-07	Healthy (84.5%)	CON	True Negative
sub-08	Healthy (100%)	CON-SIB	True Negative
sub-09	Healthy (90.6%)	CON-SIB	True Negative
sub-10	SCZ (95.7%)	CON-SIB	False Positive

Table 4 Table 3 SVM with Gaussian Kernel classification results

4.2.1.3 Support Vector Machine with Polynomial Kernel

Participant_id	Predicted Results (Confidence)	Real Condition	Evaluation
sub-01	SCZ (87.1%)	SCZ	True Positive
sub-02	SCZ (100%)	SCZ-SIB	True Positive
sub-03	SCZ (100%)	SCZ-SIB	True Positive
sub-04	SCZ (88.8%)	SCZ-SIB	True Positive
sub-05	SCZ (97.4%)	SCZ	True Positive
sub-06	Healthy (99.1%)	CON	True Negative
sub-07	Healthy (64.7%)	CON	True Negative
sub-08	Healthy (99.1%)	CON-SIB	True Negative
sub-09	SCZ (76.8%)	CON-SIB	False Positive
sub-10	SCZ (99.1%)	CON-SIB	False Positive

Table 5 SVM with Polynomial Kernel classification results

4.2.2 Classification Results Based on K-Nearest Neighbour

4.2.2.1 k-Nearest Neighbour with Standard Euclidean Distance

Participant_id	Predicted Results (Confidence)	Real Condition	Evaluation
sub-01	SCZ (70.7%)	SCZ	True Positive
sub-02	SCZ (100%)	SCZ-SIB	True Positive
sub-03	SCZ (100%)	SCZ-SIB	True Positive
sub-04	SCZ (97.4%)	SCZ-SIB	True Positive
sub-05	SCZ (97.4%)	SCZ	True Positive
sub-06	Healthy (99.1%)	CON	True Negative
sub-07	Healthy (88.8%)	CON	True Negative
sub-08	Healthy (99.1%)	CON-SIB	True Negative
sub-09	Healthy (97.4%)	CON-SIB	True Negative
sub-10	SCZ (84.4%)	CON-SIB	False Positive

Table 6 kNN with Standard Euclidean Distance classification results

4.2.2.2 k-Nearest Neighbour with Hamming Distance

Participant_id	Predicted Results (Confidence)	Real Condition	Evaluation
sub-01	SCZ (100%)	SCZ	True Positive
sub-02	SCZ (100%)	SCZ-SIB	True Positive
sub-03	SCZ (100%)	SCZ-SIB	True Positive
sub-04	SCZ (100%)	SCZ-SIB	True Positive
sub-05	SCZ (100%)	SCZ	True Positive
sub-06	SCZ (100%)	CON	False Positive
sub-07	SCZ (100%)	CON	False Positive
sub-08	SCZ (100%)	CON-SIB	False Positive
sub-09	SCZ (100%)	CON-SIB	False Positive
sub-10	SCZ (100%)	CON-SIB	False Positive

Table 7 KNN with Hamming Distance classification results

4.2.2.3 k-Nearest Neighbour with Gaussian Distance

Participant_id	Predicted Results (Confidence)	Real Condition	Evaluation
sub-01	SCZ (67.2%)	SCZ	True Positive
sub-02	SCZ (100%)	SCZ-SIB	True Positive
sub-03	SCZ (100%)	SCZ-SIB	True Positive
sub-04	SCZ (97.4%)	SCZ-SIB	True Positive
sub-05	SCZ (97.4%)	SCZ	True Positive
sub-06	Healthy (99.1%)	CON	True Negative
sub-07	Healthy (90.5%)	CON	True Negative
sub-08	Healthy (99.1%)	CON-SIB	True Negative
sub-09	Healthy (97.4%)	CON-SIB	True Negative
sub-10	SCZ (86.2%)	CON-SIB	False Positive

Table 8 KNN with Gaussian Distance classification results

4.3 Confusion Matrix

According to the classification results, the confusion matrices were obtained as **Table 9-Table 14** show:

Support Vector Machine with Linear Kernel	Schizophrenia	Healthy

Schizophrenia	4(TP)	2(FP)
Healthy	1(FN)	3(TN)

Table 9 SVM with Linear Kernel confusion matrix

Support Vector Machine with Gaussian Kernel	Schizophrenia	Healthy
Schizophrenia	4(TP)	1(FP)
Healthy	1(FN)	4(TN)

Table 10 SVM with Gaussian Kernel confusion matrix

Support Vector Machine with Polynomial Kernel	Schizophrenia	Healthy
Schizophrenia	5(TP)	2(FP)
Healthy	0(FN)	3(TN)

Table 11 SVM with Polynomial Kernel confusion matrix

k-Nearest Neighbour with Euclidean Distance	Schizophrenia	Healthy
Schizophrenia	5(TP)	1(FP)
Healthy	0(FN)	4(TN)

Table 12 kNN with Euclidean Distance confusion matrix

k-Nearest Neighbour with Hamming Distance	Schizophrenia	Healthy
Schizophrenia	5(TP)	5(FP)
Healthy	0(FN)	0(TN)

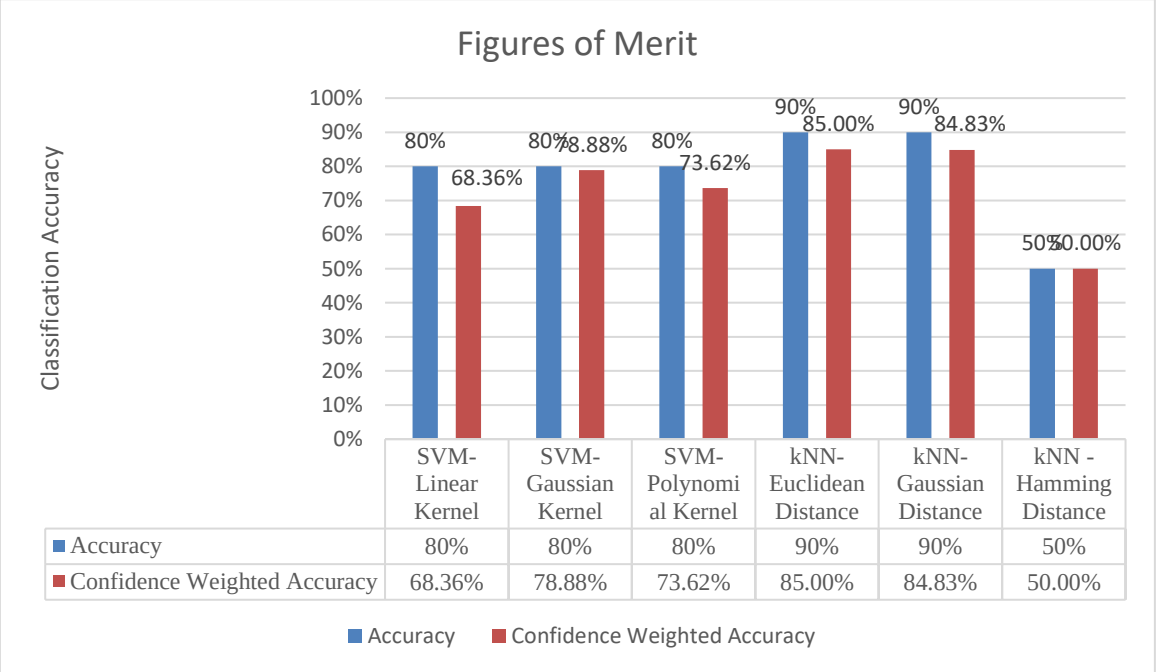
Table 13 kNN with Hamming Distance confusion matrix

k-Nearest Neighbour with Gaussian Distance	Schizophrenia	Healthy
Schizophrenia	5(TP)	1(FP)
Healthy	0(FN)	4(TN)

Table 14 kNN with Gaussian Distance confusion matrix

4.4 Classification Accuracy

Table 15 gives the figures of merit (classification accuracy). The trivial result is 50% accuracy index, which means accuracy index over 50% shows effectiveness in supporting diagnosis of schizophrenia.



	SVM-Linear Kernel	SVM-Gaussian Kernel	SVM-Polynomial Kernel	kNN-Euclidean Distance	kNN-Gaussian Distance	kNN - Hamming Distance
Accuracy	80%	80%	80%	90%	90%	50%
Confidence Weighted Accuracy	68.36%	78.88%	73.62%	85.00%	84.83%	50.00%

Table 15 Figures of merit

Based on the figures of merit, classifier performance sorting results could be expressed as follows: KNN with Euclidean Distance > KNN with Gaussian Distance > SVM with Gaussian Kernel > SVM with Polynomial Kernel > SVM with Linear Kernel > KNN with Hamming Distance.

Capítulo 5. Conclusion and Further Work

This study presented a wavelet decomposition method on fMRI to support the diagnosis of Schizophrenia based on the machine learning algorithm: SVM and KNN classifiers. The proposed multilevel decomposition on fMRI not only focused on the whole brain regions, but also exploited the frequency information for classification process. After classification, there was a problem that categorizing directly using the default SVM and kNN classifier might not yield desirable results in some cases. Therefore, SVM classifiers with various kernels and kNN with different distances were test to deal with the problem. Based on the evaluation results with the criteria of accuracy index, kNN with Standard Euclidean and Gaussian Distance demonstrated good performance on identifying whether subjects have schizophrenia disease. The next step is to train more fMRI data of subjects with Schizophrenia and Healthy Controls (after pre-processing) on the classifiers so that the results could be more general if more time can be provided.

Bibliography

[1] Power, J.D., Fair, D.A., Schlaggar, B.L. and Petersen, S.E., 2010. The development of human functional brain networks. Neuron, 67(5), pp.735-748.

[2] Friston, K.J. and Frith, C.D., 1995. Schizophrenia: a disconnection syndrome. Clin Neurosci, 3(2), pp.89-97.

[3] Lawrie, S.M., Buechel, C., Whalley, H.C., Frith, C.D., Friston, K.J. and Johnstone, E.C., 2002. Reduced frontotemporal functional connectivity in schizophrenia associated with auditory hallucinations. Biological psychiatry, 51(12), pp.1008-1011.

[4] Boksman, K., Théberge, J., Williamson, P., Drost, D.J., Malla, A., Densmore, M., Takhar, J., Pavlosky, W., Menon, R.S. and Neufeld, R.W., 2005. A 4.0-T fMRI study of brain connectivity during word fluency in first-episode schizophrenia. Schizophrenia Research, 75(2-3), pp.247-263.

[5] Honey, G.D., Pomarol-Clotet, E., Corlett, P.R., Honey, R.A., Mckenna, P.J., Bullmore, E.T. and Fletcher, P.C., 2005. Functional dysconnectivity in schizophrenia associated with attentional modulation of motor function. Brain, 128(11), pp.2597-2611.

[6] Mantini, D., Perrucci, M.G., Cugini, S., Ferretti, A., Romani, G.L. and Del Gratta, C., 2007. Complete artifact removal for EEG recorded during continuous fMRI using independent component analysis. Neuroimage, 34(2), pp.598-607.

[7] Shen, H., Wang, L., Liu, Y. and Hu, D., 2010. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. Neuroimage, 49(4), pp.3110-3121.

[8] Lee, S., Zelaya, F., Samarasinghe, Y., Amiel, S.A. and Brammer, M.J., 2011, May. Data-driven fMRI group classification using connected components and Gaussian process classifiers. In 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 717-720). IEEE.

[9] Hossein-Zadeh GA, Soltanian-Zadeh H, Ardekani BA. Multiresolution fMRI activation detection using translation invariant wavelet transform and statistical analysis based on resampling. IEEE Transactions on medical imaging. 2003 May 21;22(3):302-14.

[10] Ogawa, S., Lee, T.M., Kay, A.R. and Tank, D.W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. proceedings of the National Academy of Sciences, 87(24), pp.9868-9872.

[11] Van Den Heuvel, M.P. and Pol, H.E.H., 2010. Exploring the brain network: a review on resting-state fMRI functional connectivity. European neuropsychopharmacology, 20(8), pp.519-534.

[12] Sporns, O., 2013. Structure and function of complex brain networks. Dialogues in clinical neuroscience, 15(3), p.247.

[13] Sporns, O., 2018. Graph theory methods: applications in brain networks. Dialogues in clinical neuroscience, 20(2), p.111.

[14] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B. and Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage, 15(1), pp.273-289.

[15] Dey, A., 2016. Machine learning algorithms: a review. International Journal of Computer Science and Information Technologies, 7(3), pp.1174-1179.

[16] Repovs, G. and Barch, D.M., 2012. Working memory related brain network connectivity in individuals with schizophrenia and their siblings. Frontiers in human neuroscience, 6, p.137.

[17] Woolrich, M.W., Ripley, B.D., Brady, M. and Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage, 14(6), pp.1370-1386.

[18] Smith, S.M., 2000. BET: Brain extraction tool. FMRIB TR00SMS2b, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain), Department of Clinical Neurology, Oxford University, John Radcliffe Hospital, Headington, UK.

[19] Duarte Rosa, M.J., 2012. Development and application of model selection methods for investigating brain function (Doctoral dissertation, UCL (University College London)).

[20] Poldrack, R.A., Mumford, J.A. and Nichols, T.E., 2011. Handbook of functional MRI data analysis. Cambridge University Press.
[21] Percival, D.B. and Walden, A.T., 2000. Wavelet methods for time series analysis (Vol. 4). Cambridge university press.
[22] Mitchell, T.M., 1997. Machine learning.
[23] Mechelli, A. and Vieira, S. eds., 2019. Machine Learning: Methods and Applications to Brain Disorders. Academic Press.
[24] Townsend, J.T., 1971. Theoretical analysis of an alphabetic confusion matrix. Perception & Psychophysics, 9(1), pp.40-50.