UNIVERSITÀ DEGLI STUDI DI TORINO

SCUOLA DI SCIENZE DELLA NATURA

Corso di Laurea Magistrale in Fisica dei Sistemi Complessi

Tesi di Laurea Magistrale

Machine Learning Analysis of Resting-State Functional Connectivity Over the Lifespan

Relatore interno:
Prof.
Matteo Osella

Candidato:
Carola Maria
Caivano

Relatore esterno:
Prof.
Lauri Parkkonen

Advisor:
Ivan Zubarev

Anno Accademico 2022/2023

Abstract

This Master's thesis aims to study the change in brain connectivity among healthy adults. Brain connectivity alterations have been associated with disorders such as Alzheimer's disease, Parkinson's disease and schizophrenia. Studying brain connectivity patterns can help to find biomarkers for predicting disease progression or treatment response.
In this study, we used resting-state magnetoencephalography (MEG) recordings from 618 subjects (ages 18–88 years) with normal physical and cognitive health from the Cam-CAN repository.
We studied how functional connectivity changes over the lifespan and we used MEG features to predict brain and cognitive ageing. Before extracting features from the MEG recordings we built for each subject a connectivity matrix in each frequency band (delta, theta, alpha and beta) using as connectivity metric the Weighted Phase-Lag Index.
We decided to focus our analysis on three different sets of features: connectivity values, graph metrics and topological features. These three sets of features were used to study brain and cognitive ageing. Regarding the study of cognitive ageing, we examined three specific behavioural tasks completed by the subjects: the Cattell test, the Benton Faces test and the Motor Learning task.
For both the brain and cognitive age prediction we tested a range of machine learning algorithms focusing on support vector regression. In our study on brain age prediction, we achieved an MAE of 10 years in predicting age and we identified a correlation between age and functional connectivity features. In cognitive age prediction, incorporating MEG features did not contribute compared to using age alone, which served as the baseline prediction for cognitive ageing.

Italian Abstract

Questa tesi magistrale si occupa di studiare il cambiamento della connettività cerebrale in adulti sani. Alterazioni della connettività cerebrale sono state associate a disturbi come la malattia di Alzheimer, il morbo di Parkinson e la schizofrenia. Studiare modelli di connettività cerebrale può contribuire a trovare biomarcatori per prevedere come progredirà la malattia o quale trattamento usare come risposta.
In questo studio, sono state utilizzate registrazioni di magnetoencefalografia (MEG) in stato di riposo per 618 soggetti sani (con età tra 18 e 88 anni) prese dalla repository del Cam-CAN. E' stato studiato come la connettività funzionale cambi con l'età e come si possano utilizzare MEG features per prevedere l'invecchiamento cerebrale e cognitivo. Prima di estrarre le features dalle registrazioni MEG, abbiamo costruito per ciascun soggetto una matrice di connettività in ogni banda di frequenza (delta, theta, alfa e beta) utilizzando come metrica di connettività il Weighted Phase Lag Index.
Abbiamo deciso di concentrare la nostra analisi su tre diversi insiemi di features: i valori di connettività, metriche di grafi e caratteristiche topologiche. Questi tre insiemi di features sono stati utilizzati per studiare sia l'invecchiamento cerebrale che cognitivo. Per quanto riguarda lo studio sull'invecchiamento cognitivo, abbiamo esaminato tre specifiche attività comportamentali completate dai soggetti: il test di Cattell, il test di Benton Faces e il test sull'apprendimento motorio.
Per entrambe le predizioni, età cerebrale e cognitiva, sono stati testati una serie di algoritmi di machine-learning, focalizzandoci su support vector regression. Nella previsione dell'età cerebrale, abbiamo ottenuto una previsione dell'età con un MAE di 10 anni e abbiamo trovato una relazione tra l'età e le features che rappresentano la connettività funzionale. Nella previsione dell'età cognitiva, abbiamo scoperto che l'aggiunta di MEG features non ha aggiunto alcuna informazione rispetto all'utilizzo solo dell'età, che è stata usata come feature per ottenere una predizione da usare come baseline.

Contents

Symbols and Abbreviations

BOLD Blood-oxygenated-level-dependent

Cam-CAN Cambridge Centre for Ageing and Neuroscience

EEG Electroencephalography

fMRI Functional magnetic resonance imaging

ICA Indipendent Component Analysis

MAE Mean Absolute Error

MEG Magnetoencephalography

MSE Mean Squared Error

PCA Principal Component Analysis

PLI Phase-Lag Index

RBF Radial Basis Function

ROIs Region of Interest

SVM Support Vector Machine

SVR Support Vector Regression

TFR Time-Frequency Representation

tSSS Spatiotemporal Signal-Space Separation

WPLI Weighted Phase-Lag Index

1 Introduction

The brain can be represented as a complex network [37], with its various regions acting as nodes and the connections between them as links. Understanding the interactions between these regions typically involves examining regional activity. We can use functional connectivity, which aims to identify statistical dependencies between activities of two brain areas that are spatially separated [16]. Alterations in functional connectivity have been linked to disorders such as Alzheimer's, Parkinson's, and schizophrenia [8], making the identification of biomarkers crucial for disease prevention.
Since every person is different, it is important to define normative models that define typical patterns of brain connectivity under normal conditions to provide a baseline for detecting abnormalities [20].
While functional connectivity has been extensively studied in resting-state (a condition in which the brain exhibits spontaneous neural activity in the absence of any task) fMRI data [16], a research on it with MEG data remains limited. MEG measures the magnetic field generated by electric currents in the brain and it has a high temporal resolution. For this reason, it could be better to calculate functional connectivity because it can measure processes that occur on a fast timescale as changes in functional networks related to cognition.
As a first step to building a normative model, we studied whether there was a connection between brain ageing and changes in functional connectivity in healthy subjects. We analyzed the relation between age and functional connectivity using MEG measurements of 618 healthy subjects from the cam-CAN repository. We analyzed the functional connectivity of each subject in different frequency bands: alpha, beta, delta and theta. We developed a machine-learning pipeline to predict age based on connectivity matrices derived from these measurements (for more information see [32]). We investigated various sets of features to determine which were most effective in accurately predicting age.
In the second part of our project we replicated the procedures to examine cognitive aging.
The thesis will cover the following topics. Chapter 2 will give an overview of the background necessary to understand the analysis. It will introduce functional connectivity, MEG data and the construction of connectivity matrices from MEG data. Then it will show some principles of graph theory, topological data analysis and a description of how a machine learning model works. Chapter 3 will describe the dataset used for the analysis, how the features have been extracted from our data and the machine learning pipeline.

Finally, the last two chapters, Chapters 4 and 5, show and discuss the results.

2 Background

As the reader probably understands, this thesis studies resting-state functional connectivity with machine-learning algorithms. To understand this topic, we must first explain the basic principles of functional connectivity and the connectivity metric selected (Section 2.1), explain some basis of graph theory that can be used to inspect neurological data and extract meaningful information (Section 2.2 and 2.3) and give an overview of the theory behind the machine-learning algorithms used (Section 2.5).

2.1 Connectivity in the brain

The brain is a complex system of interconnected brain areas. One of the most important problems to be studied is understanding how these different brain areas relate. It is possible to distinguish different aspects of connectivity in the brain, the most significant distinction is between *structural connectivity* and *functional connectivity* [13].
Structural connectivity pertains to the anatomical pathways linking different brain regions, which are insufficient to characterize the interactions involved in cognitive functions [17]. Instead **functional connectivity** represents the interactions between the regional activity so it can represent the non-anatomical interactions.

2.1.1 Functional Connectivity

Functional connectivity aims to identify statistical dependencies between activities of two brain areas that are spatially separated [16].
The study of functional connectivity started using fMRI [3], which measures the correlation between BOLD time courses from the left and right motor cortex in the *resting state* (a condition in which the brain shows neural activity in the absence of any task). The study of functional connectivity based on fMRI has some limitations. fMRI has an excellent spatial resolution, but it has a limited temporal resolution because it is based on BOLD, which is an indirect measure of neural activity. This is a limitation to studying processes that occur on a fast timescale, for example, the changes in functional networks related to cognition. For these reasons, electrophysiological techniques, such as EEG or MEG with higher temporal resolution are better for calculating functional connectivity.

2.1.2 Functional Connectivity from MEG Data

MEG measures the magnetic field generated by electric currents in the brain. These electric currents are created by the synchronous activation of neurons [36].
MEG data are dominated by neuronal oscillations in the 1-200 Hz frequency range. It is possible to divide this range into five frequency bands: delta band (1-4 Hz), theta band (4-8 Hz), alpha band (8-13 Hz), beta band (13-30 Hz) and gamma band (30-200 Hz) [5].

The magnetic field is measured by sensors, that cover the whole head, these recordings can be corrupted by effects of field spread [33]. This factor has implications for interpreting connectivity measures calculated between pairs of sensors. When the activities of sources are not temporally correlated, they may exhibit high correlation values due to field spread. Shifting the connectivity analysis from the sensor level to the source level may offer a solution to this issue.
MEG source connectivity analysis begins by estimating the spatial distribution and intensity of neural activity through an inverse method. This method enables the identification of source points. Subsequently, connectivity analysis is conducted within a lower-dimensional space defined by **Regions of Interest** (ROIs).

2.1.3 Cortical parcellation

The source representation is in a high-dimensional space, there can be thousands of source points. For this reason, the functional connectivity analysis is usually performed in a lower dimensional space individued by the ROIs, which are clusters of the source points. This process is called parcellation. The selection of ROIs is a critical step because an incorrect specification can lead to erroneous results. We can choose the ROIs based on functional parcellation or anatomical parcellation. In a functional parcellation, the ROIs can be selected by looking at their involvement in a specific task or estimating coherence between a peripheral physiological signal and brain activity [33]. In an anatomical parcellation, the brain is divided into regions based on its anatomical structure. An example of anatomical parcellation is shown in Desikan's work, where 40 MRI scans were utilized to partition the human cerebral cortex into 34 Regions of Interest (ROIs) [10].

After dividing the cortex into several ROIs we want to represent the

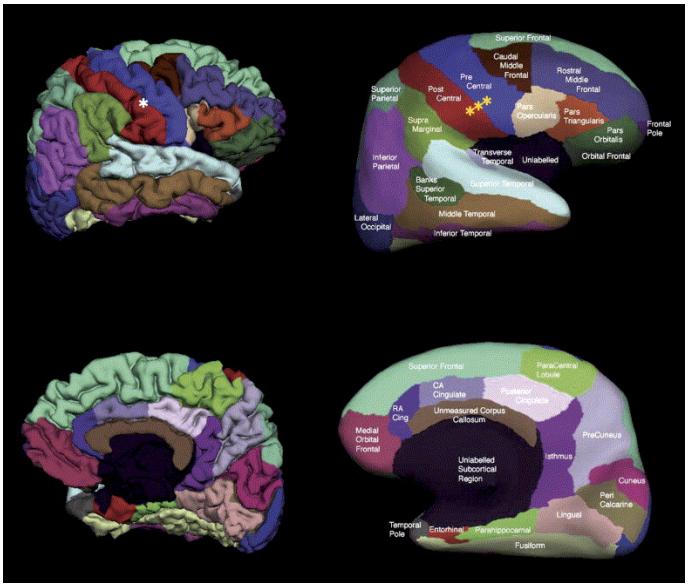

Figure 1: Cortical parcellation using Desikan-Killiany atlas with 34 ROIs [10]. The first row shows the lateral view the second the medial view.

activity of each area with one single time series [4].
Based on the parcellation each ROI can contain different source points. Then before computing connectivity between pairs of ROIs, we take a time series to represent the overall activity of the area. We can take a time series that is associated with a specific source point, for example, the time series associated with the source that is located in the centre of mass of the area. Or we can assemble different time series corresponding to different source positions, for example averaging all the source points in the area.
Since the connectivity metrics are computed in predefined frequency bands it is necessary to transform the original time series into a time-frequency representation and then is possible to measure the connection between the ROIs.

2.1.4 Connectivity metrics

There exists a variety of methods commonly utilized for estimating connectivity in resting-state analysis. These typically involve evaluating spectral coherence and assessing phase delays, as well as examining the coupling of amplitude patterns in oscillatory signals. [9]. These methods can be directed or non-directed based, we will focus on the last ones.

A good connectivity measure should satisfy repeatability and it should ensure consistency in estimating both the individual subject's outcome and the overall group-level structure. Connectivity methods are used to compute the connectivity scores between pairs of ROIs. Give n ROIs is possible to obtain a connectivity matrix C of dimension $(n \times n)$, where the c_{ij} elements are the connectivity scores between the ROIs i and j.
Below we will present two-phase estimation methods: *Phase Lag Index* and *Weighted Phase Lag Index* , that measure the phase synchrony between two signals in a specific frequency band.
The PLI [38] quantifies the asymmetry in the phase differences between two signals. This asymmetry is characterized by a phase difference lag between two-time series. A measure of the disparity in this distribution between two signals x and y can be formulated as follows:

$$PLI = | < \operatorname{sign} \sin \Delta\Phi_{xy} > | \tag{1}$$

where $\Delta\phi_{xy}$ indicates the phase difference. PLI values are contained in this range $0 \leq PLI \leq 1$. If PLI is equal to 0 that means that there is no phase locking, while a PLI of 1 indicates perfect phase locking. We can rewrite our phase difference described by PLI as the magnitude of the mean of the sign of the imaginary part of the cross-spectral density:

$$PLI_{xy} = |\mathbb{E}_t\{\operatorname{sgn}(\operatorname{Im}\{S_{xy}(f,t)\})\}| \tag{2}$$

To improve the PLI is possible to use the Weighted Phase-Lag Index [42] where we associate a weight to the phase:

$$WPLI_{xy} = \frac{|\mathbb{E}_t\{|\operatorname{Im}\{S_{xy}\}\operatorname{sgn}(\operatorname{Im}\{S_{xy}\})\}|}{\mathbb{E}_t\{|\operatorname{Im}\{S_{xy}\}|\}} = \frac{|\mathbb{E}_t\{\operatorname{Im}\{S_{xy}\}\}|}{\mathbb{E}_t\{|\operatorname{Im}\{S_{xy}\}|\}} \tag{3}$$

$S_x(f,t)$ and $S_y(f,t)$ indicate the power spectral densities of signals x and y, respectively, and $S_{xy}(f,t)$ represents the cross-spectral density between the two signals.
The PLI employs a sign function that introduces a step-like non-linearity around the real axis, potentially amplifying noise in the signal. To address this issue, the WPLI utilizes a weighting procedure, aiming to mitigate the impact of noise and introduce smoother transitions when the cross- spectra rotate across the real axis [42].

2.2 Graph Theory

In this section, we will consider the theory behind complex network analysis. We will explore the application of complex network analysis in examining brain connectivity, and we will show what are the most used measures to study local and global connectivity.
In mathematics, networks are named as *graphs*, and *graph theory* is the area that deals with the study of graphs.

Definition A graph is a mathematical object consisting of two elements: edges and vertices. Vertices symbolize objects or entities, while edges indicate their connection. We can indicate a graph as $G = (I, E)$ where I contains as elements all the *vertices* (or *nodes*) and E all the *edges* (or *links*) [11].

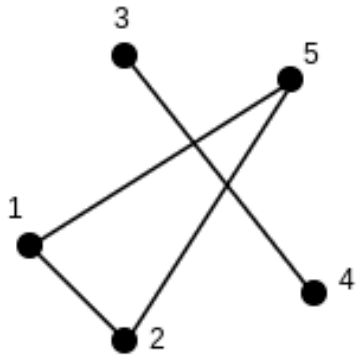

Figure 2: Graph with 5 vertices and 4 links

It is possible to associate to each of the edges a weight, that can be indicated as w_{ij}, where i and j indicate the two nodes connected by the link.

The edges can be also characterized by a directionality. In this case, we can distinguish between **directed graphs**, which are graphs where edges have a specific direction and **undirected graphs** where the edges do not have a specific direction.

In brain connectivity, the choice of nodes and links depends on the anatomical parcellation schemes and measures of connectivity used. Nodes should represent the ROIs and the links the type of connectivity (anatomical, functional or effective) [31].

In functional networks, weak links could indicate spurious connections. These connections prevent us from seeing the strong and significant connections, to avoid this we can apply a weight threshold that maximizes the separation between signal and noise.

2.2.1 Thresholding

Functional connectivity networks based on time series correlations are fully connected since they contain nonzero values in every off-diagonal element of the matrix. It's known that the brain does not constitute a completely interconnected network, so the fully connected connectivity matrices built by any method for brain reconstruction include some substantial percentage of spurious connections. To solve this problem it is possible to apply a threshold, τ to the connectivity matrix to determine which connections should be maintained [15]. Applying a threshold helps to maximize the separation between signal and noise.
The simplest method for thresholding a matrix is through *global thresholding*. In this method elements below the threshold are put equal to zero and the other elements are put equal to one if we want to have as a result a binary matrix, or we can maintain the original connectivity weights if we want to analyse weighted networks.

$$A_{ij} = \begin{cases} C_{ij} & \text{if } C_{ij} > \tau \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where A_{ij} are the elements of the thresholded matrix and C_{ij} the elements of connectivity matrix.
To establish the threshold two methods are commonly applied: *weight-based* or *density-based thresholding*.
In weight-based thresholding, the threshold value τ is determined according to the weights within the connectivity matrix. This approach may lead to varying connection densities in networks. Since most graph theoretic measures are sensitive to variations in edge numbers, density-based thresholding can be more representative.
In density-based thresholding, τ can vary from one matrix to another one, but we have a fixed connection density κ.

2.2.2 Graph metrics

Graph measures can characterize a global or local aspect of brain connectivity. In this section, we will present some of the most commonly used graph

measures.

Degree The degree of a node represents the number of links connected to it.
In a binary and undirected network, the degree can be defined as [31]:

$$k_i = \sum_{j \in N} a_{ij} \tag{5}$$

where N represents the number of nodes in the graph. a_{ij} tell us if it is present a connection between the node i and j. If $a_{ij} = 1$ i and j are connected, if $a_{ij} = 0$ they are not.
In the case of a directed network it is possible to distinguish between in-degree $k_i^{in} = \sum_{j \in N} a_{ji}$, numbers of incoming links, and out-degree $k_i^{out} = \sum_{j \in N} a_{ji}$, numbers of outgoing links. The degree gives us an idea of the importance of a node.
The degree distribution, which is an important indicator of network development, is determined but the degrees of all nodes and can be defined as $P(k) = \frac{N_k}{N}$, where N_k is the number of nodes in the network that have degree k.

Clustering Coefficient It describes the tendency of nodes in a graph to cluster together. In a binary and undirected network can be defined as [44]:

$$C = \frac{1}{n} \sum_{i \in N} C_i = \frac{1}{n} \sum_{i \in N} \frac{2t_i}{k_i(k_i - 1)} \tag{6}$$

with $t_i = \frac{1}{2} \sum_{j,h \in N} a_{ij} a_{ih} a_{jh}$ that are the numbers of triangles around node i.

Transitivity It measures the overall tendency of nodes in a network to form triangles or triplets and it can be defined in a binary undirected network as [26]:

$$T = \frac{\sum_{i \in N} 2t_i}{\sum_{i \in N} k_i(k_i - 1)} \tag{7}$$

Characteristic path length It is the average of all the shortest path lengths computed between the nodes and it can be defined as [44]:

$$L = \frac{1}{n} \sum_{i \in N} L_i = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}}{n - 1} \tag{8}$$

where L represents the number of links in the network and $d_{ij} = \sum_{a_{uv} \in gi \leftrightarrow j} a_{uv}$ represents the shortest path length.

Global Efficiency It is the average of all the inverses of shortest path lengths and gives us information about how efficiently it is exchanged information in a network. It is defined as [23]:

$$E = \frac{1}{n} \sum_{i \in N} E_i = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}^{-1}}{n-1} \tag{9}$$

We've observed that interpreting functional networks can be challenging because of the combination of complete connectedness and edge weights. This underscores the importance of employing a thresholding method before applying network measures. Disregarding weak links may involve a trade-off between information completeness and clarity [29]. To solve this problem we can consider the combined structure of connections and weights captured by topological data analysis.

2.3 Topological Data Analysis

In this section, we delve into the theoretical fundamentals of Topological Data Analysis (TDA), with a specific focus on persistent homology. Our exploration will extend to the practical utilization of persistent homology in examining neurological data.

Topology can be used to identify and understand the shape of our data. Persistent Homology characterized this shape using *n-dimensional* holes.
The first step in the computation of persistent homology begins with the representation of the data in simplicial complexes. We have seen that is possible to transform our connectivity matrices in graphs (section 2.2). The simplicial complexes are a higher-dimensional generalization of a graph.

2.3.1 Simplicial Complexes

In this paragraph, we will show how to give a shape to our data with *simplicial complexes*. We start by defining simplicial complexes and the mappings that operate between them. According to [27] we can define a simplicial complex in the following way:

Definition (*simplicial complex*) A simplicial complex, denoted as K, is a collection of non-empty subsets of a set K_0, such that for each vertex v, where $v \in K_0$, it applies that $\{v\} \in K$, and for each simplex $\tau \subset \sigma$, with $\sigma \in K$, it applies that $\tau \in K$.
We can specify the dimension of our simplicial complex such that for a cardinality of $p+1$, the dimension is p.
A *map of simplicial complexes*, $f : K \to L$ is a map on the corresponding vertices and verifies the property that $f(\sigma) \in L$ for all $\sigma \in K$.
It is possible to give a geometric definition of a simplicial complex as a collection of points, line segments, triangles, and their higher-dimensional analogues, all interconnected according to specific rules. We show below some examples of geometric simplexes:

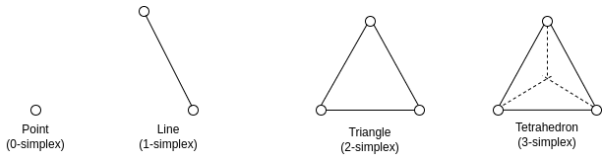

Figure 3: Examples of simplicial complexes

2.3.2 Homology for Simplicial Complexes

In this paragraph, we want to define the homology of our simplicial complexes. For each simplicial complex K it is possible to define the vector space over $\mathbb{F}_2$ as $C_p(K)$, where the basis elements are the p-simplices of K. We can define the linear maps on the basis elements where $p \in \{1, 2, ...\}$ with the rule showed by [27]:

$$\begin{aligned} d_p : C_p(K) &\to C_{p-1}(K), \\ \sigma &\to \sum_{\tau \subset \sigma, \tau \in K_{p-1}} \tau \end{aligned} \tag{10}$$

We can define the zero map d_0 for $p = 0$. d_p maps each p-simplex in its boundary. Since the boundary of a boundary is always empty we can define the following property

$$d_p \circ d_{p+1} = 0 \text{ for all } p \in \{0, 1, 2, ...\} \tag{11}$$

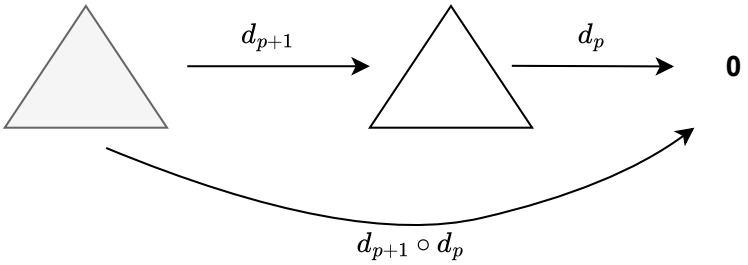

Figure 4: Illustration of $d_p \cdot d_{p+1} = 0$

From this property follow that $Imd_{p+1} \subseteq \ker d_p$. Given the previous property, it is possible to define homology groups in the following way [14, 27]

Definition (*pth homology*) For any $p \in \{0, 1, 2, ...\}$ the *pth homology* of a simplicial complex K is:

$$H_p(K) := Z_p / B_p \tag{12}$$

where $Z_p := kerd_p$ and its elements are called *p-cycles* and $B_p := imd_{p+1}$ and its elements are called *p-boundaries*. The dimension of the *pth homology*

is indicated with the *pth Betti number* of K defined as:

$$\beta_p(K) := \dim H_p(K) = \dim Z_p - \dim B_p \tag{13}$$

To give an intuitive interpretation the $p - cycles$ can represent holes of dimensionality p and the Betti number can quantify the p-dimensional holes.

2.3.3 Filtrations

To explore our data on different levels, it is important to apply a filtration process. We can define a *filtration* as a nested subcomplexes of K, $K_0 \subset K_1 \subset K_2 \subset ... \subset K_l = K$.

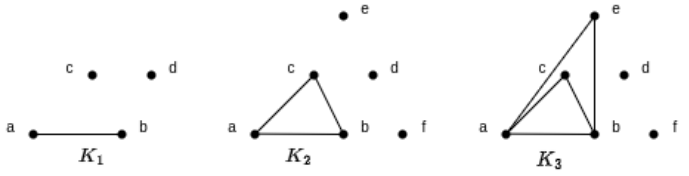

Figure 5: Example of filtered simplicial complex

Given the map between two simplicial complexes $f^{i,j} : K_i \to K_j$, it is possible to verify that it is defined as the map between the homology spaces $f_p^{i,j} : H_p(K_i) \to H_p(K_j)$. In addition if we have $f : K_i \to K_j$ and $g : K_j \to K_z$ the map $(g \circ f)_p : H_p(K_i) \to H_p(K_z)$ is equal to the compositions of the maps $(g \circ f)_p = g_p \circ f_p$ [27].
In our case, if we consider the subsets of a simplicial complex we have that:

$$f_p^{j,k} \circ f_p^{i,j} = f_p^{i,k} \text{ for all } i \leq j \leq k \tag{14}$$

Definition (*pth persistent homology*)For a filtration of a simplicial complex K, $K_0 \subset K_1 \subset K_2 \subset ... \subset K_l = K$, we can define the *pth persistent homology* of K as:

$$H_p^{i,j} := \text{im} f_p^{i,j} = Z_p(K_i)/(B_p(K_j) \cap Z_p(K_i)) \tag{15}$$

where $f_p^{i,j} : H_p(K_i) \to H_p(K_j)$ [14].

To create a filtration it is possible to use *Vietoris-Rips*. We can imagine plotting our data in an *N-dimensional* space (one dimension for each variable). Then we can define a distance $\epsilon > 0$ and build N-dimensional balls

of radius ϵ around each data point. If the balls around two data points don't overlap we have two 0-simplical complexes. If the balls overlap, we can connect the two points and create a 1-simplicial complex. If we have three balls that overlap we can create a 2-simplicial complex and so on. The filtration can be created by varying the parameter ϵ. As ϵ increases, more points become connected, resulting in a sequence of nested complexes.
We can define the Vietoris-Rips Complex formally like this [24]:

Definition (*Vietoris-Rips Filtration*) Consider a metric space X and a parameter $\epsilon > 0$. The VietorisRips complex $VR_\epsilon(X)$ of X is the simplicial complex whose vertices are the points of X and whose simplices are those finite subsets of X with diameter strictly less then ϵ. If $\epsilon \leq r$ than $VR_\epsilon(X)$ is contained in $VR_r(X)$. Hence, the family $VR_*(X)$ is a filtration, called the open Vietoris-Rips filtration of X.

2.3.4 Persistence Diagrams

In persistent homology, it is important to monitor the lifetime of homology classes. The lifetime refers to when the homology class appears, birth, and when disappears, death. To track the evolution of topological features it is possible to use persistence diagrams.

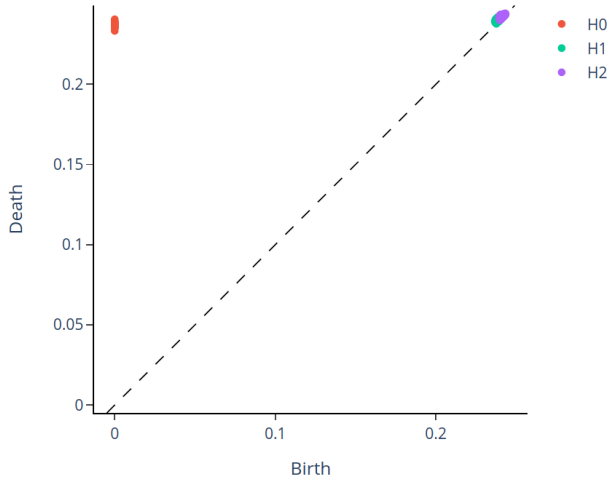

Figure 6: Example of persistence diagram

Definition (*persistence diagram*)A persistence diagramis a combination of two multisets. The first is a finite multiset in $\bar{\mathbb{R}^2}$. The second is the multiset of points on the diagonal $\Delta\{(x, y) \in \mathbb{R}^2 | x = y\}$, where each point on the diagonal has infinite multiplicity.

2.3.5 Persistent Entropy

One of the topological features that we can use is *persistent entropy* [2]. Given a filtered simplicial complex and a persistence diagram $dgm(F) = \{a_i = (x_i, y_i)/1 \leq i \leq l\}$, where x_i represent the births and y_i the deaths and l is the number of nested subcomplexes, the persistent entropy can be defined as:

$$E(F) = -\sum_{i=1}^{l} p_i \log(p_i) \tag{16}$$

with $p_i = \frac{m_i}{S_M}$, $m_i = y_i - x_i$ and $S_M = \sum_{i=1}^{l} m_i$.

2.4 Dimensionality Reduction

Neuroimaging data frequently show many more features than subject numbers. This phenomenon results in sparse data, presenting a challenge for algorithms to generalize effectively. To address this issue it is possible to use a dimensionality reduction technique to decrease the number of features in a dataset while preserving as much information as possible. We want to transform our data into a lower-dimensional representation preserving the most important properties of the original data. A lower-dimensional representation of our data means less training time, and less computational resources, avoids the problem of overfitting and can be used to remove noise in the data.

The following section will present a linear dimensionality reduction technique: *Principal Component Analysis* [35].

2.4.1 Principal Component Analysis

In the context of Machine Learning, PCA is an unsupervised machine learning algorithm that is used for dimensionality reduction.
We can consider having *p-dimensional* vectors and we want to summarize them by projecting them into a *q-dimensional* subspace. The directions in the q-subspace represent the **principal components**.
To find the principal components, we can find the projections that **maximize the variance**, so that the first principal component indicates the direction in space where the projections display the highest variance, while each subsequent component represents an orthogonal direction maximizing variance compared to the preceding ones. Maximizing the variance corresponds to minimising the projection residuals, so we can start explaining how to minimize the projection residuals.

2.4.2 Minimizing Projection Residuals

We consider a one-dimensional projection, we project our vector on a line that passes through the origin. It is possible to represent our line as a unit vector $\vec{w}$ and our data vectors as $\vec{x_i}$. If we want to compute the projection of our data vectors using coordinates in p-dimensional space we have $(\vec{x_i} \cdot \vec{w})\vec{w}$. When we substitute the original data vectors with the projected vectors an error is introduced, because the projected vectors do not perfectly coincide with the original vectors. We indicate this error as the **residual** of our

projection and we can write it as:

$$||\vec{x_i} - (\vec{w} \cdot \vec{x_i})\vec{w}||^2 = \vec{x_i} \cdot \vec{x_i} - (\vec{w} \cdot \vec{x_i})^2 \tag{17}$$

Considering the residuals across all the vectors we have

$$MSE(\vec{w}) = \frac{1}{n}\big(\sum_{i=1}^{n} ||\vec{x_i}||^2 - \sum_{i=1}^{n} (\vec{w} \cdot \vec{x_i})^2\big) \tag{18}$$

To minimize MSE we need to maximize only the second sum since the first one is not dependent on $\vec{\omega}$. It is possible to rewrite the second sum as:

$$\frac{1}{n}\sum_{i=1}^{n} (\vec{w} \cdot \vec{x_i})^2 = \big(\frac{1}{n}\sum_{i=1}^{n} \vec{x_i} \cdot \vec{w})^2 + Var[\vec{w} \cdot \vec{x_i}] \tag{19}$$

The projections' mean is zero because the vectors $\vec{x_i}$ have a mean of zero. In this condition, maximising the variance of projections or minimising the residual sum of squares is equivalent.

2.4.3 Maximizing Variance

We represent our data with a matrix $\mathbf{x}$ $n \times p$ and the projections with a matrix $\mathbf{xw}$ $n \times 1$. The variance is:

$$\sigma^2_{\vec{w}} = \frac{1}{n}\sum_{i} (\vec{x_i} \cdot \vec{w})^2 \tag{20}$$

$$= w^T \frac{x^T x}{n} w \tag{21}$$

$$= w^T v w \tag{22}$$

Our goal is to choose a vector $\vec{\omega}$ that maximize $\sigma^2_{\vec{w}}$. We need to do this using a constraint since we have $\vec{\omega} \cdot \vec{\omega} = 1$. To do this we can introduce the Lagrange multiplier λ, and we add λ times the constraint equation so that we can have an unconstrained optimization:

$$L(w, \lambda) = \sigma^2_w - \lambda(w^T w - 1) \tag{23}$$

and if we compute $\frac{\partial L}{\partial \lambda} = \frac{\partial L}{\partial \omega} = 0$ we have:

$$w^T w = 1 \tag{24}$$

$$vw = \lambda w \tag{25}$$

The vector w is an *eigenvector* of the covariance matrix $\mathbf{v}$, where $\mathbf{v}$ is a $p \times p$ matrix, with p different eigenvectors. The vector that maximizes the variance is the one corresponding to the largest eigenvalue, denoted as λ. The eigenvectors of $\mathbf{v}$ are the **principal components** of the data. Since v is symmetric the eigenvectors are orthogonal to each other. The initial principal component is identified by having the highest value of λ, and along this component, the data exhibits the greatest variance.

2.4.4 Proportion of Variance Explained

It is necessary to choose the right number of the principal components so that is preserved as much variability as possible. To do this is possible to use the proportion of **variance explained**, which tells us how much each principal component explains about the total variance of the original dataset. The variance explained by a principal component is the ratio between the variance of that principal component and the total variance. The total variance is the sum of variances of all individual principal components, while the variance of the *ith* principal component is equal to the *ith* eigenvalue.

2.5 Machine Learning

In this section, we will show some fundamentals for machine learning methods in regression tasks. We will explain later in more detail the theory behind the model used with our neurological data.

Machine learning enables computational systems to learn and adapt using experiences based on data. Machine learning algorithms, the *learners*, can find patterns in data learn from them and make their predictions. The data, which can be a collection of data points are characterized by the *features*, that quantify their properties.
The two main tasks that machine learning enables to solve are:

- **Regression** In this task a continuous value is predicted based on independent variables (features).
- **Classification** In this task the data points are assigned to some predefined classes

2.5.1 Learning

In each task, the learner is provided with a dataset $(x_1, y_1), ..., (x_n, y_n)$, where the x_i belongs to the $\mathcal{X}$ space, that represents the input space and the y_i to $\mathcal{Y}$, the output space [1]. We can indicate as $f : \mathcal{X} \to \mathcal{Y}$ the target function, that indicates us the ideal formula for our prediction so that $y_n = f(x_n)$ for $n = 1, ..., N$. The learning algorithm try to find the formula $g : \mathcal{X} \to \mathcal{Y}$, using the data set, so that g approximate f. The algorithm g is chosen from a hypothesis set H so that g best matches f looking at the training examples.

2.5.2 Error and Generalization

The function g doesn't perfectly match with f. It is possible to use an error measure to estimate the difference between the two functions. The choice of the error measure is fundamental because can lead to different outcomes. We can define the *generalization error* as the difference between E_{in} and E_{out}. E_{in}, the in-sample error, measures the training performance. E_{out}, the out-of-sample error, measures how well is the generalization to unseen data.
It presents a trade-off between approximation and generalization. The more complex the hypothesis set H is, the better chance we have of approximating f. The less complex of a hypothesis set H is, the better the chance we

have of generalizing out of the sample.
This trade-off can be studied using bias-variance analysis. The generalization error can be rewritten as:

$$E(h) = var(x) + bias(x)^2 + \epsilon^2 \tag{26}$$

The *bias* indicates the deviation of the expected output from the real value. The *variance* indicates how the model's predictions for a given data point vary as different training data subsets are used. The *noise* (ϵ) sets the lower limit for the generalization error.

2.5.3 Training Set and Test Set

It is necessary to split our dataset into different subsets to guarantee that our machine-learning model generalizes the unseen data. The three subsets in which we divided our dataset are training, validation and test set.
Training set is used to train the machine learning model, in this phase the model learns patterns and relationships between the data. *Validation set* is used to find the best hyperparameters and evaluate different versions of the model during the training process. *Testing set* is used to evaluate the final performance of the model trained with the best hyperparameters.

2.5.4 Nested Cross-Validation

When we want to solve a problem using a machine learning model, we need to choose the best algorithm for the specific dataset and the best hyperparameters for the model that we have selected. To do this we can use *nested cross-validation* [43], which is also called *double cross-validation*, because it is formed by an outer cross-validation and an inner cross-validation (figure 7). The most known cross-validation is the *k-fold* cross-validation where the dataset is divided in k-folds and at each iteration one fold is used as a validation set and the other $k - 1$ folds as the training set. In the nested cross-validation, we have an outer cross-validation that we use to select the best model evaluating the performance. Then we have in each outer cross-validation an inner validation used to tune the hyperparameters of the model.

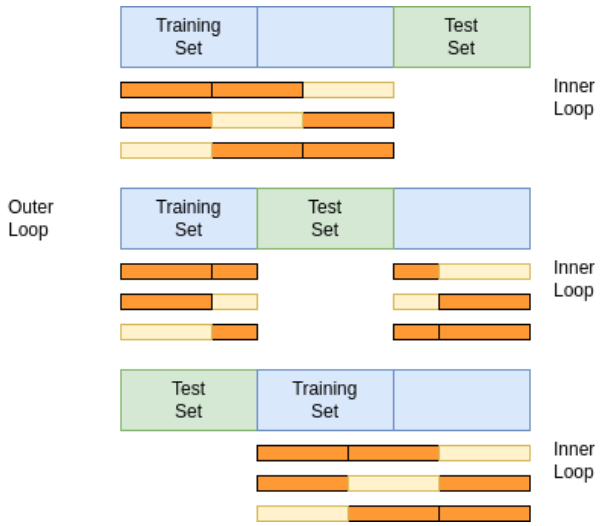

Figure 7: Illustration of Nested Cross-Validation

2.5.5 Metrics

To evaluate the performance of machine-learning models, different metrics can be used. We will show below the most used for regression tasks.
The *mean absolute error* (MAE) can be calculated as:

$$MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} \tag{27}$$

where $\hat{y}_i$ are the predicted values, y_i the real ones and n the number of samples.
The *mean squared error* (MSE) is calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{28}$$

2.5.6 SVM

We will show in more detail in this section, the theory behind the machine learning model used. Support vector regression (SVR) is a supervised machine learning model used for regression problems. Compared to other regression methods, the advantages of this model are that it can be used to study nonlinear problems, SVR can transform our problem into a linear one projecting the original data in a kernel space. The other advantage of

SVR is that can handle high-dimensional data, so it is used a lot in the neurological field.
SVR is an extension of the SVM classifier so we will briefly summarize the key principles of the SVM classifier and then we will show in more detail how SVR works [25].

SVM classifier
In SVM classifier, an **hyperplane** is used to separate and classify the data points of the training set, that are projected in a N-dimensional space, where the dimension is given by the number of features. To choose the best hyperplane, that separates our data points we need to find the one that maximizes the **margin**. The margin is defined by the distance from the decision boundary to the **support vectors** (figure 8). The support vectors are the closest data points from either class.

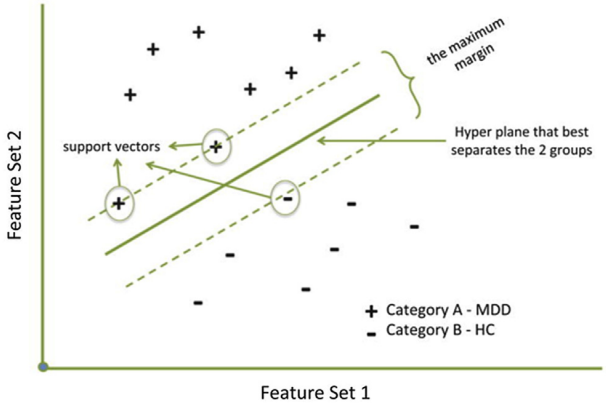

Figure 8: Illustration of the hyperplane, margin and support vectors a 2D space [25]

Support Vector Regression
In SVR the hyperplane is not used to separate the data in different classes, but to set the generalizations limits. It is defined as a parameter ϵ, such that the predicted values differ at most ϵ from the observed values. The hyperplane plus ϵ defines a tube where we want our predicted values to stay, the goal is to minimize this tube as much as possible (figure 9).

Linear SVR
The goal of SVR is to estimate a linear function $f(x)$, so that the predicted values deviate at most of ϵ from the observed values. We can write our linear

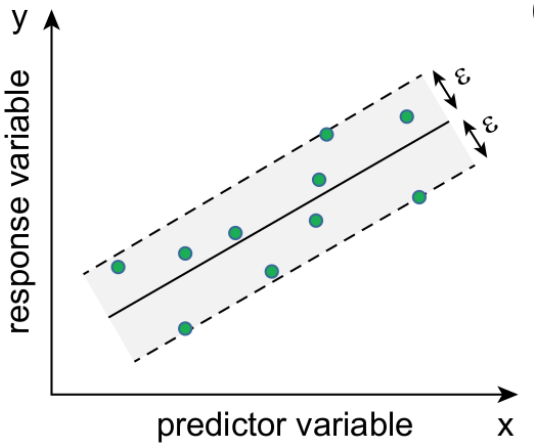

Figure 9: Illustration of the hyperplane and ϵ in a 2D space [25]

function f as follows:

$$y = f(x) = \omega,\ x + b = \omega^T x + b \tag{29}$$

where ω is the weight vector, x_i are the input data and y_i the target output. What we want to obtain is a tube (see figure 9), that is as flat as possible. The margin of the tube, and its flatness, is determined by the norm of ω, so if we want to have our tube as flat as possible we need to minimise the norm of ω. We can express this condition as follows:

$$\begin{aligned} &\min_{\omega} \frac{1}{2}||\omega||^2 \\ &\text{subject to } \begin{cases} y_i - \omega^T x_i - b \leq \epsilon \\ \omega^T x_i + b - y_i \leq \epsilon \end{cases} \end{aligned} \tag{30}$$

We can say that our SVR behaves as a linear regression but with a distinct loss function that penalizes predictions deviating by more than ϵ from the actual output. We can write our loss function as:

$$L(y, f(x)) = \begin{cases} 0 \text{ if } |y - f(x)| \geq \epsilon \\ |y - f(x)| - \epsilon \text{ otherwise} \end{cases} \tag{31}$$

What we described previously is possible only if f accurately approximates the data points with a ϵ precision. In real problems we have outliers and this

doesn't happen so we need to introduce slack variables ξ and ξ^*. These slack variables represent deviations from the ϵ-insensitive tube and determine the tolerance for prediction errors. We can rewrite our optimization problem as:

$$\min_{\omega} \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$
$$\text{subject to } \begin{cases} y_i - \omega^T x_i - b \leq \epsilon + \xi_i \\ \omega^T x_i + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{32}$$

where C is a regularization parameter.

Kernel SVR
Linear SVR is used when we have linear data, in the case of nonlinear data it is possible to use a kernel to map the data from the input space to the kernel space, where it is possible to separate them with a linear hyperplane and apply the linear optimization showed before. To do this we need a function that maps the features from the input space to the kernel space.

$$\phi : R^d \to F \tag{33}$$

and the linear function will become:

$$y = f(x) =< \omega, \phi(x) > +b = \omega^T \phi(x) + b \tag{34}$$

The optimization problem becomes

$$\min_{\omega} \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$
$$\text{subject to } \begin{cases} y_i - \omega^T \phi(x_i) - b \leq \epsilon + \xi_i \\ \omega^T \phi(x_i) + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{35}$$

we need to solve the optimization problem in the kernel space.
There are different kernel functions that we can use: linear kernels, polynomial kernels, radial basis function (RBF) kernels.

Radial Basis Function Kernel

One of the most used kernel functions is the Radial Basis Function Kernel (RBF). RBF measures the Euclidean distance between all the pair points of the training set:

$$K(\mathbf{x_i}, \mathbf{x_j}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \tag{36}$$

where x_i and x_j are data points of the dataset. All these values of distances are used to create a Kernel matrix indicating as K where all the elements represent the Euclidean distances between the data points. If K is invertible we have that $\omega = K^{-1} \cdot y$.

3 Methods

This chapter will begin by introducing the dataset used in our study, which includes resting-state MEG data. Subsequently, we will show the data preprocessing steps and the extraction of the three sets of features selected: connectivity scores, graph metrics, and topological features. Finally, we will describe the machine-learning pipeline utilized for both brain and cognitive age estimations.

3.1 Dataset

The average age is increasing, and the number of elderly people in the world is growing more than the young people [12]. For this reason, it has become important to understand the factors that influence healthy ageing. To answer this problem the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) [34] conducted measurements of neural activity in healthy subjects across various age groups using fMRI and MEG techniques, accompanied by an MRI scan for each participant.
In the first stage of the Cam-CAN project, the original dataset consisted of 700 subjects with ages between 18 and 87 years [41], in order to have ages that covered the entire lifespan. 50 females and 50 males were considered for each age decade (18-27, 28-37, 38-47, 48-57, 58-67, 68-77, 78-87), but because of some recruitment problems, only 56 subjects were considered for the youngest decade (18-27), bringing to a total number of 656 participants. The presence of fewer subjects in the younger decade could result in an unrepresentative dataset for the selected age range.
In the next stage of the project, 280 subjects were selected (taking 20 males and 20 females for each age decade) and each of them completed 14 behavioural tasks [12]. MRI, fMRI and MEG measurements were conducted for all subjects.
In our study, we used only the MEG measures in the resting state of 618 subjects (figure 10). Some participants were excluded due to the poor quality of the MEG data or to lack of some information regarding the measurements. During MEG measurements in resting state, subjects remained seated with their eyes closed for 8 minutes and 40 seconds. These measures give us information about the functional organization of the brain in the absence of any specific task or stimuli.
For the study of cognitive ageing, only 3 of the 14 behavioural tasks were analyzed: the Cattell test, the Benton test and Motor Learning. In the Cattell test the subjects had to complete sequences of non-verbal puzzles

using multiple choices. In the Benton test, participants were required to identify an unfamiliar face. They were presented with a target face and they had to locate this target face within an array of six faces. Motor learning measured motor adaptation. In this task, the participants had to hit a target present on a monitor using a stylus. During the experiment, the cursor on the screen was rotated by 30°, requiring participants to adapt to this perturbated condition in order to hit the target. We decided to focus on the behavioural tasks previously described because they exhibited greater variability in performance with age [21].

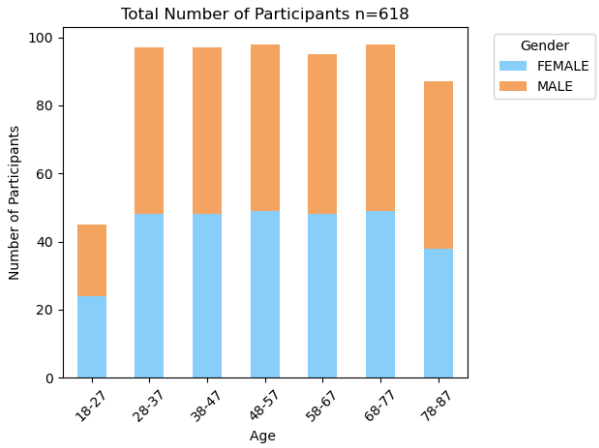

Figure 10: The histogram illustrates the distribution of participants by age decade and gender

3.2 From MEG data to Connectivity matrices

From the MEG data, it is possible to build connectivity matrices. In this paragraph, we will show the most important steps: preprocessing, source localization, cortical parcellation and connectivity estimation (for more details see [32]).
The preprocessing step is used to remove artefacts like head movement, blinks and eye movement. This is achieved using spatiotemporal signal space separation (tSS) [39] and independent component analysis (ICA) [19] techniques. The second step consists of source localization, which utilizes MRI scans for this purpose. Following this, the cortical surface of each subject's brain is reconstructed from the MRI data, enabling the division into ROIs using predefined parcellation methods. In our specific case, the cortical sur-

face was divided into 448 ROIs using parcellation developed by Khan and colleagues [22].

Due to the presence of many ROIs, it has been decided to use a sparser parcellation of 90 ROIs (figure 11).

After obtaining our ROIs, we average all the time series within each ROI to

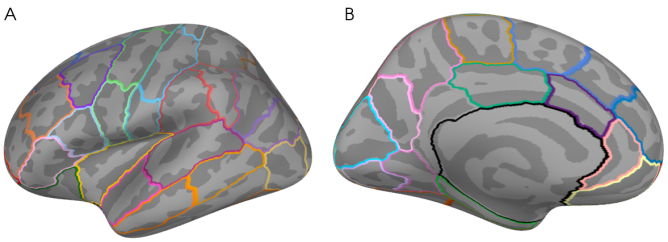

Figure 11: Cortical Parcellation with 90 ROIs. Lateral (A) and in medial (B) view. From *Resting-state functional connectivity changes over the healthy adult lifespan*, by Ruuskanen Santeri

generate a single-time series, which is then transformed into a time-frequency representation (TFR). Then using the Weighted Phase-Lag Index metric [42] on the TFRs, the connection between pairs of ROIs has been computed and a connectivity matrix of $90 \times 90 \times 29$ has been built for each subject (90 represents the ROIs and 29 the frequencies).

3.3 Data Preprocessing

In this section, we will present the data preprocessing and the different methods used to extract features.

The dataset used for this study consisted of 618 3D matrices, one for each subject. Each matrix had dimension $90 \times 90 \times 29$, where the first two dimensions represented the 90 ROIs and the last dimension the 29 frequencies (2Hz-30Hz), in which the connections have been computed.

First of all, we divided all the 3D matrices by looking at the band frequencies: delta band (1,3 Hz), theta band (4,6 Hz), alpha band (7, 13 Hz), and beta band (14, 30 Hz).

It is important to study the different frequency bands separately because different neural processes are associated with different frequency bands. In addition, neural oscillations in different frequency bands occur more in spe-

cific areas of the brain and they can be connected to different processes of the brain [7].
Subsequently, an average matrix was computed for each frequency band to obtain a 2D matrix for each frequency band across all subjects (figure 14). At the end of our preprocessing, our dataset consists of a 2D connectivity matrix for each frequency band and each subject.

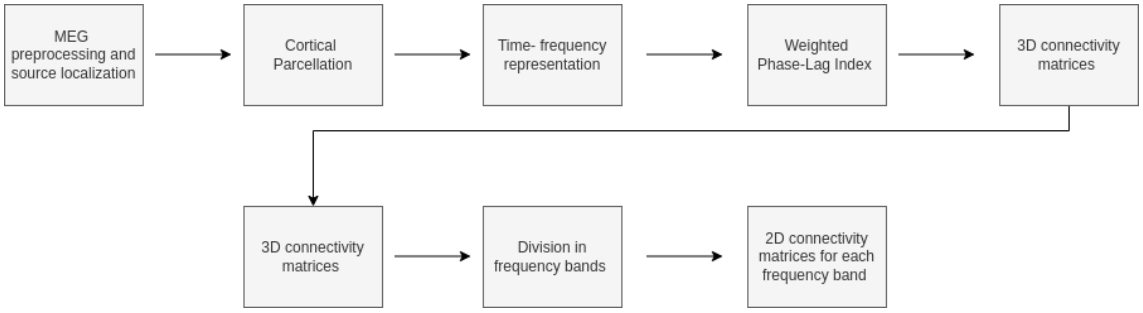

Figure 12: Flow chart illustrating the steps for the data preprocessing

3.3.1 Feature extraction

Neuroimaging data are characterized by high dimensionality, and feature extraction helps to select the most important features [6]. In addition, feature extraction can contribute to identifying biomarkers that help to diagnose disorders. This paragraph discusses the feature extraction methods used and the three sets of features chosen for the analysis: connectivity scores, graph metrics and topological features. The three sets of features have been computed for all the different frequency bands.

Connectivity scores It is possible to use directly the connections between the ROIs as features. Each connectivity matrix has been reshaped in a vector to create a feature vector. Since the connections between two brain areas don't have a direction, the connection between region A and region B is the same connection between region B and A, so our matrix is symmetric. In this case, we considered the connection between two areas only once and the feature vector created by the 90×90 connectivity matrix had 3960 elements in the case of one frequency band and 15840 when we considered all the frequency bands together for one subject.
To reduce noise and redundant information, we applied a dimensionality reduction technique, principal component analysis, to our feature vector.

For a singular frequency band, we determined the optimal number of components for the PCA to be 50. In the case of all the frequency bands considered together, we used the explained variance ratio to choose the best number of components. We used a threshold of 95 % for the explained variance ratio and we obtained the number of components that explain 95 % of the total variance of the dataset was 371 (figure 13).

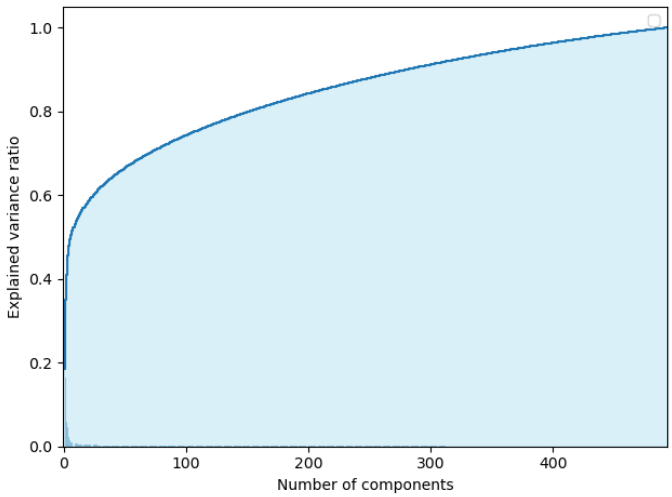

Figure 13: Proportion of the dataset explained by different numbers of components

Graph Metrics For the second set of features, we used three graph metrics: global efficiency, characteristic path length and transitivity. We have seen (section 2.2), that is possible to build a graph from a connectivity matrix. Before building our graphs, we thresholded our connectivity matrices using a density-based threshold, that fixed the density in each network at 20%. This helped to maximize the separation between signal and noise removing the weakest connections. To simplify our analysis we binarized the matrices before computing the graph metrics.

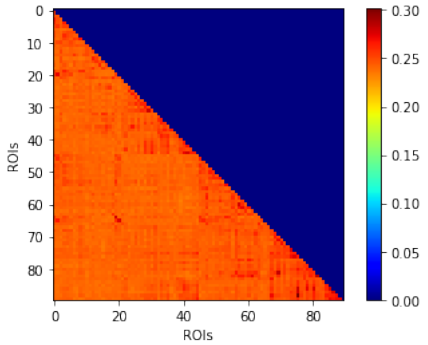

Figure 14: Connectivity matrix in the beta band

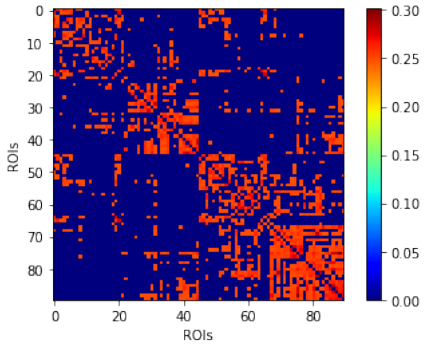

Figure 15: Binary matrix in the beta band

The graphs and the graph metrics were computed using the library NetworkX [18]. The three graphs' metrics were chosen based on their reproducibility on MEG data shown in previous works [30] and they are:
Global Efficiency

$$GE = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} \frac{1}{d_{ij}} \tag{37}$$

where N are the number of nodes d_{ij} is the shortest path length between the node i and j.
Characteristic path length:

$$CPL = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j \neq i} d_{ij} \tag{38}$$

and **Transitivity**:

$$T = \frac{\sum_{i=1}^{N} 2t_i}{\sum_{i=1}^{N} k_i(k_i - 1)} \tag{39}$$

where t_i are the triangles formed by the nodes and k_i is the degree.
These graph metrics were computed for all the frequency bands.

Topological features For the third set of features, we used some topological features more specifically *persistent entropy*. We used giotto-tda library [40] for our topological analysis.
Starting from our weighted graphs, we computed persistence diagrams using *VietorisRipsPersistence*. Persistence diagrams are triples, that contain

births, deaths and dimensionality of the topological features (figure 16). In our case, we considered H_0, H_1 and H_2 as dimensionality.

As a method to create a filtration of the dataset, we used Vietoris-Rips,

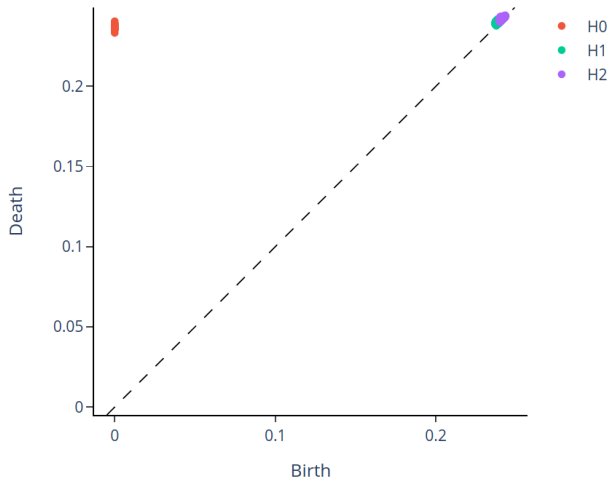

Figure 16: Persistence diagram of one subject in the beta band

(section 2.3.3). The method looks to a parameter ϵ, whose value is changed at every iteration, to create subgraphs from our original graph. Subgraphs are generalized to the simplicial complexes (section 2.3.1) and connected components (0-dimension), holes (1-dimension) and cavities (2-dimension) are counted each time and their birth and death are saved in the persistence diagram. After this, the persistent entropy is calculated from the deaths and births [40]:

$$E = -\sum_{i \in I} \frac{(d_i - b_i)}{\sum_{i \in I}(d_i - b_i)} \log\left(\frac{(d_i - b_i)}{\sum_{i \in I}(d_i - b_i)}\right) \tag{40}$$

and this has been the final topological feature used in our study, where d_i are the deaths and b_i the births of the topological features.

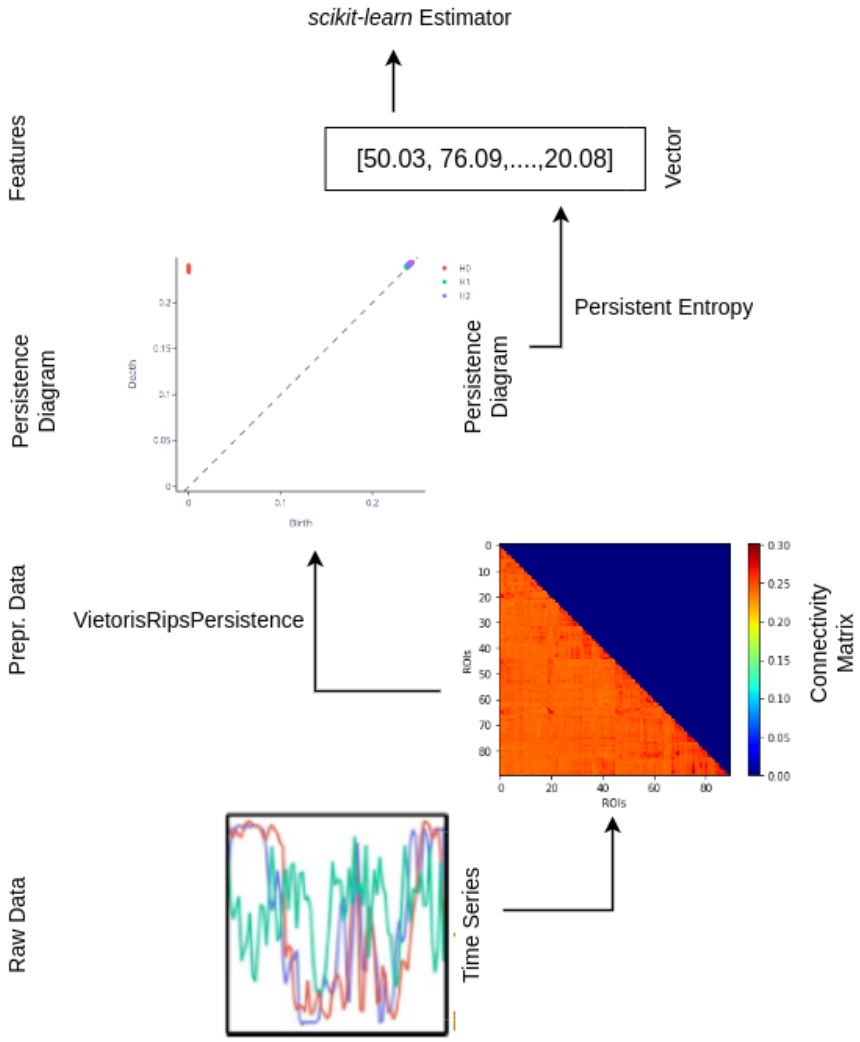

Figure 17: Steps to extract the topological features

3.4 Machine Learning Algorithm

This section shows in more detail the steps followed for the training of our model and its evaluation.

We used a nested cross-validation technique to partition our dataset into training and test sets. Initially, we divided the dataset, which consisted of the three sets of features mentioned earlier, into 10 folds. At each iteration, one fold served as test set while the remaining folds constituted the training set. Before giving the data to our model for the training we standardized them with StandardScaler using the following formula implemented in scikit-learn [28]:

$$z = \frac{x_i - \mu}{\sigma} \tag{41}$$

where z is the new standardized value, x_i the value present before the standardization, μ the mean of the training sample and σ the standard deviation of the training sample. Standardizing the data is crucial for facilitating faster model convergence, as it ensures that all features share the same scale and can lead to a better performance in the model.
Subsequently, the standardization, each training set was used for an inner cross-validation and was divided in 5-fold. At every inner iteration, a different fold was used as a validation for the research of the best hyperparameters for the model. RandomSearchCV was used for this purpose. The model with the best hyperparameters was evaluated in each outer loop using the mean absolute error (MAE).

$$MAE = \frac{\sum_{i=1}^{n} |x_i - x|}{n} \tag{42}$$

where n is the sample size, x_i the predicted value and x is the observed value. The average of the errors in the outer loop represents a measure of the model's performance, which was used to compare different sets of features and identify the most predictive features.

To estimate if the differences between the performances obtained using different sets of features were true or due to random statistical fluctuations, we used a paired samples t-test. In this test we need to evaluate the distribution of differences between performances. For each outer loop of the nested cross-validation, we calculated a difference between the two performances $X_i = P_{1i} - P_{2i}$ where P_1 is the performance of the model trained using the first set of features and P_2 the performance of the model trained using the

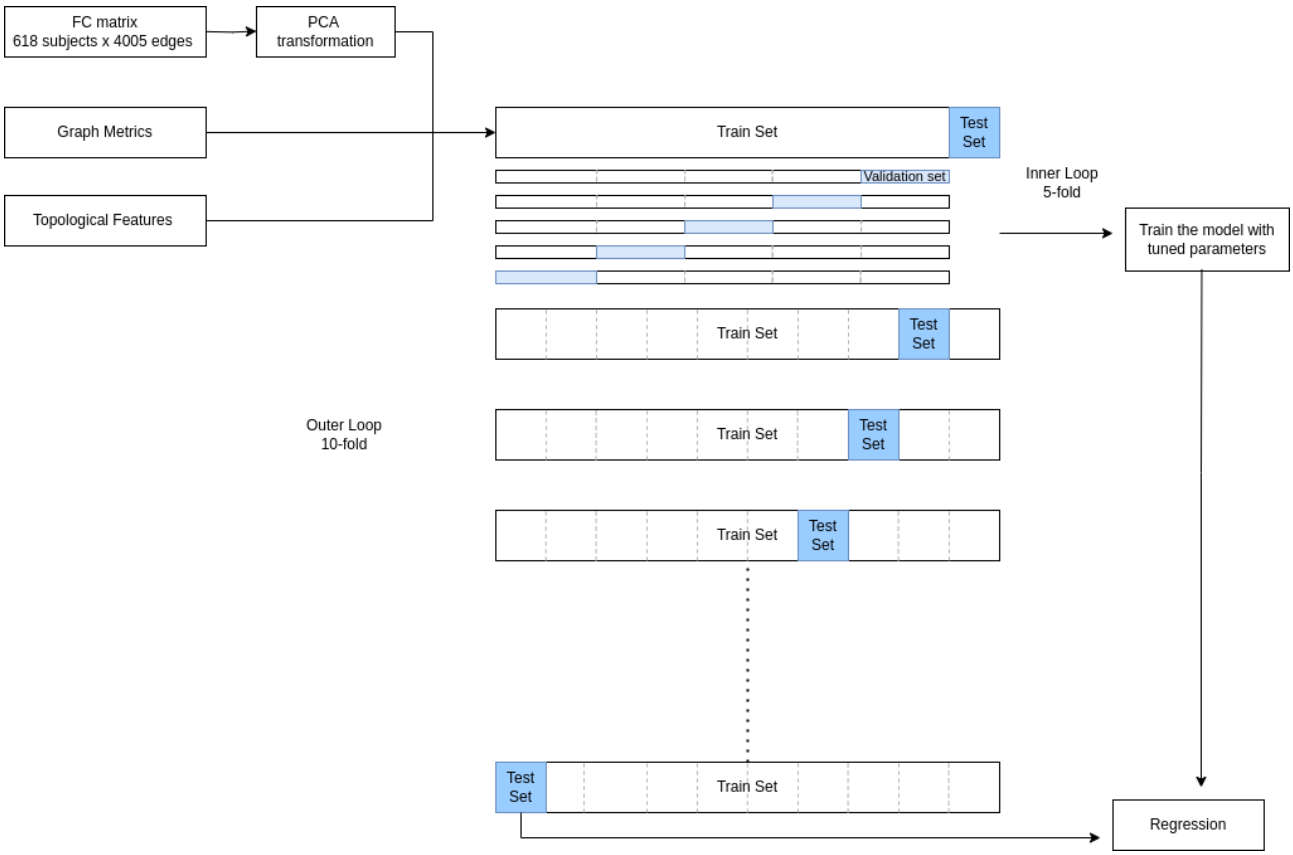

Figure 18: Flow chart of the machine learning model

second set of features. Then we defined

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{43}$$

$$\bar{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \tag{44}$$

The t-statistic is computed using the mean of the difference and the standard error of the mean $SE(\bar{X}) = \frac{\bar{\sigma}}{\sqrt{n}}$:

$$T = \frac{\bar{X}}{SE(\bar{X})} \tag{45}$$

In our study the assumption or null hypothesis for the t-test was that the two ML models perform the same. If through the t-test we reject the

null hypothesis we can prove that the difference between the two models is statistically significant.

Given the test statistic and sample size, we can calculate the *p-value*, which quantifies the probability of observing the data under the assumption of the null hypothesis. Computing the p-value involves defining the degrees of freedom, typically denoted as $df = n - 1$ for paired t-tests, and then visualizing the t-distribution based on these degrees of freedom. The p-value corresponds to the area to the right of the observed test statistic within this distribution. In hypothesis testing, we establish a hypothesis and a significance level denoted by α. If the calculated p-value is less than or equal to α, we reject the null hypothesis.

4 Results

This chapter summarizes the results obtained in the brain and cognitive age prediction. Each section shows a comparison between different models and between different sets of features.
The performances are visualized with *boxplots*, graphs used to represent the distribution of a dataset, and they are measured with the MAE.

4.1 Brain age prediction

This section illustrates the results obtained in age prediction. Initially, we explored different machine-learning models. We evaluated their performance using the mean absolute error (MAE) and we visualized the results through boxplots.
The boxplot shows 5 important values: the minimum error and the maximum error obtained, the median, the first quartile and the third quartile.

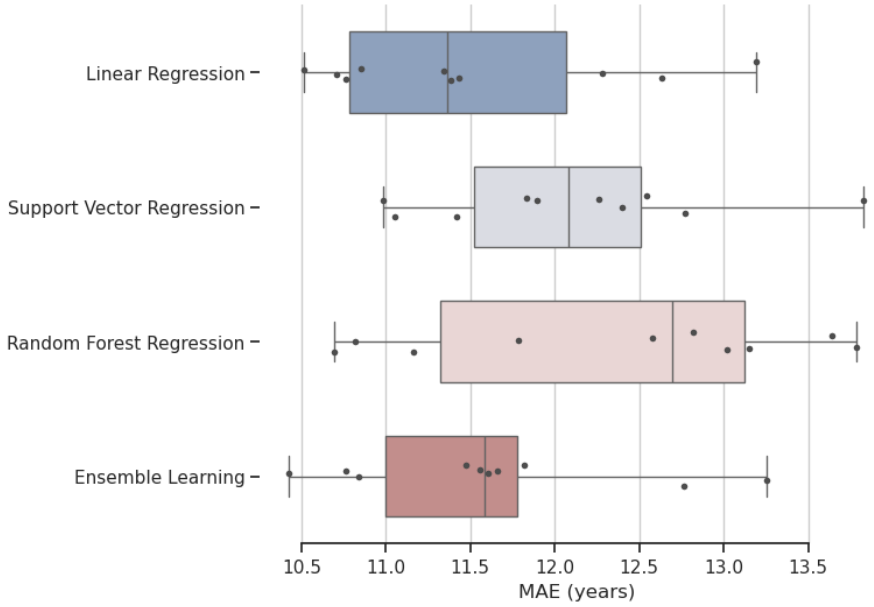

Figure 19: Boxplot representation of machine learning models trained on the set of features extracted from the beta band

In the first boxplot (figure 19) we compare the results obtained using four different models: linear regression, support vector regression, random

Model	MAE (years)
Linear Regression	11.84 ± 0.66
Support Vector Regression	11.98 ± 0.91
Random Forest Regression	12.71 ± 0.80
Ensemble Regression	11.79 ± 0.74

Table 1: MAE for the different models

forest regression, and the ensemble learning model obtained with the voting. During the voting, the models make predictions independently and then the predictions are averaged for all the models.
The feature set considered in this comparison consists only of connectivity scores in the beta band, to which we applied Principal Component Analysis with 50 components.
Since there is minimal deviation in the results among the models, we have opted to concentrate only on one model for the subsequent analysis: support vector regression (SVR).

In the second boxplot, we compare the age prediction done by SVR using 5 different sets of features.

- The first set of features consists of the connectivity scores in the beta band to which Principal Component Analysis has been applied with 50 components.
- In the second set of features we added to the first set the three graph metrics computed in the beta band, global efficiency, characteristic path length and transitivity.
- In the third set of features, we added to the previous one, the persistent entropy calculated in the three dimensions H_0, H_1, H_2 in the beta band.
- In the fourth feature set, we incorporated all the frequency bands: delta, theta, alpha, and beta. Graph metrics were computed individually for each frequency band, and the number of components (371) for Principal Component Analysis applied to connectivity scores was determined based on the explained variance ratio (selecting a ratio of 95%).

- The last set of features contains connectivity scores, graph metrics and topological features for all the frequency bands.

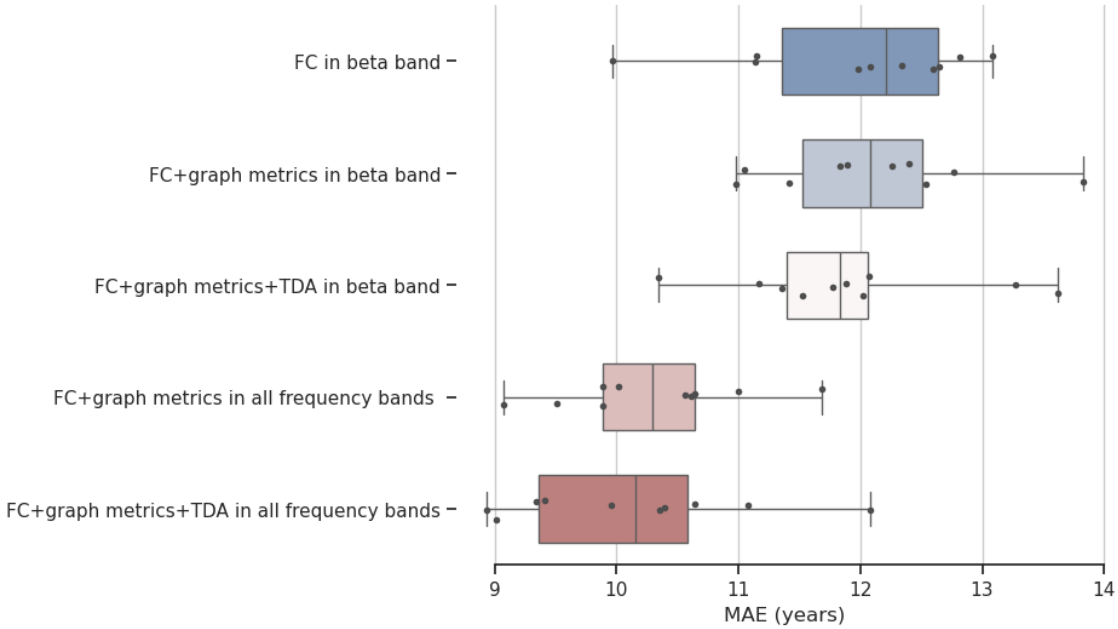

Figure 20: Boxplot representation of SVR trained on a different set of features

Model	MAE (years)
SVR	11.98 ± 0.91
SVR	11.70 ± 0.81
SVR	11.90 ± 0.91
SVR	10.28 ± 0.72
SVR	10.12 ± 0.94

Table 2: MAE for the different set of features

Looking at the boxplot we can see that there is an improvement in the prediction of our ML model from the first set of features to the last one. To verify that this is not due to the randomness with which training and testing were divided we used a t-test. Our null hypothesis was that the two ML models performed the same. We obtained a $p - value = 0.0007$, which

is less than the significance level of 0.05, we rejected the null hypothesis, indicating a statistically significant difference between the two estimations obtained.

4.2 Cognitive age prediction

This section illustrates the results obtained in the prediction of the performance of the behavioural tasks executed by the subjects: Cattell test, Benton Faces test and Motor Learning.
As in the brain age prediction first we compared different models, then we focused on one model and we compared different sets of features. The results are illustrated using as previously the boxplots and as a metric for evaluating the MAE.

4.2.1 Cattell Test

We have seen (section 3.1) that in the Cattell test the participants need to solve a series of puzzles using multiple choices. Performance is calculated on a total score of 46, for each series correctly solved is assigned one score.
In the following boxplot (figure 21) is presented the comparison between the different models

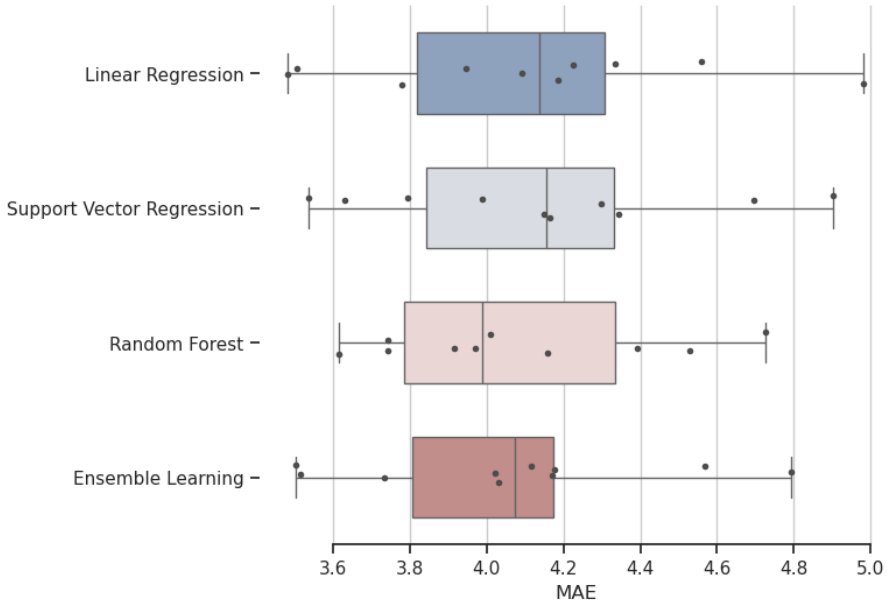

Figure 21: Boxplot representation of machine learning models trained on the set of features extracted from the beta band and the age

Model	MAE
Linear Regression	4.02 ± 0.33
Support Vector Regression	4.01 ± 0.33
Random Forest Regression	4.05 ± 0.32
Ensemble Regression	3.98 ± 0.30

Table 3: MAE for the different models

We can see the results don't deviate too much from each other, so we decided to focus only on one model and compare different sets of features (figure 22).

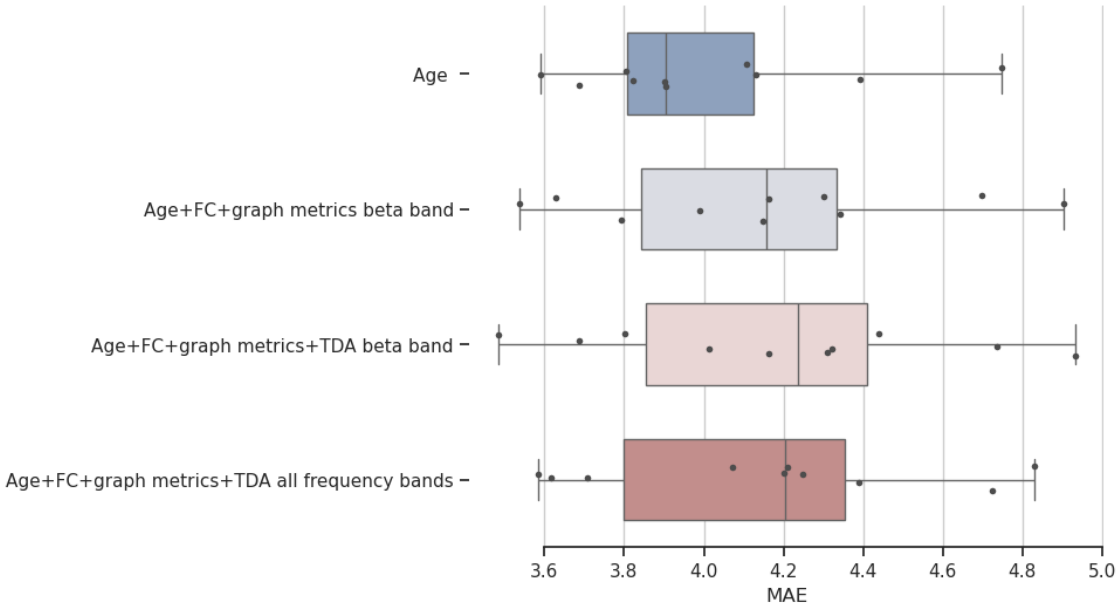

Figure 22: Boxplot representation of SVR trained on different sets of features

Model	MAE
SVR	4.01 ± 0.33
SVR	4.15 ± 0.42
SVR	4.19 ± 0.43
SVR	4.16 ± 0.41

Table 4: MAE for the different set of features

For each behavioural task, we initially used only age as a feature for performance prediction and we used the result obtained as a baseline. Subsequently, we added to the age, features extracted from connectivity matrices to see if MEG features improved prediction accuracy.
Examining the boxplot, it's visible that incorporating features from functional connectivity did not enhance the prediction accuracy. This observation was confirmed through a t-test, yielding a p-value of $p = 0.06$, indicating that $p \geq 0.05$ and thus leading us to accept the null hypothesis.

4.2.2 Benton Faces

In the second task, we saw that participants had to match pictures of unfamiliar faces (see section 3.1). The score is calculated on a total of 27, and it is due to the sum of the recognised faces in 27 trials.
Also in this case we compared the machine learning models (figure 23) and then the different set of features (figure 24).

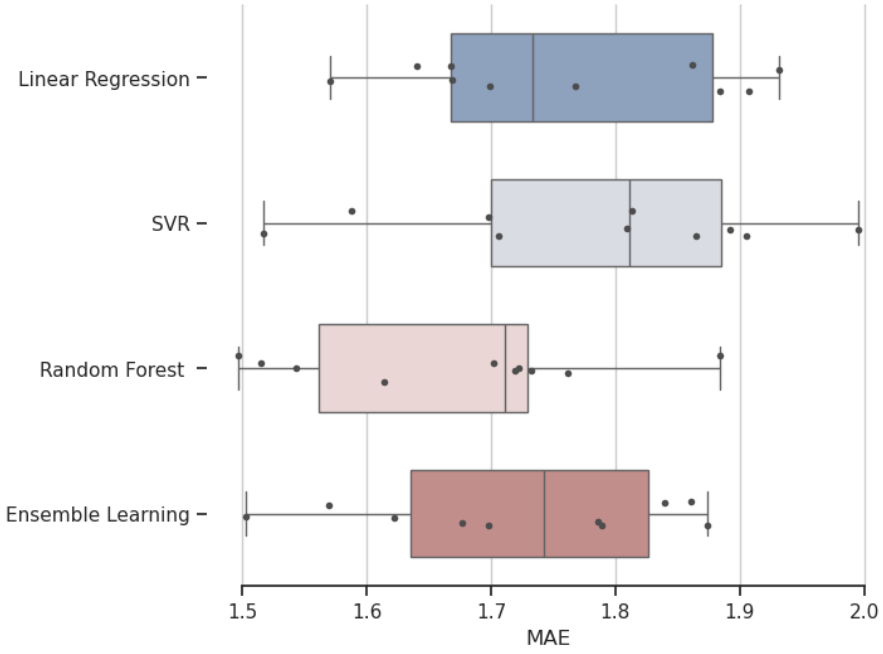

Figure 23: Boxplot representation of machine learning models trained on the set of features extracted from the beta band and the age

Model	MAE
Linear Regression	1.68 ± 0.12
Support Vector Regression	1.63 ± 0.11
Random Forest Regression	1.67 ± 0.12
Ensemble Regression	1.64 ± 0.11

Table 5: MAE for the different models

In the first boxplot (figure 23) we can see the results don't deviate too

much from each other, so we decided to focus only on one model and compare different sets of features (figure 24).

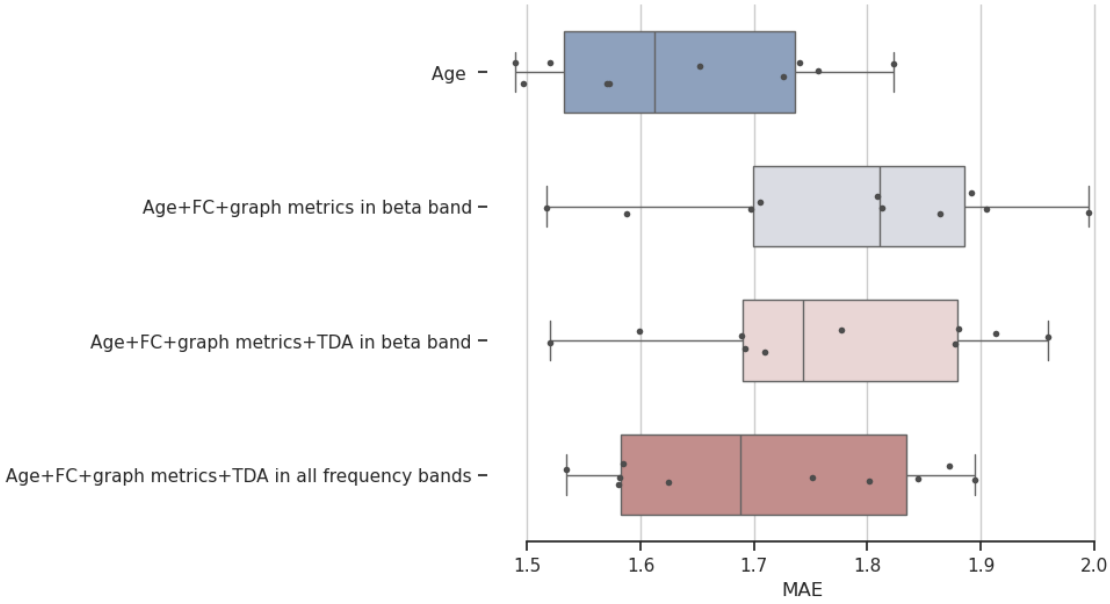

Figure 24: Boxplot representation of SVR trained on different sets of features

Model	MAE
SVR	1.63 ± 0.11
SVR	1.78 ± 0.14
SVR	1.76 ± 0.13
SVR	1.70 ± 0.33

Table 6: MAE for the set of features

Examining the boxplot, it's visible that incorporating features from functional connectivity did not enhance the prediction accuracy. This observation was confirmed through a t-test, yielding a p-value of $p = 0.06$, indicating that $p \geq 0.05$ and thus leading us to accept the null hypothesis.

4.2.3 Motor Learning

The final task examined measured motor adaption. Participants were presented with a monitor displaying a yellow disc target, and they had to hit the target using a stylus. A cursor representing the stylus position was displayed on the monitor. Throughout the experiment, the cursor position underwent rotation, and participants had to adapt to this altered condition. The performance metric utilized in this task is the final adaptation, quantified in degrees. This metric is determined by calculating the difference between the angle of rotation applied to the cursor and the fitted final error. The final error represents the trajectory deviation made by the participant after the cursor rotation.

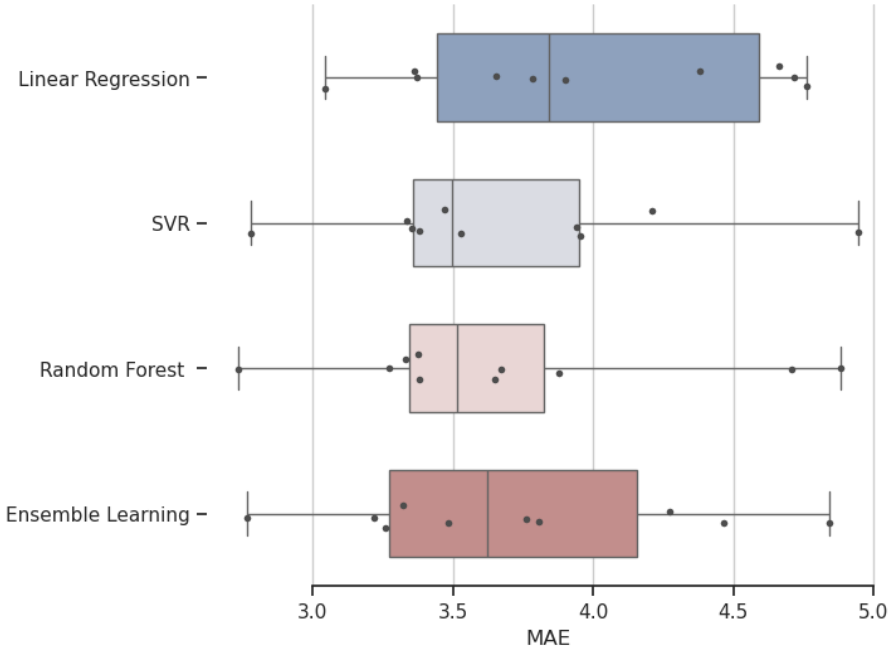

Figure 25: Boxplot representation of machine learning models trained on the set of features extracted from the beta band and the age

Model	MAE
Linear Regression	3.64 ± 0.56
Support Vector Regression	3.66 ± 0.52
Random Forest Regression	3.80 ± 0.65
Ensemble Regression	3.66 ± 0.56

Table 7: MAE for the different models

As previously we first compared different machine-learning models (figure 25) and then we compared the results obtained from SVR trained on different sets of features. The results illustrated on the first boxplot show that the predictions don't deviate too much between the models, so we focused only one machine-learning model and we compared different sets of features (figure 26)

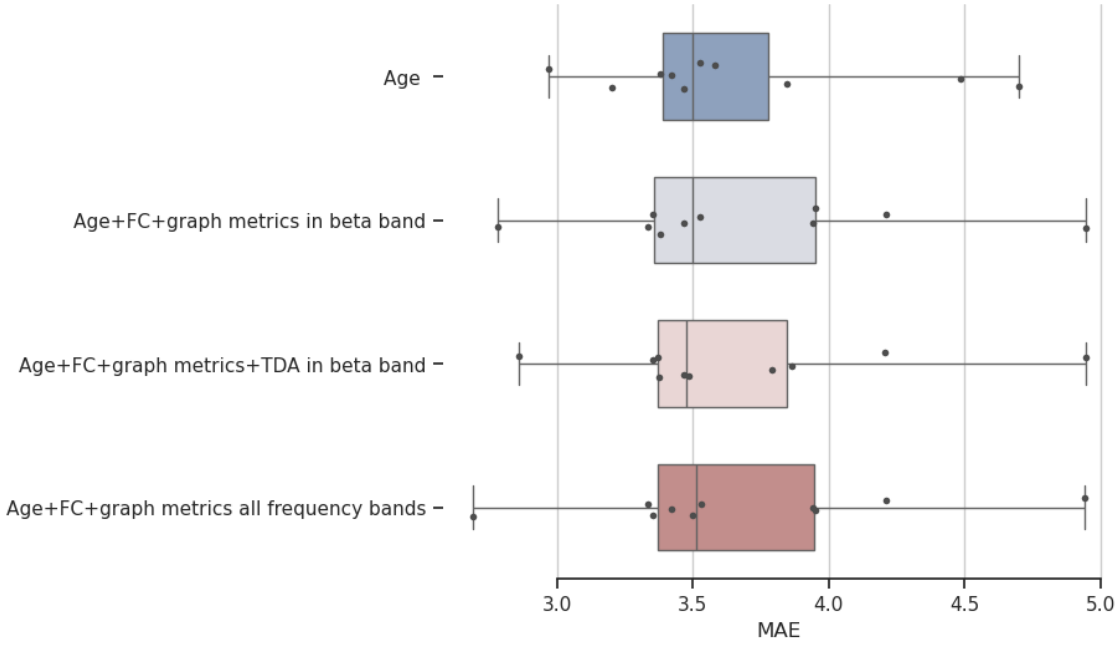

Figure 26: Boxplot representation of SVR trained on different sets of features

Model	MAE
SVR	3.65 ± 0.52
SVR	3.69 ± 0.57
SVR	3.67 ± 0.54
SVR	3.69 ± 0.58

Table 8: MAE for the set of features

Looking at the comparison of the different sets of features we can see that it is not present an improvement. We verified this using a t-test between the first set of features and the last one. We obtained a p-value of $p = 0.75$ that is $p \geq 0.05$, telling us we can accept the null hypothesis.

5 Discussion

This master's thesis aimed to study brain ageing and cognitive ageing by using machine learning models to analyze MEG data, specifically focusing on describing functional connectivity during the resting state of healthy subjects. While functional connectivity has been extensively studied in resting-state fMRI data, research on it with MEG data remains limited, so we decided to focus on MEG data, that have a better temporal resolution.
The main result of this thesis is that it is present is a correlation between age and functional connectivity. It is possible to predict the age of a healthy subject using features extracted from functional connectivity. Notably, adding features constructed through graph theory exhibits greater predictive capability than using only connectivity scores. Regarding cognitive ageing, our observations suggest that alterations in cognitive performance do not correlate with changes in functional connectivity, or at least that age contains already all the main information required for predicting cognitive ageing changes. In this chapter, we will show the results obtained, giving them an interpretation, and we will compare them to previous studies.
Before extracting the features from our MEG data, preprocessing was necessary. Decisions made during preprocessing, such as parameter selection, choice of parcellation (90 ROIs), and the connectivity metric (WPLI), influenced the resulting outcomes. Divergencies from the results obtained and results from other studies may stem from differences in data preprocessing methods. Then it is necessary to consider also the choice of set of features, the threshold used on the connectivity matrices and the choice of the machine learning model.
Support Vector Regressor have been proven to be the best model for our analysis. In previous studies, SVR outperformed other machine learning models due to its ability to handle high-dimensional data that characterize neuroimaging data [25]. We saw that the features extracted from the alpha (8-13 Hz) and beta (13-30 Hz) bands were, in general, the more predictive because in these frequency bands, the functional connectivity changes more with age (for more details see [32]). However, incorporating additional frequency bands improved the performance of our machine-learning model, likely due to the association of different frequency bands with distinct cognitive processes. Therefore the study of different frequency bands can better help understand the heterogeneity in ageing.
In terms of feature selection, incorporating graph metrics notably enhanced performance by providing additional insights into network organization, whereas connectivity scores only give us information about connections be-

tween two ROIs. The results obtained from the graph metrics depend on the threshold used on the connectivity matrices. Furthermore, the integration of topological features yielded even greater improvements. Topological features demonstrate enhanced robustness against noise and resolution changes, enabling the revelation of hidden relationships and structures within the data. In the exploration of cognitive ageing, previous studies have tried to use MEG features to predict the performance on behavioural tasks [21]. First of all, we decided to predict the performance using only the age as a feature, we used the results obtained as a baseline, and then we added the features extracted from MEG data. However, none of the feature sets demonstrated improvement over using age alone. This finding contrasts with prior studies, suggesting that age alone may suffice for prediction. This phenomenon could be attributed to our earlier observation of a correlation between age and functional connectivity in the initial phase of the project. Furthermore, the sufficient use of age could be also because variations in behavioural task performance are not solely attributable to changes in the brain but may also involve other physiological factors. For instance, in tasks involving visual processing, age-related changes in performance may arise from alterations in the retina, rather than exclusively from shifts in functional connectivity.

5.1 Possible improvements and developments

We observed that the achieved performances vary based on the parameters chosen. To improve the results, it could be useful to employ an alternative connectivity metric that contains diverse values, as feature extraction relies on matrix outcomes. Notably, superior outcomes have been obtained with topological features, though there exist various other topological features worth exploring, such as amplitude, point count, and total persistence. Finally, it would be interesting to see which features are more useful for the prediction across different scenarios.

List of Figures

List of Tables

References

[1] Yaser S Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin. *Learning from data*, volume 4. AMLBook New York, 2012.

[2] Nieves Atienza, Rocio Gonzalez-Diaz, and Matteo Rucco. Persistent entropy for separating topological features from noise in vietoris-rips complexes. *Journal of Intelligent Information Systems*, 52:637–655, 2019.

[3] Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541, 1995.

[4] Ricardo Bruña and Ernesto Pereda. Multivariate extension of phase synchronization improves the estimation of region-to-region source space functional connectivity. *Brain Multiphysics*, 2:100021, 2021.

[5] Gyorgy Buzsaki and Andreas Draguhn. Neuronal oscillations in cortical networks. *science*, 304(5679):1926–1929, 2004.

[6] Federico Calesella, Alberto Testolin, Michele De Filippo De Grazia, and Marco Zorzi. A comparison of feature extraction methods for prediction of neuropsychological scores from functional connectivity data of stroke patients. *Brain Informatics*, 8(1):1–13, 2021.

[7] Almudena Capilla, Lydia Arana, Marta García-Huéscar, María Melcón, Joachim Gross, and Pablo Campo. The natural frequencies of the resting human brain: an meg-based atlas. *NeuroImage*, 258:119373, 2022.

[8] Qian Chen, Jiaming Lu, Xin Zhang, Yi Sun, Wenqian Chen, Xin Li, Wen Zhang, Zhao Qing, and Bing Zhang. Alterations in dynamic functional connectivity in individuals with subjective cognitive decline. *Frontiers in aging neuroscience*, 13:646017, 2021.

[9] G.L. Colclough, M.W. Woolrich, P.K. Tewarie, M.J. Brookes, A.J. Quinn, and S.M. Smith. How reliable are meg resting-state connectivity metrics? *NeuroImage*, 138:284–293, 2016.

[10] Rahul S. Desikan, Florent Ségonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, Anders M. Dale, R. Paul Maguire, Bradley T. Hyman, Marilyn S. Albert, and

Ronald J. Killiany. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, 2006.

[11] Reinhard Diestel, Alexander Schrijver, and Paul Seymour. Graph theory. *Oberwolfach Reports*, 4(2):887–944, 2008.

[12] Division DoEaSA-P. World population ageing: 1950-2050. *New York: United Nations*, 2002.

[13] S.B. Eickhoff and V.I. Müller. Functional connectivity. In Arthur W. Toga, editor, *Brain Mapping*, pages 187–201. Academic Press, Waltham, 2015.

[14] Martina Flammer. Persistent homology-based classification of chaotic multi-variate time series: Application to electroencephalograms. *SN Computer Science*, 5(1):107, 2023.

[15] Alex Fornito, Andrew Zalesky, and Edward Bullmore. *Fundamentals of brain network analysis*. Academic press, 2016.

[16] Karl J Friston. Functional and effective connectivity in neuroimaging: a synthesis. *Human brain mapping*, 2(1-2):56–78, 1994.

[17] Karl J Friston. Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36, 2011.

[18] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[19] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.

[20] Veera Itälinna, Hanna Kaltiainen, Nina Forss, Mia Liljeström, and Lauri Parkkonen. Detecting mild traumatic brain injury with meg, normative modelling and machine learning. *medRxiv*, pages 2022–09, 2022.

[21] Rongtao Jiang, Dustin Scheinost, Nianming Zuo, Jing Wu, Shile Qi, Qinghao Liang, Dongmei Zhi, Na Luo, Young-Chul Chung, Sha Liu, et al. A neuroimaging signature of cognitive aging from whole-brain functional connectivity. *Advanced Science*, 9(24):2201621, 2022.

[22] Sheraz Khan, Javeria A Hashmi, Fahimeh Mamashli, Konstantinos Michmizos, Manfred G Kitzbichler, Hari Bharadwaj, Yousra Bekhti, Santosh Ganesan, Keri-Lee A Garel, Susan Whitfield-Gabrieli, et al. Maturation trajectories of cortical resting-state networks depend on the mediating frequency band. *NeuroImage*, 174:57–68, 2018.

[23] V Latora and M Marchiori. Efficient behavior of small-world networks phys rev lett. *2001*, 198701:10, 2001.

[24] Sunhyuk Lim, Facundo Memoli, and Osman Berat Okutan. Vietoris-rips persistent homology, injective metric spaces, and the filling radius. *arXiv preprint arXiv:2001.07588*, 2020.

[25] Andrea Mechelli and Sandra Vieira. *Machine learning: methods and applications to brain disorders*. Academic Press, 2019.

[26] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[27] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6:1–38, 2017.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[29] Giovanni Petri, Paul Expert, Federico Turkheimer, Robin Carhart-Harris, David Nutt, Peter J Hellyer, and Francesco Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101):20140873, 2014.

[30] Haatef Pourmotabbed, Amy L de Jongh Curry, Dave F Clarke, Elizabeth C Tyler-Kabara, and Abbas Babajani-Feremi. Reproducibility of graph measures derived from resting-state meg functional connectivity metrics in sensor and source spaces. *Human Brain Mapping*, 43(4):1342–1357, 2022.

[31] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[32] Santeri Ruuskanen et al. Resting-state functional connectivity changes over the healthy adult lifespan. 2023.

[33] Jan-Mathijs Schoffelen and Joachim Gross. Source connectivity analysis with meg and eeg. *Human brain mapping*, 30(6):1857–1865, 2009.

[34] Meredith A Shafto, Lorraine K Tyler, Marie Dixon, Jason R Taylor, James B Rowe, Rhodri Cusack, Andrew J Calder, William D Marslen-Wilson, John Duncan, Tim Dalgleish, et al. The cambridge centre for ageing and neuroscience (cam-can) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC neurology*, 14:1–25, 2014.

[35] Cosma Shalizi. Advanced data analysis from an elementary point of view. 2013.

[36] Sanjay P Singh. Magnetoencephalography: basic principles. *Annals of Indian Academy of Neurology*, 17(Suppl 1):S107, 2014.

[37] Olaf Sporns, Dante R Chialvo, Marcus Kaiser, and Claus C Hilgetag. Organization, development and function of complex brain networks. *Trends in cognitive sciences*, 8(9):418–425, 2004.

[38] Cornelis J Stam, Guido Nolte, and Andreas Daffertshofer. Phase lag index: assessment of functional connectivity from multi channel eeg and meg with diminished bias from common sources. *Human brain mapping*, 28(11):1178–1193, 2007.

[39] Samu Taulu, Juha Simola, and Matti Kajola. Applications of the signal space separation method. *IEEE transactions on signal processing*, 53(9):3359–3372, 2005.

[40] Guillaume Tauzin, Umberto Lupo, Lewis Tunstall, Julian Burella Pérez, Matteo Caorsi, Anibal M Medina-Mardones, Alberto Dassatti, and Kathryn Hess. giotto-tda:: A topological data analysis toolkit for machine learning and data exploration. *Journal of Machine Learning Research*, 22(39):1–6, 2021.

[41] Jason R Taylor, Nitin Williams, Rhodri Cusack, Tibor Auer, Meredith A Shafto, Marie Dixon, Lorraine K Tyler, Richard N Henson, et al. The cambridge centre for ageing and neuroscience (cam-can) data repository: Structural and functional mri, meg, and cognitive data from a cross-sectional adult lifespan sample. *neuroimage*, 144:262–269, 2017.

[42] Martin Vinck, Robert Oostenveld, Marijn van Wingerden, Franscesco Battaglia, and Cyriel M.A. Pennartz. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *NeuroImage*, 55(4):1548–1565, 2011.

[43] Jacques Wainer and Gavin Cawley. Nested cross-validation when selecting classifiers is overzealous for most practical applications. *Expert Systems with Applications*, 182:115222, 2021.

[44] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.