|  |  |
| --- | --- |
| **Multilayer approach to diagnose and classify Multiple Sclerosis phenotypes using graph theory measures** | |
|  | |
| Shape, rectangle  Description automatically generated | **Joan Ginard Illescas**  Master in Science in Data Science  Machine Learning in Medicine  **Project supervisor**  Eloy Martínez de las Heras  **Coordinating professor**  Ferran Prados Carrasco  **Date of submission**  XX-06-2023 |



This work is distributed under a Creative Commons [Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/3.0/) 3.0 license.

**SUMMARY OF THE FINAL PROJECT**

|  |  |
| --- | --- |
| Title of the project: | Multilayer approach to diagnose and classify Multiple Sclerosis phenotypes using graph theory measures |
| Author name: | Joan Ginard Illescas |
| Project supervisor: | Eloy Martínez de las Heras |
| Coordinating professor: | Ferran Prados Carrasco |
| Date of submission (MM/YYYY): | 06/2023 |
| Name of the degree: | Master of Science in Data Science |
| Topic of the final project: | Machine Learning in Medicine |
| Language: | English |
| Keywords: | Multiple Sclerosis, network analysis, classification |
| Abstract | |
| Multiple sclerosis (MS) is a chronic disease that affects the central nervous system and is a leading cause of disability in young adults. Magnetic resonance imaging (MRI) is a key tool for disease diagnosis, but lesions seen on an MRI do not always correlate with disease progression, known as the "clinical-radiological paradox."  Network science has proved to be a powerful tool for characterizing brain connectivity patterns, and in this work, we propose using network connectivity measures to classify MS patients. We will construct a three-layer network per subject based on MRI data and obtain a set of measures to enable the application of machine learning algorithms to differentiate between healthy subjects and MS patients and distinguish patients with worse clinical outcomes.  During this process we will find most suitable connectivity measures and determine the usefulness of considering the network layers separately or integrating them into a multilayer network. | |

Index

[1. Introduction 1](#_Toc136981661)

[1.1. Context and motivation 1](#_Toc136981662)

[1.2. Personal motivation 2](#_Toc136981663)

[1.3. Goals 2](#_Toc136981664)

[1.4. Sustainability, diversity and ethical/social challenges 2](#_Toc136981665)

[1.5. Approach and Methodology 3](#_Toc136981666)

[1.6. Schedule 4](#_Toc136981667)

[1.7. Summary of the outputs of the project 6](#_Toc136981668)

[1.8. Brief description of the remaining chapters of the report 6](#_Toc136981669)

[2. State of the art 6](#_Toc136981670)

[2.1. Brain networks 7](#_Toc136981671)

[2.2. Single layer networks applied to MS 7](#_Toc136981672)

[2.3. Brain and multilayer networks 8](#_Toc136981673)

[2.4. Multilayer networks applied to MS 9](#_Toc136981674)

[2.5. Machine Learning and MS 9](#_Toc136981675)

[2.6. Conclusions from the state-of-art 10](#_Toc136981676)

[3. Graphs and input data 10](#_Toc136981677)

[3.1. Single-layer graphs and input data form 10](#_Toc136981678)

[3.2. Multilayer networks 12](#_Toc136981679)

[3.3. Edge-colored multigraphs and multiplex networks 14](#_Toc136981680)

[4. Methods and resources 14](#_Toc136981681)

[4.1. Participants 14](#_Toc136981682)

[4.2. Brain networks and processing steps 16](#_Toc136981683)

[4.2.1. Structural brain network (FA network) 17](#_Toc136981684)

[4.2.2. Structural gray matter brain network (GM network) 17](#_Toc136981685)

[4.2.3. Resting-state functional network (fMRI network) 17](#_Toc136981686)

[4.3. Age and sex correction 18](#_Toc136981687)

[4.4. Data harmonization using ComBat 19](#_Toc136981688)

[4.5. Graphs measures 21](#_Toc136981689)

[4.5.1. SVD normalization 21](#_Toc136981690)

[4.5.2. Measures on single layers 22](#_Toc136981691)

[4.5.3. Measures on multiplex graph 23](#_Toc136981692)

[4.5.4. Interlink weights 23](#_Toc136981693)

[4.6. Statistical test and correlations 24](#_Toc136981694)

[4.7. Data augmentation 24](#_Toc136981695)

[4.8. Machine learning models 25](#_Toc136981696)

[4.8.1. Metrics 25](#_Toc136981697)

[4.8.2. Grid Search and K fold 26](#_Toc136981698)

[4.9. Feature importance 26](#_Toc136981699)

[4.10. Software and libraries 27](#_Toc136981700)

[5. Results 27](#_Toc136981701)

[5.1. Global single layer graph measures, with and without SVD correction 27](#_Toc136981702)

[5.2. Local single layer graph measures, with and without SVD correction 34](#_Toc136981703)

[5.3. Results for local multilayer measurements 35](#_Toc136981704)

[5.4. Summary of data for machine learning models 37](#_Toc136981705)

[5.4.1. Removal of correlated features 38](#_Toc136981706)

[5.5. Machine learning model results HS vs PwMS 38](#_Toc136981707)

[5.6. Machine learning model results for MS types classification 40](#_Toc136981708)

[5.7. Feature importance in HS vs PwMS 41](#_Toc136981709)

[6. Conclusions and future work 46](#_Toc136981710)

[6.1. Conclusions related to graph construction 46](#_Toc136981711)

[6.2. Conclusions from machine learning models 47](#_Toc136981712)

[6.3. Future work 48](#_Toc136981713)

[7. Glossary 49](#_Toc136981714)

[7.1. List of Abbreviations 49](#_Toc136981715)

[8. Bibliography 50](#_Toc136981716)

[9. Appendices 53](#_Toc136981717)

[9.1. Graph measures definitions 53](#_Toc136981718)

[9.2. Libraries and packages 58](#_Toc136981719)

List of Figures

[Fig 1. Project Gantt diagram 5](file:///G:\Mi%20unidad\UOC\TFM\PAC4_draft\TFM_JGinard.docx#_Toc136981720)

[Fig 2. Example of a simple graph 10](#_Toc136981721)

[Fig 3. Sample of input data. 11](#_Toc136981722)

[Fig 4. Multilayer social network. 12](#_Toc136981723)

[Fig 5. Types of mult-ilayer networks 13](#_Toc136981724)

[Fig 6. Multiplex network and Supra-adjacency matrix 14](#_Toc136981725)

[Fig 7. Participants data distribution. 15](file:///G:\Mi%20unidad\UOC\TFM\PAC4_draft\TFM_JGinard.docx#_Toc136981726)

[Fig 8. Parcellation scheme in circus plot 16](#_Toc136981727)

[Fig 9. From MRI to networks 18](#_Toc136981728)

[Fig 10. Weights distribution, before and after sex and age correction. 19](#_Toc136981729)

[Fig 11. PCA before harmonization 20](#_Toc136981730)

[Fig 12. PCA with FA data after ComBat 20](#_Toc136981731)

[Fig 13. Distribution of weights before and after SVD normalization 22](#_Toc136981732)

[Fig 14. Confusion matrix definition 26](#_Toc136981733)

[Fig 15. Boxplot graph global measures. FA Layers. With and without SVD. 29](file:///G:\Mi%20unidad\UOC\TFM\PAC4_draft\TFM_JGinard.docx#_Toc136981734)

[Fig 16. Boxplot graph global measures. GM Layers. With and without SVD. 30](file:///G:\Mi%20unidad\UOC\TFM\PAC4_draft\TFM_JGinard.docx#_Toc136981735)

[Fig 17. Boxplot graph global measures. fMRI Layers. With and without SVD. 31](file:///G:\Mi%20unidad\UOC\TFM\PAC4_draft\TFM_JGinard.docx#_Toc136981736)

[Fig 18. Correllations between global measures. Separate Layers. With and without SVD correction. 33](file:///G:\Mi%20unidad\UOC\TFM\PAC4_draft\TFM_JGinard.docx#_Toc136981737)

[Fig 19. Distribution of interlayer weights in case of averaging. 36](#_Toc136981738)

[Fig 20. Permutation Importance for Random Forest. Separate Layers 42](#_Toc136981739)

[Fig 21. Permutation Importance for XGBoost. Separate Layers 42](#_Toc136981740)

[Fig 22. Permutation Importance for Random Forest. 3-multilayer 43](#_Toc136981741)

[Fig 23. Permutation Importance for XGBoost. 3-multilayer 43](#_Toc136981742)

[Fig 24. Permutation Importance for Random Forest. 2-multilayer. 44](#_Toc136981743)

[Fig 25. Permutation Importance for XGBoost. 2-multilayer 44](#_Toc136981744)

List of Tables

[Table 1. Project Schedule 4](#_Toc136967476)

[Table 2. Graph Based Measures in literature. 8](#_Toc136967477)

[Table 3. Participans clinical, demographic and cognitive characteristics 15](#_Toc136967478)

[Table 4. Single layers results of statistical test. No SVD normalization 32](#_Toc136967479)

[Table 5 Single layers results of statistical test. SVD normalization 32](#_Toc136967480)

[Table 6. Global single layer measures that pass test 2. 33](#_Toc136967481)

[Table 7. Local single layer measures that pass statitistical tests 34](#_Toc136967482)

[Table 8. Local single layer measures and number of nodes that pass statistical test without SVD. 34](#_Toc136967483)

[Table 9. ROIs that pass both statistical tests in more than one layer. 35](#_Toc136967484)

[Table 10. Local single layer measures and number of nodes that pass statistical test witht SVD 35](#_Toc136967485)

[Table 11. # of measures that pass statistical test when varying interlayer weights 36](#_Toc136967486)

[Table 12. Summary of multi-layer measures that passed statistical test 1 37](#_Toc136967487)

[Table 13. Summary of multi-layer measures that passed statistical test 2 37](#_Toc136967488)

[Table 14. Features removed due to high correlation 38](#_Toc136967489)

[Table 15. Results from machine learning models HS vs PwMS. Separate layers. 39](#_Toc136967490)

[Table 16. Results from machine learning models HS vs PwMS. 3-multilayer. 39](#_Toc136967491)

[Table 17. . Results from machine learning models HS vs PwMS. 2-multilayer. 39](#_Toc136967492)

[Table 18. Summary of best network for each metric. HS vs PwMS 39](#_Toc136967493)

[Table 19. Results from machine learning model. MS types. Separate layers 40](#_Toc136967494)

[Table 20.Results from machine learning model. MS types. 3-multilayer 40](#_Toc136967495)

[Table 21. Results from machine learning model. MS types. 2-multilayer 40](#_Toc136967496)

[Table 22. Summary of best network for each metric. MS types 41](#_Toc136967497)

[Table 23. More frequent ROIs with positive importance. 45](#_Toc136967498)

[Table 24. More frequent ROIs with negative importance 46](#_Toc136967499)

[Table 25. R packages and versions. 58](#_Toc136967500)

[Table 26. Python libraries and versions 59](#_Toc136967501)

# Introduction

Multiple sclerosis is a chronic disease of the central nervous system and is the first non-traumatic cause of disability in young adults (D. T. Chard et al. 2021) It is characterized by inflammation, demyelination and progressive neurodegeneration (Haider et al. 2016).

Patients can evolve from a clinically isolated syndrome (CIS) into a primary progressive course (PPMS) or a relapsing-remitting course (RRMS), which will evolve into a secondary progressive course (SPMS) within in a period that may vary between 10 and 20 years (Kocevar et al. 2016).

It is of paramount importance to determine which patients will follow each disease course as early treatment can delay disease progression (Comi et al. 2009) and improve living standards of these patients.

## Context and motivation

Magnetic resonance imaging (MRI) is established as a key tool for the disease diagnosis, although lesions shown by this test seem to have no direct correlation with the evolution of the disease (Chard and Trip 2017). Those who have more lesions on an MRI do not necessarily present more symptoms, which is known as the "clinical-radiological paradox." (Barkhof 2002).

Currently, there are no other tools that can measure the progression of the disease. Therefore, it is of great interest to find a system that can classify patients based on the available diagnostic tests, primarily MRI, according to the progression of the disease.

On the other hand, increasing size and complexity of neurobiological data is met with theoretical and computational advances in data analysis (Bassett and Sporns 2017). Network science has proved to be a powerful tool to cope with this challenge and to characterize brain connectivity patterns (Fornito, et al. 2016)

In this work, we propose using network connectivity measures to classify MS patients. For this purpose, we will use data obtained from MRI resulting in the construction of a 3 layer network per subject, with 76 nodes each layer. Based on the connectivity properties of the network, we will obtain a set of measures that will enable us to apply different machine learning algorithms, which we hope will help us distinguish patients with worse clinical outcomes, or at least to differentiate between healthy subjects and MS patients.

Previous studies have explored the feasibility of several machine learning models (Kocevar et al. 2016) or (Zhao et al. 2020), but few, to the best of our knowledge, have focused on using connectivity measures (Solana et al. 2019)

## Personal motivation

One of the subjects that I have enjoyed the most during this Master Degree has been graph theory. That is why I decided I wanted to do my Master Thesis on something related to it. However, I did not want to explore typical examples in graph theory like as social networks or transportation.

While reviewing available thesis topics, my wife, a pediatrician, pointed out this one as the most original one.

I delved into the specific topic of the project and I found very attractive to be able to relate the functioning of the brain or model some aspect of it with a mathematical model such as a graph. As I learned more about this area, my interest grew, and I believe we are at an important moment in advancing our knowledge of how the brain works.

## Goals

In this work we aim to investigate the potential of using network analysis and machine learning algorithms applied to MRI data to classify MS patients.

Our main goal is to find algorithms or an ensemble of algorithms that can classify individuals into healthy subjects and MS patients and discriminate patients in different stages of the disease.

We have additional secondary objectives that are either desirable or serve as preliminary milestones. Here the most prominent ones:

• Find out most suitable network connectivity measures for the task. This is not only relevant in improving the algorithm's subsequent performance but could also assist in the research of the disease itself. It should be noted that previous studies have already explored this selection process (Solana et al. 2018) or (Casas-Roma et al. 2022)

• Determine if it is more useful to consider the layers separately, or to try to integrate them into a multilayer network.

• Obtain results with algorithms that allow for interpretation and avoid “black box” algorithms

## Sustainability, diversity and ethical/social challenges

The Ethical and Global Engagement Competence (EGEC) is defined at the Master’s level as follows: *“Act in an honest, ethical, sustainable, socially responsible and respectful manner with respect to human rights and diversity, both in academic practice and in the professional, and design solutions to improve these practices.”* It addresses three main dimensions: Sustainability, Ethical behavior and social responsibility and Diversity and Human Rights.

Our goal could potentially lead to improved living standards for those affected by MS so we can state this work main impact is on Ethical behavior and Social responsibility.

Diversity is also addressed in the sense that exists no differentiation or discrimination of patients by skin color, religion, sexual orientation or any possible source of discrimination.

However, patients sex and age certainly has to be taken into account as it is well documented that are gender and age differences in WM (Hsu et al. 2008). This does not mean we pursue results which apply to only one sex and age, in fact quite the opposite.

## Approach and Methodology

A project like this has a previous step which is reviewing available literature. A thorough literature review is a valuable tool for ensuring the accuracy of the theoretical framework adopted, refining project goals, and applying relevant findings to the present research.

Data science projects typically follow a set of common stages, such as exploratory data analysis and data processing, feature selection, model creation and model assessment. In addition, there may be additional stages that are specific to the particular topic or domain of the project. It is important to note that the data science process is not strictly linear, but rather an iterative one. Additionally, the boundaries between the different stages can be blurry at times, as there is often overlap and feedback loops between them.

Accordingly project has been divided in four major steps: Data processing, Network connectivity measures, feature selection, model creation and comparison.

**Data processing**

Project’s tutor provides the data. It is composed of a cohort of 147 patients and 18 healthy volunteers. As it has been noted before, data comprises a multi-layer network for each subject, encoded as 3 data matrix per subject. Those 3 matrices represent: structural white matter (WM) network, structural gray matter (GM) network and a resting-state functional network. For each subject some clinical information is available, including age, sex, disease duration, EDSS score, binary classification informing whether the subject is a patient or a healthy subject and the MS type they suffer.

Although data has already been processed in order to obtain the matrices, there are still some decisions to make. For instance, in GM and WM matrices there could be some connections that are not really present or in functional matrix there are negative correlations that would imply the existence of negative weights in our network.

**Network connectivity measures**

In this step, different connectivity measures for each network and patient will be calculated. Literature will serve as a guide to select the measurements to use.

Also in this step we will check whether it is convenient to work with a 3 layer network, combine it into one single and/or discard one or more layers if they are not meaningful.

**Feature Selection**

In order to optimize the performance of our models, we need to select the most relevant features. This can involve conducting statistical test on our data to determine the most informative variables, or using dimensionality reduction techniques to reduce the complexity of the dataset.

**Model creation and assessment**

This stage involves training and testing different models with the same set of train and test data. To assess model performance we will use metrics like accuracy, recall and F-scores among others.

To carry out the project, R and Python, via Jupyter notebook, will be used. In R we will use specific libraries to analyze and perform network measurements, like *igraph* and *muxViz* (De Domenico, Porter, and Arenas 2015), which is a library specially focused on multilayer networks. Besides those libraries we will also use *tidyverse* libraries.

Regarding to Python, data science most relevant libraries will be used: *Pandas, Numpy, Sckit-Learn, Matplotlib* and *Seaborn* and *Scipy*. In addition two specific libraries will also be used: *NetworkX (network measurments)* and *ComBat* to correct biases in our array due to the use of different scanners (Behdenna et al. 2021).

## Schedule

The work plan is organized around a series of milestones, which will be completed in each Continuous Assessment Test. Each milestone is then further divided into smaller steps. As shown in the table below and in the Gantt Diagram (Fig. 1), the main phase of the project (phase 3) is based on the stages outlined in the previous section.

|  |  |  |
| --- | --- | --- |
| STAGE | START DATE | END DATE |
| 1. Work planning | 01/03/2023 | 12/03/2023 |
| 1. State of the art – Bibliographic review | 08/03/2023 | 21/03/2023 |
| * 1. Literature review | 08/03/2023 | 17/03/2023 |
| * 1. Draft | 18/03/2023 | 25/03/2023 |
| 1. Work implementation | 26/03/2023 | 27/05/2023 |
| * 1. Data preprocessing | 26/03/2023 | 12/04/2023 |
| * 1. Network connectivity | 13/04/2023 | 27/04/2023 |
| * 1. Feature selection | 28/04/2023 | 12/05/2023 |
| * 1. Creation of Models | 19/05/2023 | 27/05/2023 |
| 1. Writing Report | 29/05/2023 | 25/06/2023 |
| * 1. Draft | 29/05/2023 | 11/06/2025 |
| * 1. Final version | 12/06/2025 | 25/06/2025 |
| 1. Project Defense | 26/06/2023 | 02/07/2023 |
| * 1. Slides and Video | 26/06/2023 | 02/07/2023 |
| * 1. Public Defense | Date to be determined | |

Table 1. Project Schedule

Gráfico

Descripción generada automáticamente

Fig. 1. Project Gantt Diagram

Fig 1. Project Gantt diagram

There are several reasons why planned schedules may be disrupted, including over-optimistic planning, illness, technological setbacks, and excessive workload at the workplace. That is precisely why we have allocated more time in Phase 3, as it allows for the review of the process thus far and adjustments to be made if necessary.

## Summary of the outputs of the project

The primary outputs of the project include the following: the current document, a virtual presentation discussing the key findings, and a GitHub repository. The repository contains the code files needed to reproduce the different results with the corresponding tables and images are available.

In particular in the code files, stages like data processing, network measures calculation and machine learning model development and evaluation can be found.

## Brief description of the remaining chapters of the report

In chapter 2,"State of the Art", we provides a review of the existing literature in the field. It helps to understand the current state of research.

Chapter 3, "Introduction to Graphs", offers an introduction to graph theory. Some basic concepts of both single and multilayer graphs are presented, while in chapter 4, “Methods and Resources”, we outline the methodology employed in the project, commenting the data sources, processes and some significant decisions made during the study.

In chapter 5 results obtained are presented and discussed in detail, while in chapter 6 I draw together the main conclusions and suggesting potential avenues for future research.

Final chapters, are devoted to glossary, bibliography and appendices that contains supplementary information, such as mathematical definitions for graph metrics and versions of Python and R libraries

# State of the art

As previously mentioned, Multiple Sclerosis (MS) is a neurodegenerative disease that poses a challenge when it comes to relating MRI-detected lesions to physical disability and cognitive impairment in patients. (Fleischer et al. 2016) and (Schoonheim, Meijer, and Geurts 2015) have suggested that during the early stages of the disease, a compensatory mechanism may exist that allows for the reorganization of brain functional networks to cope with disease progression. However, this mechanism is only possible when structural damage is not yet severe (Schoonheim, Broeders, and Geurts 2022). While the proposal of such a mechanism remains controversial (Schoonheim, Broeders, and Geurts 2022), it nevertheless suggests that there is something happening with brain networks in MS, which has led to the classification of MS as a network disease (Schoonheim, Broeders, and Geurts 2022; D. T. Chard et al. 2021). Hence, utilizing network analysis and graph-based measures to study MS, as we propose in this work, is a valid and appropriate approach.

## Brain networks

Although the focus of this work is on the characteristics of brain networks in individuals with MS, it is worth noting that a healthy brain is characterized by a combination of segregated and integrated processing. In a comprehensive review of the literature on brain structure and function, (Schoonheim, Broeders, and Geurts 2022) describe how a healthy brain is represented by high local clustering and short average path lengths between distant regions. The measures of network integration in this kind of network are characteristic path length and global efficiency, while segregation is quantified by modularity and clustering of local efficiency. This organization of networks is referred to as rich club organization (Heuvel and Sporns 2011; Fornito et al. 2016). In rich club networks, high-degree nodes (or network hubs) are more densely connected to each other than to lower-degree nodes.

As (Fornito, et al. 2016) point out, there are 2 main networks studied in brain connectivity: Structural and Functional networks. Structural networks are based on the anatomical connections between different regions of the brain, while functional networks are based on the patterns of synchronized activity between those regions. This means that while the structural network provides information about the anatomical pathways that connect different brain regions, the functional network provides information about the strength and efficiency of the communication between those regions.

## Single layer networks applied to MS

In their review, (Fleischer et al. 2019) enumerated all network measures found in the literature up to that point (Table 2) to distinguish between healthy subjects and MS patients or between patients in different clinical stages of the disease. Some studies have focused on studying structural network disruption, such as (Kocevar et al. 2016; Shu et al. 2016; Llufriu et al. 2016), while others have focused more on functional networks, such as (Welton et al. 2020)

|  |  |  |
| --- | --- | --- |
|  | **Measures** | **Interpretation** |
| Measures of centrality | Degree Centrality | The higher the value the higher the influence of the region |
| Eigenvector Centrality | Higher values correspond to regions which are connected to regions that are central in the network |
| Nodal Efficiency | A higher value indicates a higher ability of the region to propagate information with the other nodes |
| Measures of segregation | Clusteriing coefficient | Fraction of a node’s neighbor that also neighbors. So, it will indicate an organization principle which is cost-efficient |
| Transitivity | Variant of clustering coefficient |
| Local efficiency | It shows the capacity of the network to transfer information between neighboring regions |
| Modularity | Modules are densely connected nodes that are sparsely connected to the rest of the network. Increased values represents an optimized network in response to changing environments |
| Measures of integration | Global efficiency | Information transfer across the whole brain is more efficient |
| Path length | An increase will show a lower ability to transfer information in parallel |
| Measures of network resilience | Assortativity | Increase describes brain ability to continue functioning as response to continuous damage. |

Table 2. Graph Based Measures in literature.

Adapted from Fleischer et al. 2019

Among the first group, (Llufriu et al. 2016) observed an increase in Path Length and a decrease in Global Efficiency, which could indicate a disruption in network integration. (Shu et al. 2016) found a decrease in local and global efficiency. On the other hand, (Fleischer et al. 2016) found that, at least in the early stages of the disease, there is an increase in network clustering and modularity, which could be indicative of the compensatory mechanism mentioned previously.

According to (Schoonheim, Broeders, and Geurts 2022) we could conclude that patients tend to show more segregated and less integrated structural networks overall, particularly in patients with cognitive impairment. In the same review, they pointed out that existing studies on functional networks are more complex and that hypothetical connections between network efficiency and cognition are less clear.

Some authors, such as (Pontillo et al. 2022), have concluded that to this date, there is no "hallmark of multiple sclerosis" in the sense that conflicting results still arise when studying the brain and multiple sclerosis as a single layer network.

## Brain and multilayer networks

Multilayer networks are a relatively new approach in network analysis (Bianconi 2022), and their application to the human brain is even more recent. (Schoonheim, Broeders, and Geurts 2022) note that considering the brain as a multilayer network leads to emergent properties that cannot be fully captured by analyzing individual layers separately. (Sporns 2018) predicts that the use of a multilayer framework is likely to become more widespread.

Regarding the brain, some studies explore the application or extension of single-layer measures to a multilayer setting, such as (Vaiana and Muldoon 2020; Mandke et al. 2018). Others have proposed models, such as the core-periphery organization from a multiplex point of view (Battiston et al. 2018). With respect to disease, it is worth noting that the disruption of the core-periphery structure has been studied in Alzheimer’s disease (Guillon et al. 2019).

## Multilayer networks applied to MS

Given what we have discussed about single layer networks, it's not surprising that (Pontillo et al. 2022; Schoonheim, Broeders, and Geurts 2022) suggest that multilayer networks may provide better insights into the organization of the brain and multiple sclerosis. This approach is so new that I have only found four papers applying multilayer networks to multiple sclerosis.(Kennedy et al. 2023) used five biological layers, which are quite different from the data we have, and (Martí-Juan et al. 2023). studied the relationship between functional and structural networks using a tool called The Virtual Brain. Therefore, I will focus on the other two papers.

(Casas-Roma et al. 2022). examined a three-layer network with the same layers as in our work, including a GM morphological network, a structural brain network, and a functional network. In their approach, all nodes are the same across the layers, but each layer represents a different type of relationship between nodes. One of the main innovations in their study is the use of the WM structural network to represent interlayer connections between the other two layers. They employed global and local measures to describe the properties of the multilayer network, including Strength, Degree, Betweenness centrality, Closeness centrality, and local efficiency. The authors found that all MS patients had lower local efficiency, and most of them had lower closeness centrality and node degree.

In (Pontillo et al. 2022), the researchers also used a three-layer network similar to the one in the (Casas-Roma et al. 2022). study. However, they used a different approach by constructing a multiplex network, which is a type of multilayer network where nodes have a one-to-one correspondence between layers. This allows for the integration of different layers into a single layer. They measured coreness using the definition proposed by (Battiston et al. 2018) and also introduced a Coreness disruption index, which represents a global measure of core-periphery reorganization. They found that the weakening of the multiplex core-periphery structure depends on the disease phase and is associated with physical disability and cognition. They also noted that the modeling of different layers together is still a topic of debate, and new solutions may emerge in the future.

## Machine Learning and MS

It is interesting to note that various machine learning models have been employed to predict or classify MS patients using different types of MRI data. This can be seen in the reviews of (Nabizadeh et al. 2022; Seccia et al. 2021). However, it is surprising that there are no ML algorithms applied to graph metrics, and as far as we know, only (Kocevar et al. 2016; Solana et al. 2019) have applied SVM to connectivity measures. To our knowledge, no one has applied ML algorithms to a multilayer network in the context of MS.

## Conclusions from the state-of-art

After reviewing the literature, it is evident that analyzing MS as a network disorder and employing graph-based measures is a valid approach. Although, it appears that there is an increasing agreement that a multilayer network approach is necessary to fully capture the complexity of MS. However, there is still no consensus on the best way to model the different layers of the network.

This presents a challenge for our work, as our main goal was to explore the use of network analysis and graph-based metrics to study MS. Given the current state of the field, it may be necessary to focus on developing and comparing different approaches to modeling the multilayer network.

# Graphs and input data

## Single-layer graphs and input data form

In mathematics a graph is a structure that captures relationships between objects. It is made up of vertices and edges. Each edge connects a pair of vertices, with the vertices serving as the objects and the edges representing relations between them.

There are different types of graphs. They can be directed or undirected. In an undirected graph each edge connects two vertices without specific direction, while in a directed graph, edges point from one vertex to another in a specified direction.

Graphs can also be weighted or unweighted. In a weighted graph, each edge is assigned a numerical value, known as weight. Depending on the context of the graph, weights can represent different things, such as distances or cost.

Forma

Descripción generada automáticamente con confianza media

Fig 2. Example of a simple graph

In real-world applications, these mathematical graphs are often referred to as networks, with the vertices and edges renamed as nodes and links, respectively. However, these names are used interchangeably, and there is no strict rule dictating their usage.

Graphs can be represented by matrices. One of those matrices is the adjacency matrix. They are squared matrices representing connections between nodes, i.e. element in position i,j represents whether a relationship exists between node i and node j. Given a graph, G, its adjacency matrix is:

|  |  |  |
| --- | --- | --- |
|  |  | ( 1) |

Obviously in undirected graphs adjacency matrices are symmetric .

If graph is weighted adjacency matrix is

|  |  |  |
| --- | --- | --- |
|  |  | ( 2) |

Where is the edge of the weight connecting nodes i and j.

The input data takes the form, or could be interpreted as, weighted adjacency matrices, and thus interpreted as **undirected and weighted** networks. Each participant has one matrix per type of brain network (see section XX), 3 matrices in total.

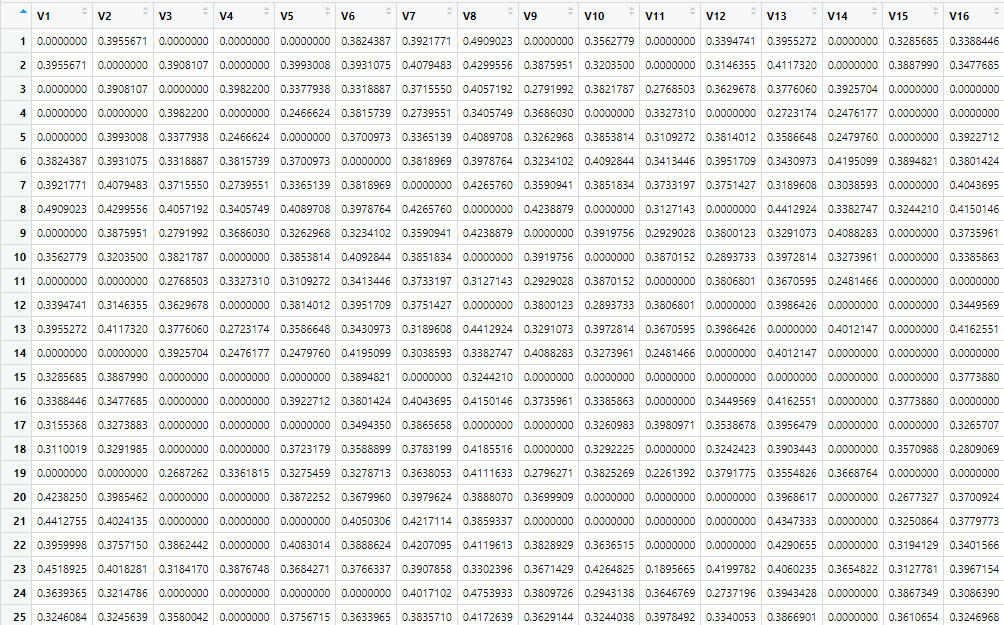


Fig 3. Sample of input data.

Image shows a sample of input data, each participant has 3 matrices of 76 rows and 76 columns (not all shown), where numbers could be interpreted edges weights. So, number in position (4, V5) is the connection between nodes 4 and V5. ROIs corresponding to the nodes are encoded in another file.

In an undirected network, the degree of a node is the number of edges incident to it. In Fig 2, node 1 has degree 2, and node 2 has degree 1. Weighted graphs have an equivalent of node degree which is node strength that instead of counting the links, sums up all weight edges connected to the node.

In network analysis, a set of metrics is employed to assess and quantify certain aspects or properties of graphs. This assessment or quantification is necessary to compare different networks in numerical terms. Depending on the aspect or property being evaluated, metrics can be divided into those associated with the entire network, referred as global measures in this work, metrics associated with the nodes of the network (local measures in this document), metrics associated with a substructure or grouping of network nodes, and metrics associated with the edges or arcs of the network.

No single metric is universally applicable in the sense that multiple metrics are typically required for a comprehensive understanding of a graph’s behavior. Metrics used in this text are defined in appendix xxx. Table 2 provides a useful summary of these measures and their meaning.

## Multilayer networks

Multilayer networks is a relatively recent development in the field (Bianconi 2022). This approach focuses on the interactions between different interconnected networks. For instance, in social interactions, each network can represent a different type of social interaction (work, friends, family, etc.), see Fig 3. Alternatively, in transportation networks, each network might represent a mean of transportation (train, bus, metro, etc.). In these examples it is obvious that there can be “jumps” from one network to another, a person can receive a message from a colleague and pass it on to a friend, moreover it is also possible that a colleague is also a friend. Each network that composes this “superior network” is called layer, hence the name multilayer network.

Diagrama

Descripción generada automáticamente

Fig 4. Multilayer social network.

Figure from (De Domenico 2020) under Creative Commons Attribution-ShareAlike 4.0 International License

An increasing number of studies have underscored the relevance of investigating the relation between structural and functional brain networks (Bullmore and Sporns 2009). As discussed in section 2.3., this can be addressed as a genuine multilayer problem where different brain networks constitute the layers of a multilayer network.

One kind of multilayer network, especially relevant to this case, is **multiplex network.** In this type of multilayer network, each layer has the same set of nodes, and these nodes are interconnected across different layers, see Fig 4. Corresponding nodes belonging to different layer are called replica nodes.

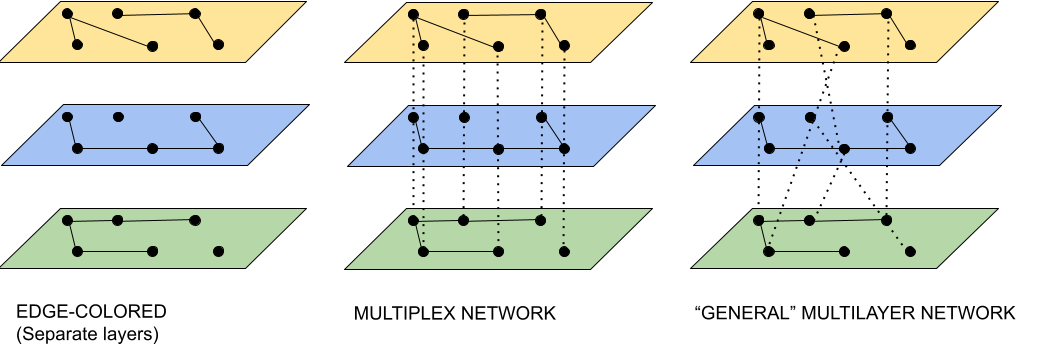


Fig 5. Types of mult-ilayer networks

Figure from (De Domenico 2020) under Creative Commons Attribution-ShareAlike 4.0 International License

Multilayer networks can be addresses in two main ways: using tensors (De Domenico 2022) or a **supra-adjacency matrix** (Bianconi 2022). In this work the latter approach has been employed.

The full information about a multiplex network composed of M layers is encoded in M distinct adjacency matrices, where the matrix for layer αis represented as a[α]. In undirected and weighted networks, these matrices take the form:

|  |  |  |
| --- | --- | --- |
|  |  | ( 3) |

As nodes are only connected to their replica nodes, supra-adjacency matrices is given by:

|  |  |  |
| --- | --- | --- |
|  |  | ( 4) |

where indicates whether node i in layer α is connected to node j in layer β.

Fig 5 provides a clear visualization of how a supra-adjacency matrix is constructed in a multiplex network.

Gráfico

Descripción generada automáticamente

Fig 6. Multiplex network and Supra-adjacency matrix

Figure from (De Domenico 2020) under Creative Commons Attribution-ShareAlike 4.0 International License

## Edge-colored multigraphs and multiplex networks

In this project, as previously introduced, the main objective is to utilize machine learning models and graph measures to distinguish between HS and PwMS, and if feasible, further differentiate between various MS types. To accomplish this, different scenarios will be considered:

1. Edge-colored multigraph: participants' matrices form a multigraph without any connection between layers (see Edge-Colored Multigraph in Fig 4)
2. Multiplex network with 3 layers: matrices are integrated into a multiplex network, creating a single, interconnected network.
3. Multiplex network with 2 layers: same case as before but considering only two layers. Preliminary results showed no real contribution to relevant measures from GM network (see section xxxx.), so we decided to compare the 3 layer approach with 2 layer case.

In this way it will not only be possible to compare the results from both approaches but also to obtain an insight into how various graph measures vary between three separate layers and a unified network (see also Fig 8)

# Methods and resources

## Participants

The dataset encompasses data from 165 subjects, with ages spanning from 22 to 72 years. Among these participants, 18 where healthy subjects (HS), who volunteered for the study, and the remaining 147 were patients with Multiple Sclerosis (PwMS). This group can be further subdivided based on the type of MS: 6 with PPMS, 16 with SPMS and the remaining 125 with RRMS. Physical disability is assessed using the EDSS.

Below is a table summarizing data from all participants:

|  |  |  |
| --- | --- | --- |
|  | PwMS ( n = 147) | HS (n = 18) |
| Female, n | 104 | 15 |
| Age years | 47.3 ± 10.1 | 36.6 ± 9.6 |
| MS type |  | ---- |
| RRMS | 125 (90 female) | ---- |
| SPMS | 16 (10 female) | ---- |
| PPMS | 6 (4 female) | ---- |
| EDSS, median (range) | 2 (0 – 7.5) | ---- |
| Disease duration, year | 14.1 ± 10.1 | ---- |

Table 3. Participans clinical, demographic and cognitive characteristics

Continuous variables are given as the mean ± standard deviation

Gráfico, Gráfico en cascada

Descripción generada automáticamente Data distribution is shown Fig 6.

Fig 7. Participants data distribution.

Color meanings are explained next to each plot. There is no particular meaning associated with color besides indicating MStype or whether a subject is HS or PwMS. From graphs a) and b), it can be inferred that HS group tends to be younger, while patients with more advanced types of MS lean towards the higher age spectrum (b). Graphs c) and d) clearly show a significantly higher representation of females. In graph e), the disease duration appears to cluster around the 10 to 20-year mark. Finally, graph f) illustrates how disability tends to exacerbate during the more advanced stages of the illness.

## Brain networks and processing steps

As it has been discussed before there are 3 matrices per subject. Scanner data acquisition and preprocessing steps needed to obtain these matrices are beyond the scope of this work, but it is important to note that these matrices still require some additional processing and verification.

During data acquisition, the brain was segmented into 76 regions of interest (ROIs), or nodes, leading to 76 x 76 matrices (see Fig 7).

Preprocessing was conducted to ensure all matrices are symmetrical. Consequently, when interpreting these matrices as graphs, they will be recognized as undirected graphs, given that the connection from node i to node j is identical to that from j to i. The main diagonal in matrices represents self-connections and will be removed if present.

Diagrama, Esquemático

Descripción generada automáticamente

Fig 8. Parcellation scheme in circus plot

Image shows relevant connections (statistically significant differences between groups), colors are associated with greater with similar significant connections.

### Structural brain network (FA network)

As described in section 2.1., this network maps the anatomical pathways connecting different regions of the brain. To generate matrices that embodies this network, the diffusion of water molecules is studied (hence the name diffusion MRI). Diffusion of water molecules in the brain is anisotropic, indicating that diffusion does not occur freely (and isotropically) but rather following pathways, running in parallel to the barriers imposed by brain structure. Fractional anisotropy (FA) quantifies how water diffusion is constraint in a given direction within a voxel.

Despite the preprocessing steps effectively minimizing or eliminating factors that could introduce noise in our data, a threshold is set to further ensure the elimination of all non-connections. This threshold is applied to data matrices to remove these spurious connections (Fornito, et al. 2016). However, since this could inadvertently remove true, albeit weak, connections, an additional criterion is used. If a connection is present in at least 60% of the healthy subjects (HS) - that is, 11 out of 18 - the connection is retained.

These conditions can be encapsulated as follows:

1. Eliminate all connections that fall below a threshold, which is set at **0.1**.
2. However, if these connections are present in at least **11 out of 18** HS, they should be retained.

The preprocessing steps proved to be precise, as no changes were observed after applying this condition.

Fig 9 shows a distribution of weights across all matrices, with a significant number of outliers at zero. This situation is partially rectified with the age and sex correction (see section XX). It is decided *not to impute* any value to remove these outliers, under the consideration that this values are likely to represent weak structural connections, thus any imputation will not greatly affect the overall data. Furthermore, brain graphs are sparse and this zeros are compatible with this kind of graph. Given that graph measures are being applied to the matrices/graphs, these values should not pose any numerical problems.

### Structural gray matter brain network (GM network)

For each participant, there is another structural network. This network is derived from the similarities in gray matter (GM) morphological patterns according to the defined parcellation scheme. As mentioned before acquisition of this data, as well as parcellation scheme fall beyond the scope of this work.

The same threshold and conditions applied to the FA networks are also implemented for these matrices, and it is observed that this processing does not induce any changes.

### Resting-state functional network (fMRI network)

Functional MRI is used to measure participants' brain activity during resting-state. After preprocessing, signal correlations between Regions of Interest (ROIs) form the matrices. As described in section 2.1. this networks provides information about the communication between nodes.

As matrices are formed by correlation coefficients, its values range from approximately -1 to 1, with diagonal values representing self-correlations. When interpreting these matrices as the graph adjacency matrices, their values will serve as network edge weights. This requires to remove negative values, as they would imply negative weights. However, discarding negative correlations could lead to a loss of significant information (Fleischer et al. 2019).

Given that negative values also suggest a relation between ROIs, absolute value is taken to retain this information while setting the diagonal values to zero.

Diagrama

Descripción generada automáticamente

Fig 9. From MRI to networks

## Age and sex correction

The brain, similar to many other biological systems or structures, undergoes changes with aging and exhibits gender-related differences (Hsu et al. 2008).

To control for age and gender, and allow comparison across participants, the matrix values are adjusted using a linear regression with age and gender as regressors. For each specific i,j position in all matrices, all 165 values are collected, one per participant, and linear regression is applied. Since the matrices are symmetric and the diagonal is zero, only one linear regression for each matrix position in the upper triangle has to be considered. Linear regression formula is:

|  |  |  |
| --- | --- | --- |
|  |  | ( 5) |

Where are predicted values, α and β are the regression coefficients. The difference between actual value (y) and predicted value, is known as residuals.

|  |  |  |
| --- | --- | --- |
|  |  | ( 6) |

These residuals account for information not explained by our regressors, i.e., sex and age. Therefore and theoretically any changes caused by the disease will be captured in these residuals. The final value of i,j position in our matrix is our residuals plus and “standard” value for this position. This standard value should represent a normal brain, corrected for age and sex. However as this value is not uniform across all brains, the mean of the healthy volunteers is considered as the standard value. Consequently, for patient m, the final value in position i,j is given by:

|  |  |  |
| --- | --- | --- |
|  |  | ( 7) |

Please note that the linear correction may result in values marginally below zero or slightly above one. Since negative values cannot represent graph weights, they are set to zero. However, values above one are retained, as they are infrequent and only marginally exceed one.

In the following figure, the differences in values distribution before and after accounting for age and sex are clearly shown:

Gráfico, Histograma

Descripción generada automáticamente

Fig 10. Weights distribution, before and after sex and age correction.

Differences are especially noticeable in FA connections (A) before, B) after), where some values of zero appear to be corrected. Visible changes are also present fMRI connections, while no substantial differences are appreciated in GM connections.

## Data harmonization using ComBat

Participants data have been collected using two different scanners. According to existing literature (Fortin et al. 2017), diversity in data acquisition methods may lead to an increase in the variance of matrix values.

To verify if there is an increase in variance, principal component analysis, PCA, is conducted on data (see Fig 10). The analysis indicate effect of data acquisition is only evident in FA data

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Fig 11. PCA before harmonization

In graph we could appreciate how only FA shows a significant effect of data acquisition.

To overcome this problem neuroCombat library (Fortin et al. 2018) is applied to FA data. And in the following figure displays data changes in PCA for FA data.

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Fig 12. PCA with FA data after ComBat

## Graphs measures

Measures mathematical definitions can be found in appendices (see section 8.1)

### SVD normalization

In order to integrate different layers in one single multi-layer network, like the multiplex network intended to use, it is necessary to correct differences in link weight across layers. According to (Mandke et al. 2018) uncorrected weights could introduce biases, and thus, it is advisable to apply a form of correction before conducting multilayer analysis

Singular Value Decomposition (SVD) technique for weight adjustment is used, as recommended by (Mandke et al. 2018) and also utilized by (Pontillo et al. 2022). If the adjacency matrix is represented as A, SVD can be applied in the following manner:

|  |  |  |
| --- | --- | --- |
|  |  | ( 8) |

In this formula, U and V hold the left and right singular vectors respectively, and Λ represents the singular values of A. For link weight correction, rescaling using the largest singular value λ1 is applied. Hence, the rescaled matrix is given by

|  |  |  |
| --- | --- | --- |
|  |  | ( 9) |

A factor of 10 is incorporated to ensure the range of values is not too small and approximately varies between 0 and 1.

The following figure shows the changes in matrices weights before and after applying this SVD “normalization”.

SVD normalization results in two sets of measures for single layers (see next section). The first set corresponds to the edge-colored multigraph where normalization is not applied because, in fact, they are “separate layers”. The second set corresponds to the case the layer layers integrate into a multiplex, in this scenario a normalization is required.

Although it could be argued that normalization is also required in the separate layers case, it is decided to proceed without SVD correction in order to compare the two situations. In section XXX differences in results obtained in graph measures with and without SVD are discussed.

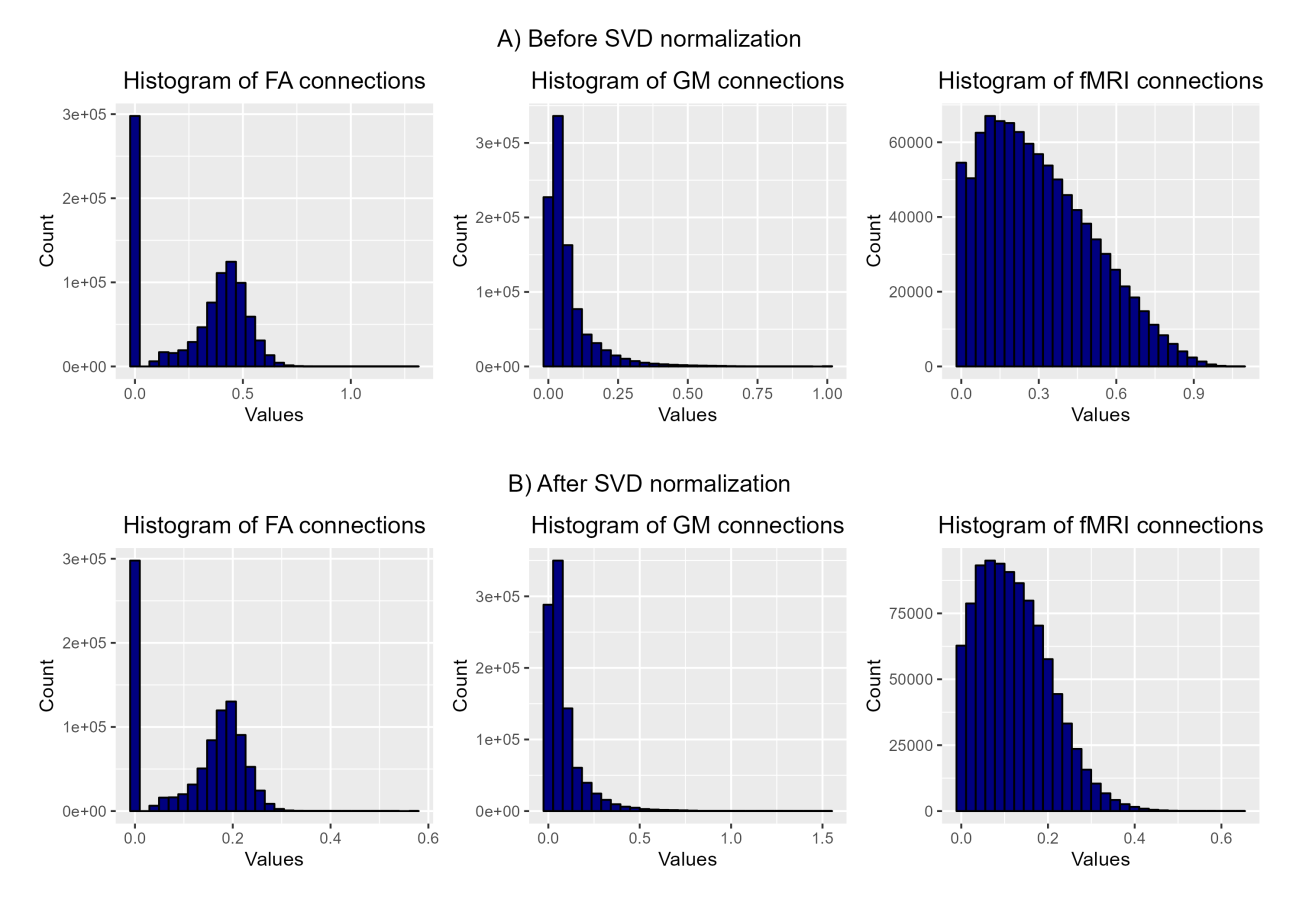


Fig 13. Distribution of weights before and after SVD normalization

We can see that the values have been compressed into a narrower range, although the overall shape remains roughly the same. It is also notable that in the GM connections, values can go up to 1.5.

### Measures on single layers

Various graph measures are carried out, both global and local, on the separate layers case. These measures will also be applied to each layer within the multilayer network. As discussed in the previous section, since matrices in the first scenario have not undergone SVD correction, this approach will show how such correction impacts the results of graph measurements.

From a practical standpoint, to perform all measures straightforwardly different programming libraries are used, eliminating the need to decide in advance which measures to include. However, it has been decided to include at least those measures that previous authors have identified as relevant (see section 2.4). It was also decided to incorporate density, even though it is less likely to yield distinctive results, given that the brain is a sparse network, making changes due to illness challenging to detect using this metric.

For global measures following measures have been considered:

1. Global Efficiency
2. Mean Path Length
3. Transitivity, global clustering
4. Diameter
5. Modularity
6. Global Strength
7. Density

In the case of local measures the following have been taking into account:

1. Efficiency
2. Closeness Centrality
3. Strength

Many measures are somehow related. Global efficiency is approximately the inverse of the average shortest path, and local efficiency and closeness centrality are clearly related (see mathematical definitions in the appendices). As a result it is likely that there will be some highly correlated graph measures, which will need to be removed prior to applying machine learning models. However, before that step, statistical analysis will be conducted to identify which measures exhibit statistical differences between the groups (see section x.x.)

### Measures on multiplex graph

The measures that consider the entire graph and not just individual layers as in the previous section, are referred generally as measures on multiplex or multilayer graph

In this case, preliminary results showed that global measures were not statistically significant. Thus, to reduce code execution time, it was decided to focus solely on local measures.

Three measures of centrality have been applied

1. MultiPageRank centrality
2. Sum of total multi-Strength (a multilayer “version” of strength )
3. Closeness Versatility (closeness centrality for multilayers on some plots or tables “MultiPath” due to the name programming library muxVix gives to the functions that performs the measure)

These three measures have been chosen, while disregarding others, for two main reasons: to my knowledge they have not been used in previous studies, and they showed variability as interlink strength changed (see following section).

### Interlink weights

In a multiplex network, all nodes in one layer are connected to their replica nodes in other layers. However, what should be the weight assigned to these interlinks?

Literature does not provide standard methods for choosing these interlink weights: most of the techniques use some sort of *a posteriori* method, choosing weights that maximize an output of interest. Only in Mandke et al. 2018, where the interlink weight is chosen as the average of weights in layers, an explicit a priori method can be found.

Given this, both approaches are tested:

1. Interlink weight as the average of weights
2. A comprehensive search of links that maximize statistically significant measures.

## Statistical test and correlations

Statistical tests on measures are conducted to keep only those measures that show statistically significant differences between groups. This step is not strictly required but assists in reducing the number of features fed to the machine learning models. Furthermore, it helps to identify measures and nodes that are more important to differentiate between groups.

The statistical testing process is as follows:

1. The Shapiro-Wilk test is applied to ascertain if the data follows a normal distribution.
2. If data does not conform a normal distribution the Mann-Whitney-Wilcoxon test is applied.
3. If data does follow a normal distribution, Bartlett test is uses to check for homoscedasticity before proceeding to the t-test
4. The previous test were conducted to distinguish between HS and PwMS. To further refine the feature selection, Bonferroni correction is employed to identify measures that show statistical differences between the MS types.

The initial three steps form what is referred to throughout the document as the 'first test' or 'test 1' in reference to the statistical test. Meanwhile, step 4 will be referred to as ‘second test’, ‘test 2’ or simply as ‘Bonferroni’.

After this step the number of features is further reduced by removing highly correlated measures.

Results are shown in section XXX

## Data augmentation

One well known problem in classification task is the issue of imbalanced datasets, where there is a majority class significantly more represented that the other(s). In this project, the majority class is PwMS when differentiating control subjects from participants with MS, and RRMS when differentiating MS types. This imbalance can lead to scenarios where a model primarily predicts the majority class, resulting in artificially high accuracy despite poor predictive power for the minority class. This is because the model may struggle to identify patterns in the minority class due to insufficient sample size.

In order to solve this, augmentation techniques that overrepresents the minority class are used. One such method is Synthetic Minority Oversampling Technique (SMOTE) proposed by (Chawla et al. 2002). Briefly, SMOTE algorithm works by randomly selecting an example from the minority class and then one of its nearest neighbors. It then draws a line in the feature space between these two samples and creates a new instance in a point along the line (Brownlee 2021).

One drawback of this method is that since it generates new samples based on existing ones, it must be applied after partitioning the dataset into training and testing sets. This is to avoid having very similar samples in both sets, which could affect model performance. However, applying SMOTE after this split further limits the availability of minority class instances for the algorithm. This can potentially compromise the diversity of the minority instances, and subsequently the robustness of the model.

Another problem is that SMOTE works better when used in conjunction with under-sampling of majority class (Brownlee 2021). This approach might be suitable for large datasets, but it seems impractical for our case with just 147 instances of the majority class. Consequently, it was considered that the application of under-sampling will further reduce the already limited data.

## Machine learning models

One of the major goals in this work is the interpretability of the machine learning models outcomes. As such, this study consciously avoids “black box” algorithms that do not provide insights into the significance attributed to different features.

In recent years ensemble algorithms like *Random Forest* (Ho 1995) or boosting algorithms like *Gradient Boosting* had garnered significant attention. These methods enhance the results of individual algorithms by combining several simple standalone models to yield improved predictions (Gironés Roig, Casas Roma, and Minguillón Alfonso 2017).

Another prevalent method in literature is *Support Vector Machines* (SVM or SVC for the case of a classifier), including numerous studies on MS (Zurita et al. 2018, Kocevar et al. 2016, Solana et al. 2019).

In addition to the aforementioned methods, the decision was made to apply a relatively simple algorithm, K-Nearest Neighbors (KNN). While not widely cited in MS literature, the intention is to explore the performance of a different model.

To recap, 4 models have been utilized:

1. Random Forest Classifier
2. Support Vector Classifier
3. KNN Classifier
4. Extreme Gradient Boosting Classifier (XGBoost)

### Metrics

To evaluate machine learning models performance metrics that are not only the most sensitive for our case but also the most commonly used in literature (Gironés Roig et al. 2017) have been employed:

|  |  |  |
| --- | --- | --- |
| Accuracy |  | ( 10) |
| Precision |  | ( 11) |
| Recall |  | ( 12) |
| F1-Score |  | ( 13) |

Where TP, TN, FP, FN stand for True Positive, True Negative, False Positive and False Negative, respectively.

All these metrics are derived from the confusion matrix which is a contingency table representing true values and predicted ones, as illustrated in the following image.

Imagen que contiene Gráfico

Descripción generada automáticamente

Fig 14. Confusion matrix definition

Adapted from (Gironés Roig, Casas Roma, and Minguillón Alfonso 2017)

### Grid Search and K fold

The identification of the best hyperparameters for each model was facilitated through a grid search. This method evaluates the model by testing every combination of possible parameters provided and returns the combination that yields the best value for the chosen metric - in this case, the F1 score, see definition in the Metrics section.

The model's training utilizes a K-fold approach. In this procedure training set is randomly divided into k number of observations of approximately equal size and the model is trained on k-1 folds and validated using the remaining one, and this process is repeated k times; each time a different fold is treated as a validation set. (James et al. 2013)

## Feature importance

Constructing a model that makes accurate predictions is important, but identifying which measures and nodes play a key role in the model is equally significant. Therefore, the final part of model construction is dedicated to explore feature importance, which pinpoint the nodes and measures most essential for prediction.

To identify the importance of the features a technique implemented in Scikit-Learn, called *Permutation Feature Importance* will be used. Initially, the model is trained using the training set, and a score with set data is obtained. Then one feature within the model is shuffled and the test data is passed to get a new score. If the feature is important, the model performance will deteriorate and the score will drop. Otherwise the score will remain fairly the same.

Owing to time constraints, a comprehensive analysis of all models and scenarios is not feasible. Consequently, attention will be concentrated on ensembles and boosting, specifically Random Forest and XgBoost models.

## Software and libraries

The initial part of the project which encompasses preprocessing, age and sex correction and graph measures was carried out using R, version 4.2.3. The igraph package was employed for single layer measures, whereas muxViz was used for multilayer measures. However there was one notable exception in this part, to conduct data harmonization, I used Neurocombat library for Python

The final part of the project, involving machine learning models, was conducted using Python, version 3.10.10, and Scikit-Learn library. Additionally, customary libraries like Pandas, Numpy, Matplotlib and Seaborn were also utilized. Some other specific libraries used in this part were imblearn for data augmentation and xgboost for this classifier

In order to perform other specific tasks, like plotting results other libraries were needed. Complete list of libraries and versions can be found in appendices XXXX

# Results

## Global single layer graph measures, with and without SVD correction

In this section we will examine the result of applying the network measures to the separated layers, while comparing how the result depends on whether we apply SVD normalization or not. We have already seen (see Fig 12) how the range of values of the weights change noticeably when the SVD correction is applied. Now we will check if this translates into a difference in the measurements of the graphs.

In Fig 14, Fig 15 and Fig 16 we can see boxplots with the local graph measures and a comparison between case a) without normalization and case b) with SVD correction. Diameter and average path length are related and approximately inversely proportional to efficiency so this measures trends will be approximately correlated (see section XXX and appendix XXX).

In the case of **FA layer,** images for the case without SVD, show a decrease in average path length and diameter and an increase of efficiency for PwMS, We can appreciate also a clear decrease in global strength, modularity shows an slightly increase, while transitivity and density show no differences between groups. If we compare this sets of boxplots with the case where we apply SVD, we can see how differences smooth out.

For the **GM layer** (Fig 15) we see that there are no clear differences between groups in either case. Here we are anticipating what will be a trend along this project, GM layer seem to provide no discriminating information.

In the plots corresponding to the functional layer (fig 16) we see that its behavior with SVD is the opposite of the FA layer. In this case the correction process increases the differences between groups and makes them more evident. Another really interesting differences between FA layer and fMRI layer is that we find opposite situation in almost all measures, if in FA layer we see an increase for PwMS in fMRI it turns out to be a decrease.

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamenteGráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Fig 15. Boxplot graph global measures. FA Layers. With and without SVD.

In a) we see how the transitivity and density measures show no differences between groups, modularity only minimal differences. The rest of the measures do, at least visually, exhibit differences. On the other hand, in b) we see how these differences are smoothed or even disappear, especially in the case of efficiency and mean path length. Note how in the case of strength, the boxes visually show differences but the mean is basically the same.

**FA LAYER**

1. SVD NORMALIZATION
2. WITHOUT SVD NORMALIZATION

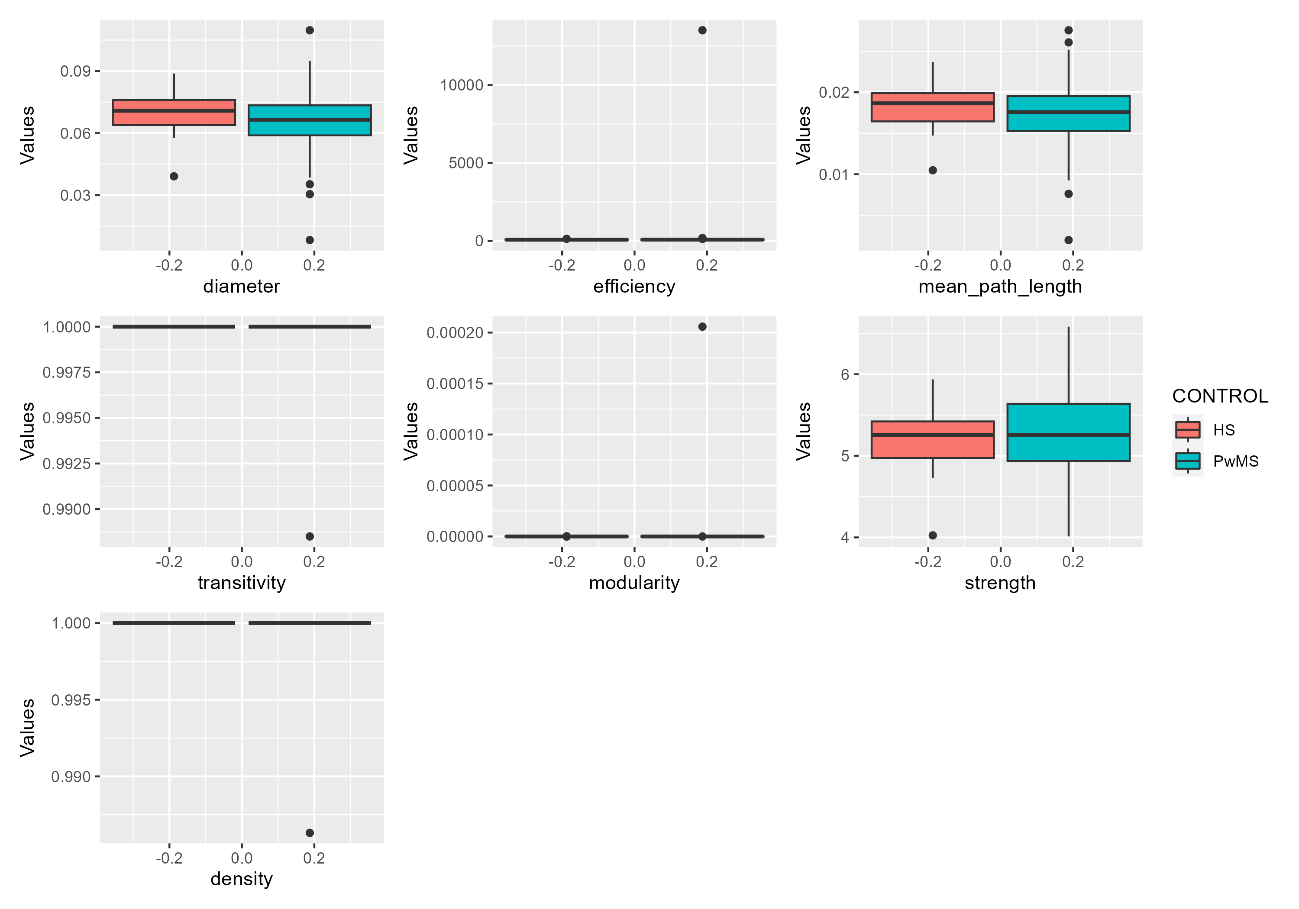
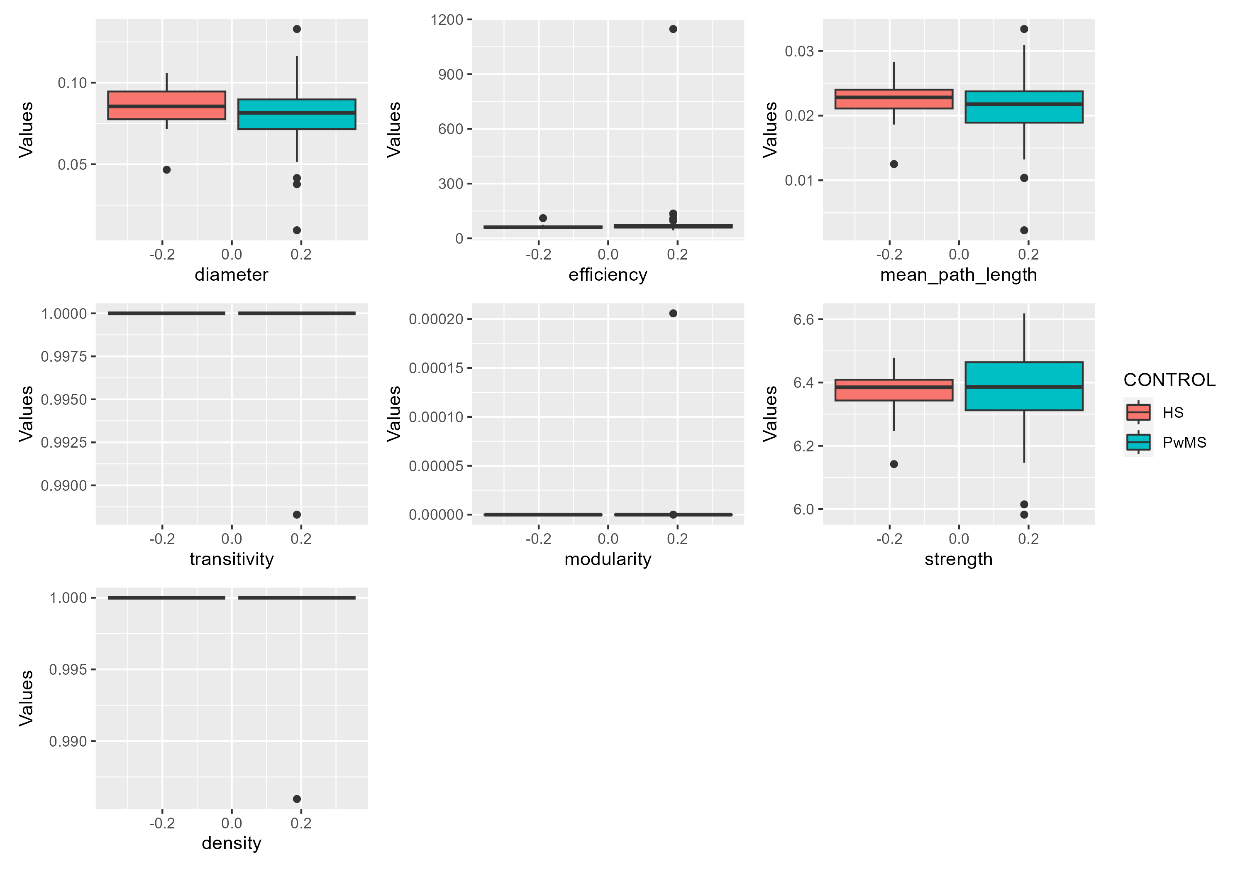


Fig 16. Boxplot graph global measures. GM Layers. With and without SVD.

In the case of GM layers we see no differences between groups in either case.

**GM LAYER**

1. SVD NORMALIZATION
2. WITHOUT SVD NORMALIZATION

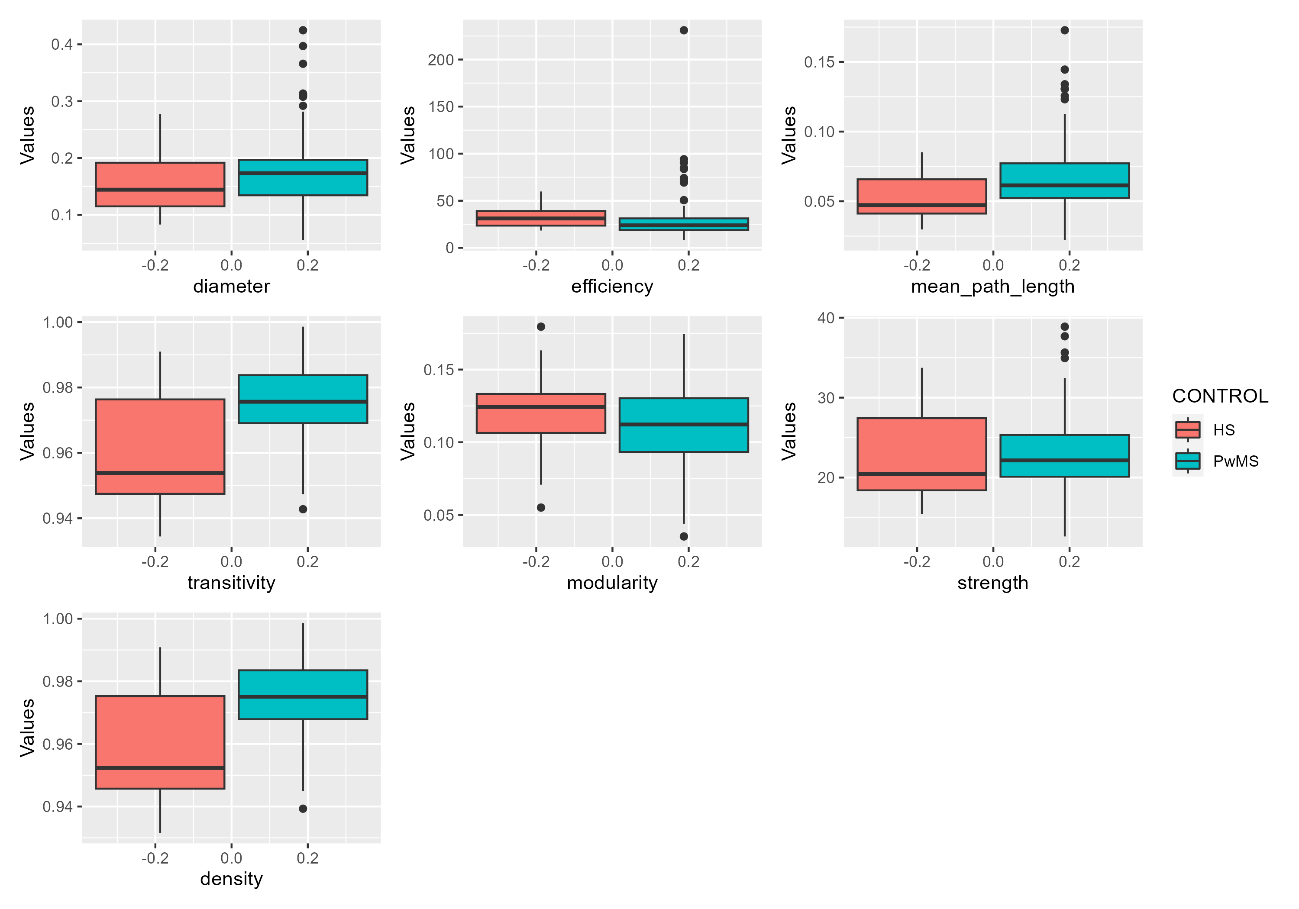
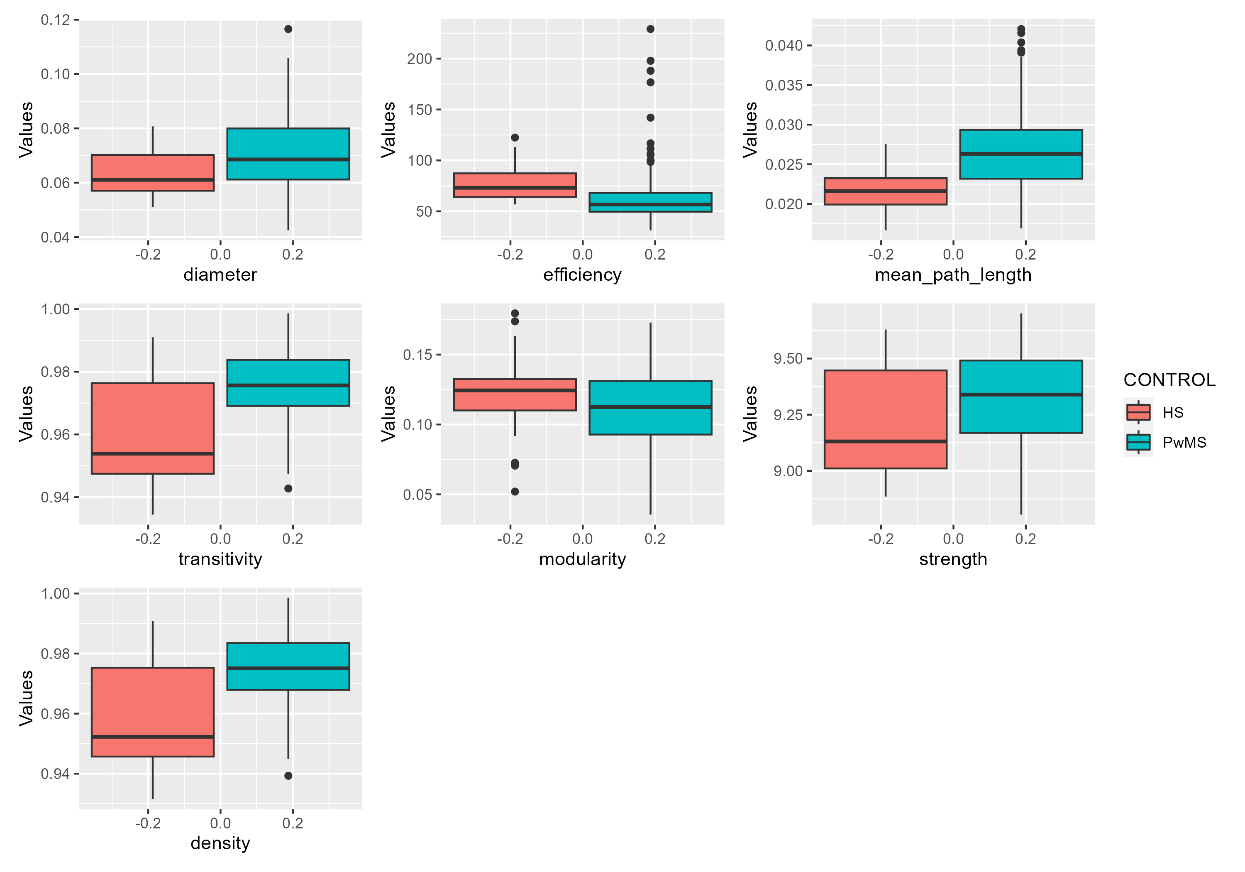


Fig 17. Boxplot graph global measures. fMRI Layers. With and without SVD.

In this case we can see how the opposite happens to the FA layer, the differences between groups are greater in b), especially in the case of efficiency, mean path length and strength.

**fMRI LAYER**

1. SVD NORMALIZATION
2. WITHOUT SVD NORMALIZATION

#### Statistical differences between global measures

Once we have conducted statistical test on our global measures we obtain results shown in table xxx (we only show measures with p < 0.05 for brevity reasons).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **WITHOUT SVD NOMALIZATION** | | | | |
| LAYERS | MEASURE | HS | PwMS | p value |
| FA | diameter | 0.82 ± 0.09 | 0.72 ± 0.1 | 0.000 |
| efficiency | 2.56 ± 0.46 | 2.81 ± 0.46 | 0.009 |
| mean path length | 0.44 ± 0.07 | 0.4 ± 0.06 | 0.007 |
| strength | 23.06 ± 1.13 | 21.42 ± 1.9 | 0.000 |
| fMRI | efficiency | 34.12 ± 13.85 | 28.64 ± 22.36 | 0.018 |
| mean path length | 0.05 ± 0.02 | 0.07 ± 0.02 | 0.011 |
| transitivity | 0.96 ± 0.02 | 0.98 ± 0.01 | 0.001 |
| density | 0.96 ± 0.02 | 0.98 ± 0.01 | 0.001 |

Table 4. Single layers results of statistical test. No SVD normalization

Continuous variables are given as the mean ± standard deviation. Colored measures are the ones present in both cases, Table 4 and 5, with and without SVD.

It is noteworthy that, in the case without normalization, we have four statistically significant measures each from the white matter layer (FA) and the functional layer (fMRI).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SVD NORMALIZATION** | | | | |
| LAYERS | MEASURE | HS | PwMS | p value |
| FA | diameter | 0.33 ± 0.03 | 0.31 ± 0.04 | 0.041 |
| fMRI | diameter | 0.06 ± 0.01 | 0.07 ± 0.01 | 0.015 |
| efficiency | 77.67 ± 19.56 | 63.61 ± 28.36 | 0.000 |
| mean path length | 0.02 ± 0.00 | 0.03 ± 0.00 | 0.000 |
| transitivity | 0.96 ± 0.02 | 0.98 ± 0.01 | 0.001 |
| density | 0.96 ± 0.02 | 0.98 ± 0.01 | 0.001 |

Table 5 Single layers results of statistical test. SVD normalization

Continuous variables are given as the mean ± standard deviation. Colored measures are the ones present in both cases, Table 4 and 5, with and without SVD.

However, upon applying normalization, while the functional layer (fMRI) yields one additional significant measure, the FA layer is left with just one. Interestingly, in both scenarios, we do not observe any significant measures from the grey matter (GM) layer

It is also important to point out that density yield distinctive results in the case of the functional layer.

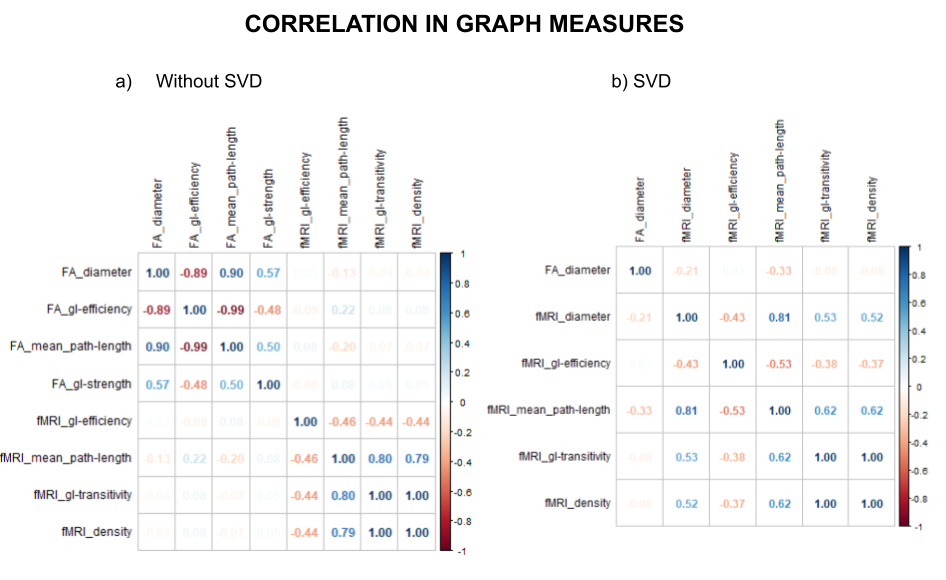
It is also interesting to examine the correlations between graph measures and compare the difference between both cases. In the case involving SVD, as depicted in Figure 3, the diameter, efficiency, and mean path length in the WM layer display a strong correlation, either positive or negative, as predicted. In contrast, efficiency and mean path length in the functional layer do not exhibit significant correlation. Simultaneously, for both scenarios—SVD and non-SVD—the functional layer's transitivity and density display an extremely high correlation, so much so that they appear as 1.00 in the figure

Fig 18. Correllations between global measures. Separate Layers. With and without SVD correction.

After conducting statistical tests with Bonferroni correction for distinctions among MS types, the number of measures that pass the test shrinks to just 2 and 1, for cases with and without SVD, respectively. The limited number of available PPMS and SPMS participants must be taken into consideration and could explain why there are so few measures

|  |  |  |
| --- | --- | --- |
| LAYERS | Without SVD | With SVD |
| FA | Diameter | Diameter |
| Efficiency | ----- |

Table 6. Global single layer measures that pass test 2.

For the machine learning models in this study, the decision was made to incorporate all measures that successfully passed the statistical tests for group differentiation.

## Local single layer graph measures, with and without SVD correction

As it is explained in section xxx 3 local measures will be made, this means 3 measures x 76 nodes x 3 layers for a total of 684.

#### Statistical differences between local measures

In this case the following results are obtained

|  |  |  |
| --- | --- | --- |
| 684 total measures | **Without SVD** | **With SVD** |
| # measures | # measures |
| HS vs PwMS | 326 | 191 |
| Between MS types  (Bonferroni correction) | 89 | 9 |

Table 7. Local single layer measures that pass statitistical tests

The most notable aspect of these results lies in the considerable disparity between the outcomes with and without the application of SVD. As discussed in Section XXX, applying this correction tends to even out differences in weight distribution. Therefore, it seems a logical consequence that fewer measures will be statistically different as range of weights is reduced.

Results for each case after Bonferroni examined more closely in the following tables. Main measures and layer for the case without SVD are shown in table XXX

|  |  |  |
| --- | --- | --- |
| **WITHOUT SVD** | | |
| LAYERS | MEASURE | TOTAL NODES |
| FA | Closeness centrality | 28 |
| strength | 56 |
| GM | Strength | 1 |
| fMRI | Closeness centrality | 4 |

Table 8. Local single layer measures and number of nodes that pass statistical test without SVD.

The following table shows the nodes that have statistically significant measures in more than one layer. Only the right hemisphere’s paracentral is statistically significant in all layers

|  |  |
| --- | --- |
| ROI | # of layers |
| ctx-rh-paracentral | 3 |
| ctx-lh-caudalmiddlefrontal | 2 |
| ctx-lh-entorhinal | 2 |
| ctx-lh-fusiform | 2 |
| ctx-lh-inferiorparietal | 2 |
| ctx-lh-isthmuscingulate | 2 |
| ctx-lh-lingual | 2 |
| ctx-lh-middletemporal | 2 |
| ctx-lh-paracentral | 2 |
| ctx-lh-postcentral | 2 |
| ctx-lh-precentral | 2 |
| ctx-lh-rostralmiddlefrontal | 2 |
| ctx-rh-caudalanteriorcingulate | 2 |
| ctx-rh-cuneus | 2 |
| ctx-rh-entorhinal | 2 |
| ctx-rh-fusiform | 2 |
| ctx-rh-lateraloccipital | 2 |
| ctx-rh-parstriangularis | 2 |
| ctx-rh-precentral | 2 |
| ctx-rh-superiorfrontal | 2 |
| ctx-rh-supramarginal | 2 |
| Left-Pallidum | 2 |
| Right-Amygdala | 2 |

Table 9. ROIs that pass both statistical tests in more than one layer.

Next table shows results for the case where SVD is applied

|  |  |  |
| --- | --- | --- |
| **WITH SVD** | | |
| LAYERS | MEASURE | NODE |
| FA | Closeness centrality | lh-isthmuscingulate |
| Left-Hippocampus |
| ctx-rh-caudalanteriorcingulate |
| ctx-rh-lateralorbitofrontal |
| ctx-rh-precentral |
| strength | Right-Caudate |
| fMRI | Closeness centrality | ctx-rh-entorhinal |
| ctx-rh-parahippocampal |
| ctx-rh-paracentral |

Table 10. Local single layer measures and number of nodes that pass statistical test witht SVD

Again it should be noted that only one measurement from GM layer shows a statistically significant differences.

## Results for local multilayer measurements

In light of the aforementioned results, where the GM layer does not yield significant results, two cases were considered: One multilayer network with all three layers, another multilayer network with only two layers (FA and fMRI)

The first decision to be made concerns the weight to be assigned to inter-layer links. As previously mentioned (Mandke) uses the average of all weights. A close examination of these averages show that they are quite similar, moving within a tight range, as can be seen in fig XXX. The range is approximately 0.0125 and 0.025 for 3 and 2 layers respectively, moreover range does not seem different for HS and PwMS. Although the range is larger for the latter group as it includes more subjects

Gráfico, Histograma

Descripción generada automáticamente

Fig 19. Distribution of interlayer weights in case of averaging.

Following these findings, another option was considered, selecting the interlayer weight that maximizes the number of measures that passes the statistical tests. A range of weights from 0.1 to 1 was evaluated, as presented in table xxx

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 3 LAYERS | | 2 LAYERS | |
| interlayer weights | Statistical test | Bonferroni | Statistical test | Bonferroni |
| 0.1 | 28 | 3 | 52 | 2 |
| 0.2 | 29 | 4 | 52 | 2 |
| 0.3 | 29 | 4 | 52 | 2 |
| 0.4 | 29 | 4 | 52 | 2 |
| 0.5 | 29 | 4 | 52 | 2 |
| 0.6 | 31 | 4 | 52 | 2 |
| 0.7 | 31 | 4 | 52 | 2 |
| 0.8 | 31 | 4 | 52 | 2 |
| 0.9 | 31 | 4 | 52 | 2 |
| 1.0 | 31 | 4 | 52 | 2 |

Table 11. # of measures that pass statistical test when varying interlayer weights

In the scenario with two layers, the variation in interlayer weights makes no difference, and somewhat surprisingly, although more measures pass Test 1 compared to the three-layer case, fewer measures pass Test 2. When considering three layers, minor differences emerge between weights, indicating that an increase in weight results in an increased number of measures. Ultimately, an interlayer weight of 1 was selected.

The following table provides brief summary of those measures that passed the initial statistical test in either case:

|  |  |  |
| --- | --- | --- |
|  | **2 layers** | **3 layers** |
| # measures | # measures |
| MultiPageRank | 7 | 15 |
| Sum Multi-Strength | 8 | 9 |
| Closeness centrality | 37 | 7 |

Table 12. Summary of multi-layer measures that passed statistical test 1

There is an important difference in closeness centrality between both cases.

After test with Bonferroni correction we obtain the following results:

|  |  |  |
| --- | --- | --- |
|  | **2 layers** | **3 layers** |
| # Nodes | # Nodes |
| MultiPageRank |  | ctx-lh-precentral |
|  | Left Accumbens area |
|  | Ctx-rh-lateraloccipital |
| Sum Multi-Strength |  | Ctx-rh-superiorfrontal |
| Closeness centrality | Ctx-lh-medialorbitofrontal |  |
| Ctx-lh-parsorbitalis |  |

Table 13. Summary of multi-layer measures that passed statistical test 2

Following the second statistical test, it is worth mentioning that statistical significant measures and nodes differ between the two scenarios.

Upon comparing all measures from single layer (with and without SVD) and multilayer (with 2 and 3 layers), the node 'ctx-lh-precentral' stands out as the most frequently appearing.

While the GM layer might not seem to contribute significantly, its presence still influences the results of multilayer measurements

## Summary of data for machine learning models

For the machine learning models 3 cases where considered: single layer, multilayer with 3 layers (3-multilayer), and multilayer with 2 layers (2-multilayer). Final data was selected in the following manner:

* Single layer: The data from measurements **without SVD correction** was considered. For local measures only those that passed the 2nd test were considered, while global ones were selected from those that passed the first test.
* Multilayer: From the single layer measurements only data with SVD correction was considered, following the same criterion as before for global and local measurements. From the multilayer measures all those that passes first statistical test were selected.

The motivation to conduct these tests is to reduce the sheer number of features to those that may yield to better results. The reason for including features that passed first and/or second statistical test lies in the quantity of remaining features.

This leaves us with the following number of features

1. Separate Layers: 97 features
2. Mulitlayer 3 layers: 53 features
3. Multilayer 2 layers: 74 features

### Removal of correlated features

Machine learning models perform better when features are not correlated. Given the substantial number of features in each case, there is a strong likelihood that some may indeed be correlated.

A threshold of ±0.8 is established and measures correlating above (or below) this threshold are removed. The numbers of measures taken out are shown in the table below

|  |  |
| --- | --- |
| CASE | Features removed |
| Separate layers | 86 |
| 3 Multiplex | 9 |
| 2 Multiplex | 10 |

Table 14. Features removed due to high correlation

The influence of the SVD correction on the results is evident in the count of correlated features that are eliminated. Barely any measure is removed in scenarios where the correction has been applied.

Though this is a necessary step for model construction, it is important to note that correlations may well vary with another dataset and potentially lead to alternate features. Moreover, this approach can complicate model interpretation. When a feature is identified as important in the final model, it raises a question: Is it always this feature that holds importance, or it could be one of its highly correlated counterparts?

## Machine learning model results HS vs PwMS

Subsequent table shows results for separate layers, 3-multilayer and 2-multilayer, green background color highlights best model for that metric and scenario.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Separate layers** | | | | | | | | | | | |
|  | |  |  | |  | | |  | | **# Errors out of 33** | |
|  | **accuracy** | | | **precision** | | **recall** | **F1 score** | | **FP** | | **FN** |
| **RF** | **0.901** | | | **0.943** | | **0.949** | **0.945** | | **0** | | **3** |
| **SVC** | **0.871** | | | **0.878** | | **0.991** | **0.931** | | **2** | | **0** |
| **KNN** | **0.818** | | | **0.953** | | **0.836** | **0.888** | | **0** | | **4** |
| **XGBoost** | **0.894** | | | **0.943** | | **0.940** | **0.940** | | **0** | | **2** |

Table 15. Results from machine learning models HS vs PwMS. Separate layers.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **3- multilayer** | | | | | | |
|  | **accuracy** | **precision** | **recall** | **F1 score** | **# Errors out of 33** | |
| **FP** | **FN** |
| **RF** | **0.894** | **0.926** | **0.957** | **0.940** | **1** | **1** |
| **SVC** | **0.872** | **0.878** | **0.992** | **0.931** | **2** | **0** |
| **KNN** | **0.749** | **0.953** | **0.749** | **0.839** | **1** | **8** |
| **XGBoost** | **0.901** | **0.949** | **0.939** | **0.944** | **0** | **2** |

Table 16. Results from machine learning models HS vs PwMS. 3-multilayer.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **2-multilayer** | | | | | | |
|  | **accuracy** | **precision** | **recall** | **F1 score** | **# Errors out of 33** | |
| **FP** | **FN** |
| **RF** | **0.856** | **0.894** | **0.949** | **0.920** | **2** | **3** |
| **SVC** | **0.879** | **0.879** | **1.000** | **0.936** | **2** | **0** |
| **KNN** | **0.787** | **0.924** | **0.827** | **0.871** | **2** | **8** |
| **XGBoost** | **0.841** | **0.907** | **0.914** | **0.910** | **1** | **2** |

Table 17. . Results from machine learning models HS vs PwMS. 2-multilayer.

With some exceptions, results exhibit a pattern that is a model that perform better in one metric, tend to perform better in that metric in all scenarios, this could mean that one must choose one model over another depending on the target. For instance, if goal is to minimize false negatives, clearly SVC is a better choice and KNN is the worst

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **BEST METRIC RESULT FOR EACH MODEL** | | | | | | |
|  | **Accuracy** | **Precision** | **Recall** | **F1 score** | **FP** | **FN** |
| **RF** | **2-multi** | **Sep-layers** | **3-multi** | **Sep-layers** | **Sep-layers** | **3-multi** |
| **SVC** | **2-multi** | **2-multi** | **2-multi** | **2-multi** | **ALL** | **ALL** |
| **KNN** | **Sep-layers** | **Sep/3-multi** | **3-multi** | **3-multi** | **Sep-layers** | **Sep-layers** |
| **XGBoost** | **3-multi** | **3-multi** | **Sep-layers** | **3-multi** | **Sep/3-multi** | **ALL** |

Table 18. Summary of best network for each metric. HS vs PwMS

On each cell is displayed the network that performs better on that metric; 2-multi for 2-multilayer, 3-multi for 3-multilayer and Sep-layers for separate layers. Sep/3-multilayers indicate that both, separate-layer and 3-multilayer performance is equally good. ALL indicates that 3 layers performs equally well.

Previous table shows which scenario performs better on each model. Differences between metrics is tiny, 0.001 in some case. It seems no scenario performs better than any other across all models. However it must be noted that SVC has all its better metrics in 2-multilayer (with the caveat that differences are minimal)

## Machine learning model results for MS types classification

For classification between MS types (RRMS, SPMS, PPMS), only PwMS participants were selected, as in most diagnostic situations, general diagnosis (MS) comes first, and then diving into details is possible.

As has already been discussed limitation in the number of patients with PPMS (6), potentially limits results of this work and generalization.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Separate layers** | | | | | |
|  | **accuracy** | **precision** | **recall** | **F1 score** | **# Errors out of 30** |
| **RF** | **0.855** | **0.804** | **0.855** | **0.823** | **6** |
| **SVC** | **0.855** | **0.744** | **0.855** | **0.795** | **6** |
| **KNN** | **0.727** | **0.818** | **0.727** | **0.758** | **9** |
| **XGBoost** | **0.846** | **0.803** | **0.846** | **0.819** | **5** |

Table 19. Results from machine learning model. MS types. Separate layers

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **3- multilayer** | | | | | |
|  | **accuracy** | **precision** | **recall** | **F1 score** | **# Errors out of 30** |
| **RF** | **0.863** | **0.824** | **0.863** | **0.838** | **5** |
| **SVC** | **0.863** | **0.745** | **0.863** | **0.800** | **6** |
| **KNN** | **0.693** | **0.829** | **0.693** | **0.739** | **8** |
| **XGBoost** | **0.803** | **0.764** | **0.803** | **0.783** | **3** |

Table 20.Results from machine learning model. MS types. 3-multilayer

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2-multilayer** | | | | | |
|  | **accuracy** | **precision** | **recall** | **F1 score** | **# Errors out of 30** |
| **RF** | **0.855** | **0.804** | **0.855** | **0.823** | **6** |
| **SVC** | **0.854** | **0.744** | **0.855** | **0.796** | **6** |
| **KNN** | **0.727** | **0.818** | **0.727** | **0.759** | **9** |
| **XGBoost** | **0.846** | **0.803** | **0.846** | **0.819** | **5** |

Table 21. Results from machine learning model. MS types. 2-multilayer

A surprising finding is that XGBoost does not perform better in any metric, but is the model that commit less mistakes in all scenarios. Again we find that models that perform better in one metric in one case, does so in all scenarios.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **BEST METRIC RESULT FOR EACH MODEL** | | | | | |
|  | **Accuracy** | **Precision** | **Recall** | **F1 score** | **FP** |
| **RF** | **3-multi** | **3-multi** | **3-multi** | **3-multi** | **3-multi** |
| **SVC** | **3-multi** | **3-multi** | **3-multi** | **3-multi** | **ALL** |
| **KNN** | **Sep/2-multi** | **3-multi** | **Sep/2-multi** | **2-multi** | **3-multi** |
| **XGBoost** | **Sep/2-multi** | **Sep/2-multi** | **Sep/2-multi** | **Sep/2-multi** | **3-multi** |

Table 22. Summary of best network for each metric. MS types

On each cell is displayed the network that performs better on that metric; 2-multi for 2-multilayer, 3-multi for 3-multilayer and Sep-layers for separate layers. Sep/2-multil indicate that both, separate-layer and 2-multilayer performance is equally good. ALL indicates that 3 layers performs equally well.

In multiclass classification, though differences are still minimal, 3-multilayer performs better in the metrics of 2 out of 4 modes, and makes fewer errors in all models. This strongly suggest that this is the path to follow and that GM layer though it does not yield to most relevant measures in single layer, it plays a role in the overall multilayer network.

## Feature importance in HS vs PwMS

Following figures show feature importance for Random Forest and XGBoost for separate layers and 2-multilayer and 3-multilayer

**Keys to read the figures.**

To interpret the figures, the following conventions should be understood as they are consistent across all figures:

1. Features that show an average importance of zero have been removed
2. Global measures of single layers, follow this convention “Layer\_measure”. For instance, “FA\_diameter” refers to the diameter of FA layer or “fMRI\_gl\_efficiency” for global efficiency of fMRI layer
3. Local measures follow this convention: “Measure\_ROI”, for instance “MultiPRCent\_ctx-lh\_inferiorparietal”, where abbreviations indicate the following:

|  |  |
| --- | --- |
| Abbreviation | Measure |
| MultiPRCent | MultiPageRank Centrality |
| MultiPath | Closeness centrality for multilayers |
| MultiSTSum | Strength sum in multilayers |
| loc-clos-cent | Local closeness centrality (single layers) |
| loc-strength | Strength |

|  |  |
| --- | --- |
| Abbreviation in ROI | meaning |
| lh | Left hemisphere |
| rh | Right hemisphere |
| ctx | cortex |

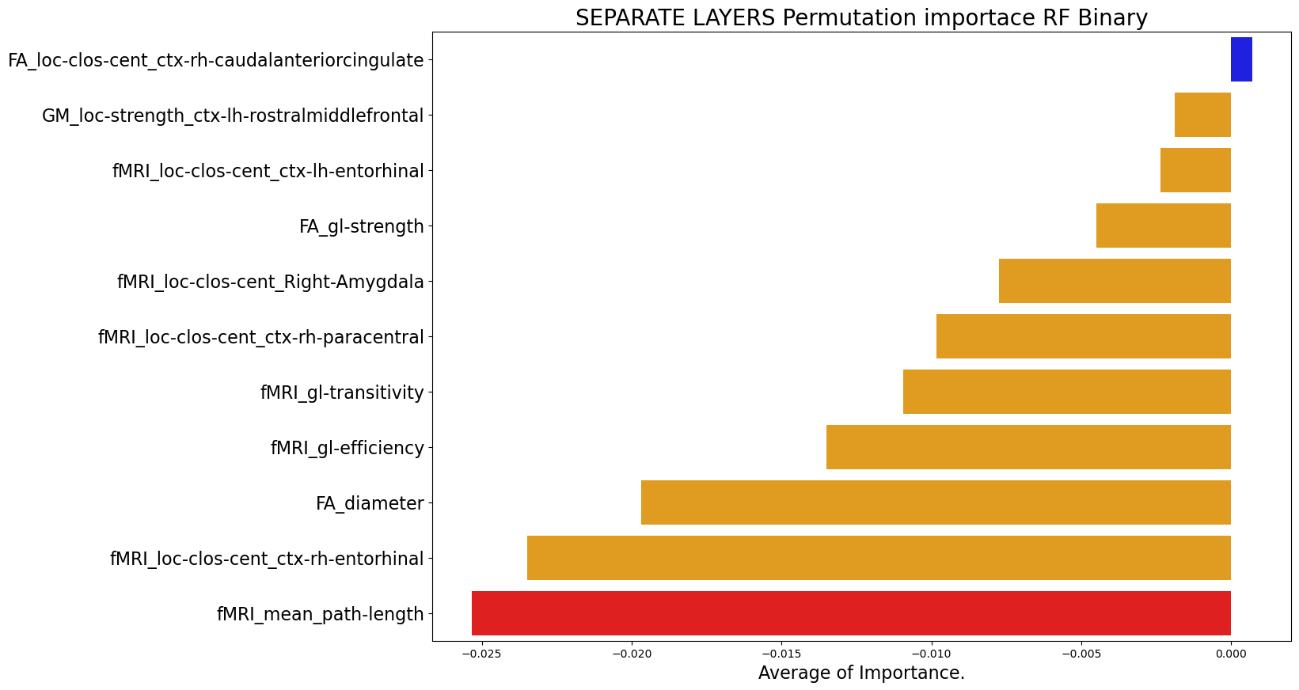


Fig 20. Permutation Importance for Random Forest. Separate Layers

Gráfico, Gráfico de embudo

Descripción generada automáticamente

Fig 21. Permutation Importance for XGBoost. Separate Layers

Figures displays a curious pattern. In RF all features except one have a negative impact on the model (upper image), whereas all features except one have a positive impact on XGBoost model (bottom image). Positive feature in RF is closeness centrality for ctx-rh-cuadalanteriorcingulate which is also positive for XGBoost, and negative feature in both models is global strength for FA layer Blue color indicates most positive impact and red color the most negative.

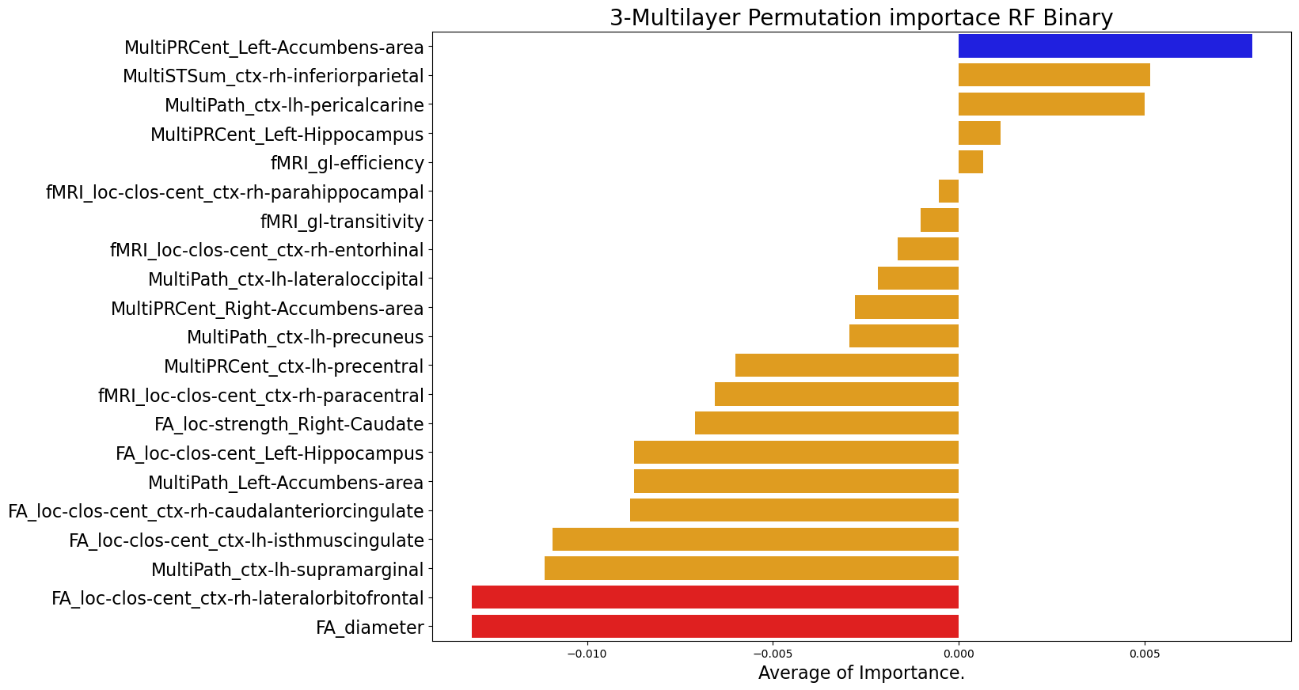


Fig 22. Permutation Importance for Random Forest. 3-multilayer

Gráfico, Gráfico de barras

Descripción generada automáticamente

Fig 23. Permutation Importance for XGBoost. 3-multilayer

In this case, the figure shows how RF (upper image) and XGBoost (bottom image) display more or less the same number of feature over and below zero. MultiPageRank closeness centrality for Left Accumbens is the feature with higher impact in both models, whereas among the most negative both have multilayer closeness centrality for lh-supramarginal ROI. Blue color in plot indicate most positive feature and red one, the two most negative ones.

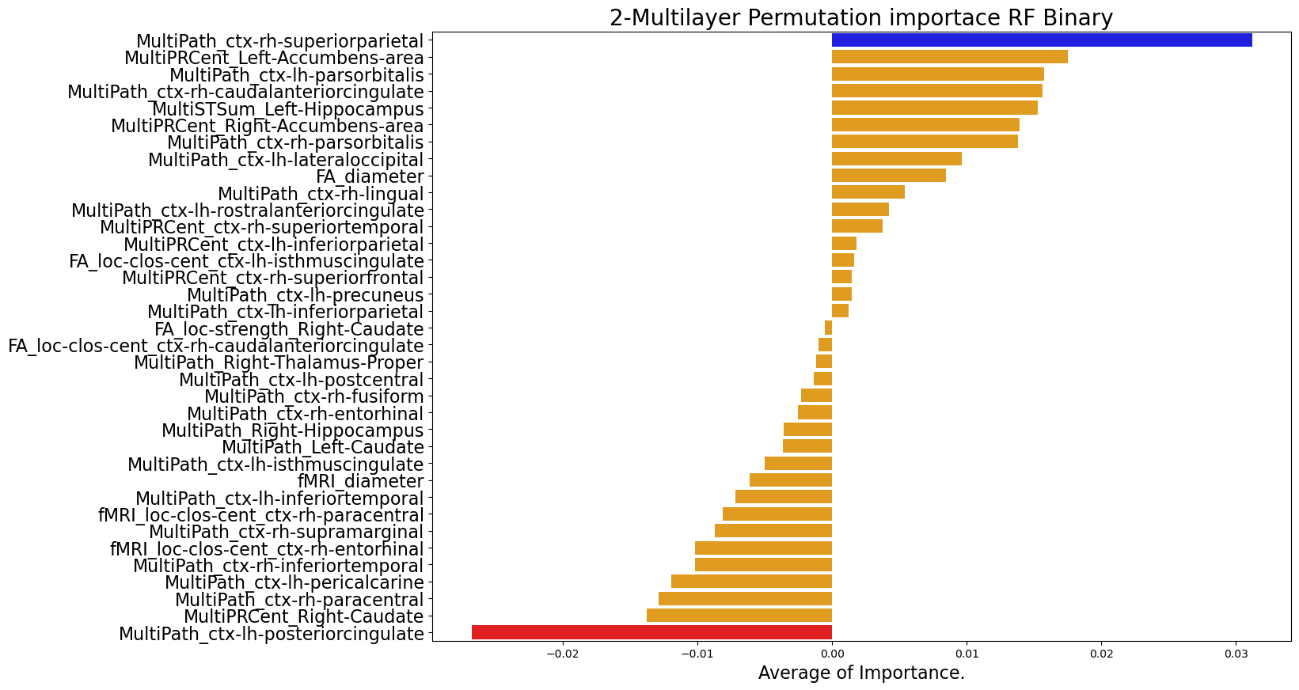


Fig 24. Permutation Importance for Random Forest. 2-multilayer.

Gráfico, Gráfico de embudo

Descripción generada automáticamente

Fig 25. Permutation Importance for XGBoost. 2-multilayer

For RF (upper image) we have a lot of features that either show positive or negative impact, while XGBoost, (bottom image) show only one feature below zero (closeness centrality for multilayers in lh-posteriorcingulate) which is also the most negative in RF. On the other hand 3 most positive ones in one model are not among the top three in the other model. Although if more features (5 or 6) are considered, Right Accumbens MultiPageRank centrality and lh-parsobitalis for multilayer closeness centrality are found in both. Plots show in blue color most positive feature and in red most negative.

While some common features above or below zero with similar importance do appear between models within the same scenario, no consistent pattern emerges across different scenarios or even within the same case. Especially because some models show only one positive or negative feature reducing and, in fact, eliminating the possibility of identifying a common node and measure. Nonetheless, a summary across all scenarios and models reveals that certain ROIs and measures do appear more frequently than others. The subsequent tables provide a summary of these frequently occurring ROIs and measures in both cases - values above zero and below zero:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **POSITIVE IMPORTANCE (ABOVE ZERO)** | | | | |
| ROI | # of times present | SCENARIOs | MODELS | MEASURES |
| ctx-rh-caudalanteriorcingulate | 4 | Separate layers | RF, XGB | Closeness centrality FA layer (3) |
| 2-multilayer | RF, XGB | Multilayer closeness centrality (1) |
| Left Accumbens Area | 4 | 2-multilayer | RF, XGB | MultiPageRank centrality (4) |
| 3-multilayer | RF, XGB |
| ctx-rh-entorhinal | 4 | Separate layers | XGB | Closeness centrality fMRI layer (3)  Multilayer closeness centrality(1) |
| 2-multilayer | XGB |
| 3-multilayer | RF, XGB |
| ctx-rh-paracentral | 3 | Separate layers | XGB | Closeness centrality fMRI layer (2)  Multilayer closeness centrality (1) |
| 2-multilayer | XGB |
| ctx-rh-superiorparietal | 3 | 2-multilayer | RF | Multilayer closeness centrality(3) |
|  |  | 3.multilayer | RF, XGB |

Table 23. More frequent ROIs with positive importance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **NEGATIVE IMPORTANCE (BELOW ZERO)** | | | | |
| ROI | # of times present | SCENARIOs | MODELS | MEASURES |
| Right Caudate | 6 | 2-multilayer | RF | Local strength Fa layer (3) |
| 3-multilayer | RF, XGB | MultiPageRank centrality (3) |
| ctx-rh-paracentral | 5 | Separate layers | RF | Closeness centrality fMRI layer (4)  Multilayer closeness centrality (1) |
| 2-multilayer | RF |
| 3-multilayer | RF, XGB |
| ctx-lh-pericalcarine | 3 | 2-multilayer | RF | Multilayer closeness centrality(3) |
| 3-multilayer | RF, XGB |
| ctx-rh-entorhinal | 3 | Separate layers | RF | Closeness centrality fMRI layer (2)  Multilayer closeness centrality (1) |
| 2-multilayer | RF |

Table 24. More frequent ROIs with negative importance

It is of no surprise that 2-multilayer appears more frequently than the rest, given that it shows more measures above and below zero, see fig 21xxx (it also has more available features table xxx). Although, as it has been discussed, there are no nodes that consistently appear in all models, scenarios and measures, it is indeed the case that some can be regarded as more important to differentiate between HS and PwMS. Table XXX show how cortex, right hemisphere caudalanteriorcingulate left Accumbens area show importance above zero, while right caudate and cortex, left hemisphere paracentral zone show importance below zero. It it noteworthy how cortex right hemisphere entorhinal zone seems to play opposite rols depending on the case (see both tables XXX)

# Conclusions and future work

The final conclusion is divided into two section: Conclusions related to the graph construction, and conclusions derived from the machine learning models. However, there are not the only areas of focus in this discussion.

## Conclusions related to graph construction

#### SVD correction and single-layer results

This study presented the opportunity to compare two approaches to single layer measures, one applying an SVD correction and another without. The SVD correction reduces the range of values of the graph weights, which appears to dilute differences among graphs (from the perspective of the measures applied). This is evidenced by the fact fewer *local* measures passes both statistical tests, 9 with SVD vs 89 without. Conversely, the measures with SVD normalization that pass the tests show no significant correlation, while 86 out of 97 single layer measures, both global and local, show high correlation in the case without SVD.

This observation that SVD reduces the range of weights and statistically significant measures could potentially be interpreted as a reduction of noise or bias as suggested by Mandke et al. 2018. Whether this approach should be used when considering three separate layers, or even just a single layer, remains a question for future research.

#### Inter-layer weights

Another important conclusion is the lack of clear criteria for selecting interlayer weights. Indeed, setting these values presents a challenge due to the distinct biological attributes each layer holds, making it difficult to propose a theoretical approach.

Considering the primary objective of this project, which is to differentiate between participant classes, it make sense to select values that maximize features passing statistical tests, thereby providing more features available for machine learning models. However, it has been observed that changes in weights (ranging from 0.1 to 1) barely affect the number of features passing one or both statistical tests. It may be beneficial to extend the range of values beyond 1, or developing a theoretical framework to establish this value.

#### GM layer

The finding that only one measurement out of 235 from the GM layer passed the statistical tests raises the question of whether this layer provides information. This realization emerged during the project’s development, led to the consideration of a multilayer with only FA and fMRI layers (2-multilayer).

A retrospective analysis reveals 2 conflicting tendencies. On one hand 3-multilayer models seem to outperform 2-multilayer models in distinguishing between MS types suggesting that this layer indeed plays a role in the global multiplex network. On the other hand, in the analysis of feature importance, 2-multilayer show more features with average importance above and below zero, possibly as a consequence of having more available features after the statistical tests. The former is more important for model construction, while the latter helps in identifying potentially clinical important ROIs.

#### Multilayer results

The difference between the 2-multilayer and 3-multilayer scenarios is not as significant as it is in the single-layer measurements. In the 2-multilayer case, 52 measures pass the first statistical test and 2 measures pass the second, while in the 3-multilayer scenario, 31 and 4 measures pass the first and second statistical tests respectively (see table XXX). When examining correlations, a similar number of features (10 and 9) are removed due to high correlation in both scenarios.

From the point of view of the graph construction, it does not appear to be an objective reason to prefer one network over the other.

## Conclusions from machine learning models

#### Shortage of data

One of the major challenges encountered during this project is the limited size of the dataset, with only 165 participants and particularly low representation of certain classes, such as PPMS (6). Although data augmentation technique like SMOTE have been used, it cannot really solve the problem only mitigates it.

Data augmentation techniques “artificially” inflates minority classes by creating new samples from the existing ones. This could potentially introduce inaccurate new instances. Furthermore, the division of the dataset in into train and test sets reduces the available samples for SMOTE to operate. It also downsize to the minimum the number of samples from the minority class in the test dataset, thus making it difficult to evaluate the model and reducing its validity.

The scarcity of data presents a challenge not only in the construction of machine learning models but also during feature selection, as statistical tests would undoubtedly benefit from a larger dataset.

#### HS vs PwMS

The application of machine learning models to distinguish between HS and PwMS yield mixed results across algorithms and network types. SVM shows a slight advantage as it shows the highest recall and fewest number of false negatives across all networks. SVM also the highest accuracy and F1 score in the 2-multilayer case. However differences in scores are minimal as well as differences in the number of errors between SVM, RF and XGBoost (see tables XXX, XXX, XXX)

Separate layers and 3-multlayer network, both, perform better in 7 metrics across models, when networks are compared (see table XXX), the 2-multilayer scenario performs better in 5 metrics, and all of them or two of them are matched in a few others. Despite these observations, the differences in values are minimal, and therefore, it remains challenging to conclusively determine which network is superior.

#### RRMS vs SPMS vs PPMS

When it comes to differentiating between types of MS, SVM and Random Forest models generally outperform other models in most cases. However, somewhat surprisingly, XGBoost, despite having slightly lower scores, makes the fewest errors. This counterintuitive result could be attributed to the minimal differences in the scores between these models (tables XXX, XXX, XXX)

Upon analyzing the performance of the network across models, 3-multilayer completely outperforms the other two networks in two models (RF and SVM) while separate layers and 2-multilayer are better in XGBoost. KNN model presents mixed results, with no clear superior network (Table XXX)

#### Feature importance between HS and PwMS

Probably as a consequence of all mixed results there is no single ROI or measure that completely stands out when comparing their importance in the different models. However, upon closer examination, certain ROIs like the right hemisphere’s caudalanteriorcingulate and entorhinal and left Accumbens area seem to slightly exceed others in positive importance. Conversely right caudate and paracentral of the right hemisphere surface more often as negatively important.

Regarding to the measures, the situation is even more mixed, and no single clearly distinguishing itself from the rest.

It is important to recall that a significant number of measures were eliminated from the separate layers network due to high correlation. This makes difficult to infer which nodes could be important as their measures are all related. This may be a clear sign that multilayer networks should be preferred.

## Future work

To enhance the outcomes of the current project it is of paramount importance to increase the number of subjects. Some efforts has been done in this direction with the creation of SUMMIT (Bove et al. 2018) or the use of Virtual Brain for modeling (Martí-Juan et al. 2023).

In relation to the creation and analysis of graphs, additional measures that haven not been considered in this study such as k-coreness (Pontillo et al. 2022) could be incorporated. Additionally, to further optimize the creation of machine learning models, the exploration of different thresholds (Zhao et al. 2020), could prove useful.

# Glossary

* *Expanded Disability Status Scale* (EDSS) is a test used to assess physical disability in MS patients. It ranges from 0, no disability, to 10, deceased.

Voxel?

## List of Abbreviations

Terms in alphabetical order

|  |  |
| --- | --- |
| TERM | ABBREVIATION |
| Clinically Isolated Syndrome | CIS |
| Expanded Disability Status Scale | EDSS |
| Fractional Anisotropy | FA |
| Functional MRI | fMRI |
| Gray Matter | GM |
| Healthy Subject | HS |
| K-Nearest Neighbor | KNN |
| Left hemisphere | lh |
| Multiple Sclerosis | MS |
| Magnetic Resonance Imaging | MRI |
| Patient with MS | PwMS |
| Primary progressive course | PPMS |
| Principal Component Analysis | PCA |
| Region(s) of interest | ROI(s) |
| Relapsing-Remitting Course | RRMS |
| Right Hemisphere | rh |
| Secondary progressive course | SPMS |
| Singular vector decomposition | SVD |
| Support vector classification | SVC |
| Support Vector Machine | SVM |
| Extreme Gradient Boosting | XGBoost, XGB |

# Bibliography

Barabási, Albert-László. 2016. *Network Science*. Cambridge University Press. http://networksciencebook.com/.

Barkhof, Frederik. 2002. “The Clinico-Radiological Paradox in Multiple Sclerosis Revisited.” *Current Opinion in Neurology* 15 (3): 239–45. https://doi.org/10.1097/00019052-200206000-00003.

Bassett, Danielle S., and Olaf Sporns. 2017. “Network Neuroscience.” *Nature Neuroscience* 20 (3): 353–64. https://doi.org/10.1038/nn.4502.

Battiston, Federico, Jeremy Guillon, Mario Chavez, Vito Latora, and Fabrizio De Vico Fallani. 2018. “Multiplex Core–Periphery Organization of the Human Connectome.” *Journal of The Royal Society Interface* 15 (146): 20180514. https://doi.org/10.1098/rsif.2018.0514.

Behdenna, Abdelkader, Julien Haziza, Chloé-Agathe Azencott, and Akpéli Nordor. 2021. “PyComBat, a Python Tool for Batch Effects Correction in High-Throughput Molecular Data Using Empirical Bayes Methods.” bioRxiv. https://doi.org/10.1101/2020.03.17.995431.

Bianconi, Ginestra. 2022. *Multilayer Networks: Structure and Function*. Oxford, New York: Oxford University Press.

Bove, Riley, Tanuja Chitnis, Bruce AC Cree, Mar Tintoré, Yvonne Naegelin, Bernard MJ Uitdehaag, Ludwig Kappos, et al. 2018. “SUMMIT (Serially Unified Multicenter Multiple Sclerosis Investigation): Creating a Repository of Deeply Phenotyped Contemporary Multiple Sclerosis Cohorts.” *Multiple Sclerosis Journal* 24 (11): 1485–98. https://doi.org/10.1177/1352458517726657.

Brownlee, Jason. 2021. *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning*.

Bullmore, Edward, and Olaf Sporns. 2009. “Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems.” *Nature Reviews. Neuroscience* 10 (March): 186–98. https://doi.org/10.1038/nrn2575.

Casas-Roma, Jordi, Eloy Martinez-Heras, Albert Solé-Ribalta, Elisabeth Solana, Elisabet Lopez-Soley, Francesc Vivó, Marcos Diaz-Hurtado, et al. 2022. “Applying Multilayer Analysis to Morphological, Structural, and Functional Brain Networks to Identify Relevant Dysfunction Patterns.” *Network Neuroscience* 6 (3): 916–33. https://doi.org/10.1162/netn\_a\_00258.

Chard, Declan T., Adnan A. S. Alahmadi, Bertrand Audoin, Thalis Charalambous, Christian Enzinger, Hanneke E. Hulst, Maria A. Rocca, et al. 2021. “Mind the Gap: From Neurons to Networks to Outcomes in Multiple Sclerosis.” *Nature Reviews Neurology* 17 (3): 173–84. https://doi.org/10.1038/s41582-020-00439-8.

Chard, Declan, and S Anand Trip. 2017. “Resolving the Clinico-Radiological Paradox in Multiple Sclerosis.” *F1000Research* 6 (October): 1828. https://doi.org/10.12688/f1000research.11932.1.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. “SMOTE: Synthetic Minority Over-Sampling Technique.” *Journal of Artificial Intelligence Research* 16 (June): 321–57. https://doi.org/10.1613/jair.953.

De Domenico, Manlio. 2020. “Multilayer Networks Illustrated,” November. https://doi.org/10.17605/OSF.IO/GY53K.

———. 2022. *Multilayer Networks: Analysis and Visualization: Introduction to MuxViz with R*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-75718-2.

De Domenico, Manlio, Mason A. Porter, and Alex Arenas. 2015. “MuxViz: A Tool for Multilayer Analysis and Visualization of Networks.” *Journal of Complex Networks* 3 (2): 159–76. https://doi.org/10.1093/comnet/cnu038.

Fleischer, Vinzenz, Adriane Gröger, Nabin Koirala, Amgad Droby, Muthuraman Muthuraman, Pierre Kolber, Eva Reuter, Sven Meuth, Frauke Zipp, and Sergiu Groppa. 2016. “Increased Structural White and Grey Matter Network Connectivity Compensates for Functional Decline in Early Multiple Sclerosis.” *Multiple Sclerosis* 23 (May). https://doi.org/10.1177/1352458516651503.

Fleischer, Vinzenz, Angela Radetz, Dumitru Ciolac, Muthuraman Muthuraman, Gabriel Gonzalez-Escamilla, Frauke Zipp, and Sergiu Groppa. 2019. “Graph Theoretical Framework of Brain Networks in Multiple Sclerosis: A Review of Concepts.” *Neuroscience*, Non-invasive MRI windows on brain inflammation, 403 (April): 35–53. https://doi.org/10.1016/j.neuroscience.2017.10.033.

Fornito, Alex, Zalesky, Andrew, and Bullmore, Edward. n.d. *Fundamentals of brain network analysis*. 2016th ed. Elsevier Academic Press.

Fortin, Jean-Philippe, Nicholas Cullen, Yvette I. Sheline, Warren D. Taylor, Irem Aselcioglu, Philip A. Cook, Phil Adams, et al. 2018. “Harmonization of Cortical Thickness Measurements across Scanners and Sites.” *NeuroImage* 167 (February): 104–20. https://doi.org/10.1016/j.neuroimage.2017.11.024.

Fortin, Jean-Philippe, Drew Parker, Birkan Tunç, Takanori Watanabe, Mark A. Elliott, Kosha Ruparel, David R. Roalf, et al. 2017. “Harmonization of Multi-Site Diffusion Tensor Imaging Data.” *NeuroImage* 161 (November): 149–70. https://doi.org/10.1016/j.neuroimage.2017.08.047.

Gironés Roig, Jordi, Jordi Casas Roma, and Julià Minguillón Alfonso. 2017. *Minería de datos: modelos y algoritmos*. Tecnología. Barcelona: Editorial UOC.

Guillon, Jeremy, Mario Chavez, Federico Battiston, Yohan Attal, Valentina La Corte, Michel Thiebaut de Schotten, Bruno Dubois, Denis Schwartz, Olivier Colliot, and Fabrizio De Vico Fallani. 2019. “Disrupted Core-Periphery Structure of Multimodal Brain Networks in Alzheimer’s Disease.” *Network Neuroscience* 3 (2): 635–52. https://doi.org/10.1162/netn\_a\_00087.

Haider, Lukas, Tobias Zrzavy, Simon Hametner, Romana Höftberger, Francesca Bagnato, Günther Grabner, Siegfried Trattnig, Sabine Pfeifenbring, Wolfgang Brück, and Hans Lassmann. 2016. “The Topograpy of Demyelination and Neurodegeneration in the Multiple Sclerosis Brain.” *Brain* 139 (3): 807–15. https://doi.org/10.1093/brain/awv398.

Heuvel, Martijn P. van den, and Olaf Sporns. 2011. “Rich-Club Organization of the Human Connectome.” *Journal of Neuroscience* 31 (44): 15775–86. https://doi.org/10.1523/JNEUROSCI.3539-11.2011.

Ho, Tin Kam. 1995. “Random Decision Forests.” In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1:278–82 vol.1. https://doi.org/10.1109/ICDAR.1995.598994.

Hsu, Jung-Lung, Alexander Leemans, Chyi-Huey Bai, Cheng-Hui Lee, Yuh-Feng Tsai, Hou-Chang Chiu, and Wei-Hung Chen. 2008a. “Gender Differences and Age-Related White Matter Changes of the Human Brain: A Diffusion Tensor Imaging Study.” *NeuroImage* 39 (2): 566–77. https://doi.org/10.1016/j.neuroimage.2007.09.017.

———. 2008b. “Gender Differences and Age-Related White Matter Changes of the Human Brain: A Diffusion Tensor Imaging Study.” *NeuroImage* 39 (2): 566–77. https://doi.org/10.1016/j.neuroimage.2007.09.017.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-1418-1.

Kennedy, Keith E., Nicole Kerlero de Rosbo, Antonio Uccelli, Maria Cellerino, Federico Ivaldi, Paola Contini, Raffaele De Palma, et al. 2023. “Multiscale Networks in Multiple Sclerosis.” bioRxiv. https://doi.org/10.1101/2023.02.26.530153.

Kocevar, Gabriel, Claudio Stamile, Salem Hannoun, François Cotton, Sandra Vukusic, Françoise Durand-Dubief, and Dominique Sappey-Marinier. 2016. “Graph Theory-Based Brain Connectivity for Automatic Classification of Multiple Sclerosis Clinical Courses.” *Frontiers in Neuroscience* 10. https://www.frontiersin.org/articles/10.3389/fnins.2016.00478.

Llufriu, Sara, Eloy Martinez-Heras, Elisabeth Solana, Nuria Sola-Valls, Maria Sepulveda, Yolanda Blanco, Elena H. Martinez-Lapiscina, et al. 2016. “Structural Networks Involved in Attention and Executive Functions in Multiple Sclerosis.” *NeuroImage : Clinical* 13 (December): 288–96. https://doi.org/10.1016/j.nicl.2016.11.026.

Mandke, Kanad, Jil Meier, Matthew J. Brookes, Reuben D. O’Dea, Piet Van Mieghem, Cornelis J. Stam, Arjan Hillebrand, and Prejaas Tewarie. 2018. “Comparing Multilayer Brain Networks between Groups: Introducing Graph Metrics and Recommendations.” *NeuroImage* 166 (February): 371–84. https://doi.org/10.1016/j.neuroimage.2017.11.016.

Martí-Juan, Gerard, Jaume Sastre-Garriga, Eloy Martinez-Heras, Angela Vidal-Jordana, Sara Llufriu, Sergiu Groppa, Gabriel Gonzalez-Escamilla, et al. 2023. “Using The Virtual Brain to Study the Relationship between Structural and Functional Connectivity in Patients with Multiple Sclerosis: A Multicenter Study.” *Cerebral Cortex*, February, bhad041. https://doi.org/10.1093/cercor/bhad041.

Nabizadeh, Fardin, Soroush Masrouri, Elham Ramezannezhad, Ali Ghaderi, Amir Mohammad Sharafi, Soroush Soraneh, and Abdorreza Naser Moghadasi. 2022. “Artificial Intelligence in the Diagnosis of Multiple Sclerosis: A Systematic Review.” *Multiple Sclerosis and Related Disorders* 59 (March): 103673. https://doi.org/10.1016/j.msard.2022.103673.

Pontillo, Giuseppe, Ferran Prados, Alle Meije Wink, Baris Kanber, Alvino Bisecco, Tommy A. A. Broeders, Arturo Brunetti, et al. 2022. “More than the Sum of Its Parts: Disrupted Core-Periphery of Multiplex Networks in Multiple Sclerosis.” medRxiv. https://doi.org/10.1101/2022.12.17.22283623.

Schoonheim, Menno M., Tommy A.A. Broeders, and Jeroen J.G. Geurts. 2022. “The Network Collapse in Multiple Sclerosis: An Overview of Novel Concepts to Address Disease Dynamics.” *NeuroImage : Clinical* 35 (July): 103108. https://doi.org/10.1016/j.nicl.2022.103108.

Schoonheim, Menno M., Kim A. Meijer, and Jeroen J. G. Geurts. 2015. “Network Collapse and Cognitive Impairment in Multiple Sclerosis.” *Frontiers in Neurology* 6 (April): 82. https://doi.org/10.3389/fneur.2015.00082.

Seccia, Ruggiero, Silvia Romano, Marco Salvetti, Andrea Crisanti, Laura Palagi, and Francesca Grassi. 2021. “Machine Learning Use for Prognostic Purposes in Multiple Sclerosis.” *Life* 11 (2): 122. https://doi.org/10.3390/life11020122.

Shu, Ni, Yunyun Duan, Mingrui Xia, Menno M. Schoonheim, Jing Huang, Zhuoqiong Ren, Zheng Sun, et al. 2016. “Disrupted Topological Organization of Structural and Functional Brain Connectomes in Clinically Isolated Syndrome and Multiple Sclerosis.” *Scientific Reports* 6 (July): 29383. https://doi.org/10.1038/srep29383.

Solana, Elisabeth, Eloy Martinez-Heras, Jordi Casas-Roma, Laura Calvet, Elisabet Lopez-Soley, Maria Sepulveda, Nuria Sola-Valls, et al. 2019. “Modified Connectivity of Vulnerable Brain Nodes in Multiple Sclerosis, Their Impact on Cognition and Their Discriminative Value.” *Scientific Reports* 9 (1): 20172. https://doi.org/10.1038/s41598-019-56806-z.

Solana, Elisabeth, Eloy Martinez-Heras, Elena H. Martinez-Lapiscina, Maria Sepulveda, Nuria Sola-Valls, Nuria Bargalló, Joan Berenguer, et al. 2018. “Magnetic Resonance Markers of Tissue Damage Related to Connectivity Disruption in Multiple Sclerosis.” *NeuroImage: Clinical* 20 (January): 161–68. https://doi.org/10.1016/j.nicl.2018.07.012.

Sporns, Olaf. 2018. “Graph Theory Methods: Applications in Brain Networks.” *Dialogues in Clinical Neuroscience* 20 (2): 111–21. https://doi.org/10.31887/DCNS.2018.20.2/osporns.

Vaiana, Michael, and Sarah Feldt Muldoon. 2020. “Multilayer Brain Networks.” *Journal of Nonlinear Science* 30 (5): 2147–69. https://doi.org/10.1007/s00332-017-9436-8.

Welton, Thomas, Cris S. Constantinescu, Dorothee P. Auer, and Rob A. Dineen. 2020. “Graph Theoretic Analysis of Brain Connectomics in Multiple Sclerosis: Reliability and Relationship with Cognition.” *Brain Connectivity* 10 (2): 95–104. https://doi.org/10.1089/brain.2019.0717.

Zhao, Yijun, Tong Wang, Riley Bove, Bruce Cree, Roland Henry, Hrishikesh Lokhande, Mariann Polgar-Turcsanyi, et al. 2020. “Ensemble Learning Predicts Multiple Sclerosis Disease Course in the SUMMIT Study.” *Npj Digital Medicine* 3 (1): 1–8. https://doi.org/10.1038/s41746-020-00338-8.

Zurita, Mariana, Cristian Montalba, Tomás Labbé, Juan Pablo Cruz, Josué Dalboni da Rocha, Cristián Tejos, Ethel Ciampi, Claudia Cárcamo, Ranganatha Sitaram, and Sergio Uribe. 2018. “Characterization of Relapsing-Remitting Multiple Sclerosis Patients Using Support Vector Machine Classifications of Functional and Diffusion MRI Data.” *NeuroImage. Clinical* 20: 724–30. https://doi.org/10.1016/j.nicl.2018.09.002.

# Appendices

## Graph measures definitions

We will only review measures used in this work. All definitions are from this references (Fornito et al. 2016), (Bianconi 2022) and (Barabási, Albert-László 2016), though no explicit reference will be made in each definition

#### Strength

In an undirected network, the degree of a node is the number of edges incident to it. When network is weighted, we have an analog of node degree which is **node strength, ,** defined as the sum of the weights of the edges incident to node

|  |  |  |
| --- | --- | --- |
|  |  | ( 14) |

Where is the weight of the edge connecting nodes and .

In directed networks, we can differentiate between edges leaving the node and edges arriving to the node, and this way define an out-degree and in-degree:

#### Clustering coefficient

The **local clustering** coefficient of a node of degree measures the probability of that two neighbors of node are also connected. It is defined counting the number of triangles that node is part of

|  |  |  |
| --- | --- | --- |
|  |  | ( 15) |

**Global clustering C** is the average of the local clustering over all the nodes. In a network with N nodes:

|  |  |  |
| --- | --- | --- |
|  |  | ( 16) |

An alternative definitions of global clustering is found in literature, quantity defined this way Is called **transitivity**:

|  |  |  |
| --- | --- | --- |
|  |  | ( 17) |

Both transitivity and clustering values range from zero to one. Values close to one indicate a network with large number of triangles and thus heavily interconnected.

#### Shortest Path

The shortest path between two nodes is the path with the fewest number of edges, and is often called distance and denoted by .

The average or **mean shortest path** is the average of shortest paths between any two nodes of the network. Therefore, in a network with N nodes:

|  |  |  |
| --- | --- | --- |
|  |  | ( 18) |

We can extract another important measure from distance, the **diameter of the network**, which is the maximum of the shortest distances between any two nodes of the network.

|  |  |  |
| --- | --- | --- |
|  |  | ( 19) |

Path length in a *weighted* network is calculated as the *sum of the weights* of the edges in the path.

#### Efficiency

**Global Efficiency** is closely related to mean path length as is defined as the inverse of distances

|  |  |  |
| --- | --- | --- |
|  |  | ( 20) |

We can define a local efficiency. For a node

|  |  |  |
| --- | --- | --- |
|  |  | ( 21) |

Where denotes the subgraph that comprises all nodes connected to , but once removed node and all its incident edges.

#### Modularity

In a network, a community is formed by a set of nodes that are more densely connected to each other than to the rest of the network.

Modularity defines a way of comparing the density of the edges within each community to the expected value in the hypothesis that a link between a two generic nodes and occur with probability . Null hypothesis is that network is completely random, while preserving same degree distribution as the real graph. Taking all of this into account, modularity is thus defined:

|  |  |  |
| --- | --- | --- |
|  |  | ( 22) |

Where is the Kronecker delta function and are the elements of the adjacency matrix.

This definition can be extended to weighted networks

|  |  |  |
| --- | --- | --- |
|  |  | ( 23) |

Where W is the total weight of the unique edges of the network, and now p is the probability but given by strength ().

Modularity is a measure that can evaluate the significance of a community.

#### Density

Density has a quite intuitive definition, is the ratio between the number of edges in a graph and the largest possible number of edges in that graph. In a undirected graph, maximum number of edges is N(N-1)/2 where N is the number of nodes, so density is:

|  |  |  |
| --- | --- | --- |
|  |  | ( 24) |

This a concept not defined for weighted graphs.

#### Centrality (Closeness Centrality and PageRank)

In a common network, different nodes have different number of links that connect them to the rest of the network, therefore playing different roles in that network. It is reasonable to assume that highly connected vertices exert a more important role in the network.

We have introduced the strength, but if we want to rank the nodes in order of their “importance”, strength only gives a partial account of the picture. There are different measures aimed to this purpose, two of them are closeness centrality and PageRank

**Closeness centrality** measures how close a node is from the rest of the nodes in a graph. To do this one computes shortest distance from that node to the rest of nodes, a node with short average path length is able to interact efficiently with many nodes, and also non central nodes will be able to reach that node easily. Formal definition of closeness centrality of a node :

|  |  |  |
| --- | --- | --- |
|  |  | ( 25) |

It is important to note how similar is this definition to the one for efficiency. In weighted networks, this measure can be calculated using shortest weighted path. It is also important to note that if two nodes are disconnected,

Closeness centrality can be applied to multilayer networks, in **multilayer closeness centrality**, instead of distances confined to one layer, multilayer shortest-path is considered.

Although it seems fair to consider a node more important (i.e.: it has more centrality) if it has a link to a very important node, we have to take into account that it is likely high centrality nodes may have many links. As this measure was developed for webpages it is easy to understand with an example from that field: If a very important webpage contains a lot of URL links, the prestige of the pointed webpages is not the same as if the webpage contained only a very few and selected links. **PageRank centrality** of node satisfies:

|  |  |  |
| --- | --- | --- |
|  |  | ( 26) |

Where , this way if a node has no out-degree we do not have a divide by zero problem[[1]](#footnote-1), and is a parameter between 0 and 1. Last term is given by

|  |  |  |
| --- | --- | --- |
|  |  | ( 27) |

With indicating the Kronecker delta.

#### Multilayer PageRank Centrality

We can extend PageRank centrality to multilayer networks in order to evaluate the centrality of nodes considering all the layers across the network. It captures the effect of the centrality of a node in one layer on its centrality in another layer. A fine example is when a famous actor takes part in a political cause, his centrality in one layer (movies) influences his centrality in another layer (political). The **multilayer PageRank** of node depends on two parameters and satisfies

|  |  |  |
| --- | --- | --- |
|  |  | ( 28) |

Where and are given by

|  |  |  |
| --- | --- | --- |
|  |  | ( 29) |

|  |  |  |
| --- | --- | --- |
|  |  | ( 30) |

## Libraries and packages

In the following tables all packages imported during this project are shown, along with their respective versions. It is worth noting that these table do not include dependent libraries that may have been automatically imported as requirements.

|  |  |
| --- | --- |
| R 4.2.3 | |
| Package | Version |
| igraph | 1.4.2 |
| ggplot2 | 3.4.2 |
| muxViz | 3.1 |
| patchwork | 1.1.2 |
| tidyverse | 2.0.0 |

Table 25. R packages and versions.

|  |  |
| --- | --- |
| Python 3.10.10 | |
| library | Version |
| Imbalanced-learn | 0.10.1 |
| Matplotlib | 3.7.1 |
| networkX | 2.8.4 |
| neuroCombat | 0.12.12 |
| numpy | 1.23.5 |
| nxviz | 0.7.4 |
| Pandas | 1.5.3 |
| scikit-learn | 1.2.2 |
| Scipy | 1.10.0 |
| Seaborn | 0.12.2 |
| statsmodels | 0.13.5 |
| xgboost | 1.7.3 |

Table 26. Python libraries and versions

1. In this case will be zero and result will be zero but we avoid having a zero in the denominator [↑](#footnote-ref-1)