BACHELOR' S THESIS / BIOMEDICAL ENGINEERING 2023

The use of whole-brain models and variational autoencoders for the low-dimensional representation of psychosis and its perturbational landscape

Iraïs Garcés de Marcilla Lappin

upf. Universitat Pompeu Fabra Barcelona

The Use of Whole-Brain Models and Variational Autoencoders for the Low-Dimensional Representation of Psychosis and its Perturbational Landscape

Iraïs Garcés de Marcilla Lappin

Bachelor's Thesis UPF 2022/2023

Thesis Supervisor(s)

Dr. Gustavo Deco , (Department: Computational Neuroscience)

Yonatan Sanz , (Department: Computational Neuroscience)

Universitat Pompeu Fabra Barcelona

Acknowledgments

I would like to thank my tutors, Yonatan Sanz and Gustavo Deco for their help and guidance throughout this project. Also, express my gratitude to Ludovica Mana for the help in the acquisition and understanding of the data set.

Summary/Abstract

Psychosis can be described as an alteration in brain connectivity that leads to an impairment of cognition and the speed at which the information gets processed, what causes a diversity of psychiatric symptoms. This symptomatology is characterized by changes in the brain activity in certain areas, which can be detected by Functional Magnetic Resonance Imaging (fMRI) as it registers changes in the brain associated with blood flow, and this allows us to measure brain activity and connectivity between regions. Furthermore, the state of these alterations may differ between patients depending on the severity of their condition and the number of episodes they have had or may suffer. This study focuses on the use of the connectivity and structural information extracted from fMRIs and a whole-brain model to generate synthetic data with enough resemblance to the original dataset cases to train a Variational Autoencoder architecture for the creation of a low dimensional space in which the cases where patients have had one psychotic episode (remitting) or multiple (relapsing) are represented, and therefore a classification model can be trained to distinguish them. A dimensionality analysis has been performed to find the most optimal dimension of this space that allow us to distinguish between remitting and relapsing cases with high enough accuracy. Moreover, perturbations were introduced in the original model to generate new data which was reclassified in the low dimensional space to find which alterations could produce changes in the classification of the psychotic stage.

Keywords

Psychosis, Psychosis relapsing, Deep Learning, Variational Autoencoder, Whole-Brain Dynamics, Classification

Preface or prologue

Psychosis can arise from many different cause and conditions, as a symptom of neurodegenerative diseases or even as a consequence of the withdrawal of addictive substances. Usually, it is described as an alteration in brain connectivity that leads to an impairment in cognition and the speed at which the information gets processed. This symptomatology can be related to changes in the brain activity and degradation of certain brain areas, which can be detected via Functional Magnetic Resonance Imaging (fMRI) as it registers changes in the brain associated with blood flow, and this allows us to measure brain activity and connectivity between regions. Therefore, the state of this connection between different brain areas and its alteration in diagnosed psychotic patients can be studied. Furthermore, the state of these alterations differs between patients depending on the severity of their condition and the number of episodes they have had or may suffer.

Knowing this, the following Bachelor's Thesis focuses on the use of the connectivity and structural information extracted from fMRI and a whole-brain model to generate synthetic data with enough resemblance to the original dataset cases to train a Variational Autoencoder (VAE) architecture for the creation of a low dimensional space in which different classes of psychotic patients will be classified regarding the severity of their condition. Three classes are defined for this classification: control cases, cases where patients have had one psychotic episode (remitting) or multiple episodes (relapsing). This way, we can generate a low dimensional space that show us if there are any relevant differences between those conditions regarding the information encoded in their brain connectivity. The evaluation of the performance of the VAEs is assessed via a classification model. Furthermore, this low dimensional space and classification model can be used to perceive variation in the state of patients. After the proper training of the deep learning architecture, the perturbation of remitting and relapsing cases was performed to see if there were any alterations that could lead to the reclassification of the patients into another class.

With this methodology, a series of brain areas were found that produced improvement and worsening of the conditions for some values of the perturbation. These areas were then related to the literature, as they coincide with some of the brain regions that show excessive degradation in psychotic patients.

Therefore, this Bachelor's Thesis proves that the use of the pipeline proposed in other works for the classification of consciousness states or for the classification of Alzheimer disease against other types of dementia, can be also useful when applied to psychiatric disease, such as psychosis.

Universitat Pompeu Fabra Barcelona

Index

List of Figures

List of Tables

1 Introduction

1.1 Psychosis

Psychosis can be vaguely defined as the loss of capacity to distinguish between what is real and what is not. The World Health Organization defines it as the presence of hallucinations and/or delusions, which lead to the impairment in reality testing and could result in the incapacity to meet ordinary demands of life [1]. Furthermore, in the Diagnostic and statistical Manual of Mental Disorders, Fifth Edition (DSM-5), the diagnosis of this condition further requires the presence of one or more symptoms such as disorganized behaviour, catatonia (motor anomalies) or negative symptoms (which refer to a state of reduced emotional expression, decreased motivation, inability to feel pleasure and reduced spontaneous speech) [2] [3].

Furthermore, psychosis is a feature of multiple psychiatric disorders, it is a defining characteristic of schizophrenia spectrum disorders (such as personality disorder, delusional disorder, schizophreniform disorder, schizophrenia, or schizoaffective disorder). This condition can also occasionally occur in patients with bipolar disorder during a manic or depressive episode, or in patients with major depressive disorder. However, it can also derive from the withdrawal from substances in addicted patients, or be associated with obsessive-compulsive disorder as the obsessions that characterize them are usually classified as delusions [1] [4]. Moreover, psychotic symptoms have been also observed in neurocognitive disorders like Alzheimer's disease or other dementias, Parkinson's disease, Huntington's disease, traumatic brain injury, as a side effect of some psychoactive medications, or be triggered by a traumatic experience or stress [1].

1.1.1 Causes and Prevalence

As stated, psychosis can arise from, or be associated with a wide range of conditions which makes the worldwide prevalence of psychotic disorders to be around 3% [3] [5]. This positions psychosis among the world's leading causes of disability [6]. Furthermore, psychosis is associated with high levels of social disconnection (absence of family or social relationships and marginal participation in social activities) that leads to a reduction in employment and educational opportunities, affecting the individual's mental and physical health [7].

However, the etiology regarding psychotic disorders is still undetermined [5]. Some models postulate that the cause of primary psychotic disorders such as schizophrenia, which are not related to other neurocognitive illnesses, are related to the interaction of genetic and environmental risk factors [8]. In [6] the authors imply that from an etiopathological level, psychotic disorders could be related to the patient's social context, as exposure to certain situations during critical developmental periods impact neurocognition and affect social development. However, they also remark that the use of polygenetic risk scores to assess the status of the patient after first-episode psychosis would not be enough for the justification of this condition. And, for an accurate assessment, a wide range of factors such as genetic risk profiling,

familiar antecedents, and socio-demographic factors besides other indicators need to be considered before diagnosis.

It has also been seen that the monoamine system in the brain (dopamine, serotonin, and norepinephrine neurotransmitter networks) plays a major role in normal behavioural patterns and controls different functions of the brain and its disruption can be related to the pathophysiology of several psychiatric and psychotic disorders [9].

1.1.2 Current Screening Methods

However, currently, there is no specific neurobiological marker that helps us define the transition from a risk state to established psychosis or the relapsing risk after diagnosis with enough accuracy to be relevant. Neuroimaging techniques have tried to address this issue, as psychotic disorders have been observed in association with alterations of some brain structures, their function, and connectivity, in addition to neurochemical functions of the brain [10] [11]. In [10] the authors saw that grey matter volumes (GMV) decrease in those patients suffering from a psychotic disorder when compared with healthy controls. These reductions were mainly observed in the right temporal and left anterior cingulate, and cerebellar, and insular regions. Furthermore, temporal gyrus alterations have been related to the generation of hallucinations and thought disorders, while deformations in the left anterior cingulate area are related to emotional processing and executive performances and its alteration leads to difficulties in cognitive and emotional integration [10] [12].

All of this, indicating that the onset of first-episode psychosis is linked to GMV loss, and it is independent of illness duration [10] [12]. However, diagnosis is still purely based on psychopathological criteria [11].

1.1.3 Current Treatment and procedures

Usually, the treatment for psychotic disorders highly depends on the cause of the psychotic episode [13]. And different treatments and strategies are followed in each case. However, some aspects of the treatment are similar between those different cases:

- Medication such as antipsychotics are usually used [13]. Even if, when treating older patients this has to be carefully considered as some cardiovascular complications may arise due to the use of this medication. Furthermore, dopamine receptor antagonist have shown effectiveness in treating the symptoms of psychosis [3].

- Complementing the use of medication, psychotherapy is also recurred to for mitigating the symptoms of psychosis, as some patients may be resistant to the medication or stop taking it due to the side effects [14]. And Cognitive Behavioural Therapy (CBT) is usually the main course of action.

Early intervention can also be key in the treatment of psychosis. As it conditions

the access to better treatment and prognosis [15]. However, psychosis can only be diagnosed by the apparition of its symptomatology and, for now, its appearance cannot be predicted. Therefore, its early intervention can only occur after the first psychotic episode and the monitoring of the patients becomes essential for preventing and controlling possible relapses [16].

1.2 Psychosis Remission and Relapsing

However, even with treatment and monitoring, 40% of patients with a psychotic disorder do not achieve symptomatic remission, and long-term outcomes after a first psychotic episode (first episode psychosis, FEP) are poorly predicted [17]. In the case of schizophrenia, the cumulative relapse rate after five years after initial recovery is around 80% [18] [19]. As with every relapse, it has been observed that more severe cognitive deterioration occurs in the brain, the recovery possibilities decrease [18], and the risk of developing persistent psychotic symptoms increases [19]. This highlights the need for early intervention and the use of a predictive model to distinguish those cases with higher remission probability after a FEP. As predictive models would aid in the planing of treatments and focus the resources on those patients with worse prognosis [17].

1.3 State of the Art

The use of models to forecast the probability of a diagnosis, the outcomes of a certain condition (prognosis), or the response to an intervention (prediction), at an individual level, are being developed for many conditions as they can be useful tools in medical research [20]. And, even in the psychiatric field, emerging researchers are starting to develop prognostic and predictive models to lessen the personal, clinical, and societal burden associated with psychotic disorders [20] [21]. The mentioned reviews, state that even if these kinds of models are being developed the biggest challenges they have to face before their implementation in real-world scenarios are related to validation, as there is a lack of frameworks or methodological infrastructures to follow for this process and the literature in the field is still small, even if some models are already available. Studerus and colleagues state in their review [22] that models focused on the prediction of psychosis lack big sample sizes for their development, usually they do not perform internal or external cross-validation, and they use poor model development strategies and therefore most of them are probably overfitted and have low accuracy.

Following this line of research, transition models based on different strategies (using neurocognitive and neurobiological data) have been proposed for the prediction of clinical high-risk patients after a first psychotic episode [21]. And algorithms using clinical and cognitive data such as [23] [24] have also been developed and validated to predict psychosis in critical high-risk (CHR) patients, even if their results show low accuracy. These models are focused on improving prognosis, reducing unnecessary treatment and identify decisive factors for transition to remission [21]. The authors focused on regression models with already obtained data (clinical and demographic information, neuropsychological data) and performed a systematic external

validation for their proposed model. Demonstrating that these kinds of models are potentially feasible.

In [18] two models were found with relevant accuracy to predict relapse in people with psychosis, being [25] the most relevant as it had moderate discriminatory power and low-risk bias as the authors counted with a large dataset. But, their research is focused on using population-level sociodemographic and health administrative data to develop a model to predict the readmission among adults discharged from an acute psychiatric unit in the short term.

Moreover, it has also been stated that in the context of psychosis treatment, remission does not have a well-defined criterion. This has been challenging, as the severity and functional outcome of the symptoms differ significantly between patients [26]. In 2005, the Remission in Schizophrenia Working Group (RSWG) proposed a remission criterion that has been widely accepted. This criterion focuses on three dimensions of the disorder: negative symptoms, disorganization, and psychoticism. And the initiative focused on symptomatic remission, however, the disappearance of symptomatology may not always be a good indicator of improving condition in schizophrenic patients. But still, no consensus on its functional remission is still reached, nor for schizophrenia and neither for other psychotic-related disorders [26].

Recent research has been done on biomarker identification to understand psychosis onset regarding electrophysiological, neuroimaging, and other biological indicators [27]. The identification of biomarkers can not only aid in the prediction of psychotic disorders remission but also further the understanding of them.

Regarding the use of neuroimaging for identifying biomarkers related to psychotic disorders, these technics have been used to study the relevant structural changes in GMV, these changes occur mainly in the temporal and frontal regions in patients with psychosis [10] [12]. Furthermore, recurrent psychotic episodes are associated with progressive loss of GMV and reduced effect of antipsychotic medications [19]. Furthermore, it is believed that schizophrenia and psychotic disorders are primarily disorders of connectivity, and evidence from functional connectivity studies indicates that alteration in multiple brain systems characterize the dysfunctionality associated with this kind of disorders. Meaning that altered connectivity in these areas and between them can be related to psychosis [28] [29]. And not only regarding GMV alterations, but other imaging studies have found a relation between white matter (WM) alterations and the transitions between a preclinical and chronic state of psychotic disorders. These alterations have been found in the corticospinal tracts, interhemispheric connections, cerebello-thalamo-cortical circuits, and limbic systems. And it has been hypothesized that there is a strong relation between these alterations and the pathological state of the patient and their evolution [30] [31].

In [15] the authors have tried to examine relationships between biomarkers and syndromes to distinguish disease markers, vulnerability markers, and risk factors across stages. They have tried to generate a clinical staging model for different mental disorders taking into account their symptomatology to classify patients with

different mental illnesses such as mania or psychosis. They also state that GMV decrease, mainly in the frontal cortex, and this is more pronounced in chronic cases of psychosis than in those patients that just experienced one episode. Furthermore, similar evidence has been found in other brain areas that correlate the worsening of the deformations with the illness progression.

Nowadays, the most common approach to improve the prognosis of patients after FEP is the use of antipsychotic medication, as it is associated with the fast improvement of symptoms in the majority of patients and with a reduced number of relapses [19]. The authors of [19] propose the use of pharmacological and nonpharmacological interventions to prevent relapse in FEP patients, as they remark that the first years after the first episode are crucial for long-term recovery.

Regarding computational models to predict the progression of psychotic disorders after a FEP, we can find machine learning models based on patient data and their current treatment to predict outcome trajectories for treatment optimization in the research performed in [32]. The authors of [32] also found indicators (such as previous depressive episodes or suicidality) that could lead to symptom persistence and non-adherence to treatment. However, this approach was not directly focused on the prediction of psychosis relapsing.

In [33] the authors try to identify cognitive subgroups in Schizophrenia Spectrum Disorders to compare their functional profiles. They used the Partition Around Medoids algorithm with cognitive variables and logistic regression for the cluster classification. With this, they distinguished two groups: those patients with cognitive impairment and those with relatively intact cognition. Being the first subgroup at more risk of early age onset, severe symptoms, and more probability of relapsing.

Even some other technologies are being addressed for monitoring and relapsing prevention in patients with mental disorders. The design proposed by [34] is based on the idea that the biological processes that lead to relapse in psychosis are being developed over time, and therefore they can be anticipated by observing changes in the behaviour of the patient. Some warnings of the worsening of the condition can be physically observable, such as rigidity and atypical motion or withdrawal from outdoor activities. Therefore, they propose the analysis of physiological or wearable sensors to try a machine or deep learning algorithm for the prediction of relapsing.

Advances in imaging and its analysis allow us to map the neural network anatomy and the dynamics in different scales of the brain [35]. This enables us to generate connectomes (neural wiring diagrams), and with the help of the advances in networks and complex systems understanding we can determine an unified framework for the analysis and interpretation of large datasets of brain connectomic information [35]. These developments could be the base of understanding the psychopathology of these conditions in new ways, as models generated through connectomes can allow us to identify essential mechanisms to comprehend these conditions and their evolution. Regarding psychosis, abnormal network wiring has been observed in the dysfunction of brain activity [35].

In [36] the authors studied the disconnection in the thalamocortical regions related to the risk of psychosis from functional magnetic resonance images (fMRI). And found that the reduced connection in those areas and between them was more prominent in those patients that ended up developing the illness in its severe and chronic states.

In addition, previous research has demonstrated that the use of computational models, based on functional connectivity and/or structural information, can be used to determine physical and biological mechanisms of brain activity, such as consciousness [37] [38] or neurodegeneration [39]. As whole-brain network modelling trained on virtual connectomes has been proved to achieve discrimination permeance in the classification of different brain states [40].

1.4 Objectives and Experimental Design

This work is aimed to find the optimal dimension of a latent space generated via a Variational Autoencoder (VAE) architecture for the clustering and classification of the different stages of psychosis. As will be further explained, we have subdivided a given dataset into healthy controls and two categories of psychotic patients: remitting and relapsing, depending on how many psychotic episodes the patients have had (a single episode or multiple). This original dataset will be used to generate a whole brain model for each of the defined classes, and these models will help us generate in-silico functional connectivity (FC) matrices to have enough individuals of each category for the proper training of the deep learning architectures. All of this information is generated via a genetic algorithm and data augmentation procedures to maximize the resemblance between simulated and empirical FC matrices for all the groups in the dataset. Next, a series of VAEs will be trained to generate models to adapt to this data with different dimensions of their middle layer. These models will be studied to find the optimal one for the classification of the given conditions, with a Neuronal Network (NN) as a classifier to measure their performance. The classification of these stages in a low dimensional latent space will help us investigate the functional connectivity alterations related to biophysical and anatomical changes in patients with different stages of psychosis. Furthermore, this work was extended by perturbing a simulated FC matrix of each of these classes to find those perturbations that can alternate the classification of the sample into another area of the latent space, which will help us relate the connectivity alterations these perturbations represent to biophysical and anatomical changes. And, finally, find those brain areas more prone to lead to changes in the condition of psychotic patients during treatment.

Knowing this, we can hypothesize that a latent space will be created where the different stages of psychosis and the healthy controls will be properly clustered, and therefore a good classificatory model would be trained. Furthermore, this will help us reclassify cases after different sets of perturbations to see which ones can be related to changes in the condition of psychotic patients.

This framework has been previously helpful for distinguishing functional connectivity patterns (mapped clearly in different regions of the latent space) in cases of

Alzheimer disease patients and classify them against other dementias. Furthermore, it was also seen that this kind of encoding in a latent space can be used to reflect the severity of those illnesses regarding the degradation of white matter volumes [39]. However, from what has been found in the literature there is no similar approach regarding the distinction between psychotic remission and relapsing, even if machine learning algorithms and other ways to predict relapsing in psychosis have been tried.

2 Methods

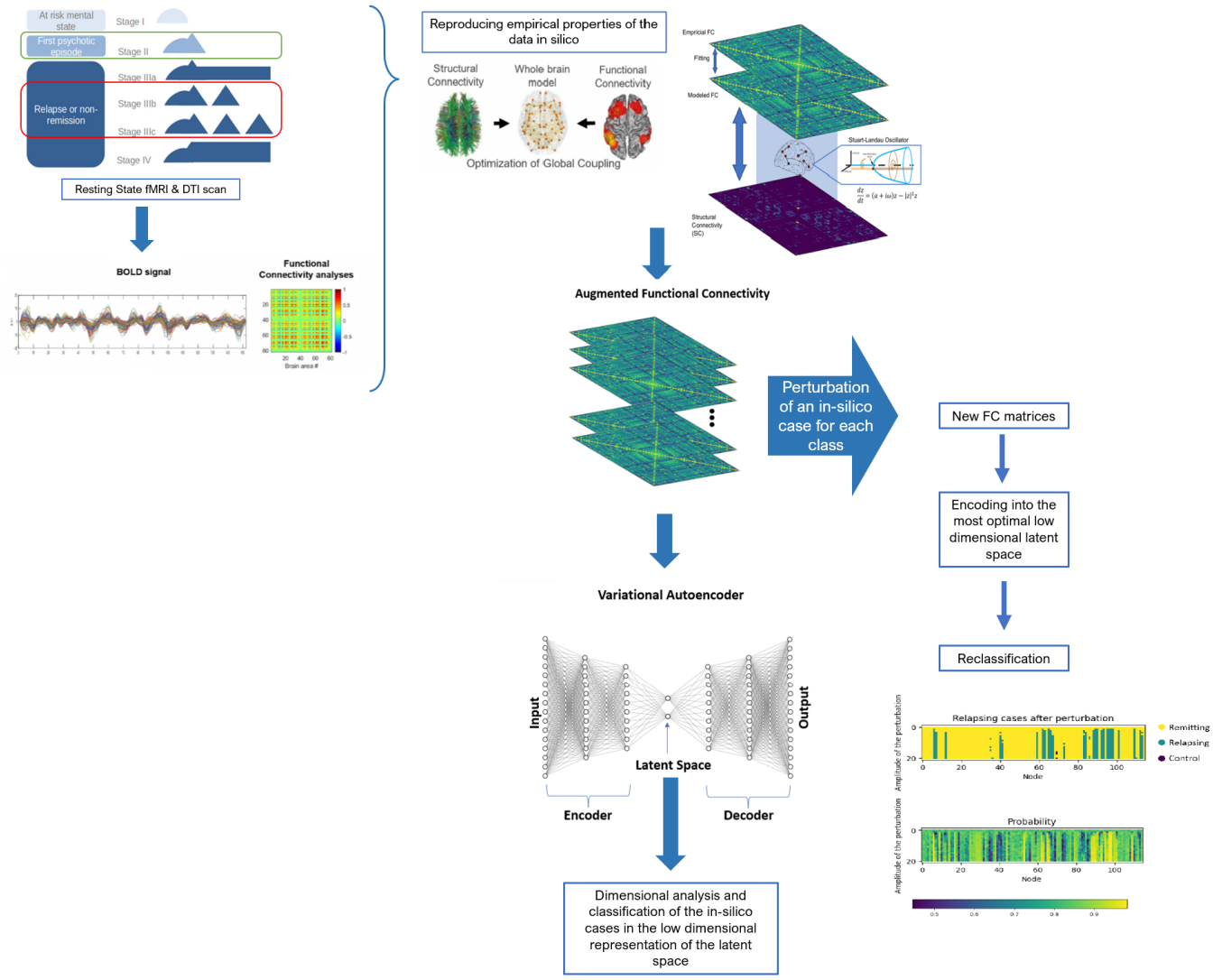

Figure 1: Overview of the methodology followed in this bachelor's thesis.

2.1 Dataset

The used dataset for this work is provided by [41] and consists of connectivity information processed from magnetic resonance images (MRIs) of 216 subjects. The participants are classified as 88 early psychosis patients (EPPs) and 128 healthy

control subjects (HCs). More detail regarding the selection process and its criteria is given in the reference.

Moreover, the EEP group goes through further subclassification to distinguish patients in different stages of psychosis:

- 37 Stage II patients: which are those patients with FEP and no other previous psychotic episodes. In these patients, the psychotic symptomatology has remitted completely.
- 19 Stage IIIa patients: meaning they present incomplete remission from stage II at 12 months after observations and the illness has not been present for longer than 5 years.
- 22 Stage IIIb patients: regarding the patients with a single recurrence or relapse from a psychotic episode.
- 9 Stage IIIc patients: which defines those patients with two or more relapses after stage II with remission between episodes.
- 1 Stage IV patient: meaning no remission from stage II at 12 months after observations and the illness has not been present for longer than 5 years.

For the purpose of this study, we will classify the cohort of EEPs into two subgroups: stage II will be classified as remitting and stages IIIb and IIIc will be grouped together and classified as relapsing, as the only difference between them is the number of relapses. While those patients classified in Stage IIIa will not be taken into account for the training of the VAE, as they correspond to an intermediate stage between remitting and relapsing that could cause misleading results in the classification. Neither the Stage IV is going to be used in this study as there is only one case in the dataset under this classification, and this would not be optimal for the generation of synthetic data.

Further information like age, biological sex, and severity assessment of each cohort of patients can be obtained in [41].

2.2 MRI Acquisition and Connectome Generation

The connectivity information given was extracted from MRIs on a 3-Tesla Siemens scanner, including a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence with 1 mm in-plane resolution, 1.2 mm slice thickness and diffusion spectrum imaging (DSI). The combination of this data gave the authors of [41] the connectomes for each patient. And, from this information, they estimated the structural connectivity between regions as the number of streamlines connecting those areas.

2.3 Mathematical Model

The whole-brain model used in this study consists of 115 coupled brain areas or nodes derived from the parcellation provided by [41]. The global dynamics of the brain network model results from the computation of mutual interactions of local node dynamics coupled through the underlying empirical anatomical structural connectivity matrix (Cij) which encodes the density of fibres between the i and the j areas given by the fMRIs.

The local dynamics of each node are described by the normal form of a supercritical Hopf bifurcation, which allows us to model the transition from asynchronous noisy behaviour to oscillations and which describes the following equations at each node:

$$\frac{dz_j}{dt} = z[a_j + i\omega_j - |z_j^2|] + \beta\eta_j(t) \tag{1}$$

$$z_j = \rho_j e^{i\theta_j} = x_j + iy_j \tag{2}$$

Where η represents an additive Gaussian noise with a standard deviation equal to $\beta = 0.02$. The supercritical Hopf bifurcation occurs at $a_j = 0$, as for lower values of a_j the local dynamics have a stable fixed point at $z_j = 0$ corresponding to a low activity asynchronous state, while for values of $a_j > 0$ it exist a stable limit cycle oscillation with frequency $f_j = \omega/2\pi$ [42].

Therefore, we can define the whole-brain dynamics as the following by separating the imaginary and real part of Equation 1:

$$\frac{dx_j}{dt} = [a_j - x_j^2 - y_j^2]x_j - \omega_j y_j + G\sum_i C_{ij}(x_i - x_j)\beta\eta_j(t) \tag{3}$$

$$\frac{dy_j}{dt} = [a_j - x_j^2 - y_j^2]y_j + \omega_j x_j + G\sum_i C_{ij}(y_i - y_j)\beta\eta_j(t) \tag{4}$$

Following the methodology proposed by [42], Equation 3 and Equation 4 will be coupled using the common difference coupling for approximating their linear part, which allows sharing the connectivity information between the different nodes.

In these latter equations, the parameter G represents a global scaling factor that defines the global conductivity parameter scaling equally all synaptic connections.

2.4 Genetic Algorithm

As we got the data ready to be processed, the first thing we needed to do was to ensure that the given grouping of the data met the parcellation requirements of the codes provided by Gustavo Deco and Yonatan Sanz, and that the needed changes were done to adjust this information.

The code provided implements a genetic algorithm that allows us to generate empirical data for the future training of a deep learning algorithm. This process is

based on the identification of the characteristics (genes) of the individuals that provide the greatest fitness values and allows them to be passed down to the next generation of individuals. This process starts with the generation of 10 sets of individuals randomly generated with values close to the lowest fitness. Then the outputs of this generation are processed with their corresponding fitness (Goodness of fit, GoF). The parameters that define this fitness are the global scaling factor G and the bifurcation parameter a. Then those individuals with the greatest results (lower GoF values) are chosen to generate the parent generation and go under different operations:

- Crossover: where two selected parents get combined to obtain a new individual to carry the information from them to the next generation.
- Mutation: where one of the selected parent's gene is changed randomly, and this produces a new individual for the next generation.
- Elite Selection: when a selected parent has an extraordinarily good GoF the solution is replicated without changes in the next generation.

These processes generate a new generation of offsprings, that will act as parents for the next iteration of the process. This is repeated iteratively and the genetic algorithm only stops when one of these criteria is met: the algorithm has reached 200 iterations, the best solution of the population does not change for 50 generations or the average GoF across the last 50 generations is less than 10^{-6}.

Furthermore, in this step some optimization was required, as the original code did not adjust to the data properly. The optimization function, defined originally as a function of the structural similarity index (SSIM), now was conditioned by the correlation between the empirical and simulated matrices, as indicated in Equation 5. This gave us a more precise indicator of the similarity between the empirical data and the new synthetic information that was being generated.

$$GoF = 1 - corr(FC_{empirical}, FC_{simulated}) \tag{5}$$

In addition, in order to correctly generate the empirical FC matrices, the data needs to be filtered by a high pass temporal filter in order to focus on the most relevant frequencies. Therefore, the band pass of the filter was adjusted to 0.04-0.07 Hz.

2.5 Data Augmentation

Afterward, once the final generation has been reached in the genetic algorithm, we used this information to generate a higher number of individuals to train the following deep learning algorithm better.

This process of data augmentation consists of choosing from the individuals, previously generated, the one with the greatest fitness, this is defined by the lowest value of the optimization function seen in Equation 5.

The defined genetic algorithm was repeated 10 times for each of the groups that the dataset was divided in (controls, stage II, stage IIIb and stage IIIc), from each of these 10 repetitions the individual with the greatest fitness was saved. Once the individual with the highest fitness value were identified for each subgroup, new individuals were generated from it using the same function that created the offsprings in the previous step. This way we got a greater number of samples with similar characteristics, corresponding to each subgroup to train the Variational Autoencoder with. And this allowed us to get a better classification of the patient's stages. More information and the validation of this process can be found in the Supplementary Material for [43].

Therefore, at the end of this step our dataset was conformed by: 1000 healthy control cases, 1000 remitting cases (generated from the stage II provided in the original dataset), and 1000 relapsing cases (500 generated from stage IIIb provided by the original dataset, and the remaining 500 generated from stage IIIc).

2.6 Variational Autoencoder Training

As stated, we need to encode the information of the FC matrices in a low-dimensional representation that allows us to classify the different stages of psychosis and, ideally, distinguish three classes conditioned by the previously stated criteria of remission or relapsing.

VAEs are able to learn a representation of the data in an arbitrary nonlinear manifold, therefore this kind of model can adjust better to the information we are working with. VAEs are autoencoders trained to map inputs to probability distributions in a latent space that can be regularized during the training process to produce meaningful outputs after the decoding.

The architecture of these models is subdivided into three parts:

- Encoder network: which is a deep neural network with rectified linear units as activation functions and dense layers that bottle-necks into an n-dimensional variational layer defined by z_n parameters, which are the variables that bound the generated latent space. In this part, a nonlinear transformation is done to map the C_{ij} into a Gaussian probability distribution.

- Middle variational layer: where the latent space variables z_n are defined. Different VAE architectures can be created by changing the number of z_n variables that form this layer. For this study, the dimensional analysis will be performed from latent spaces of 2 dimensions till 12 dimensions.

- Decoder network: in this part, the processes of the encoder network are mirrored to reconstruct the matrices in a C^*_{ij} form.

This architecture was trained by the backpropagation of errors via a gradient descent to minimize a loss of function composed by:

- A standard reconstruction error term: resulting from the output layer of the decoder.

- And a regularization term: being the difference between the distribution in the latent space and a standard Gaussian distribution. This term ensures continuity and completeness of the latent space, meaning that similar values will be decoded into similar outputs and that those represent meaningful combinations of the encoded inputs.

Regarding the training of the VAE we split the generated dataset via data augmentation consisting of 3000 cases randomly, using the 80% of this data to train the model and 20% to test it for optimization of the VAE weights. This training process consisted of batches with 128 samples and 20 training epochs, the loss function previously described, and an Adam optimizer (which is a gradient descent with a parameter-specific learning rate and a running average of gradients and their second moments to attenuate the effects of noise) with a learning rate of 0.004.

This generates the latent spaces where the differences between the psychosis stages will be encoded regarding the chosen z_n parameters. This space can be used to generate a classification model via Neuronal Network (NN) architectures to evaluate the precision of the VAE to distinguish between the remitting cases, those who relapse and the healthy controls. The neuronal networks designed for the classification were trained with a learning rate of 0.001, 30 epochs and a batch size of 32, and with the same training and test set than the VAE architecture.

2.7 Perturbation of original dynamics

This methodology can further be used to investigate how each of the defined stages of psychosis responds to different perturbations. The perturbations can be applied at different intensities and at all the different nodes of our parcellation, resulting in new FC matrices depending on those variations. Then, they can be mapped again in the previously generated latent space and reclassified into a new category using a previously trained neuronal network to see if any changes in the condition have occurred after the perturbation.

The chosen perturbation has previously been use for similar proposes in [39]. It is defined as a Wave simulation which incorporates a periodic term into the equation of a given node. This perturbation is described as $F_0 cos(\omega_0 t)$, F_0 being the amplitude of the perturbation and ω_0 the frequency of the node computed directly from the whole-brain model. In this case, amplitudes from 0 to 2 in 0.1 steps will be used to obtain the results.

This will allow us to bring together combinations of brain biomarkers, as it provides us with information on how certain connectivity alterations can make psychotic patients' condition change. The idea behind the performed perturbations is to find those brain areas more prone to produce changes in the brain dynamic of psychotic patients once disturbed.

3 Results

After performing the steps detailed in the methodology, the following results were obtained:

3.1 Genetic Algorithm and Data Augmentation

	CNT	Remitting (Stage II)	Relapsing	
			Stage IIIb	Stage IIIc
GoF	0.4218	0.4765	0.5293	0.4674
$\mu \pm \sigma$	0.4570 $\pm$0.0154	0.5214 $\pm$0.0182	0.5576 $\pm$0.0155	0.4913 $\pm$0.0136

Table 1: Results of the genetic algorithm for each class after ten repetitions. In the first row the GoF of the best individual of those ten repetitions is indicated, while the second row shows the mean and standard deviation of the GoF obtained by comparing all the best obtained individuals for each class within the ten repetitions

Regarding the genetic algorithm, the results detailed in Table 1 indicate the best fitness value obtained in each of the subgroups after ten repetitions, which represents the GoF reached by the most optimal individual for each subgroup and characterizes the information that most resembles the characteristics of the empirical data and that will be used for the generation of the in-silico dataset. As lower values of this indicator mean higher correlation between the empirical and simulated FC matrices.

Furthermore, we can also observe the mean and standard deviation of all the repetitions for each subgroup. It must be highlighted that the low values for the standard deviation stand for consistency in the results of the genetic algorithm along the different repetitions.

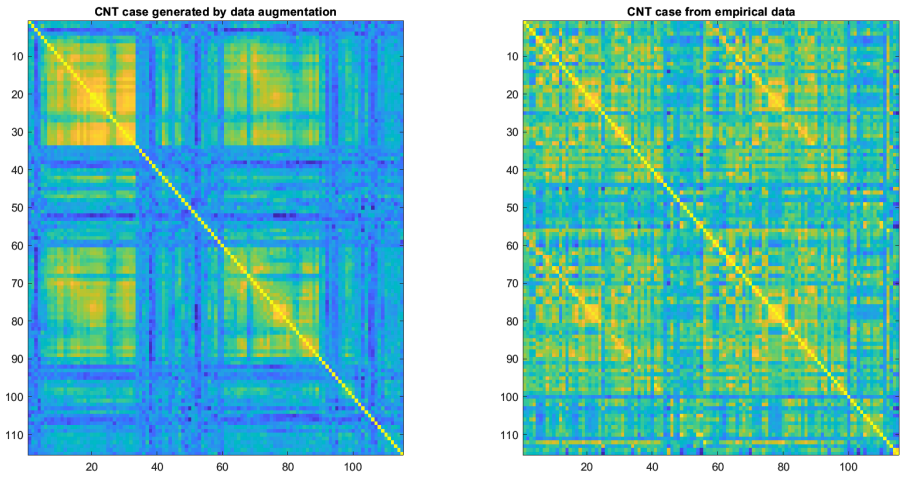

Figure 2: a) FC matrix generated from a control (CNT) case; b) FC matrix generated via data augmentation representing a CNT case.

In Figure 2 we see an example of a FC matrix generated via data augmentation for the control cases against one of the empirical cases it tries to resemble.

3.2 VAE Training

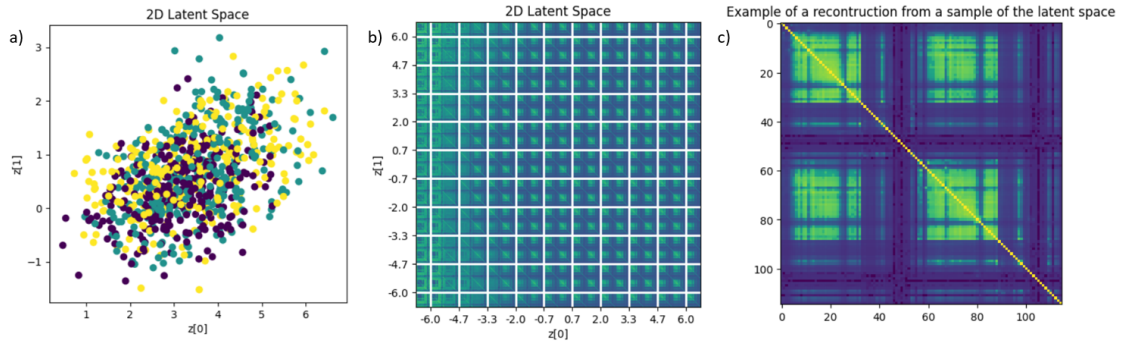

Figure 3: a) Example of the resulting Latent Space with a trained 2D VAE architecture. Where yellow cases represent relapsing cases, green remitting cases and dark blue symbolize control cases; b) Representation of the variation of FC matrices in a sample of the Latent Space; c) FC matrix reconstructed from a random coordinate of the example Latent Space.

In Figure 3 a) we can observe the results of training the VAE model with the found parameters after optimization. The VAE was trained to avoid overfitting of the data and lose of precision. This example shows the resulting latent space of a 2D VAE, and we can see that 2 dimensions do not extract enough information of our data to distinguish proper regions for each class in the latent space, highlighting the need to perform a dimensional analysis. However, for visualization purposes a 2D latent space was easier to represent and show, that indeed, this methodology may behave as expected. On the other hand, in Figure 3 b) a different representation of the previous latent space can be observed. Here, some of the different coordinates in the latent space are decoded, and we can see the resulting variation and progression between different regions of the latent space, and how this is translated to information that can be decoded as FC matrices. Showing us that we are able to map this kind of information into low dimensional spaces and then retrieve it. Finally, Figure 3 c) shows us one of these reconstructions and how, at some degree, it still resembles the initial information that the algorithm was fed with.

Then, after the training of the VAE for all studied dimensions of the intermediate layer (from 2 to 12), the classification models based on NN architecture were tested a 100 times for each dimension with shuffling of the data set each time a new classification model was trained. This generated a different test and training set for each of the 100 classification models, testing the classification accuracy for each of the 11 trained VAE architectures with all the possible data used as training and/or test set. Furthermore, the null hypothesis was tested by shuffling the labels in the data randomly to see if the obtained results were due to chance. This incorrect dataset was used to train a VAE for each studied dimension and 100 classification models as a reference of how these behaved when faced with incorrectly classified data.

In Figure 4 we observe the results of this process. In Figure 4 b) we see that for all dimensions the results corresponding to the incorrect labelling remains around the 0.33 value of accuracy for all tested dimensions. This was expected as we are

trying to classify 3 classes, and this is the probability of correctly classifying an individual by chance with these conditions. On the other hand, the results with the proper labelling go from values around 0.6 till almost 0.8, with a clear progression and better performance for those latent spaces with higher dimensions. Here we see that from dimension 2 till 7 the accuracy gets higher for each new dimension that is added to the VAE. However, once dimension 8 is reached, the precision of this classification remains very similar between dimensions. And even if in some particular cases values of almost 0.8 are reached, the mean for dimensions 8 to 12 is very similar, leading us to believe that by this methodology higher accuracy values cannot be reached even as we continue increasing the number of dimensions of the latent space.

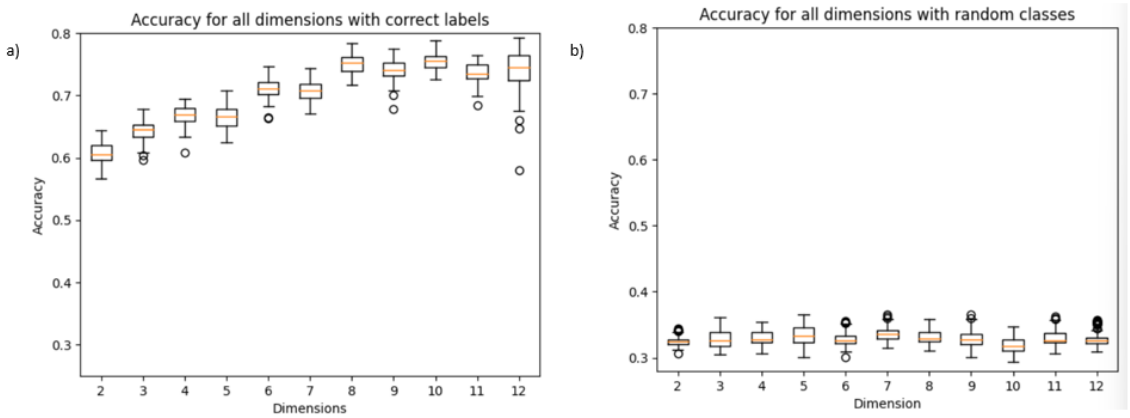

Figure 4: a) Box plot resulting of the k-cross validation with 100 classifiers for all the studied dimensions and correct labelling; b) Box plot resulting of computing all the accuracies for each dimension with 100 repetitions for the classifier with the incorrect labelling of the dataset.

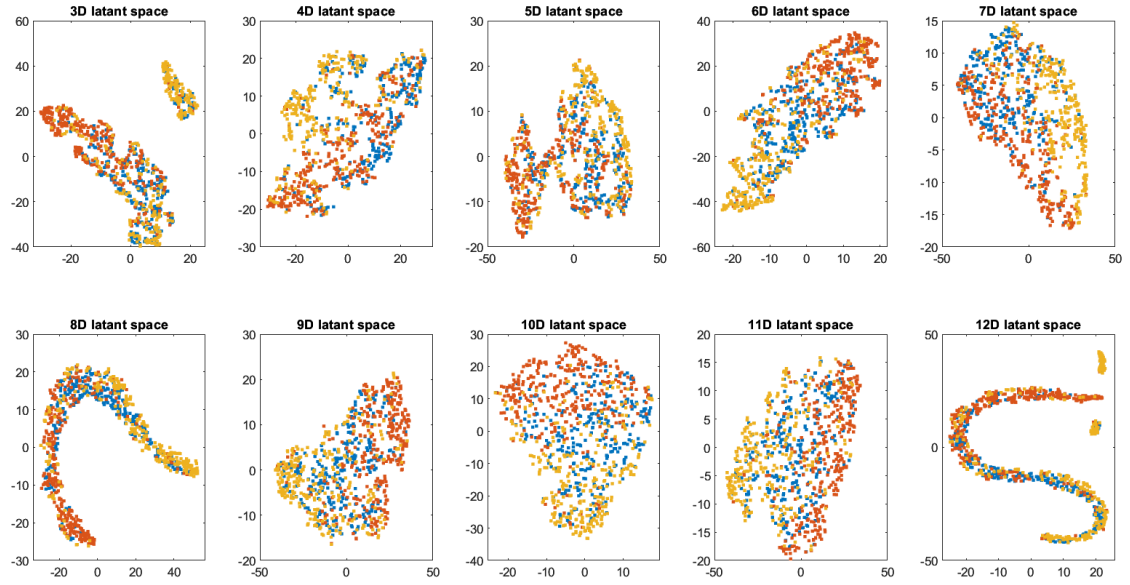

Figure 5: 2D representations of the latent spaces for dimensions 3 to 12. Blue cases represent CNT patients, while the orange ones represent remitting patients, and yellow cases the relapsing ones.

3.3 Latent Space Results

In Figure 5 we can observe the representation of the different latent spaces (dimensions 3 to 12) in a 2D space. This was computed using the *tsne* function in MATLAB [44] that allows to reduce the dimensionality of our data to two dimensions for easier representation. We can see that in some of those spaces we can distinguish some regions separating the different classes, while in others not all the latent space is occupied by our data. This could be due to the way the function used reduces the dimensionality of the data.

None of these spaces will be used to train further classification models, as this is just a way to visualize our results.

3.4 Perturbation Results

After the obtention of the previous result, the pipeline was extended to investigate how relapsing and remitting patients would respond to a series of perturbations, and if any of them could lead to a reclassification of their original condition.

These perturbations were applied at different intensities and at each of the 115 nodes individually, giving us a series of FC matrices that depend on these two variables. The case used for these perturbations was the same individual as the one used for the data augmentation, meaning that it had the greatest fitness of each class. After obtaining this information, the FC matrices obtained were encoded in the 8D VAE previously trained and reclassified by a NN with an accuracy of 0.753 and characterized by the Confusion Matrix in Table 2.

		Real Value		
		CNT	Remitting	Relapsing
Predicted Values	CNT	718	123	168
	Remitting	102	773	48
	Relapsing	180	104	784

Table 2: Confusion Matrix of the neuronal network trained for the classification of the perturbed FC matrices. These values were obtained by training a new NN with the original unperturbed dataset in the 8D latent space.

In Table 2 we can see that all three classes get classified correctly at similar rates, even if the control class is the one with fewer cases correctly classified. Furthermore, we can observe a tendency for both, remitting and relapsing patients, to get wrongly classified as controls at higher rates than to be wrongly classified as other psychotic stages.

In Figure 6 a) we can observe that after the indicated perturbations, most of the new generated FC matrices from a remitting case get reclassified as this same class. However, we can also observe some nodes that at any intensity of the perturbation or for most of its values accomplish a reclassification of the condition into a relapsing case, which indicates a worsening of the condition. While for other nodes, even if this appears in lower cases, a reclassification of the remitting cases after perturbation

into control ones is also accomplished, indicating an improvement from the original condition. Moreover, Figure 6 b) shows us that for most of this newly classified cases, the probability for the chosen label to be the correct one is very high.

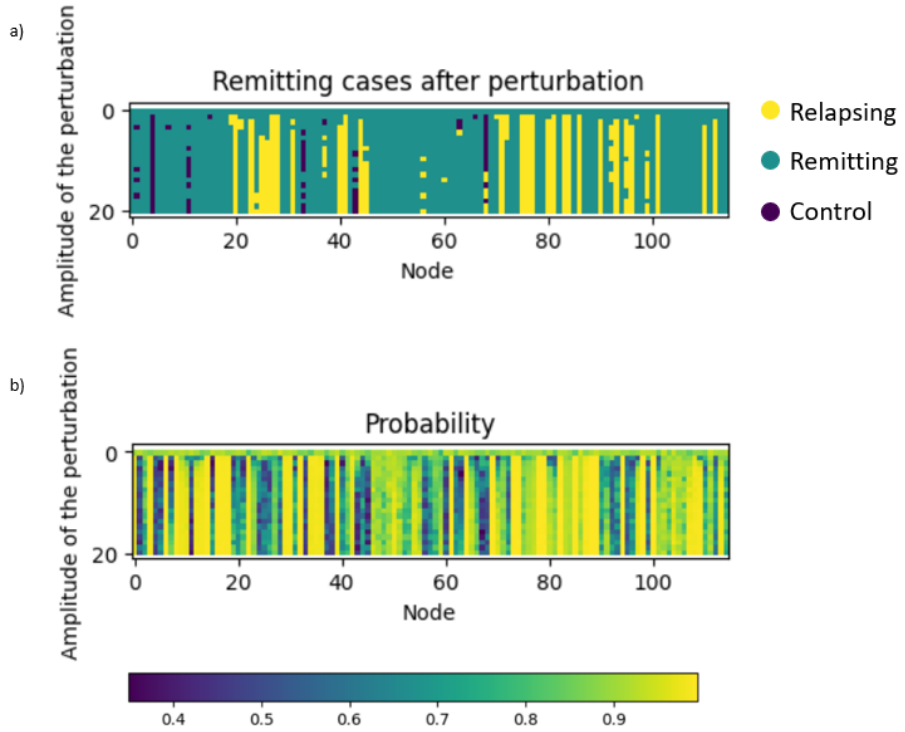

Figure 6: a) Resulting classification of the FC matrices after the perturbation of an original Remitting case, distributed by node and amplitude of said perturbation. Yellow cases indicated FC matrices reclassified as relapsing cases, green cases indicate FC matrices classified again as remitting, and dark blue cases indicated FC matrices that after the perturbation have been classified as control cases; b) Confidence associated to each new classified value after perturbation for each amplitude and node.

In Figure 7 a) we can observe that after the indicated perturbations, most of the new generated FC matrices coming from a relapsing case get reclassified as this same class. Moreover, we can observe some nodes that at any intensity of the perturbation or for most of its values they accomplish a reclassification of this condition into a remitting case, which indicates an improvement in the condition. We can even observe that for one of the nodes at some specific amplitude of the perturbation, a reclassification of the relapsing condition into control is also accomplished. Furthermore, in Figure 7 b) we can observe the confidence associated to each of this new labels after the perturbation. In this case, we observe lower probabilities of the new label being the correct one after the perturbation, however still high enough to be considered relevant.

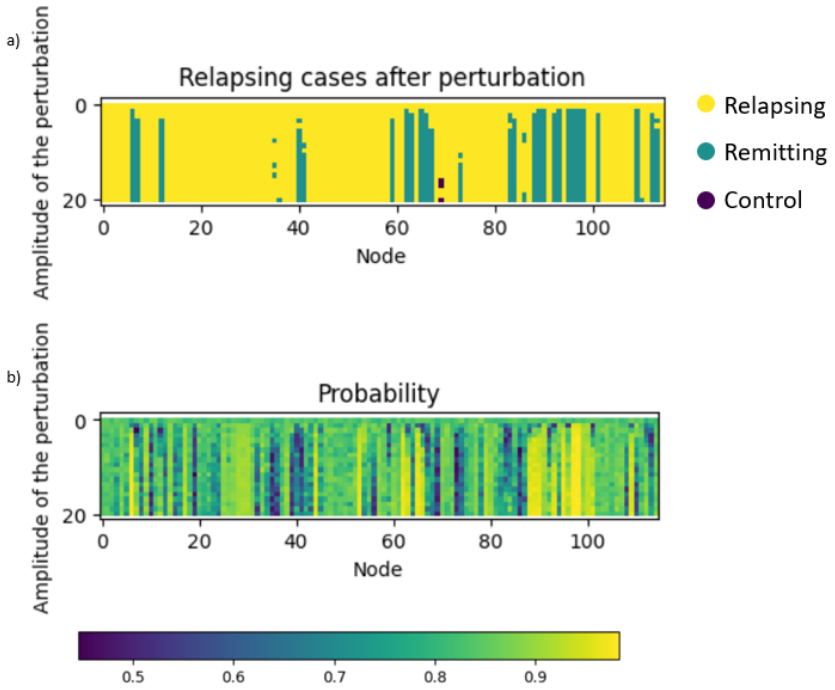

Figure 7: a) Resulting classification of the FC matrices after the perturbation of an original relapsing case, distributed by node and amplitude of said perturbation. Yellow cases indicated FC matrices classified again as relapsing cases, green cases indicate FC matrices reclassified as remitting, and dark blue cases indicated FC matrices that after the perturbation have been classified as control cases; b) Probability associated to each new classified value after perturbation for each amplitude and node.

In Figure 8 we can observe more clearly the brain areas associated with each of the mentioned changes. This visualization highlights the relation between the perturbations applied to different regions of the left hemisphere of the brain and an improvement of the psychotic condition in both originally remitting and relapsing patients. Furthermore, we can appreciate that there is no overlapping between the regions that have been associated with improvements in the psychotic condition and those indicated with changes from remitting to relapsing conditions. As the changes associated with a worsening of the original case are focused on areas in the right hemisphere. The relevance of this brain regions and its relation to the psychotic condition will be discussed in the following section.

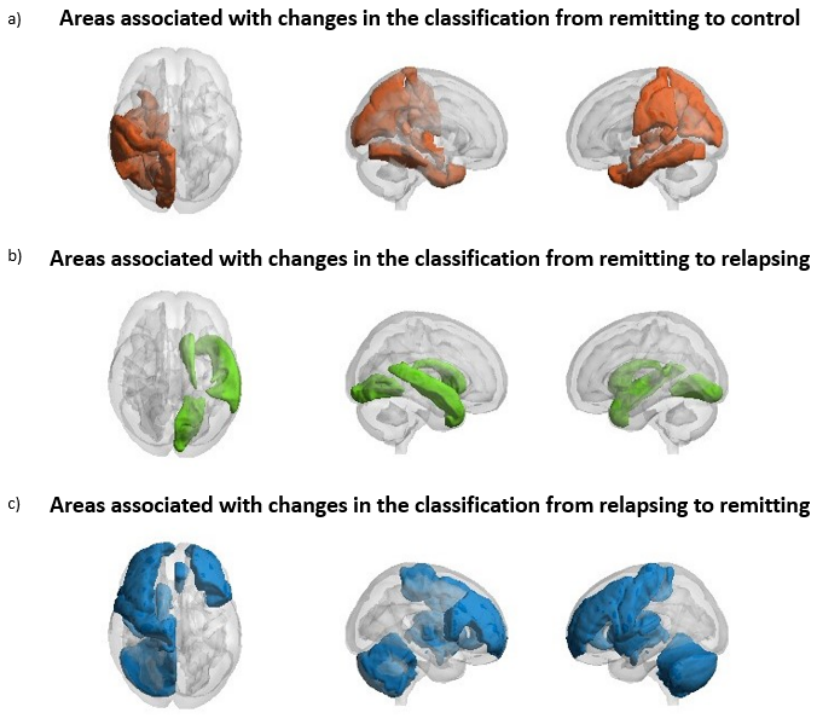

Figure 8: Representation of the brain areas associated with the different detected changes.

4 Discussion

This study is based on the use of phenomenological model from fMRI dynamics classified regarding the stage of psychosis for each individual in: remitting, relapsing or healthy controls. This was performed with the aim to train a VAE for the generation of a low dimensional latent space in which the classification of this different psychotic stages could be performed with high accuracy. Furthermore, the use of this deep learning architecture allowed the exploration of how FC matrices behaved after certain perturbations were applied to the original dynamics for each defined psychotic class. This allowed us to identify the key regions to possibly induce a change in the condition of psychotic patients.

As seen in the results, VAE architectures with different dimension can be successfully trained to classify the different defined classes regarding psychotic patients with high accuracy and enough statistical significance. After the training of various VAE models with different dimensions of their middle layers, we can conclude that the VAE model with the most optimal classification was the one generating a latent space of eight dimensions. This can be seen in Figure 4, and it leads us to conclude that an 8D latent space is good enough to perform the given task, as lower dimensions do not perform as accurate and the addition of more dimensions would consume more computational resources without this being translated to a better accuracy.

Then, the study of how the different psychotic cases would behave as perturbations were introduced in the model was studied, the results shown in Figure 6 and Figure 7. For both cases, we can appreciate that most of the applied perturbations do not

lead into a reclassification of the initial stage. However, if we focus on the originally relapsing case, the perturbation of specific nodes translates for most amplitudes into an improvement of the condition, reclassifying the case as a remitting one. There are also a couple of cases, in the originally relapsing example, in which the perturbation lead to a reclassification into control cases, however the confidence associated to these results was not high enough to consider them relevant. This shows us that there are some perturbations that could induce a remitting or healthier stated in the brain of those patients that have had multiple psychotic episodes.

On the other hand, in the originally classified remitting case, we can see two different kinds of behaviour after the perturbation:

- An improvement of the condition, some of the perturbations lead to a reclassification of remitting cases into a control cases, which could be translated as a disappearance of the symptomatology and leading towards a healthier state.

- And a worsening of the condition leading to its reclassification into a relapsing case after the perturbation. This shows us that there are some perturbations that lead to undesired results and could worsen the condition of patients.

Both changes happened associated with a confidence in this classification high enough to consider them relevant.

Knowing this information and the brain structures to which each of the 115 nodes of the brain parcellation are associated to, we can find those areas in the brain that are more prone to induce changes in the condition of patients and lead to healthier states.

- The areas found that are related to a change from a relapsing to a remitting state are mostly associated with structures in the cortex such as some frontal and left temporal areas, a region of the cingulate and the left insula. Outside the cortex, the hippocampal tail and the left cerebellum are also associated with these changes too. In [10] and [12] the authors relate degeneration of the frontal and temporal areas to higher vulnerability to psychosis, and full onset of the disease and worse prognosis when accompanied also with alterations in the cingulate and cerebellar areas. Therefore, certain stimulation in those areas could mean a reversing of the condition, as seen by the obtained results.

- Furthermore, perturbations in the node corresponding to the left amygdala and thalamus have been associated with improvement in the condition to those remitting and relapsing cases. Leading remitting cases to healthy dynamics, and relapsing cases into remitting ones. This matches what is found in the literature where deteriorations in the limbic system, which the amygdala is part of, are associated with chronic stages of psychosis [30].

- There were also found some areas in which perturbations lead to control dynamics from remitting cases, those were areas in the cortex in the left temporal lobe and near the hippocampus. Which, as stated, have been previously asso-

ciated with psychosis though the literature.

- However, some changes also were found that worsen the condition after the application of the perturbations. These changes were associated to nodes in the right temporal regions, leading from an original remitting case to a relapsing one. In [10] it is seen that the degeneration of the temporal region is one of the main indicators associated with the psychotic conditions, therefore it is logical to believe that its prolonged and severe perturbation could lead to a worsening of the condition, even if some changes could produce sporadic improvement.

- And, it has to be highlighted that no brain area has been found to produce both worsening and improvement of the psychotic condition simultaneously. It is true, that the temporal lobe has been related to both changes, however the perturbations in the left hemisphere indicated changes towards healthier states, while perturbations in the right one produced a reclassification into the relapsing category from remitting cases.

In conclusion, this work proves that the methodology proposed in [43] and [45] for the classification of consciousness states and also used in the distinction between different types of dementia in [39] can further be adapted for classifying psychotic stages depending on the number of episodes patients have had. Furthermore, it also demonstrates that the generation of a low dimensional space based on VAE architecture is possible for the classification of control patients, remitting psychotic cases with just one episode and relapsing psychotic cases with several episodes. And that the addition of perturbations to these models can help us identify several areas in the brain that are heavily linked with the severity of the psychotic condition and could lead to changes in it if perturbed.

4.1 Limitations and Future Work

The main limitations of this research are regarding the use of computational resources and synthetic data. As stated in the methodology, both the VAE and classification model were trained with 3000 synthetic cases. This could mean that the dataset used may not be as good of a representation for real cases as expected, because of its synthetic nature. Even if this methodology has been proven useful for other conditions, having access to more empirical cases would be ideal to further validate the results.

Moreover, the dataset originally used to generate the synthetic data comes from patients between 20 and 30 years old, with higher representation of male participants and from the same hospital in Switzerland as seen in [41]. All this factors lead to bias in the creation of the synthetic dataset and the training of the VAE and NN for classification, meaning that the results are more representative for patients with the same age range, biological sex and from the same region than for others with different characteristics.

Also, the accuracy of the prediction is not only affected by these biases in the dataset,

but also by the number of cases in which it was trained with. Due to computational resources, both the VAE and the NN models for the creation of the latent space and the classification of the data were trained with 3000 case, as the use of a larger dataset would have not been viable due to the time and computational limitations. Moreover, this also conditioned that more repetitions of the classification of the perturbed results could not be performed. These would have been proven useful to see possible mistakes in the classification of the perturbed FC matrices, however due to the time these calculations took it was not possible to perform such confirmation. Therefore, further validation with bigger datasets and more repetitions after the perturbation would have help us to prove the consistency in the results.

Furthermore, this work could be extended to also study the effect that other kind of perturbations have over this kind of models. Or, even, find perturbation models that resemble the pharmacological modulation or other lines of treatment in a more precise manner to enhance the precision of this models in the prediction of changes in the psychotic condition. Future work could also be focused on the search for other approaches in the classification of psychotic stages, for example, using Stage IIIa of the original dataset to study how cases where the symptomatology persist, but no second episode occurs after the FEP condition the creation of the latent space, affect the classification of the classes and how they would behave after perturbations are introduced.

Bibliography

1. Arciniegas, D. B. Psychosis. en. *Continuum (Minneap. Minn.)* **21,** 715–736 (2015).
2. World Health Organization. *The ICD-10 classification of mental and behavioural disorders : clinical descriptions and diagnostic guidelines* en (World Health Organization, Genève, Switzerland, 1992).
3. Griswold, K. S., Del Regno, P. A. & Berger, R. C. Recognition and differential diagnosis of psychosis in primary care. en. *Am. Fam. Physician* **91,** 856–863 (2015).
4. *Overview - psychosis* en. https://www.nhs.uk/mental-health/conditions/psychosis/overview/. Accessed: 2023-1-19.
5. McCleery, A. & Nuechterlein, K. H. Cognitive impairment in psychotic illness: prevalence, profile of impairment, developmental course, and treatment considerations. en. *Dialogues Clin. Neurosci.* **21,** 239–248 (2019).
6. Radua, J. *et al.* What causes psychosis? An umbrella review of risk and protective factors. en. *World Psychiatry* **17,** 49–66 (2018).
7. Green, M. F. *et al.* Social disconnection in schizophrenia and the general community. en. *Schizophr. Bull.* **44,** 242–249 (2018).
8. Van Os, J., Rutten, B. P. & Poulton, R. Gene-environment interactions in schizophrenia: review of epidemiological findings and future directions. en. *Schizophr. Bull.* **34,** 1066–1082 (2008).
9. Grace, A. A. Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. en. *Nat. Rev. Neurosci.* **17,** 524–532 (2016).
10. Fusar-Poli, P., Radua, J., McGuire, P. & Borgwardt, S. Neuroanatomical maps of psychosis onset: voxel-wise meta-analysis of antipsychotic-naive VBM studies. en. *Schizophr. Bull.* **38,** 1297–1307 (2012).
11. Bartholomeusz, C. F. *et al.* Structural neuroimaging across early-stage psychosis: Aberrations in neurobiological trajectories and implications for the staging model. en. *Aust. N. Z. J. Psychiatry* **51,** 455–476 (2017).
12. Niznikiewicz, M. A. Neurobiological approaches to the study of clinical and genetic high risk for developing psychosis. en. *Psychiatry Res.* **277,** 17–22 (2019).
13. Joyce, E. M. Organic psychosis: The pathobiology and treatment of delusions. en. *CNS Neurosci. Ther.* **24,** 598–603 (2018).
14. Cooper, R. E., Laxhman, N., Crellin, N., Moncrieff, J. & Priebe, S. Psychosocial interventions for people with schizophrenia or psychosis on minimal or no antipsychotic medication: A systematic review. en. *Schizophr. Res.* **225,** 15–30 (2020).
15. McGorry, P. *et al.* Biomarkers and clinical staging in psychiatry. en. *World Psychiatry* **13,** 211–223 (2014).
16. McGorry, P. D., Killackey, E. & Yung, A. Early intervention in psychosis: concepts, evidence and future directions. en. *World Psychiatry* **7,** 148–156 (2008).

17. Lee, R. *et al.* Prediction models in first-episode psychosis: systematic review and critical appraisal. en. *Br. J. Psychiatry* **220,** 1–13 (2022).

18. Sullivan, S. *et al.* Models to predict relapse in psychosis: A systematic review. en. *PLoS One* **12,** e0183998 (2017).

19. Alvarez-Jiménez, M., Parker, A. G., Hetrick, S. E., McGorry, P. D. & Gleeson, J. F. Preventing the second episode: a systematic review and meta-analysis of psychosocial and pharmacological trials in first-episode psychosis. en. *Schizophr. Bull.* **37,** 619–630 (2011).

20. Salazar de Pablo, G. *et al.* Implementing precision psychiatry: A systematic review of individualized prediction models for clinical practice. en. *Schizophr. Bull.* **47,** 284–297 (2021).

21. Rosen, M. *et al.* Towards clinical application of prediction models for transition to psychosis: A systematic review and external validation study in the PRONIA sample. en. *Neurosci. Biobehav. Rev.* **125,** 478–492 (2021).

22. Studerus, E., Ramyead, A. & Riecher-Rössler, A. Prediction of transition to psychosis in patients with a clinical high risk for psychosis: a systematic review of methodology and reporting. en. *Psychol. Med.* **47,** 1163–1178 (2017).

23. Cannon, T. D. *et al.* An individualized risk calculator for research in prodromal psychosis. en. *Am. J. Psychiatry* **173,** 980–988 (2016).

24. Fusar-Poli, P. *et al.* Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. en. *JAMA Psychiatry* **74,** 493–500 (2017).

25. Vigod, S. N. *et al.* READMIT: a clinical risk index to predict 30-day readmission after discharge from acute psychiatric units. en. *J. Psychiatr. Res.* **61,** 205–213 (2015).

26. Bınbay, T., Ergül, C. & van Os, J. Symptomatic remission along the clinical psychosis spectrum: A historical and conceptual review. en. *Noro Psikiyatr. Ars.* **58,** S3–S6 (2021).

27. Hamilton, H. K., Roach, B. J. & Mathalon, D. H. Forecasting remission from the psychosis risk syndrome with mismatch negativity and P300: Potentials and pitfalls. en. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **6,** 178–187 (2021).

28. Bois, C., Whalley, H. C., McIntosh, A. M. & Lawrie, S. M. Structural magnetic resonance imaging markers of susceptibility and transition to schizophrenia: a review of familial and clinical high risk population studies. en. *J. Psychopharmacol.* **29,** 144–154 (2015).

29. Dandash, O. *et al.* Altered striatal functional connectivity in subjects with an at-risk mental state for psychosis. en. *Schizophr. Bull.* **40,** 904–913 (2014).

30. Canu, E., Agosta, F. & Filippi, M. A selective review of structural connectivity abnormalities of schizophrenic patients at different stages of the disease. en. *Schizophr. Res.* **161,** 19–28 (2015).

31. Wood, S. J., Yung, A. R., McGorry, P. D. & Pantelis, C. Neuroimaging and treatment evidence for clinical staging in psychotic disorders: from the at-risk mental state to chronic schizophrenia. en. *Biol. Psychiatry* **70,** 619–625 (2011).

32. Koutsouleris, N. *et al.* Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. en. *Lancet Psychiatry* **3,** 935–946 (2016).

33. Amoretti, S. *et al.* Cognitive clusters in first-episode psychosis. en. *Schizophr. Res.* **237,** 31–39 (2021).

34. Zlatintsi, A. *et al.* E-prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures. en. *Sensors (Basel)* **22,** 7544 (2022).

35. Fornito, A., Bullmore, E. T. & Zalesky, A. Opportunities and challenges for psychiatry in the connectomic era. en. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **2,** 9–19 (2017).

36. Anticevic, A. *et al.* Association of thalamic dysconnectivity and conversion to psychosis in youth and young adults at elevated clinical risk. en. *JAMA Psychiatry* **72,** 882–891 (2015).

37. Cofré, R. *et al.* Whole-brain models to explore altered states of consciousness from the bottom up. en. *Brain Sci.* **10,** 626 (2020).

38. Ipiña, I. P. *et al.* Modeling regional changes in dynamic stability during sleep and wakefulness. en. *Neuroimage* **215,** 116833 (2020).

39. Perl, Y. S. *et al. Model-based whole-brain perturbational landscape of neurodegenerative diseases* 2022.

40. Arbabyazd, L. *et al. Virtual connectomic datasets in Alzheimer's Disease and aging using whole-brain network dynamics modelling* 2020.

41. Griffa, A. *et al.* Brain connectivity alterations in early psychosis: from clinical to neuroimaging staging. en. *Transl. Psychiatry* **9,** 62 (2019).

42. Deco, G., Kringelbach, M. L., Jirsa, V. K. & Ritter, P. The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. en. *Sci. Rep.* **7,** 3095 (2017).

43. Perl, Y. S. *et al.* Generative embeddings of brain collective dynamics using variational autoencoders. en. *Phys. Rev. Lett.* **125,** 238101 (2020).

44. Inc., T. M. *MATLAB version: 9.13.0 (R2022b)* Natick, Massachusetts, United States, 2022. https://www.mathworks.com.

45. Perl, Y. S. *et al.* Data augmentation based on dynamical systems for the classification of brain states. en. *Chaos Solitons Fractals* **139,** 110069 (2020).

Appendix A

Detailed information of which brain area each node represents and if its perturbation has been associated to changes in the condition of remitting and relapsing patients.

				Area found in changes in classification from		
Node	Hemisphere	Region	Area	Remitting to Control	Remitting to Relapsing	Relapsing to Remitting
1	Right	Cortex	lateral orbitofrontal			
2	Right	Cortex	parsorbitalis			
3	Right	Cortex	frontal pole			
4	Right	Cortex	medial orbitofrontal			
5	Right	Cortex	parstriangularis			
6	Right	Cortex	parsopercularis			x
7	Right	Cortex	rostral middle frontal			x
8	Right	Cortex	superior frontal			
9	Right	Cortex	caudal middle frontal			
10	Right	Cortex	precentral			
11	Right	Cortex	paracentral			
12	Right	Cortex	rostral anterior cingulate			x
13	Right	Cortex	caudal anterior cingulate			
14	Right	Cortex	posterior cingulate			
15	Right	Cortex	postcentral			
16	Right	Cortex	supramarginal			
17	Right	Cortex	superior parietal			
18	Right	Cortex	inferior parietal			
19	Right	Cortex	precuneus			
20	Right	Cortex	cuneus			
21	Right	Cortex	pericalcarine			
22	Right	Cortex	lateraloccipital			
23	Right	Cortex	lingual		x	
24	Right	Cortex	fusiform			
25	Right	Cortex	parahippocampal			
26	Right	Cortex	entorhinal			
27	Right	Cortex	temporal pole		x	
28	Right	Cortex	inferior temporal			
29	Right	Cortex	middle temporal			
30	Right	Cortex	bankssts			
31	Right	Cortex	superior temporal		x	
32	Right	Cortex	transversetemporal			
33	Right	Cortex	insula			
34	Right	Thalamo	pulvinar			
35	Right	Thalamo	anterior			
36	Right	Thalamo	mediodorsal			

				Area found in changes in classification from		
Node	Hemisphere	Region	Area	Remitting to Control	Remitting to Relapsing	Relapsing to Remitting
37	Right	Thalamo	ventralaterodorsal			
38	Right	Thalamo	intralaminar nuclei medial pulvinar			
39	Right	Thalamo	ventral anterior			
40	Right	Thalamo	ventral_latero_ventral			
41	Right	Subcortical	caudate		x	
42	Right	Subcortical	putamen			
43	Right	Subcortical	pallidum			
44	Right	Subcortical	accumbens area			
45	Right	Subcortical	amygdala			
46	Right	Subcortical	hippocampus			
47	Right	Hippocampus	presubiculum			
48	Right	Hippocampus	subiculum			
49	Right	Hippocampus	Cornu Ammonis 1 (CA1)			
50	Right	Hippocampus	Cornu Ammonis 4 (CA4)			
51	Right	Hippocampus	Granule Cell layer of the Dentate Gyrus (GCDG)			
52	Right	Hippocampus	molecular layer			
53	Right	Hippocampus	hippocampal tail			
54	Right	Vental Diencephalon (VDC)	VDC			
55	Right	Hypothalamus	hypothalamus			
56	Right	Cerebellum	cerebellum			
57	Left	Cortex	lateral orbitofrontal			
58	Left	Cortex	parsorbitalis			
59	Left	Cortex	frontal pole			x
60	Left	Cortex	medial orbitofrontal			
61	Left	Cortex	parstriangularis			
62	Left	Cortex	parsopercularis			x
63	Left	Cortex	rostral middle frontal			x
64	Left	Cortex	superiorfrontal			
65	Left	Cortex	caudal middle frontal			x
66	Left	Cortex	precentral			x
67	Left	Cortex	paracentral			x
68	Left	Cortex	rostral anterior cingulate			
69	Left	Cortex	caudal anterior cingulate			
70	Left	Cortex	posterior cingulate			
71	Left	Cortex	postcentral	x		
72	Left	Cortex	supramarginal	x		
73	Left	Cortex	superior parietal			
74	Left	Cortex	inferior parietal			

				Area found in changes in classification from		
Node	Hemisphere	Region	Area	Remitting to Control	Remitting to Relapsing	Relapsing to Remitting
75	Left	Cortex	precuneus	x		
76	Left	Cortex	cuneus	x		
77	Left	Cortex	pericalcarine	x		
78	Left	Cortex	lateral occipital			
79	Left	Cortex	lingual			
80	Left	Cortex	fusiform	x		
81	Left	Cortex	parahippocampal	x		
82	Left	Cortex	entorhinal			
83	Left	Cortex	temporal pole	x		
84	Left	Cortex	inferior temporal	x		
85	Left	Cortex	middle temporal			
86	Left	Cortex	bankssts	x		
87	Left	Cortex	superior temporal			
88	Left	Cortex	transversetemporal			x
89	Left	Cortex	insula			x
90	Left	Thalamo	pulvinar	x		x
91	Left	Thalamo	anterior			
92	Left	Thalamo	medio dorsal			x
93	Left	Thalamo	ventral_latero_dorsal			x
94	Left	Thalamo	intrlaminar_nuclei_ medial			
95	Left	Thalamo	ventral_anterior			x
96	Left	Thalamo	ventral_latero_ventral	x		x
97	Left	Subcortical	caudate			x
98	Left	Subcortical	putamen			x
99	Left	Subcortical	pallidum			
100	Left	Subcortical	accumbens area			
101	Left	Subcortical	amygdala	x		x
102	Left	Subcortical	hippocampus			
103	Left	Hippocampus	presubiculum			
104	Left	Hippocampus	subiculum			
105	Left	Hippocampus	CA1			
106	Left	Hippocampus	CA4			
107	Left	Hippocampus	GCDG			
108	Left	Hippocampus	molecular layer			
109	Left	Hippocampus	hippocampal tail			x
110	Left	VDC	VDC	x		
111	Left	Hypothalamus	hypothalamus			
112	Left	Cerebellum	cerebellum			x

				Area found in changes in classification from		
Node	Hemisphere	Region	Area	Remitting to Control	Remitting to Relapsing	Relapsing to Remitting
113			brain-stem-midbrain			
114			brain-stem-pons			
115			brain-stem-scp			