POLITECNICO DI MILANO

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

TESI DI LAUREA MAGISTRALE IN INGEGNERIA BIOMEDICA

METODI STATISTICI DI ANALISI DELL'IMPATTO GENETICO E AMBIENTALE SUL NEUROSVILUPPO. STUDIO MRI A 3T SU COPPIE DI GEMELLI

Relatore: Prof. Sergio CERUTTI

Correlatori: Prof.ssa Anna Maria BIANCHI,

Prof. Paolo BRAMBILLA

Dott.ssa Ing. Eleonora MAGGIONI, Ph.D.

Tesi di laurea di:
Adriana AMBROSINO,
Matr. 884189

Anno accademico 2018/2019

A mia madre, a mio padre
e a mia sorella Federica.

Ringraziamenti

La presente tesi è stata svolta presso il Dipartimento di Elettronica, Informatica e Bioingegneria B^3 Lab (Biosegnali-Bioimmagini-Bioinformatica) del Politecnico di Milano, in collaborazione con la Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano.

Vorrei esprimere sentiti ringraziamenti al Professor Sergio Cerutti, relatore della presente tesi, che mi ha dato l'opportunità di svolgere questo lavoro nell'affascinante ambito delle immagini biomediche affidandomi alla splendida Dott.ssa Ing. Eleonora Maggioni, correlatrice di questa tesi, alla quale sono immensamente grata. Non finirò mai di ringraziarla per la sua disponibilità, per il suo aiuto, per le sue rassicurazioni e soprattutto per le opportunità di crescita che mi ha dato durante tutti questi mesi. Vorrei, inoltre, ringraziare la Professoressa Anna Maria Bianchi e il Professor Paolo Brambilla, anche loro correlatori di questa tesi, per i loro preziosi consigli.

Infine, mi sento di ringraziare la mia famiglia alla quale devo la possibilità di aver studiato presso questo brillante Ateneo. Vi ringrazio dal profondo del mio cuore per tutto l'amore che mi donate quotidianamente, per aver sempre creduto in me ed essermi sempre stati vicini.

Abstract

In the last decades, the research for the neurobiological bases of mental illnesses has involved multidisciplinary fields, including psychopathology, genetics, functional and structural *brain imaging* as well as their integration (e.g. imaging-genetics). *Brain imaging* has undoubtedly played a leading role in allowing the study of functional and structural aspects of brain structures in vivo, making use of numerous methods of image acquisition including the powerful and non-invasive Magnetic Resonance Imaging (MRI). MRI is a radiological technique that in medicine allows to study anatomical and functional aspects of the human body by exploiting the principle of Nuclear Magnetic Resonance. MRI represents the main method of investigation in *brain imaging*, with a history dating back to the early 1980s, when for the first time MRI studies were performed on patients suffering from bipolar disorder (Rangel- Guerra et al. ,1983) and on schizophrenic patients (Smith et al. ,1984), and which continues with constant technological improvements and increasingly complex fields of use.

In the present experimental thesis, 3T MRI has been used to investigate the complex determinants of structural brain development. The MRI study involved a unique sample of 59 young couples of twins, of which 25 monozygotic (MZ) and 34 dizygotic (DZ) from Northern Italy. The experimental work concerned 1) the development of algorithms for the estimation of brain morphological measures and 2) the implementation of statistical methods for twin designs, with the ultimate goal of estimating the genetic and environmental impacts on a set of brain morphological measures during neurodevelopment. To this end, sophisticated *region-based* and *voxel-based* neuroanatomical analysis techniques were applied, which allowed to extract the morphological measures of interest. These measures (phenotypes) were then compared within each pair of twins and between MZ and DZ couples to estimate genetic and environmental effects.

The structural MRI images were processed using powerful and widely used neuroimaging softwares, FreeSurfer and SPM 12.

The first branch of investigation involved, through a *region-based* approach, the estimation of heritability, i.e. the portion of inter-individual variance linked to genetic differences, of regional neuroanatomical parameters. Furthermore, the univariate ACE twin model was

developed to identify the extent to which genetic factors and environmental factors (shared and unique) influenced the investigated parameter, respectively. The neuroanatomical similarity / diversity of twin couples was investigated using an alternative approach to the classical method, based on multiple linear regression. This phase of analysis has involved the creation of the multiple generalized linear regression model in order to evaluate the relative influence of zygosity, age and sex (independent variables) on intra-couple differences in terms of neuroanatomy (dependent variables).

Furthermore, a complementary *voxel-based* approach was used for automatic classification of zygosity. This analysis involved the implementation of Machine Learning algorithms, in the software NeuroMiner, capable of predicting the twins' zygosity starting from the intra-couple punctual difference (voxel by voxel) of gray and white matter tissue maps. This method identifies the brain clusters that can help differentiate between MZ and DZ twins based on the similarity in tissue volume, thus providing information on genetic influences on the local volumes of gray and white matter. Such investigations represent an additional piece to the understanding of the relative role of genes and environment on intra-couple neuroanatomical similarity during development.

Another research topic of the thesis regarded the preliminary identification of brain regional *markers* of psychotic risk. Indeed, twins represent the ideal sample for exploring genetic and environmental effects on such markers. Information concerning the psychotic risk of each subject was obtained on the basis of the ERIraos-CL scale. We identified 13 couples discordant for psychotic risk and selected 13 age-matched singletons, chosen from the twin pairs without psychotic risk, considered as healthy controls. Discordant twins and healthy subjects were compared in terms of neuroanatomical parameters in order to 1) identify the brain regions characterized by anatomical alterations due to the risk of manifesting psychosis, 2) speculating about the genetic / environmental determinants of these alterations. Algorithms have been created to carry out the following statistical comparisons for each measure: (1) comparison between high-risk twins and low-risk twins of discordant couples (2) comparison between high-risk twins and control subjects (3) comparison between low-risk twins and control subjects.

Heritability analyses showed that the *regional volume* is significantly heritable (*p-value* <0.05) in a larger number of brain regions (equal to 19) compared to the other morphological parameters analyzed. Therefore, it is reasonable to hypothesize that volume is the

morphological measure most widely influenced by genetic factors. On the contrary, the *mean curvature* emerged as the least heritable measure, in fact it is significantly inherited only in three cortical regions (frontal pole of the left hemisphere, lingual cortex of the right hemisphere and triangular part of the left inferior frontal gyrus). Thus, except in these three brain areas, it is likely that the mean curvature is less influenced by genetic factors and more by environmental factors. Furthermore, correlation analysis for *surface area*, *cortical thickness* and *volume* parameters showed a significant correlation in both groups of zygosity greater in the brain regions of the right hemisphere than the brain regions of the left hemisphere, while regarding *mean curvature* correlation analysis showed for MZ pairs a greater correlation in brain areas of the right hemisphere than left hemisphere and for DZ pairs a greater correlation in brain areas of the left hemisphere than the right one. The ACE model highlighted a contribution of common environment, shared by the twins on the *volume* of many *subcortical regions*, which unlike the cortical regions seem to be less affected by genetic and unique environmental factors.

The results obtained by the different approaches show that *the volume of white matter in the cerebellar cortex* is defined by genetic factors rather than by environmental factors. In addition, the *surface area value of the mean temporal gyrus of the left hemisphere* is also largely attributable to genetic factors. Furthermore, the analyses revealed an important aspect regarding the impact of development on intra-couple similarity. In fact, irrespective of the zygosity of the couple, some brain regions showed increased neuroanatomical similarity with age, other brain regions showed reduced similarity with age. This result could presumably be due to genetic or environmental factors that come into play during growth and do influence the twins in a similar way causing greater similarity, or influence them differently, accentuating diversity.

The preliminary analyses on neuroanatomical alterations in the presence of psychotic risk allowed to identify the *mean curvature* and the *volume* as the parameters most affected by alterations in the presence of nascent psychosis. Given that for these two parameters heritability scores were found to be low and high respectively, this result corroborates the thesis that psychopathology is largely influenced by both environmental and genetic factors. Indeed, it has been observed that in at-risk twins these parameters have higher values than in the control twins. Differences in the values were evident both in the twins that are considered to be at psychotic risk for the ERIraos-CL scale and in their siblings who form

each discordant couple. The obtained results could represent neurobiological markers for the risk of developing psychosis, allowing the early recognition of subjects at risk and the planning of adequate preventive strategies.

Regarding the analyses implemented in NeuroMiner (a Machine Learning software), from the comparison between the results obtained using the intra-couple punctual difference (voxel by voxel) of gray and white matter tissue maps emerges a higher accuracy of the classifier that bases the prediction of zygosity on the punctual difference of white matter tissue maps than the classifier that bases the prediction of zygosity on the punctual difference of gray matter tissue maps. The classifier appeared able to identify similarities / differences in white matter values and to attribute the correct zygosity to new images. The complementary *voxel-based* classification analysis suggested the presence of significant genetic influences in the *local white matter volumes*, on the basis of which it was possible to predict the zygosity of a new pair. In particular, the *volume of white matter in the cerebellar cortex* was more similar in the MZ twins than in the DZ twins and determinant in the prediction.

The results obtained in this thesis work provide quantitative information regarding the extent to which genes and environment influence brain morphology in the delicate phase of development. This discrimination is of fundamental importance in the clinical context, for the understanding of pathologies associated with brain morphological alterations. The presence of a relevant environmental impact on the expression of the parameter can be useful for the development of preventive, personalized and innovative therapeutic strategies.

Sommario

Negli ultimi decenni la ricerca per le basi neurobiologiche sulle malattie mentali ha coinvolto ambiti multidisciplinari, che comprendono la psicopatologia, la genetica, il *brain imaging* funzionale e strutturale e la loro integrazione (ad esempio, il *genetic imaging*). Il *brain imaging* ha avuto senza dubbio un ruolo di prim'ordine consentendo lo studio di aspetti funzionali e strutturali di strutture cerebrali in vivo avvalendosi di numerosi metodi di acquisizione delle immagini tra cui il potente e non invasivo *Imaging* a Risonanza Magnetica (MRI). L'MRI è una tecnica radiologica che in medicina consente di studiare aspetti anatomici e funzionali del corpo umano sfruttando il principio della Risonanza Magnetica Nucleare. L'MRI rappresenta il principale metodo di indagine in *brain imaging*, con una storia che risale agli inizi degli anni '80, quando per la prima volta furono eseguiti studi MRI su pazienti affetti da disturbo bipolare (Rangel- Guerra et al. ,1983) e su pazienti schizofrenici (Smith et al. ,1984), e che continua con costanti miglioramenti tecnologici e campi di utilizzo sempre più complessi.

Nella presente tesi sperimentale, l'MRI a 3T è stata utilizzata per investigare i complessi determinanti dello sviluppo della struttura celebrale. Lo studio MRI ha coinvolto un campione unico composto da ben 59 giovani coppie di gemelli, dei quali 25 monozigoti (MZ) e 34 dizigoti (DZ) provenienti dal Nord Italia. Il lavoro sperimentale ha riguardato 1) l'implementazione di algoritmi per la stima di parametri morfologici celebrali e 2) l'implementazione di metodi statistici gemellari, con l'obiettivo ultimo di stimare l'impatto genetico ed ambientale su un certo numero di parametri morfologici celebrali nel neurosviluppo. A tale scopo sono state applicate sofisticate tecniche di analisi neuroanatomica *region-based* e *voxel-based*, che hanno consentito di estrarre le misure morfologiche di interesse. Queste misure (fenotipi) sono state poi confrontate in ogni coppia di gemelli e tra coppie di gemelli MZ e DZ per stimare gli effetti genetici ed ambientali.

Le immagini MRI strutturali sono state analizzate usando software ampiamente utilizzati in neuroimaging quali FreeSurfer e SPM 12.

Il primo ramo di indagine ha previsto, mediante un approccio *region-based*, la stima dell'ereditabilità, intesa come porzione di varianza inter-individuale legata a differenze genetiche, di parametri neuroanatomici regionali. Inoltre, è stato sviluppato il modello

gemellare univariato ACE al fine di identificare in che misura, rispettivamente, i fattori genetici e i fattori ambientali (condivisi e unici) influenzino i valori del parametro investigato. La similarità / diversità neuroanatomica delle coppie è stata analizzata applicando un approccio regionale alternativo al metodo classico, basato sulla regressione lineare multipla. Questa fase di analisi ha previsto la creazione del modello di regressione lineare generalizzata multipla per valutare l'influenza relativa delle variabili zigosità, età e sesso (variabili indipendenti) sulle differenze intra-coppia in termini di neuroanatomia (variabili dipendenti).

Inoltre, è stato usato un approccio complementare *voxel-based* per la classificazione automatica della zigosità. Questa analisi ha previsto l'implementazione di algoritmi di Machine Learning, nel software NeuroMiner, in grado di predire la zigosità dei gemelli a partire da un'immagine ottenuta dalla differenza puntuale (voxel per voxel) intra-coppia delle mappe tessutali di materia grigia e bianca. Tale metodo identifica cluster celebrali che svolgono un ruolo determinante nella differenziazione tra gemelli MZ e DZ basandosi sulla similarità nei volumi locali di materia grigia e bianca, quindi forniscono informazioni sulle influenze genetiche su tali volumi. Tali indagini rappresentano un tassello aggiuntivo alla comprensione del ruolo relativo di geni e ambiente sulla similarità neuroanatomica intra-coppia durante lo sviluppo.

Un altro tema di ricerca di questa tesi ha riguardato l'identificazione preliminare di *marker* celebrali di rischio psicotico. Infatti, i gemelli rappresentano il campione ideale per esplorare gli effetti genetici e ambientali su tali marker. Le informazioni riguardanti il rischio psicotico di ciascun soggetto sono state ottenute sulla base della scala *ERIraos-CL*. Sono state identificate 13 coppie discordanti per rischio psicotico e 13 gemelli di età comparabile, scelti tra le coppie concordanti per l'assenza di rischio, considerati come soggetti di controllo. I gemelli discordanti e i soggetti di controllo sono stati confrontati in termini di parametri neuroanatomici per 1) identificare le regioni celebrali caratterizzate da alterazioni anatomiche dovute al rischio di manifestare psicosi, 2) fare ipotesi circa i determinanti genetici / ambientali di queste alterazioni. Sono stati sviluppati algoritmi per effettuare i seguenti confronti statistici per ogni parametro: (1) Confronto tra i valori nei gemelli ad alto rischio e nei gemelli a basso rischio delle coppie discordanti (2) Confronto tra i valori nei gemelli ad alto rischio e nei soggetti di controllo (3) Confronto tra i valori nei gemelli a basso rischio e nei soggetti di controllo.

Le analisi di ereditabilità hanno mostrato che il *volume regionale* è significativamente ereditabile (*p-value<0.05*) in un numero maggiore di regioni celebrali (pari a 19) rispetto agli altri parametri morfologici analizzati. Pertanto, è ragionevole ipotizzare che il volume sia il parametro morfologico più largamente influenzato da fattori genetici. Al contrario, la *curvatura media* è risultato essere il parametro meno ereditabile infatti è significativamente (*p-value<0.05*) ereditabile solo in tre regioni corticali (polo frontale dell'emisfero sinistro, corteccia linguale dell'emisfero destro e parte triangolare della circonvoluzione frontale inferiore sinistra). Infatti, eccetto che in queste tre aree celebrali, è presumibile che la curvatura media assuma valori meno influenzati da fattori genetici e maggiormente da fattori ambientali. Inoltre, l'analisi di correlazione per i parametri *area superficiale, spessore corticale e volume,* ha mostrato una correlazione significativa in entrambi i gruppi di zigosità maggiore nelle regioni celebrali dell'emisfero destro rispetto alle regioni celebrali dell'emisfero sinistro, mentre l'analisi di correlazione della *curvatura media* ha mostrato per i MZ una maggiore correlazione nelle aree celebrali dell'emisfero destro rispetto all'emisfero sinistro e per i DZ una maggiore correlazione nelle aree celebrali dell'emisfero sinistro rispetto al destro. Il modello ACE ha messo in luce un contributo dell'ambiente condiviso tra i gemelli sul *volume* di molte *regioni sottocorticali* che, diversamente dalle regioni corticali, sembrano essere meno affette da fattori genetici e dai fattori ambientali unici.

I risultati ottenuti mediante i diversi approcci mostrano che il *volume di materia bianca nella corteccia cerebellare* è definito dal patrimonio genetico più che da fattori ambientali. In aggiunta, anche il valore di *area superficiale della circonvoluzione temporale media dell'emisfero sinistro* è attribuibile in larga parte a fattori genetici. Inoltre, le analisi hanno messo in luce un aspetto importante riguardo l'impatto dello sviluppo sulla similarità intra-coppia. Infatti, indipendentemente dalla zigosità della coppia, alcune regioni celebrali hanno mostrato un aumento della similarità neuroanatomica con l'età, altre regioni celebrali hanno mostrato una riduzione della similarità con l'età. Questo risultato, potrebbe presumibilmente essere attribuibile a fattori di tipo genetico oppure a fattori di tipo ambientale che entrano in gioco durante la crescita che potrebbero condizionare i gemelli allo stesso modo causando una maggiore similarità, oppure condizionarli in maniera differente accentuandone la diversità.

Le analisi preliminari sulle alterazioni neuroanatomiche in presenza di rischio psicotico, hanno permesso di individuare la *curvatura media* ed il *volume* come parametri maggiormente affetti da alterazioni in presenza di psicosi nascente. Dato che per questi due parametri i punteggi di ereditabilità sono risultati essere rispettivamente bassi e alti, questo risultato avvalora la tesi che la psicopatologia è largamente influenzata sia da fattori ambientali che da fattori genetici. Infatti, è stato osservato che nei gemelli a rischio, questi parametri assumono valori maggiori rispetto ai valori nei gemelli di controllo. Le anomalie nei valori sono risultate evidenti sia nei gemelli che sono considerati a rischio psicotico per la scala ERIraos-CL che nei fratelli gemelli di questi soggetti che costituiscono ciascuna coppia discordante. I risultati ottenuti potrebbero rappresentare marker neurobiologici per il rischio di sviluppare psicosi, permettendo il riconoscimento precoce di soggetti a rischio e la pianificazione di strategie preventive adeguate.

Nell'ambito delle analisi implementate nel software di Machine Learning NeuroMiner, dal confronto tra i risultati ottenuti utilizzando le immagini di differenza puntuale intra-coppia di mappe tessutali di materia grigia e di materia bianca emerge un'accuratezza maggiore del classificatore che basa la predizione di zigosità sulle immagini di differenza puntuale intra-coppia di mappe tessutali di materia bianca. Il classificatore è apparso in grado di individuare similarità / differenze nei valori di materia bianca e di attribuire a nuove immagini la corretta zigosità. Questa complementare analisi di classificazione di tipo *voxel-based* ha suggerito la presenza di influenze genetiche significative nei *volumi locali* di materia bianca, sulla base dei quali è stato possibile predire la zigosità di una nuova coppia. In particolare, il *volume* di *materia bianca* nella *corteccia cerebellare* è risultato più simile nei gemelli MZ rispetto ai gemelli DZ e determinante nella predizione.

I risultati ottenuti nel presente lavoro di tesi forniscono informazioni di tipo quantitativo riguardanti la misura in cui geni e ambiente influenzano la morfologia nella delicata fase dello sviluppo celebrale. Tale discriminazione è di fondamentale importanza in ambito clinico, per la comprensione di patologie associate ad alterazioni morfologiche celebrali. La presenza di un impatto ambientale rilevante sull'espressione del parametro può essere utile per lo sviluppo di strategie terapeutiche preventive, personalizzate e innovative.

Indice

METODI STATISTICI DI ANALISI DELL'IMPATTO GENETICO E AMBIENTALE SUL NEUROSVILUPPO. STUDIO MRI A 3T SU COPPIE DI GEMELLI .....

ABSTRACT ..... **VI**

SOMMARIO ..... **XI**

LISTA DELLE FIGURE ..... **XX**

LISTA DELLE TABELLE ..... **XXV**

CAPITOLO 1 ..... **1**

INTRODUZIONE 1

SCOPO DELLA RICERCA 3

1. L'*IMAGING* A RISONANZA MAGNETICA 5

1.1 Hardware di un'apparecchiatura di Risonanza Magnetica Nucleare *6*

1.2 Principi fisici *7*

1.2.1 Momento angolare e momento magnetico *7*

1.2.2 Generazione della magnetizzazione macroscopica *8*

1.2.3 Magnetizzazione macroscopica netta *9*

1.2.4 Equazione di Larmor *9*

1.2.5 Caratteristiche di un impulso in radiofrequenza *11*

1.2.6 Processo di rilassamento *11*

1.2.7 Segnale Free Induction Decay *12*

1.3 Sequenze *14*

2. METODO GEMELLARE 15

2.1 Applicazione del metodo gemellare per lo studio dell'ereditabilità dell'anatomia celebrale *16*

3. IL RISCHIO PSICOTICO 17

4. IL MACHINE LEARNING IN MEDICINA 18

4.1 Diverse metodologie di apprendimento per una macchina *20*

4.2 La classificazione *21*

4.2.1 Fasi di sviluppo di un modello di classificazione *21*

CAPITOLO 2 ..... **22**

STATO DELL'ARTE 22

1. STORIA DEL METODO GEMELLARE 22

2. STORIA DELL'*IMAGING* A RISONANZA MAGNETICA 23

2.1 Ambiti di applicazione dell'imaging a risonanza magnetica *24*

2.2 L' imaging a risonanza magnetica a 3 Tesla *25*

2.3 Esempi di studi MRI condotti su gemelli *26*

CAPITOLO 3 ..... **29**

MATERIALI E METODI 29

1. RECLUTAMENTO DEI DATI DEI SOGGETTI E DEFINIZIONE DEL DATSET 29

2. ACQUISIZIONE DEI DATI SPERIMENTALI 30

2.1 Sequenza MPRAGE *31*

3. SOFTWARE PER LE ANALISI DELLE IMMAGINI NEUROANATOMICHE 33

4. ANALISI REGION-BASED 34

4.1 Pre-processing delle immagini con FreeSurfer *34*

4.1.1 Analisi qualitativa dell'elaborazione delle immagini con FreeSurfer *37*

4.1.2 Estrazione dei parametri di interesse da FreeSurfer *38*

4.2 Analisi statistiche *39*

4.2.1 Analisi di correlazione fenotipica intra-coppia *40*

4.2.2 Stima dell'ereditarietà dei parametri morfologici nelle diverse regioni celebrali *41*

4.2.3 Modello gemellare univariato: ACE *43*

4.2.4 Regressione lineare multipla *46*

4.2.4.1 Effetto della zigosità sulla differenza fenotipica intra-coppia *48*

4.2.4.2 Effetto dell'età sulla differenza fenotipica intra-coppia *49*

4.2.4.3 Effetto del sesso sulla differenza fenotipica intra-coppia *49*

4.2.5 Analisi pilota dell'impatto del rischio psicotico sulla neuroanatomia *50*

4.2.5.1 Selezione del campione di studio *50*

4.2.5.2 Metodi di analisi dell'effetto del rischio psicotico sulla neuroanatomia *50*

5. ANALISI VOXEL-BASED 52

5.1 Pre-elaborazione delle immagini *53*

5.1.1 Visualizzazione e riallineamento delle immagini con SPM 12 ..... *53*

5.1.2 Segmentazione dei tessuti celebrali con SPM 12 ..... *55*

5.1.3 Generazione del template anatomico con l'algoritmo DARTEL (Diffeomorphic Anatomical Registration using Exponentiated Lie algebra) *58*

5.1.4 Normalizzazione in spazio MNI (Montreal Neurological Istitute) del template DARTEL con SPM 12 61

5.2 Calcolo delle mappe di differenza tessutale intra-coppia *62*

5.2.1 Mappa di differenza tessutale intra-coppia (immagine delta) *62*

5.3 Analisi di Machine Learning *65*

5.3.1 Applicazione del Machine Learning nel metodo gemellare *65*

5.3.2 Inizializzazione in NeuroMiner *67*

5.3.3 LINEAR SUPPORT VECTOR MACHINES 68

5.3.4 Creazione dei modelli di classificazione delle immagini di materia bianca e grigia *73*

5.3.4.1 Calcolo dei pesi delle features nella procedura di classificazione *75*

5.3.4.2 Valutazione del modello di classificazione: Grafici ROC e Tabelle di *76*

4.3.5.3 Validazione esterna: predizione della classe target per una nuova immagine *77*

CAPITOLO 4 ..... **79**

RISULTATI E DISCUSSIONE 79

1. RISULTATI DELL'ANALISI DI CORRELAZIONE 79

2. ANALISI DI EREDITABILITÀ DEI PARAMETRI NEUROANATOMICI 94

3. RISULTATI MODELLO ACE 99

3.1 Discussione dei risultati *103*

4. RISULTATI DELL'ANALISI DI REGRESSIONE LINEARE 104

4.1 Effetto della zigosità sulla similarità intra-coppia *104*

4.2 Effetto dell'età sulla similarità intra-coppia *105*

4.3 Effetto della variabile sesso sulla similarità intra-coppia *107*

5. IMPATTO DEL RISCHIO PSICOTICO SULLA NEUROANATOMIA 108

5.1 Discussione dei risultati *117*

6. ANALISI DEI MODELLI DI CLASSIFICAZIONE OTTENUTI CON NEUROMINER 119

6.1 Pesi dei voxel nelle immagini di materia bianca e nelle immagini di materia grigia *121*

6.2 Predizione della zigosità di una nuova coppia *126*

6.3 Discussione dei risultati *127*

CAPITOLO 5 ..... **128**

CONCLUSIONI E SVILUPPI FUTURI 128

BIBLIOGRAFIA **134**

Lista delle figure

FIGURA 1.1: ESEMPI DI GRADIENTE RISPETTIVAMENTE IN (A) DIREZIONE X, (B) DIREZIONE Y, (C) DIREZIONE Z. 6

FIGURA 1.2: VETTORI DI MOMENTO MAGNETICO NUCLEARE (A) CHE PUNTANO IN DIREZIONI RANDOM E (B) ALLINEATI IN DIREZIONE DEL CAMPO MAGNETICO ESTERNO **[1]**. 9

FIGURA 1.3: AMBITI DI UTILIZZO DELL'INTELLIGENZA ARTIFICIALE IN MEDICINA **[5]**. 19

FIGURA 2.1: GENETIC INFLUENCE ON BRAIN STRUCTURE, THOMPSON ET AL. 2001. NOTA: F CORTECCIA FRONTALE, S/M SENSORIMOTORIA, W CORTECCIA DI WERNICKE **[10]**. 28

FIGURA 3.1: SEQUENZA GRADIENT ECHO **[1]**. 31

FIGURA 3.2: SEQUENZA INVERSION-RECOVERY **[1]**. 32

FIGURA 3.3: SEQUENZA MPRAGE. 32

FIGURA 3.4: PIPELINE DI ANALISI DELLE IMMAGINI NEL SOFTWARE *FREESURFER*. 37

FIGURA 3.5: MODELLO ACE: P= FENOTIPO OSSERVATO SU OGNI GEMELLO (I NUMERI 1 E 2 SI RIFERISCONO AI DUE GEMELLI DI UNA STESSA COPPIA); A=FATTORI GENETICI; C= FATTORI AMBIENTALI CONDIVISI DAI GEMELLI NELLA COPPIA; E= FATTORI AMBIENTALI NON CONDIVISI DAI GEMELLI; V=VARIANZA FENOTIPICA TOTALE; A $= 1$ PER I MZ OPPURE 0,5 PER I DZ; B $= 1$ PER I MZ E PER I DZ. 44

FIGURA 3.6: SONO RAPPRESENTATI I DUE VETTORI DI 13 CELLE, CORRISPONDENTI AL NUMERO DI COPPIE DISCORDANTI PER RISCHIO PSICOTICO, CHE RISPETTIVAMENTE CONTENGONO I VALORI DEL PARAMETRO ANALIZZATO NEI SOGGETTI AD ALTO RISCHIO E NEI SOGGETTI A BASSO RISCHIO. 51

FIGURA 3.7: INTERFACCIA GRAFICA SPM12. 54

FIGURA 3.8: INTERFACCIA FUNZIONE DISPLAY. 55

FIGURA 3.9: *BATCH* DI SPM 12 PER LA SEGMENTAZIONE. 56

FIGURA 3.10: (A) ESEMPIO DI IMMAGINE C1*.NII. (B) ESEMPIO DI IMMAGINE C2*.NII. SCELTA UNA *SLICE* DI ESEMPIO, NELL'IMMAGINE (A), SONO PRESENTI I VOXEL CHE CON UN'ELEVATA PROBABILITÀ APPARTENGONO ALLA MATERIA GRIGIA (MAPPA DI DENSITÀ DI MATERIA GRIGIA) MENTRE, NELL'IMMAGINE (B), SONO RAPPRESENTATI I VOXEL CHE CON UN'ELEVATA PROBABILITÀ APPARTENGONO ALLA MATERIA BIANCA (MAPPA DI DENSITÀ DI MATERIA BIANCA). 57

FIGURA 3.11: (A) ESEMPIO DI IMMAGINE C1*.NII. (B) ESEMPIO DI IMMAGINE RC1*.NII. SCELTA UNA *SLICE* DI ESEMPIO, NELL'IMMAGINE (A), SONO PRESENTI I VOXEL CHE CON UN'ELEVATA PROBABILITÀ APPARTENGONO ALLA MATERIA GRIGIA ALLA RISOLUZIONE ORIGINALE MENTRE, NELL'IMMAGINE (B), SONO RAPPRESENTATI I VOXEL CHE CON UN'ELEVATA PROBABILITÀ APPARTENGONO ALLA MATERIA GRIGIA ALLA RISOLUZIONE *DARTEL IMPORTED*. 58

FIGURA 3.12: INTERFACCIA GRAFICA DELLA FUNZIONE *RUN DARTEL (CREATE TEMPLATES)*. 59

FIGURA 3.13: TEMPLATE6.NII. 60

FIGURA 3.14: ESEMPIO DI IMMAGINE SMWC1*.NII SOVRAPPOSTA ALL'ATLANTE AAL (*AUTOMATED ANATOMICAL LABELING*), DEFINITO IN SPAZIO MNI. RAPPRESENTAZIONE DI UNA *SLICE* NEL PIANO CORONALE, SAGITTALE E ORIZZONTALE. 61

FIGURA 3.15: *IMMAGINE DELTA* DI MATERIA BIANCA DI UNA COPPIA DI GEMELLI..... 63
FIGURA 3.16: *IMMAGINE DELTA* DI MATERIA BIANCA DI DUE SOGGETTI SCELTI IN MANIERA CASUALE DAL CAMPIONE..... 64
FIGURA 3.17: SOVRAPPOSIZIONE *IMMAGINE DELTA* DI MATERIA BIANCA CON *IMMAGINE DELTA* DI MATERIA GRIGIA DI UNA COPPIA DI GEMELLI 64
FIGURA 3.18: INTERFACCIA GRAFICA NEUROMINER..... 66
FIGURA 3.19: LE RETTE RAPPRESENTANO POSSIBILI SUPERFICI DI SEPARAZIONE LINEARE PER I PUNTI DEL DIAGRAMMA..... 69
FIGURA 3.20: SUPERFICIE DI SEPARAZIONE DI MASSIMO MARGINE, IPERPIANI CANONICI E VETTORI DI SUPPORTO..... 69
FIGURA 3.21: SUPERFICIE DI SEPARAZIONE DI MASSIMO MARGINE E IPERPIANI CANONICI PER UN TRAINING SET NON SEPARABILE LINEARMENTE..... 71
FIGURA 3.22: RAPPRESENTAZIONE SCHEMATICA DELLA *DOUBLE NESTED CROSSVALIDATION*..... 75
FIGURA 3.23: DIAGRAMMA DELLA CURVA ROC PER UN CLASSIFICATORE..... 76
FIGURA 3.24: TABELLA DI CONTINGENZA. 77
FIGURA 3.25: INTERFACCIA GRAFICA *INDIPENDENT TEST DATA MANAGER* DI NEUROMINER. 78
FIGURA 4.1: REGIONI CORTICALI IN CUI È EMERSA UNA CORRELAZIONE SIGNIFICATIVA (*P-VALUE<0.05)* DEL VALORE ASSUNTO DAL PARAMETRO *AREA* NELLE COPPIE MZ E CONFRONTO CON IL COEFFICIENTE DI CORRELAZIONE NELLE COPPIE DZ QUANDO ANCHE QUESTO È SIGNIFICATIVO (*P-VALUE<0.05*) NELLE STESSE REGIONI. 82
FIGURA 4.2: REGIONI CORTICALI IN CUI È EMERSA UNA CORRELAZIONE SIGNIFICATIVA (*P-VALUE<0.05)* DEL VALORE ASSUNTO DAL PARAMETRO *SPESSORE CORTICALE* NELLE COPPIE MZ E CONFRONTO CON IL COEFFICIENTE DI CORRELAZIONE NELLE COPPIE DZ QUANDO ANCHE QUESTO È SIGNIFICATIVO (*P-VALUE<0.05*) NELLE STESSE REGIONI..... 84
FIGURA 4.3: REGIONI CORTICALI IN CUI È EMERSA UNA CORRELAZIONE SIGNIFICATIVA (*P-VALUE<0.05)* DEL VALORE ASSUNTO DAL PARAMETRO *CURVATURA MEDIA* NELLE COPPIE MZ E CONFRONTO CON IL COEFFICIENTE DI CORRELAZIONE NELLE COPPIE DZ QUANDO ANCHE QUESTO È SIGNIFICATIVO (*P-VALUE<0.05*) NELLE STESSE REGIONI..... 86
FIGURA 4.4: REGIONI CORTICALI IN CUI È EMERSA UNA CORRELAZIONE SIGNIFICATIVA (*P-VALUE<0.05)* DEL VALORE ASSUNTO DAL PARAMETRO *VOLUME* NELLE COPPIE MZ E CONFRONTO CON IL COEFFICIENTE DI CORRELAZIONE NELLE COPPIE DZ QUANDO ANCHE QUESTO È SIGNIFICATIVO (*P-VALUE<0.05*) NELLE STESSE REGIONI. 89
FIGURA 4.5: REGIONI SOTTOCORTICALI IN CUI È EMERSA UNA CORRELAZIONE SIGNIFICATIVA (*P-VALUE<0.05)* DEL VALORE ASSUNTO DAL PARAMETRO *VOLUME* NELLE COPPIE MZ E CONFRONTO CON IL COEFFICIENTE DI CORRELAZIONE NELLE COPPIE DZ QUANDO ANCHE QUESTO È SIGNIFICATIVO (*P-VALUE<0.05*) NELLE STESSE REGIONI 93
FIGURA 4.6: STIMA DELL'EREDITARIETÀ (ORDINE CRESCENTE DI PUNTEGGIO) DEL PARAMETRO *AREA SUPERFICIALE* NELLE REGIONI CELEBRALI IN CUI È EMERSA UNA SIGNIFICATIVITÀ STATISTICA IN ACCORDO AL TEST DI FISHER SULLA DIFFERENZA DI CORRELAZIONE. 94

FIGURA 4.7: VALORI DEL *P-VALUE* DELLA DIFFERENZA DI CORRELAZIONE OTTENUTI MEDIANTE IL TEST DI FISHER PER IL PARAMETRO *AREA SUPERFICIALE*. 94
FIGURA 4.8: STIMA DELL'EREDITARIETÀ (ORDINE CRESCENTE DI PUNTEGGIO) DEL PARAMETRO *SPESSORE CORTICALE* NELLE REGIONI CELEBRALI IN CUI È EMERSA UNA SIGNIFICATIVITÀ STATISTICA IN ACCORDO AL TEST DI FISHER SULLA DIFFERENZA DI CORRELAZIONE. 95
FIGURA 4.9: STIMA DELL'EREDITARIETÀ (ORDINE CRESCENTE DI PUNTEGGIO) DEL PARAMETRO *CURVATURA MEDIA* NELLE REGIONI CELEBRALI IN CUI È EMERSA UNA SIGNIFICATIVITÀ STATISTICA IN ACCORDO AL TEST DI FISHER SULLA DIFFERENZA DI CORRELAZIONE. 95
FIGURA 4.10: STIMA DELL'EREDITARIETÀ (ORDINE CRESCENTE DI PUNTEGGIO) DEL PARAMETRO *VOLUME* NELLE REGIONI CELEBRALI IN CUI È EMERSA UNA SIGNIFICATIVITÀ STATISTICA IN ACCORDO AL TEST DI FISHER SULLA DIFFERENZA DI CORRELAZIONE. 96
FIGURA 4.11: STIME SIGNIFICATIVE (*P-VALUE<0.05)* DI EREDITABILITÀ DEL *VOLUME* DELLE REGIONI SOTTOCORTICALI. 96
FIGURA 4.12: VALORI DI *AREA* NEL SOLCO OCCIPITALE LATERALE DELL'EMISFERO SINISTRO NEI GEMELLI AD ALTO RISCHIO (IN BLUE), NEI GEMELLI A BASSO RISCHIO (IN ARANCIONE) E NEI GEMELLI DI CONTROLLO (IN GRIGIO). 109
FIGURA 4.13: VALORI DI *SPESSORE CORTICALE* NELLA REGIONE ROSTRALE DELLA CORTECCIA CINGOLATA ANTERIORE DESTRA NEI GEMELLI AD ALTO RISCHIO (IN BLUE), NEI GEMELLI A BASSO RISCHIO (IN ARANCIONE) E NEI GEMELLI DI CONTROLLO (IN GRIGIO). 110
FIGURA 4.14: VALORI DI *SPESSORE CORTICALE* NELLA PARTE TRIANGOLARE DELLA CIRCONVOLUZIONE FRONTALE INFERIORE DESTRA NEI GEMELLI AD ALTO RISCHIO (IN BLUE), NEI GEMELLI A BASSO RISCHIO (IN ARANCIONE) E NEI GEMELLI DI CONTROLLO (IN GRIGIO). 110
FIGURA 4.15: CONFRONTO DEI VALORI DI *CURVATURA MEDIA* NELLA (A) REGIONE ROSTRALE DELLA CIRCONVOLUZIONE FRONTALE MEDIA, (B) NELLA CORTECCIA CINGOLATA POSTERIORE DESTRA, (C) NELLA REGIONE CAUDALE DELLA CIRCONVOLUZIONE FRONTALE MEDIA DESTRA, (D) NELLA CIRCONVOLUZIONE SOPRAMARGINALE DESTRA, NEI GEMELLI AD ALTO RISCHIO (IN BLUE), NEI GEMELLI A BASSO RISCHIO (IN ARANCIONE) E NEI GEMELLI DI CONTROLLO (IN GRIGIO). 113
FIGURA 4.16: CONFRONTO DEI VALORI DI *VOLUME* NEL (A) CUNEO SINISTRO, (B) NELLA CIRCONVOLUZIONE OCCIPITALE LATERALE SINISTRA, (C) NELLA PARTE TRIANGOLARE DELLA CIRCONVOLUZIONE FRONTALE INFERIORE SINISTRA, (D) NELLA CIRCONVOLUZIONE PARAIPPOCAMPALE DESTRA, NEI GEMELLI AD ALTO RISCHIO (IN BLUE), NEI GEMELLI A BASSO RISCHIO (IN ARANCIONE) E NEI GEMELLI DI CONTROLLO (IN GRIGIO). 116
FIGURA 4.17: VALORI DI *VOLUME* NEL SOLCO SUPERIORE TEMPORALE SINISTRO NEI GEMELLI AD ALTO RISCHIO (IN BLUE), NEI GEMELLI A BASSO RISCHIO (IN ARANCIONE) E NEI GEMELLI DI CONTROLLO (IN GRIGIO). 117
FIGURA 4.18: INTERFACCIA GRAFICA DEI RISULTATI, OTTENUTI CON LE IMMAGINI DI MATERIA BIANCA, DEL SOFTWARE *NEUROMINER*. 119
FIGURA 4.19: INTERFACCIA GRAFICA DEI RISULTATI, OTTENUTI CON LE IMMAGINI DI MATERIA GRIGIA, DEL SOFTWARE *NEUROMINER*. 120

FIGURA 4.20: IMMAGINE IN USCITA DAL SOFTWARE *NEUROMINER* CONTENETE IL PESO DEI VOXEL NELLE (A) IMMAGINI DI MATERIA GRIGIA E (B) IMMAGINI DI MATERIA BIANCA. 121

FIGURA 4.21: RAPPRESENTAZIONE NEL PIANO CORONALE (A) E SAGITTALE (B) DEI PESI SIGNIFICATIVI (*P-VALUE<0.05)* DEI VOXEL DELL'IMMAGINE DI MATERIA BIANCA SULL'ATLANTE *CH2BETTER.NII.* IN FUXIA SONO RAPPRESENTATI I VOXEL IN CUI LA QUANTITÀ DI MATERIA BIANCA È PIÙ SIMILE NELLE COPPIE DZ. IN BLUE SONO RAPPRESENTATI I VOXEL IN CUI LA QUANTITÀ DI MATERIA BIANCA È PIÙ SIMILE NELLE COPPIE MZ. 122

FIGURA 4.22: SOVRAPPOSIZIONE DEI VOXEL CHE ASSUMONO VALORI SIGNIFICATIVAMENTE DIVERSI TRA LE COPPIE MZ E LE COPPIE DZ, ALL'ATLANTE AAL. IN ROSSO SONO RAPPRESENTATI I VOXEL IN CUI LA DIFFERENZA INTRA-COPPIA, NEL VALORE DI MATERIA BIANCA, È MINORE NELLE COPPIE DZ RISPETTO ALLE COPPIE MZ. IN VIOLA È RAPPRESENTATA LA REGIONE APPARTENENTE ALLA CIRCONVOLUZIONE POSTCENTRALE. 123

FIGURA 4.23: SOVRAPPOSIZIONE DEI VOXEL CHE ASSUMONO VALORI SIGNIFICATIVAMENTE DIVERSI TRA LE COPPIE MZ E LE COPPIE DZ, ALL'ATLANTE AAL. IN ROSSO SONO RAPPRESENTATI I VOXEL IN CUI LA DIFFERENZA INTRA-COPPIA, NEL VALORE DI MATERIA BIANCA, È MINORE NELLE COPPIE DZ RISPETTO ALLE COPPIE MZ. IN GRIGIO È RAPPRESENTATA LA REGIONE APPARTENENTE ALLA CIRCONVOLUZIONE PRECENTRALE SINISTRA. 123

FIGURA 4.24: SOVRAPPOSIZIONE DEI VOXEL CHE ASSUMONO VALORI SIGNIFICATIVAMENTE DIVERSI TRA LE COPPIE MZ E LE COPPIE DZ, ALL'ATLANTE AAL. IN ROSSO SONO RAPPRESENTATI I VOXEL IN CUI LA DIFFERENZA INTRA-COPPIA, NEL VALORE DI MATERIA BIANCA, È MINORE NELLE COPPIE MZ RISPETTO ALLE COPPIE DZ. IN AZZURRO È RAPPRESENTATA LA REGIONE APPARTENENTE ALLA CORTECCIA LINGUALE DESTRA. 125

FIGURA 4.25: SOVRAPPOSIZIONE DEI VOXEL CHE ASSUMONO VALORI SIGNIFICATIVAMENTE DIVERSI TRA LE COPPIE MZ E LE COPPIE DZ, ALL'ATLANTE AAL. IN ROSSO SONO RAPPRESENTATI I VOXEL IN CUI LA DIFFERENZA INTRA-COPPIA, NEL VALORE DI MATERIA BIANCA, È MINORE NELLE COPPIE MZ RISPETTO ALLE COPPIE DZ. IN NERO È RAPPRESENTATA LA REGIONE APPARTENENTE ALLA CORTECCIA SOPRAMARGINALE EMISFERO DESTRO. 125

FIGURA 4.26: SOVRAPPOSIZIONE DEI VOXEL CHE ASSUMONO VALORI SIGNIFICATIVAMENTE DIVERSI TRA LE COPPIE MZ E LE COPPIE DZ, ALL'ATLANTE AAL. IN ROSSO SONO RAPPRESENTATI I VOXEL IN CUI LA DIFFERENZA INTRA-COPPIA, NEL VALORE DI MATERIA BIANCA, È MINORE NELLE COPPIE MZ RISPETTO ALLE COPPIE DZ. IN FUXIA È RAPPRESENTATA LA REGIONE POZIONE DI MATERIA BIANCA APPARTENENTE AL CERVELLETTO NELL'EMISFERO DESTRO. 125

FIGURA 4.27: SCHERMATA OTTENUTA NELLA COMMAND WINDOW DI MATLAB DOPO L'INSERIMENTO DEL COMANDO PER VISUALIZZARE I RISULTATI DELLA PREDIZIONE DELLA ZIGOSITÀ DELLA NUOVA IMMAGINE. IN CORRISPONDENZA DI *BINPROBPREDICTIONS* È POSSIBILE LEGGERE LA CLASSE PREDETTA. IN QUESTO CASO SI LEGGE { -1} OVVERO DZ. 126

Lista delle tabelle

TABELLA 4.1: REGIONI CORTICALI DEI DUE EMISFERI IN CUI IL COEFFICIENTE DI CORRELAZIONE R_{MZ} È STATISTICAMENTE SIGNIFICATIVO (*P-VALUE<0.05)* PER IL PARAMETRO *AREA SUPERFICIALE*. 80

TABELLA 4.2: REGIONI CORTICALI DEI DUE EMISFERI IN CUI IL COEFFICIENTE DI CORRELAZIONE R_{DZ} È STATISTICAMENTE SIGNIFICATIVO (*P-VALUE<0.05)* PER IL PARAMETRO *AREA SUPERFICIALE*. 81

TABELLA 4.3: REGIONI CORTICALI DEI DUE EMISFERI IN CUI IL COEFFICIENTE DI CORRELAZIONE R_{MZ} È STATISTICAMENTE SIGNIFICATIVO (*P-VALUE<0.05)* PER IL PARAMETRO *SPESSORE CORTICALE*. 83

TABELLA 4.4: REGIONI CORTICALI DEI DUE EMISFERI IN CUI IL COEFFICIENTE DI CORRELAZIONE R_{DZ} È STATISTICAMENTE SIGNIFICATIVO (*P-VALUE<0.05)* PER IL PARAMETRO *SPESSORE CORTICALE*. 84

TABELLA 4.5: REGIONI CORTICALI DEI DUE EMISFERI IN CUI IL COEFFICIENTE DI CORRELAZIONE R_{MZ} È STATISTICAMENTE SIGNIFICATIVO (*P-VALUE<0.05)* PER IL PARAMETRO *CURVATURA MEDIA*..... 85

TABELLA 4.6: REGIONI CORTICALI DEI DUE EMISFERI IN CUI IL COEFFICIENTE DI CORRELAZIONE R_{DZ} È STATISTICAMENTE SIGNIFICATIVO (*P-VALUE<0.05)* PER IL PARAMETRO *CURVATURA MEDIA*..... 86

TABELLA 4.7: REGIONI CORTICALI DEI DUE EMISFERI IN CUI IL COEFFICIENTE DI CORRELAZIONE R_{MZ} È STATISTICAMENTE SIGNIFICATIVO (*P-VALUE<0.05)* PER IL PARAMETRO *VOLUME*. 88

TABELLA 4.8: REGIONI CORTICALI DEI DUE EMISFERI IN CUI IL COEFFICIENTE DI CORRELAZIONE R_{DZ} È STATISTICAMENTE SIGNIFICATIVO (*P-VALUE<0.05)* PER IL PARAMETRO *VOLUME*. 88

TABELLA 4.9: REGIONI IN CUI IL COEFFICIENTE DI CORRELAZIONE R_{MZ} È STATISTICAMENTE SIGNIFICATIVO (*P-VALUE<0.05)* PER IL PARAMETRO *VOLUME*. 91

TABELLA 4.10: REGIONI IN CUI IL COEFFICIENTE DI CORRELAZIONE R_{DZ} È STATISTICAMENTE SIGNIFICATIVO (*P-VALUE<0.05)* PER IL PARAMETRO *VOLUME*. 92

TABELLA 4.11: MODELLO ACE, PER IL PARAMETRO *AREA*, NELLE REGIONI CORTICALI IN CUI I COEFFICIENTI DI CORRELAZIONE SONO STATISTICAMENTE SIGNIFICATIVI (*P-VALUE<0.05)*. A: EFFETTI GENETICI ADDITIVI; C: EFFETTI AMBIENTALI CONDIVISI; E: EFFETTI AMBIENTALI NON CONDIVISI 99

TABELLA 4.12: MODELLO ACE, PER IL PARAMETRO *SPESSORE CORTICALE*, NELLE REGIONI CORTICALI IN CUI I COEFFICIENTI DI CORRELAZIONE SONO STATISTICAMENTE SIGNIFICATIVI (*P-VALUE<0.05)*. A: EFFETTI GENETICI ADDITIVE; C: EFFETTI AMBIENTALI CONDIVISI; E: EFFETTI AMBIENTALI NON CONDIVISI..... 100

TABELLA 4.13: MODELLO ACE, PER IL PARAMETRO *CURVATURA MEDIA*, NELLE REGIONI CORTICALI IN CUI I COEFFICIENTI DI CORRELAZIONE SONO STATISTICAMENTE SIGNIFICATIVI (*P-VALUE<0.05)*. A: EFFETTI GENETICI ADDITIVE; C: EFFETTI AMBIENTALI CONDIVISI; E: EFFETTI AMBIENTALI NON CONDIVISI..... 100

TABELLA 4.14: MODELLO ACE, PER IL PARAMETRO *VOLUME*, NELLE REGIONI CORTICALI IN CUI I COEFFICIENTI DI CORRELAZIONE SONO STATISTICAMENTE SIGNIFICATIVI (*P-VALUE<0.05)*. A: EFFETTI GENETICI ADDITIVE; C: EFFETTI AMBIENTALI CONDIVISI; E: EFFETTI AMBIENTALI NON CONDIVISI. 101

TABELLA 4.15: MODELLO ACE, PER IL PARAMETRO *VOLUME*, NELLE REGIONI SOTTOCORTICALI IN CUI I COEFFICIENTI DI CORRELAZIONE SONO STATISTICAMENTE SIGNIFICATIVI (*P-VALUE<0.05)*. A: EFFETTI GENETICI ADDITIVE; C: EFFETTI AMBIENTALI CONDIVISI; E: EFFETTI AMBIENTALI NON CONDIVISI..... 102

TABELLA 4.16: REGIONI CELEBRALI IN CUI SONO EMERSE SIGNIFICATIVITÀ STATISTICHE DEL MODELLO DI REGRESSIONE LINEARE MULTIPLA E DEI COEFFICIENTI DELLA VARIABILE *ZIGOSITÀ*. 104
TABELLA 4.17: REGIONI CELEBRALI IN CUI SONO EMERSE SIGNIFICATIVITÀ STATISTICHE DEL MODELLO DI REGRESSIONE LINEARE MULTIPLA E DEI COEFFICIENTI DELLA VARIABILE *ETÀ*. 105
TABELLA 4.18: REGIONI CELEBRALI IN CUI SONO EMERSE SIGNIFICATIVITÀ STATISTICHE DEL MODELLO DI REGRESSIONE LINEARE MULTIPLA E DEI COEFFICIENTI DELLA VARIABILE *SESSO*..... 107
TABELLA 4.19: REGIONI CELEBRALI IN CUI SONO EMERSE SIGNIFICATIVITÀ STATISTICHE DAL *TTEST* TRA I GEMELLI AD ALTO RISCHIO E I GEMELLI A BASSO RISCHIO. 111
TABELLA 4.20: REGIONI CELEBRALI IN CUI SONO EMERSE SIGNIFICATIVITÀ STATISTICHE DAL *TTEST2* TRA I GEMELLI AD ALTO RISCHIO E I GEMELLI DI CONTROLLO..... 111
TABELLA 4.21: REGIONI CELEBRALI IN CUI SONO EMERSE SIGNIFICATIVITÀ STATISTICHE DAL *TTEST2* TRA I GEMELLI A BASSO RISCHIO E I GEMELLI DI CONTROLLO. 112
TABELLA 4.22: REGIONI CELEBRALI IN CUI SONO EMERSE SIGNIFICATIVITÀ STATISTICHE DAL TTEST TRA I GEMELLI AD ALTO RISCHIO E I GEMELLI A BASSO RISCHIO. 114
TABELLA 4.23: REGIONI CELEBRALI IN CUI SONO EMERSE SIGNIFICATIVITÀ STATISTICHE DAL TTEST2 TRA I GEMELLI AD ALTO RISCHIO E I GEMELLI DI CONTROLLO..... 115
TABELLA 4.24: REGIONI CELEBRALI IN CUI SONO EMERSE SIGNIFICATIVITÀ STATISTICHE DAL TTEST2 TRA I GEMELLI A BASSO RISCHIO E I GEMELLI DI CONTROLLO. 115
TABELLA 4.25: REGIONI CELEBRALI IN CUI È EMERSA UNA SIMILARITÀ MAGGIORE, IN TERMINI DI MATERIA BIANCA, NELLE COPPIE DZ. 122
TABELLA 4.26: REGIONI CELEBRALI IN CUI È EMERSA UNA SIMILARITÀ MAGGIORE, IN TERMINI DI MATERIA BIANCA, NELLE COPPIE MZ..... 124

Capitolo 1

Introduzione

Negli ultimi decenni, *l'Imaging* a Risonanza Magnetica (MRI) ha avuto una grande diffusione e un grande sviluppo sia per la qualità delle immagini acquisite, sia per la versatilità di applicazione in ambito clinico e di ricerca. In particolare, dal momento che, diversamente da altre condizioni patologiche, i disturbi psichiatrici sono raramente diagnosticati sulla base di segni obiettivi o di test laboratoristici o di *imaging*, è nata la necessità di ricercare *markers* validati ed obiettivabili che forniscano un ausilio diagnostico, una migliore comprensione della patogenesi, un supporto alla quantificazione e al monitoraggio degli interventi terapeutici. Quindi, si sono ricercati nuovi metodi d'indagine per lo studio del cervello in vivo. In particolare, si è diffuso molto il *brain imaging* in cui l'MRI rappresenta il principale metodo di indagine, con una storia che risale agli inizi degli anni '80, quando per la prima volta furono eseguiti studi MRI su pazienti affetti da disturbo bipolare (Rangel- Guerra et al. ,1983) e su pazienti schizofrenici (Smith et al. ,1984), e che continua con costanti miglioramenti tecnologici e campi di utilizzo sempre più complessi.

Fino ad ora, la maggior parte degli studi sullo sviluppo celebrale sono stati confinati ai meccanismi celebrali di interesse senza un approfondimento sul diverso ruolo dei fattori genetici ed ambientali. Gli studi sui gemelli permettono di far luce sui fattori intrinseci (genetici) ed estrinseci (ambientali) che modellano le differenze inter-individuali nello sviluppo del cervello, fornendo così spunti rilevanti sulle traiettorie fisiologiche e patologiche. Pertanto, il metodo gemellare consente di ottenere delle stime dei contributi genetici ed ambientali alle caratteristiche morfologiche celebrali, fornendo risposte importanti per la comprensione dei complessi determinanti dello stato di salute. Questa informazione è di fondamentale importanza affinchè si possano ricercare ed attuare strategie preventive (nell'ambito dei comportamenti e degli stili di vita) e terapie adeguate, innovative e personalizzate.

Le potenzialità della gemellarità come strumento di ricerca furono intuite la prima volta nella seconda metà dell'Ottocento da Francis Galton. Egli riteneva che studiare coppie di gemelli permettesse di effettuare una distinzione fra le tendenze ricevute alla nascita e quelle

acquisite con l'esperienza di vita, consentendo quindi di valutare gli effetti dell'ereditarietà (nature) e dell'ambiente (nurture) nel plasmare le caratteristiche umane.

Il metodo gemellare si fonda essenzialmente sul confronto tra il grado di similarità delle caratteristiche in studio tra i gemelli monozigoti (monovulari, 100% di DNA uguale, fisicamente identici o quasi, sempre dello stesso sesso) e i gemelli dizigoti (biovulari, circa il 50% di DNA in comune, che si somigliano come due normali fratelli e possono essere anche di sesso opposto). Se si osserva che, per certe caratteristiche i gemelli monozigoti (MZ) sono più simili tra di loro all'interno della coppia rispetto ai gemelli delle coppie dizigoti (DZ), è possibile dedurre che la componente genetica di quella caratteristica sia piuttosto importante (i MZ hanno molto più DNA in comune dei DZ). Se invece, sempre per la caratteristica presa in considerazione, i gemelli delle coppie MZ non sono particolarmente più simili tra di loro rispetto ai gemelli delle coppie DZ, è lecito supporre che i geni abbiano poca importanza, e che siano invece i fattori ambientali a spiegare prevalentemente la caratteristica. Quindi, attraverso il confronto tra le coppie di gemelli MZ e DZ si può stimare in quale misura il patrimonio genetico ed i fattori ambientali contribuiscano all'espressione di caratteristiche morfologiche e psicologiche, sia normali che di malattia.

Lo stato dell'arte mostra un quadro in cui sono presenti molti studi basati sui gemelli ma ancora non è stata raggiunta un'elevata robustezza dei risultati in termini di quantificazione del contributo dei fattori genetici ed ambientali alle caratteristiche neuroanatomiche. Questo lavoro di tesi sperimentale si inserisce in questo ambito di ricerca con uno studio MRI a 3T con lo scopo principale di stimare l'impatto genetico e ambientale su una serie di caratteri morfologici celebrali in coppie di gemelli in età evolutiva. La corte gemellare analizzata nel presente lavoro rappresenta un contesto ideale per investigare le interazioni geni-ambiente e identificare marker predittivi del rischio di psicosi, mediante l'applicazione del metodo gemellare e sfruttando i notevoli vantaggi epidemiologici e statistici di questo particolare disegno di studio.

Scopo della ricerca

Il presente lavoro di tesi sperimentale è stato condotto in collaborazione tra il Politecnico di Milano e la Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico. Le analisi sono state condotte su dati acquisiti nel corso di due progetti di ricerca finanziati dal Ministero della Salute tramite il bando di Ricerca Finalizzata, il progetto SPES, "**Service for** detection of **Pre-psychosES**: from gene-environment interactions to phenomenology" e il *"*WHY ME?" PROJECT (WM?P), studio della connettività della sostanza bianca nella schizofrenia a livello di imaging, cellulare e genomico in gemelli monozigoti discordanti, condotti dal 2012 al 2018 presso l'IRCSS "E. Medea"- Ass. la nostra famiglia di Bosisio Parini (Lc). Il responsabile principale del progetto è Paolo Brambilla, Professore di Psichiatria presso l'Università degli Studi di Milano e Direttore dell'U.O.C. Psichiatra presso la Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, che vanta una collaborazione decennale con l'IRCCS Medea. Le unità operative che hanno partecipato a ai progetti SPES e WM?P sono l'Azienda ospedaliera universitaria di Udine, gli ospedali Niguarda e San Raffaele di Milano, il Registro Nazionale dei Gemelli dell'Istituto Superiore di Sanità, la Fondazione Imago7 di Pisa, l'Istituto Italiano di Tecnologie di Genova e la Scuola internazionale superiore di studi avanzati di Trieste.

I progetti riguardano l'analisi di dati di soggetti gemelli individuati grazie al prezioso contributo del Registro Nazionale dei Gemelli (RNG) (http://old.iss.it/gemelli/), che ha fornito le informazioni di coppie di gemelli tra i 9 e i 30 anni iscritti al Registro. Nello specifico, il reclutamento della coorte gemellare è avvenuto tramite le procedure standardizzate del RNG, gestito dal Reparto di Epidemiologia Genetica del Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute (CNESPS) dell'Istituto Superiore di Sanità (ISS) di Roma (Resp. Dott.ssa Maria Antonietta Stazi).

Uno degli scopi principali dei due progetti, che sono stati svolti in continuità, è di verificare se gemelli a rischio per psicosi maggiore (schizofrenia, disturbo bipolare, schizotipia) presentino, rispetto ai rispettivi gemelli non a rischio (1) alterazioni nella connettività tra corteccia dorsolaterale prefrontale e ippocampo e nella neurotrasmissione glutamatergica (2) alterazioni nell'interazione neuroni-glia (3) alterazioni nel numero di copie dei retrotransposoni del DNA. Tali valutazioni sono utili per la comprensione del funzionamento

del cervello umano e dei fattori predisponenti a disturbi mentali, e possono consentire di migliorare sia le capacità diagnostiche - al fine di riconoscere precocemente i soggetti a rischio - sia le capacità riabilitative e terapeutiche per la schizofrenia in età infantile e adulta.

La presente tesi si inserisce in questo framework di ricerca e riguarda l'analisi delle immagini neuroanatomiche T1-pesate acquisite da coppie di gemelli con lo scanner MRI a 3 Tesla installato presso l'IRCCS Medea di Bosisio Parini. Il lavoro sperimentale ha avuto lo scopo principale di quantificare l'impatto genetico e ambientale sull'anatomia celebrale in coppie di gemelli di età compresa tra 9 e 30 anni. A tale scopo, sono state applicate sofisticate tecniche di analisi neuroanatomica (di tipo *region-based* e *voxel-based*) che hanno permesso di estrarre le misure morfologiche di interesse. Tali misure (fenotipi) sono state confrontate all'interno di ciascuna coppia di gemelli e tra coppie MZ e DZ utilizzando metodi statistici gemellari classici, analisi di regressione e di classificazione al fine di stimare i gradi di influenza genetica e ambientale sui tratti di interesse.

A partire dai valori di correlazione fenotipica in coppie MZ e DZ e dalle stime di ereditabilità è possibile quantificare l'effetto del patrimonio genetico e l'influenza dei fattori ambientali su una caratteristica morfologica. Tali informazioni risultano essenziali per la comprensione di patologie associate ad alterazioni della caratteristica morfologica in analisi. Infatti, a partire dalle stime di ereditabilità è possibile quantificare in che misura la modifica del carattere neuroanatomico è legata ai fattori genetici e in che misura è legata a fattori ambientali. La presenza di un impatto ambientale importante sull'espressione del parametro può essere un'informazione fondamentale per applicare misure preventive come, ad esempio, modifiche negli stili di vita.

1. L'*Imaging* a Risonanza Magnetica

L'*Imaging* a Risonanza Magnetica (o Magnetic Resonance Imaging, MRI) è una tecnica radiologica che in medicina consente di studiare aspetti anatomici e funzionali del corpo umano sfruttando il principio della Risonanza Magnetica Nucleare (RMN). Come la Tomografia Computerizzata (TC) è in grado di fornire immagini tomografiche cross sezionali ma, diversamente da questa, consente una maggiore specificità nello studio dei diversi tessuti molli, grazie all'elevata risoluzione di contrasto intrinseca. Inoltre, sfruttando radiazioni non ionizzanti, consente un approccio non invasivo e applicazioni diagnostiche che interessano molte specialità mediche.

Il segnale di Risonanza Magnetica dipende dalle proprietà di risonanza nucleare dei diversi tessuti. Il fenomeno della risonanza si basa sulla capacità dei protoni, contenuti nel nucleo dell'atomo, di assorbire o emettere energia elettromagnetica quando, posti in un forte campo magnetico, vengono eccitati da onde radio di opportuna frequenza. Nel corpo umano l'*Imaging* a Risonanza Magnetica indaga prevalentemente le proprietà degli atomi d'idrogeno. Quest'ultimo si configura come l'elemento più semplice, essendo costituito da un protone e da un elettrone, e il più rappresentato grazie all'elevato contenuto di acqua e grasso dell'organismo; per tali ragioni esso è quindi in grado di fornire un forte segnale. La diversa composizione molecolare, e quindi il diverso contenuto di idrogeno dei diversi tessuti, fa sì che essi assorbano e rilascino energia in maniera distinta.

Uno dei motivi principali che spinge verso il forte utilizzo dell'MRI è la non invasività. Questa caratteristica è legata al non utilizzo delle radiazioni ionizzanti oltre che all'esecuzione dell'esame: possono essere generate immagini di elevata qualità anche senza l'utilizzo di mezzo di contrasto. Sebbene talvolta si ricorra alla somministrazione del mezzo di contrasto, questo ha un potere allergizzante molto basso e quindi espone il paziente a minori rischi di reazione avversa rispetto ai mezzi di contrasto iodati per angiografia convenzionale e TC. Il 'discomfort' del paziente è legato alla durata dell'esame (circa mezz'ora) che necessita della immobilità, e alla geometria delle apparecchiature, che sono imponenti e possono creare problemi di adattamento. Sebbene il forte campo magnetico non sia di per sé pericoloso, gli strumenti medici che contengono metallo o parti elettroniche impiantati nel corpo del paziente possono presentare malfunzionamenti o causare problemi durante l'esame. Per tali motivi non possono essere sottoposti a risonanza magnetica i soggetti portatori di pacemaker, neurostimolatore, impianto cocleare oppure con particolari

tipi di clip usate negli aneurismi cerebrali o di cui non si conoscano le caratteristiche e la sicura compatibilità magnetica.

1.1 Hardware di un'apparecchiatura di Risonanza Magnetica Nucleare

L' hardware di un'apparecchiatura di Risonanza Magnetica è composto dai seguenti componenti:

- Un *magnete principale*, che genera un forte campo magnetico statico ad alta intensità (B_0). Il magnete principale può essere di tipo superconduttivo (permette di raggiungere campi di elevata intensità, maggiori di 11 Tesla), resistivo (fino a 0,15 Tesla), o permanente (fino a 0,30 Tesla). Il magnete più comunemente usato è quello di tipo superconduttivo: è composto da una bobina che è resa superconduttiva mediante raffreddamento con elio liquido. I fili di questa bobina hanno una resistenza approssimativamente nulla alle temperature vicine allo 0 K (precisamente si arriva a 4.2 K, immergendo la bobina nell'elio liquido). I magneti superconduttivi sono continuamente in funzione, quindi per limitare i loro effetti all'esterno, si ricorre a opportuni sistemi di schermatura.

- Tre *coppie di bobine di Maxwell* che producono una variazione lineare nell'intensità del campo magnetico lungo le tre direzioni che definiscono lo spazio 3D. I campi magnetici spazio varianti si vanno a sovrapporre a B_0. Tale sistema è importante per localizzare le sorgenti del segnale.

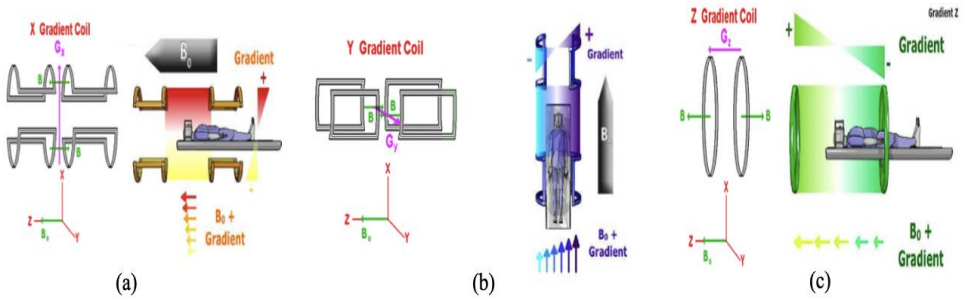

Figura 1.1: Esempi di gradiente rispettivamente in (a) direzione x, (b) direzione y, (c) direzione z.

- Le *bobine di radiofrequenza*, che generano un campo magnetico in radiofrequenza (B_1), ortogonale a B_0. Tali bobine possono essere sia trasmittenti che riceventi, o solo trasmittenti, o solo riceventi, possono essere volumetriche (se circondano il paziente), superficiali

(adiacenti al paziente) o interne (progettate per ottenere segnali provenienti dall'esterno di essa, per esempio nel caso di una bobina in un catetere); comprendono elementi induttivi e capacitivi che determinano la frequenza di risonanza.

-Un *computer*, che controlla tutte le componenti, tra cui il generatore delle radiofrequenze, il programmatore degli impulsi in radiofrequenza, e il programmatore degli impulsi di gradiente. Il generatore delle radiofrequenze dà luogo a una sinusoide con una determinata frequenza (la frequenza di Larmor), il programmatore degli impulsi determina l'inviluppo dell'impulso in radiofrequenza **[1]**.

1.2 Principi fisici

La Risonanza Magnetica Nucleare è un fenomeno che avviene quando i nuclei di certuni atomi sono immersi in un campo magnetico statico (B_0) e vengono esposti ad un secondo campo magnetico oscillante (B_1). Tali nuclei rispondono producendo un segnale elettromagnetico con una frequenza caratteristica. Ciò avviene vicino alla frequenza di risonanza cioè quando la frequenza di precessione dei nuclei, che dipende dal campo magnetico statico (B_0), dall'ambiente chimico, dalle caratteristiche magnetiche dell'isotopo considerato, coincide con quella del campo magnetico oscillante (B_1).

Il segnale di Risonanza Magnetica è dunque un fenomeno fisico macroscopico, ma ha la sua origine nel comportamento dei nuclei degli atomi di alcuni elementi presenti nella materia vivente.

1.2.1 Momento angolare e momento magnetico

È noto dalla chimica che ogni oggetto fisico, può essere scomposto in molecole, e successivamente in atomi. Ogni atomo è costituito da un nucleo con protoni e neutroni, e da elettroni orbitali. Il nucleo di ciascun atomo possiede una massa, una carica elettrica e un raggio finito. I nuclei con peso atomico dispari e numero atomico dispari possiedono un momento angolare $\vec{J}$, detto *spin*. Anche in assenza di un campo magnetico esterno, i nuclei che hanno uno spin diverso da zero, essendo dotati di carica elettrica, ruoteranno intorno al proprio asse. Ognuno di questi nuclei ruotando genera un campo magnetico intorno ad esso

definito momento magnetico $\vec{\mu}$. La relazione fondamentale della fisica delle particelle, che lega il momento magnetico e il momento angolare è:

$$\vec{\mu} = \gamma \vec{J} \qquad (1.1)$$

dove γ è una costante fisica chiamata rapporto gyromagnetico, ed è specifica per ogni materiale. Il momento magnetico è una grandezza vettoriale, definita da modulo, direzione e verso. L'orientazione di $\vec{\mu}$ in assenza di un campo magnetico esterno è casuale ed è legata al movimento di agitazione termica.

Tra i tanti elementi dotati di spin diverso da zero presenti nei tessuti biologici (1H, 13C, 19F, 23Na, 31P), l'idrogeno è scelto come elemento rispetto al quale produrre l'immagine di Risonanza Magnetica poiché oltre ad essere l'elemento più semplice, essendo costituito da un protone e da un elettrone, e il più approfonditamente studiato da un punto di vista fisico e il più abbondante nel corpo umano (è presente con una concentrazione di 1019 atomi per ogni mm^3 di tessuto) ed è dotato di un momento dipolare magnetico più intenso rispetto a quello degli altri elementi.

1.2.2 Generazione della magnetizzazione macroscopica

All'equilibrio termico ogni atomo presenta un orientamento degli spin casuale, motivo per cui in natura e quindi in ogni tessuto del corpo umano non è presente un orientamento preferenziale dei nuclei e non è evidenziabile una magnetizzazione macroscopica. Per attivare la magnetizzazione macroscopica è necessario che tutti i vettori spin siano allineati e questo può essere ottenuto esponendo l'oggetto macroscopico di cui si vuole acquisire l'immagine, ovvero il paziente, ad un forte campo magnetico esterno.

Per convenzione si considera che il campo magnetico esterno B_0 sia applicato in direzione z rispetto al sistema di riferimento del laboratorio, cioè:

$$\overrightarrow{B_0} = B_0 \vec{k} \qquad (1.2)$$

Quando vengono collocati all'interno di un campo magnetico la maggioranza dei protoni di un tessuto si allinea nella direzione del campo magnetico principale, con verso parallelo, a minor dispendio energetico, o con verso antiparallelo, a maggior richiesta energetica.

All'aumentare della forza del campo magnetico aumenta il numero di protoni allineati alla direzione del campo e soprattutto di quelli di verso concorde al campo.

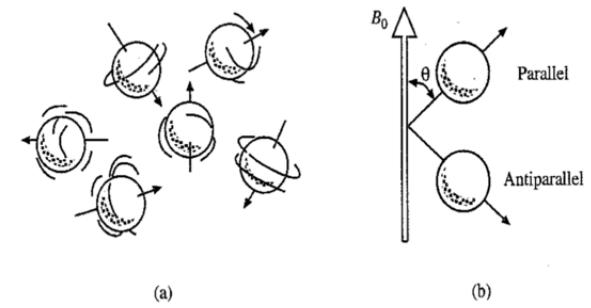

Figura 1.2: Vettori di momento magnetico nucleare (a) che puntano in direzioni random e (b) allineati in direzione del campo magnetico esterno **[1]**.

1.2.3 Magnetizzazione macroscopica netta

Si definisce vettore di *magnetizzazione macroscopica netta* ($\vec{M}$), la somma vettoriale di tutti i momenti magnetici microscopici che avrà stessa direzione e verso di B_0 e modulo definito dall'Eq. (1.3).

$$\vec{\mathrm{M}} = \sum_{\mathrm{n}=1}^{\mathrm{N_s}} \vec{\mu}_{\mathrm{n}} \qquad (1.3)$$

dove N_s è il numero totale di spin presenti nell'oggetto di cui si vuole acquisire l'immagine, cioè nel paziente. Dato che in assenza di un campo magnetico esterno tutti gli spin si dispongono in maniera casule, il modulo di $\vec{M}$ è pari a zero, mentre sarà diverso da zero in presenza di B_0. In particolare, la magnetizzazione macroscopica netta è detta *magnetizzazione longitudinale* perché la direzione è la stessa rispetto all'asse del campo magnetico principale.

1.2.4 Equazione di Larmor

Quando i protoni sono sottoposti al campo magnetico iniziano un nuovo movimento, il movimento di precessione: si tratta di una rotazione attorno all'asse del campo magnetico, la cui frequenza è direttamente proporzionale all'intensità del campo magnetico, secondo la legge di Larmor (Eq. (1.4)).

$$\omega_0 = \gamma B_0 \qquad (1.4)$$

Questa equazione mostra che la frequenza del moto di precessione di un sistema spin (ω_0) dipende linearmente sia dal rapporto giromagnetico che dal campo magnetico statico esterno. In tale condizione il vettore di *magnetizzazione netta* ($\vec{M}$) punta lungo la direzione di B_0 (cioè lungo z) quindi seppur fosse presente a livello microscopico una componente trasversa per ogni vettore di momento magnetico, la componente trasversa di $\vec{M}$ (cioè della magnetizzazione netta) sarebbe comunque zero all'equilibrio perché i momenti di precessione magnetica hanno fase random **[1]**.

Al fine di ottenere il fenomeno di risonanza è necessario fornire uno stimolo che abbia la stessa frequenza del moto di precessione; tale stimolo viene attivato per pochi microsecondi, ed è molto meno intenso di B_0, infatti l'attivazione di B_1 viene comunemente definita impulso in radiofrequenza (RF). Si inviano quindi uno o più impulsi RF, separati da piccoli intervalli, che vanno a disturbare l'equilibrio del campo magnetico e determinano due effetti contemporanei: la riduzione della magnetizzazione longitudinale e la formazione di una magnetizzazione trasversale. I protoni ricevono energia che spendono ponendosi nella posizione energeticamente più dispendiosa e cambiando verso (antiparallelo), in una percentuale che dipende dalle caratteristiche dell'impulso: come conseguenza, la magnetizzazione longitudinale diminuisce. L'impulso RF permette, inoltre, di sincronizzare la fase del movimento di precessione, facendo in modo che in uno stesso istante tutti i protoni siano sullo stesso punto dell'orbita, creando una forza magnetica sull'asse trasversale detta *magnetizzazione trasversale*. La trasmissione dell'impulso RF determina uno scambio di energia il cui effetto è detto appunto *risonanza*: quando, in un sistema spin magnetizzato, è raggiunta la coerenza di fase. L'impulso dura pochi millisecondi e nel momento in cui cessa si assiste ad un ritorno alla condizione antecedente, con cessione dell'energia ricevuta all'ambiente. I protoni che erano sincronizzati in fase tornano a ruotare ognuno per conto proprio e quindi la magnetizzazione trasversale scompare. La forza magnetica longitudinale invece torna a regime al termine dell'impulso.

La *condizione di risonanza* stabilisce che B_1 debba ruotare con la stessa frequenza di precessione degli spin:

$$\omega_{rf} = \omega_0 \qquad (1.5)$$

1.2.5 Caratteristiche di un impulso in radiofrequenza

I parametri fondamentali che caratterizzano un impulso in radiofrequenza sono: la funzione di inviluppo ($B_1^e(t)$), la frequenza di eccitazione del carrier (determinata dalla condizione di risonanza, ovvero deve essere pari alla frequenza di Larmor), e la fase iniziale (φ). L'impulso di radiofrequenza ha la forma

$$B_1(t) = B_{1,x}(t) + iB_{1,y}(t) = B_1^e(t)e^{-i(\omega_{rf}t+\varphi)} \tag{1.6}$$

($B_1^e(t)$) è la funzione di inviluppo che determina la forma e la durata dell'impulso e quindi le proprietà dell'eccitazione **[1]**. Per una descrizione quantitativa degli effetti degli impulsi RF sul vettore magnetizzazione netta si ricorre all'*equazione di Bloch* (Eq. (1.7)).

$$\frac{d\vec{M}}{dt} = \gamma\vec{M} \times \vec{B} - \frac{M_x\vec{i} + M_y\vec{j}}{T_2} - \frac{(M_z - M_z^0)\vec{k}}{T_1} \tag{1.7}$$

Tale equazione descrive il comportamento di $\vec{M}$ in funzione del tempo. L'utilizzo di tale relazione nel sistema di riferimento rotante (alla frequenza di Larmor) permette di descrivere e comprendere efficacemente la *condizione di risonanza*. Dall'applicazione di tale equazione, si dimostra che il vettore $\vec{M}$, nella suddetta condizione, assume un movimento di precessione attorno B_1, con frequenza pari a γB_1. Nel caso in cui non si è in condizione di risonanza, il campo B_1 non ruota alla frequenza di precessione dei nuclei, e il vettore di magnetizzazione $\vec{M}$, mantiene la sua frequenza di precessione attorno a B_0. L'applicazione di un impulso RF ha come effetto lo spostamento graduale della direzione del vettore $\vec{M}$ che assume una componente trasversale (*magnetizzazione trasversale*). L'angolo tra il vettore $\vec{M}$, e la direzione longitudinale, generatosi a causa di questo spostamento graduale, viene comunemente definito *flip angle* (α) **[1]**.

1.2.6 Processo di rilassamento

Come accennato nel paragrafo 1.2.4 al termine della perturbazione fornita dall'impulso RF il sistema ritorna al suo stato di equilibrio, secondo le leggi della termodinamica. La

componente trasversale (M_{xy}) decade a zero, mentre la componente longitudinale (M_z) tende a recuperare il valore M_0. L'evoluzione temporale di tali processi è descritta dalle seguenti relazioni (ottenute applicando appropriatamente l'*equazione di Bloch*):

$$\begin{cases} M_{x'y'}(t) = M_{x'y'}(0_+)e^{-t/T_2} \\ M_{z'}(t) = M_z^0\left(1 - e^{-t/T_1}\right) + M_{z'}(0_+)e^{-t/T_1} \end{cases} \tag{1.8}$$

Questo fenomeno viene chiamato *rilassamento* e comprende due fasi principali:

- la prima, definita da una costante di tempo T2, è determinata dalle interazioni spin-spin. T2 è il tempo necessario affinché venga annullato il 63 % della magnetizzazione trasversale;
- la seconda, definita da una costante T1, è costituita dalle interazioni spin-reticolo. Esprime il tempo necessario per recuperare il 63 % della magnetizzazione longitudinale.

L'energia che viene rilasciata al cessare dell'impulso costituisce il segnale di risonanza magnetica, che permette di generare un'immagine tomografica cross sezionale. L'intensità di tale segnale è direttamente proporzionale alla densità protonica e alla differenza tra protoni disposti in verso parallelo e antiparallelo, esprimendo così le caratteristiche tissutali dell'organismo, rilevabili nell'immagine di risonanza magnetica. Per discriminare il contributo dei diversi nuclei di idrogeno nello spazio è necessario applicare dei gradienti di campo magnetico, ossia dei campi magnetici aggiuntivi, sovrapposti al campo principale e caratterizzati da intensità crescenti lungo gli assi. Le immagini sono ottenibili su ogni piano anatomico aggiustando l'orientamento dei gradienti dei campi magnetici sull'asse x, sull'asse y e sull'asse z e possono essere ricostruite attraverso l'applicazione di più gradienti.

1.2.7 Segnale Free Induction Decay

Il punto di partenza per comprendere la generazione e la rilevazione del segnale è la legge di Faraday dell'induzione elettromagnetica, che si verifica quando il flusso del campo magnetico attraverso la superficie delimitata da un circuito elettrico è variabile nel tempo. La legge impone che nel circuito (bobina ricevente) si generi una forza elettromotrice indotta

pari all'opposto della variazione temporale del flusso (Eq. (1.10)). Se il flusso magnetico attraverso la bobina è dato da

$$\Phi(t) = \int_{\text{object}} \vec{B}_r(\boldsymbol{r}) \cdot \vec{M}(\boldsymbol{r}, t) d\boldsymbol{r} \tag{1.9}$$

applicando la legge di Faraday dell'induzione si ottiene

$$V(t) = -\frac{\partial \Phi(t)}{\partial t} = -\frac{\partial}{\partial t} \int_{\text{object}} \vec{B}_r(\boldsymbol{r}) \cdot \vec{M}(\boldsymbol{r}, t) d\boldsymbol{r} \tag{1.10}$$

Dove V(t) è la tensione indotta nella bobina ricevente, tale segnale è identificato come il segnale di risonanza magnetica 'grezzo' **[1]**. Da questa formula è possibile osservare che a seconda della variazione spaziale e temporale del vettore di magnetizzazione globale $\vec{M}$, varia la tensione nella bobina ricevente.

Dato che il segnale V(t) è un segnale alle alte frequenze, esso viene demodulato sia in fase che in quadratura: si moltiplica V(t) per un segnale sinusoidale di riferimento (ottenendo così l'inviluppo del segnale) e poi si applica un filtro passa basso per rimuovere le componenti alle alte frequenze. Se la bobina ricevente ha un campo recettivo omogeneo nella regione di interesse, si assume che il segnale possa essere semplificato ed avere la seguente forma

$$S(t) = \int_{\text{object}} M_{xy}(\boldsymbol{r}, 0) e^{-i\Delta\omega(\boldsymbol{r})t} d\boldsymbol{r} \tag{1.1}$$

Inoltre, facendo le seguenti assunzioni:

- L'impulso RF sia applicato istantaneamente in modo che l'intervallo dell'impluso sia nullo.
- Si ignora ogni imperfezione nell'eccitazione e nella ricezione in modo che ogni sfasamento legato alle agli effetti delle *off-resonance* e a tutte le non uniformità di *Br* che pesano sul segnare ricevuto siano omesse **[1]**.
- Se il sistema spin è eterogeneo, si caratterizza il contenuto in frequenza mediante la funzione densità spettrale di spin ($\rho(\omega)$),

il segnale può essere riscritto mediante l'Eq. (1.12).

$$S(t) = \int_{-\infty}^{\infty} \rho(\omega) e^{-t/T_2(\omega)} e^{-i\omega t} d\omega \tag{1.2}$$

Il segnale *Free Induction Decay* (FID) è una risposta transiente del sistema spin a seguito di un impulso di eccitazione ed ha la seguente forma

$$S(t) = \sin\alpha \int_{-\infty}^{\infty} \rho(\omega) e^{-t/T_2(\omega)} e^{-i\omega t} d\omega \qquad t \geq 0 \tag{1.13}$$

L'ampiezza di questo segnale decresce con il tempo e l'ampiezza iniziale dipende dal numero di nuclei reclutati ovvero dal numero di nuclei che "risuonano" con una certa frequenza nell'oggetto di cui si vuole ottenere l'immagine, pari alla frequenza di Larmor **[1]**.

1.3 Sequenze

Gli impulsi RF vengono emessi in serie dette *sequenze*. Variando le sequenze di impulsi a radiofrequenza, la loro durata e gli intervalli di tempo tra questi, si possono ottenere immagini differenti. I tempi di rilassamento dipendono dalla composizione tissutale della struttura anatomica esaminata e conseguentemente consentono di ottenere immagini differenti a seconda della sequenza utilizzata. Si progettano sequenze di impulsi specifiche, volte ad enfatizzare la dipendenza del segnale FID da un determinato parametro. Alcune delle sequenze più comuni sono:

- *T1 pesata*: è una delle più comuni nelle scansioni a scopo clinico e di ricerca. Nelle immagini T1 pesate i tessuti costituiti da lipidi sono bianchi, quelli composti da acqua sono grigi e il fluido cerebrospinale è scuro, generando un buon contrasto tra tipi di tessuti diversi. Questa scansione può essere facilmente impiegata per analisi strutturali, in quanto consente di raccogliere set di dati ad alta risoluzione.
- *T2 pesata*: si tratta di una sequenza ottimizzata per lo studio di fluidi ad alto contenuto di acqua. I tessuti appaiono scuri e il fluido cerebrospinale appare chiaro. Applicate all'imaging cerebrale, queste sequenze definiscono chiaramente i

ventricoli ma mettono in evidenza anche le lesioni cerebrali grazie all'elevato contenuto acquoso dell'edema perilesionale.

- *Inversion Recovery*: permette di ottenere immagini T2 pesate dove è possibile annullare il segnale dell'acqua (*FLAIR – Fluid Attenuated Inversion Recovery*) o del grasso (*STIR – Short Time Inversion Recovery*), rendendo evidenti lesioni difficili da individuare, per sede o conformazione, nelle immagini T2 pesate standard.
- *DWI* o *imaging pesato in diffusione (Diffusion-Weighted Imaging)*: permette la misurazione della distanza di diffusione delle molecole d'acqua. Più breve è questa distanza, più chiara appare la regione considerata. Con questa sequenza è possibile ottenere, tramite opportune post-elaborazioni, immagini di trattografia, ovvero stimare la posizione dei tratti assonali consentendo la visualizzazione architetturale della sostanza bianca.

2. Metodo gemellare

Il metodo gemellare è un affascinante disegno di studio in epidemiologia che permette applicazioni efficaci in tutte le fasi del ragionamento epidemiologico, offrendo soluzioni uniche nel processo di inferenza causale. È fondamentale sottolineare subito che i risultati derivanti dagli studi gemellari sono ampiamente trasferibili alla popolazione generale, della quale i gemelli assicurano la rappresentatività rispetto a numerose caratteristiche di interesse nella ricerca biomedica e psicosociale **[2]**. Più dettagliatamente, il metodo gemellare, consiste nel confronto tra gemelli monozigoti (MZ) e dizigoti (DZ) allo scopo di investigare l'influenza che geni ed ambiente esercitano su un certo carattere di interesse, normale o patologico. Tale confronto riguarda il livello di somiglianza fenotipica, ed è basato sulla diversa correlazione genetica nelle coppie delle due zigosità. Dato che i gemelli MZ sono geneticamente identici, mentre i gemelli DZ condividono in media il 50% del loro patrimonio genetico, al pari di normali fratelli, è chiaro che un'eventuale maggiore somiglianza osservata tra i primi può essere assunta come indice di influenze genetiche sul carattere in studio. Questo ragionamento, però, è valido solo se si assume che i gemelli MZ condividano le esperienze ambientali, rilevanti per la caratteristica in esame, nella stessa misura dei gemelli DZ (*equal environmental assumption*). Per molte variabili tale assunzione può essere ritenuta valida poichè sia i MZ che i DZ condividono i fattori della vita

intrauterina, nonché le esposizioni dell'ambiente familiare e domestico nell'infanzia. Se l'assunzione non è valida, ovvero se una più forte condivisione di esperienze ambientali contribuisce all'eccesso di somiglianza dei MZ rispetto ai DZ, allora il semplice confronto tra MZ e DZ porta ad una sovrastima degli effetti genetici sulla caratteristica in studio **[3]**.

Il metodo gemellare prevede confronti tra coppie di gemelli MZ e coppie di gemelli DZ piuttosto che confronti tra coppie di gemelli MZ e coppie composte da normali fratelli, poiché i gemelli DZ condividono maggiormente i fattori ambientali (docenti, classe scolastica, amicizie, ambiente domestico etc..) rispetto ai normali fratelli i quali, avendo età differente, sicuramente li condivideranno in una misura minore, pertanto l'*equal environmental assumption* sarebbe meno plausibile.

Gli studi classici sui gemelli permettono di stimare l'*heritability* ("ereditabilità"), definita come proporzione della varianza fenotipica spiegata dalla componente genetica, e misurare così il contributo dei fattori genetici alle differenze inter-individuali osservate per la caratteristica in studio. Analogamente, questi studi permettono di stimare la proporzione di varianza spiegata dalle componenti ambientali, che costituisce un'informazione utile per individuare programmi di intervento (es. prevenzione e promozione della salute) basati su fattori modificabili, ad esempio nell'ambito dei comportamenti e degli stili di vita **[2]**.

2.1 Applicazione del metodo gemellare per lo studio dell'ereditabilità dell'anatomia celebrale

Il metodo gemellare può essere utilizzato per stimare l'ereditabilità (h) della morfologia celebrale, in particolare di parametri neuroanatomici locali e di regioni celebrali di interesse. L'ereditabilità si sofferma sulle differenze di uno specifico carattere morfologico presenti tra gli individui e cerca di spiegare quanto di questa diversità sia legata a differenze genetiche. In generale, la variabilità dell'espressione di una determinata caratteristica (fenotipo) tra gli individui, è attribuibile in una percentuale a fattori genetici e in una percentuale a fattori ambientali (che includono errori di misurazione del fenotipo). Quando la percentuale di varianza attribuita a fattori genetici risulta esser nulla, i geni non influenzano la diversità riscontrata per quel fenotipo, ovvero h=0. Viceversa, quando la

percentuale di varianza attribuita a fattori ambientali risulta esser nulla, l'ambiente non influenza il fenotipo considerato, ovvero *h*=1.

La situazione ideale per lo studio dell'ereditabilità di un qualsiasi carattere è quella in cui si riescono a separare i fattori genetici da quelli ambientali; ciò accade in due casi ipotetici:

A) Gemelli MZ che crescono separati in due ambienti completamente diversi (ad esempio, per motivi di adozione): in questa situazione, dato che i gemelli MZ condividono il 100% dei geni, ogni differenza intra-coppia è legata a fattori ambientali esterni non condivisi dai cogemelli. Dunque, i fattori genetici sono del tutto controllati, *h*=0 ovvero la varianza tra i due individui è attribuita ai soli fattori ambientali.
B) Soggetti con genoma differente che conducono la stessa identica vita: in questo caso, poiché l'ambiente è ipoteticamente identico e del tutto controllato, a contribuire alle differenze tra i soggetti sono al 100% i fattori genetici, *h*=1 ovvero le differenze tra gli individui sono legate solo a fattori genetici.

3. Il Rischio Psicotico

Lo Stato Mentale a Rischio o Sindrome di Rischio Psicotico è una condizione caratterizzata dalla presenza di sintomi psicotici (es. deliri, allucinazioni, comportamento disorganizzato) di breve durata e/o di bassa intensità, associati ad una marcata compromissione del rapporto con la realtà e, dunque, del livello di funzionamento sociale (ritiro sociale) e scolastico/lavorativo.

(https://www.terzocentro.it/disturbi-psicotici/esordio-psicotico/)

La schizofrenia è un disordine mentale cronico debilitante preceduto da una lunga fase prodromica durante la quale il paziente sperimenta un progressivo decadere del proprio funzionamento socio-relazionale, o una rallentata acquisizione delle tappe maturative attese per l'età. Questo aspetto, che costituisce clinicamente il nucleo esistenziale della psicosi nascente, è stato ampiamente documentato dalla ricerca, in particolare quella Europea sui sintomi di base, e costituisce il presupposto clinico-psicopatologico della definizione degli stati mentali a rischio come formalizzati nella ricerca contemporanea **[4]**. *L'ERIraos-CL* (Early Recognition Inventory for the Retrospective Assessment of the Onset of

Schizophrenia CheckList) è una scala di valutazione clinica che combina gli indicatori aspecifici di disagio che nella maggior parte dei casi accompagnano i prodromi della schizofrenia con quei sintomi più caratterizzati in senso psicopatologico e che sono indicativi di aumentato rischio di transizione a psicosi **[4]**. È stato dimostrato che nella sua versione Italiana l'*ERIraos-CL* è uno strumento che possiede buone proprietà psicometriche, che lo rendono adatto all'uso routinario come scala per l'identificazione di casi che possono beneficiare di una valutazione più approfondita in merito al rischio di psicosi. Lo strumento ha una capacità discriminante che appare massima non solo per quei sintomi che qualificano un episodio di psicosi come pienamente attivo (dispercezioni, interferenza del pensiero, ideazione paranoide e allucinazioni), ma anche per quelle anomalie dell'esperienza soggettiva indicative di vulnerabilità schizotropica **[4]**.

4. Il Machine Learning in medicina

Il Machine Learning (ML) o Apprendimento Automatico, sotto-area fondamentale dell'Intelligenza Artificiale, è una disciplina scientifica che si occupa dell'implementazione di algoritmi che permettono ai calcolatori (learners) di sviluppare delle decisioni intelligenti e di creare in maniera automatica modelli analitici complessi in base a dati empirici. Si basa sull'idea che i sistemi possono imparare dai dati e prendere decisioni con un intervento umano ridotto al minimo. Questo avviene attraverso il riconoscimento di *pattern*, ovvero schemi ricorrenti nei dati che permettono di classificare determinate situazioni e di ricondurle a specifici esiti.

Schematicamente si possono definire 4 ambiti di applicazione dell'Intelligenza Artificiale in medicina (Figura 1.3):

- Cura del paziente
- Diagnostica generale e per immagini
- Gestione
- Ricerca e sviluppo

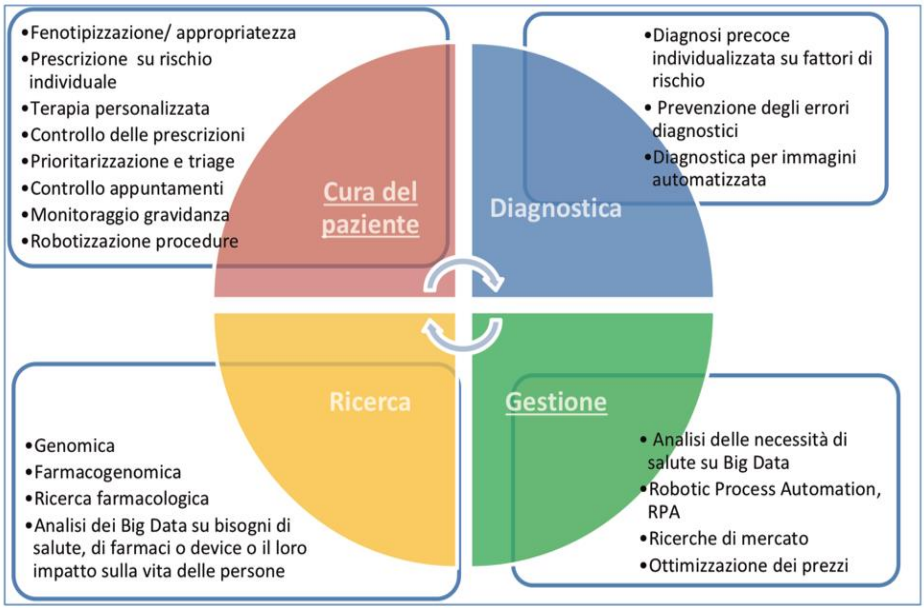

Figura 1.3: Ambiti di utilizzo dell'intelligenza artificiale in medicina **[5]**.

In ambito clinico, attraverso l'Intelligenza Artificiale si possono analizzare ed interpretare i dati disponibili per formulare ipotesi ed ottenere risposte utili per una diagnosi precoce ed un'adeguata pianificazione del trattamento terapeutico, aprendo le porte a strategie di medicina di precisione che tengano conto delle caratteristiche peculiari dei singoli pazienti. In altri termini, è possibile perseguire gli obiettivi di efficacia, efficienza ed appropriatezza tipici della medicina attuale, attraverso l'utilizzo di tecnologie capaci di affinare ed innovare le classiche metodologie epidemiologiche su quantità di dati molto vaste **[5]**.

Altre applicazioni riguardano l'automatizzazione della diagnostica, la gestione economica e la ricerca. Questa rivoluzione tecnologica è alla base della realizzazione della medicina delle *4P* volta a superare il concetto della "cura delle malattie" per rivolgersi alla "gestione complessiva del benessere degli individui". Il termine *4P Medicine* è stato coniato dallo scienziato Leroy Hood. Le 4P stanno per prevenzione, predizione, personalizzazione e partecipazione, una rivoluzionaria architettura clinica basata su una visione generale e d'insieme del paziente. Le attese sono il miglioramento delle cure, la loro estensione ad un numero più grande di malati, la riduzione dei costi, lo stimolo all'innovazione e quindi alla creazione di nuove attività **[5]**.

Il Machine Learning permette di svolgere alcuni compiti che prevedono procedure di pattern recognition nelle immagini, in maniera uguale o migliore dei clinici, come ad esempio la detezione del tumore della pelle (Esteva et al., 2017), tumore del polmone (Yu et al., 2016),

o danni agli occhi (Long et al., 2017). Le aziende di imaging medicale hanno già integrato algoritmi di Machine Learning sui dispositivi ad ultrasuoni per rilevare il cancro al seno (e.g., Samsung, RS80A).

4.1 Diverse metodologie di apprendimento per una macchina

In generale gli algoritmi di Machine Learning possono utilizzare tre metodologie per l'apprendimento:

• Apprendimento supervisionato: sono forniti in ingresso alla macchina un certo numero di osservazioni, dette esempi, composte da una coppia di dati contenenti il dato originale (input) e il risultato atteso (output o classe). Compito della macchina è quello di trovare la regola (funzione o modello) con cui creare una relazione tra l'input e la classe in modo che, al presentarsi di un nuovo dato di cui non si conosce il risultato atteso, essa possa predirlo correttamente. In questa categoria ricadono i problemi di classificazione e regressione.

• Apprendimento non supervisionato: prevede che le informazioni inserite all'interno della macchina non siano codificate, ossia la macchina ha la possibilità di attingere a determinate informazioni senza avere conoscenza dei risultati attesi. Viene fornita alla macchina solo una serie di input da cui essa deve estrarre pattern che li collegano, senza avere alcuna conoscenza della classe. Dovrà essere la macchina stessa, quindi, a catalogare tutte le informazioni in proprio possesso secondo caratteristiche che ricava dai dati stessi. In questa categoria ricadono problemi di clusterizzazione, il cui obiettivo è raggruppare una serie di input in classi, non note in precedenza.

• Apprendimento per rinforzo rappresenta probabilmente il sistema di apprendimento più complesso, che prevede che la macchina sia dotata di sistemi e strumenti in grado di migliorare il proprio apprendimento e, soprattutto, di comprendere le caratteristiche dell'ambiente circostante. In questo caso, alla macchina vengono forniti una serie di elementi di supporto (ad esempio sensori, telecamere, etc) che permettono di rilevare quanto avviene nell'ambiente circostante ed effettuare scelte per un migliore adattamento all'ambiente intorno a loro. Questo tipo di apprendimento è tipico delle auto senza pilota.

4.2 La classificazione

Nell'ambito del Machine Learning e del Pattern Recognition, con il termine *classificazione* ci si riferisce ad una procedura algoritmica che assegna a ogni nuovo dato in ingresso una delle categorie possibili (dette *classi*). Ogni osservazione in ingresso è descritta attraverso un vettore di caratteristiche (*features*) che insieme costituiscono una descrizione globale del dato da classificare. Se le classi di distinzione sono solo due si parla di classificazione binaria, altrimenti di parla di classificazione multi-classe. A partire da un insieme di osservazioni riferite al passato per le quali è nota la classe di appartenenza, i *modelli di classificazione* si propongono di generare un insieme di regole che consentono di predire la classe di osservazioni future e, per questo motivo, i modelli di classificazione fanno parte della categoria dell'apprendimento supervisionato.

4.2.1 Fasi di sviluppo di un modello di classificazione

Lo sviluppo di un modello di classificazione si compone di 3 fasi principali:

1) Fase di training: la macchina analizza solo un sottoinsieme del dataset, denominato *training set*, allo scopo di individuare le regole di classificazione che consentono di attribuire a ciascun dato la corrispondente classe nota.

2) Fase di test: le regole prodotte nel corso della fase precedente vengono impiegate per classificare il resto delle osservazioni presenti nel dataset, detto *test set*, che non erano state incluse nel *training set*, per le quali è nota la classe di appartenenza. In questa fase si valuta la capacità dell'algoritmo di classificare gli esempi facendo dei confronti tra la classe predetta e la classe di appartenenza.

3) Fase di predizione: corrisponde all'effettivo utilizzo del modello di classificazione per assegnare la classe alle nuove osservazioni che si presenteranno nel futuro di cui l'algoritmo non conosce a priori la classe di appartenenza. La predizione è ottenuta applicando le regole generate in fase di training agli attributi esplicativi (*features*) che descrivono la nuova osservazione.

Capitolo 2

Stato dell'arte

1. Storia del metodo gemellare

Nella seconda metà dell'Ottocento Francis Galton, psicologo e naturalista inglese cugino di Darwin, pubblicò il suo più famoso scritto sui gemelli intitolato «*The history of twins as a criterion of the relative powers of nature and nurture*». Egli aveva intuito le potenzialità della gemellarità come strumento di ricerca. Galton riteneva che studiare coppie di gemelli permettesse di effettuare una distinzione fra le tendenze ricevute alla nascita e quelle acquisite con l'esperienza di vita, consentendo quindi di valutare gli effetti dell'ereditarietà (*nature*) e dell'ambiente (*nurture*) nel plasmare le caratteristiche umane. Grande intuizione, che però ha trovato una vera e propria applicazione scientifica solo molti anni dopo, quando è entrato a far parte del quadro un pezzo essenziale di informazione: la distinzione tra gemelli monozigoti e gemelli dizigoti, su cui si baserà l'intero impianto del metodo gemellare **[6]**. II lavoro di Galton ha notevoli limiti metodologici e statistici ma raggiunge comunque il risultato di attribuire ai gemelli il valore di una disposizione sperimentale naturale. Tale disegno sperimentale utilizzato da Galton diventò la base del metodo gemellare e della determinazione delle componenti genetiche ed ambientali di un carattere ereditario. Nel 1929 Holzinger utilizzando sia la varianza che le correlazioni calcolò l'indice di ereditabilità di un carattere, *h*. La formalizzazione è data dalla trascrizione formale delle osservazioni di Galton e suppone che le fonti di variabilità tra cogemelli siano per i monozigoti solo quelle dovute a fattori ambientali, mentre nelle coppie dizigotiche ai fattori ambientali si aggiungono come fonte di variabilità anche i fattori genetici. La formulazione dell'indice di ereditabilità risultava quindi la seguente

$$h = \frac{Var_{DZ} - Var_{MZ}}{Var_{DZ}} \qquad (2.1)$$

nelle ipotesi che Var_{DZ} sia in funzione dei fattori ambientali e genetici, mentre Var_{MZ} sia in funzione dei soli fattori ambientali.

Usando l'indice di correlazione, l'Eq. (2.1) diventava

$$h = \frac{r_{MZ} - r_{DZ}}{1 - r_{DZ}} \qquad (2.2)$$

dove *h* rappresenta una stima della ereditabilità tra le più agevoli, ma la sua stessa formulazione ne determina i notevoli limiti. Infatti, in questa disposizione gemellare si considera la distribuzione dei fattori ambientali nei dizigoti coincidente con quella dei monozigoti **[7]**.

Nel 1934 in uno studio condotto per identificare i fattori causali della schizofrenia, Kallman propone un ampliamento del metodo gemellare classico nel senso di considerare nel disegno sperimentale oltre alle correlazioni intra-coppia tra i gemelli dei due zigotismi, le correlazioni intra-coppia tra coppie di fratelli, di figliastri, di cugini, etc... Allargando cosi il disegno sperimentale, Kallman si propone, ed in parte attua, la riduzione dei limiti imposti al metodo dagli assunti previ e non dimostrati ed in particolare della presupposta uguaglianza degli ambienti nei cogemelli MZ e DZ. II vero superamento della difficoltà è ottenuto soprattutto limitando le informazioni che si richiedono dall'esperimento in quanto Kallman si limita a richiedere la prova di un andamento dei dati compatibile con l'ipotesi di un determinismo genetico. Egli afferma infatti: «Se un supposto fattore genetico esiste, le percentuali di concordanza dovrebbero essere analoghe nei cogemelli DZ e nelle coppie fraterne, decrescere in qualsiasi altro tipo di coppie eccezione fatta per le coppie MZ dove dovrebbero aumentare» **[7]**.

2. Storia dell'*imaging* a risonanza magnetica

La scoperta del principio di risonanza magnetica si deve a Felix Bloch ed Edward Purcell, che, nel 1946, scoprirono questo fenomeno indipendentemente l'uno dall'altro. Ad entrambi fu assegnato il premio Nobel per la Fisica nel 1952.

Nel 1971 Raymond Damadian dimostrò, con esperimenti su cavie da laboratorio, che i tempi di rilassamento magnetico-nucleari dei tessuti sani erano differenti da quelli dei tessuti tumorali, stimolando così i ricercatori a prendere in considerazione la risonanza magnetica come strumento per la rivelazione delle malattie. Nel 1973 Paul Lauterbur sperimentò per primo, e con successo, la possibilità di fare *imaging* con risonanza magnetica utilizzando

come "oggetto test" delle provette contenenti acqua. Per la ricostruzione dell'immagine egli utilizzò una tecnica di retroproiezione simile a quella usata nella TC. Fu Richard Ernst nel 1975 a proporre nell'MRI l'utilizzo di un processo di codifica di fase e frequenza e l'impiego della trasformata di Fourier, elementi, questi, ancora oggi alla base delle moderne tecniche di MRI. Un paio di anni dopo, nel 1977, Raymond Damadian sperimentò l'MRI sull'intero corpo umano. Nel 1980, Edelstein ed i suoi collaboratori sperimentarono l'imaging del corpo usando la tecnica di Ernst, grazie alla quale una singola immagine poteva essere acquisita in circa cinque minuti. Il tempo di imaging verrà drasticamente ridotto a circa cinque secondi, senza significativi cambiamenti della qualità dell'immagine, a partire dal 1986. L'MRI è dunque una scienza giovane, ma in rapida evoluzione.

2.1 Ambiti di applicazione dell'*imaging* a risonanza magnetica

L'MRI viene utilizzata in moltissimi campi: in ambito neurologico, neurochirurgico, traumatologico, oncologico, ortopedico, cardiologico, gastroenterologico. Questo esame è più utile della Tomografia Assiale Computerizzata (TAC) nella patologia muscolo-scheletrica e in ambito neurologico, per lo studio dell'encefalo e del midollo spinale. In altri ambiti clinici, come nel caso dell'epatologia, l'MRI è da considerare complementare alla TAC.

Le macchine utilizzate per l'MRI hanno subito negli ultimi anni un rapido sviluppo, che ha consentito l'allargamento del campo di applicazione della metodica diagnostica. L'utilizzo dell'MRI per patologie cardiologiche e gastroenterologiche, ad esempio, è reso possibile solo dalle apparecchiature più recenti e ad alto campo magnetico, che consentono di eseguire esami completi in tempi significativamente ridotti, con sequenze di immagine ottenibili con tecnica 'a respiro sospeso', che supera il problema dei movimenti involontari, quali sono ad esempio la peristalsi intestinale, il battito cardiaco o la stessa attività respiratoria, aprendo così nuovi orizzonti di applicazione clinica a organi e apparati prima difficilmente studiabili. L' MRI oggi trova così applicazione alternativa a esami più invasivi.

Prima di trovare ampie applicazioni in ambito clinico, l'MRI è stata largamente utilizzata nell'ambito della ricerca scientifica: per esempio nell'articolo *The impact of psychosis on brain anatomy in bipolar disorder: a structural MRI study* **[8]**, è stata effettuata un'analisi dell'anatomia del cervello in pazienti affetti da psicosi. Dalle acquisizioni di immagini

neuroanatomiche mediante *imaging* a risonanza magnetica a 3 Tesla su pazienti di questo tipo, si è dimostrato che essi presentano un deficit di materia grigia nella corteccia sinistra frontale rispetto ai soggetti sani.

2.2 L' *imaging* a risonanza magnetica a 3 Tesla

La qualità delle immagini di risonanza magnetica è direttamente proporzionale alla intensità del campo magnetico che in genere varia negli ambiti di applicazione diagnostica radiologica da 0.5 a 1.5 Tesla (unità di misura di intensità del campo).

L'Imaging a Risonanza Magnetica a 3 Tesla (o *MRI a 3T*) trova utilizzo soprattutto nell'ambito di protocolli di ricerca, sia in campo neuro-radiologico che in radiologia generale, con applicazioni elettive in ambito oncologico per lo studio della vascolarizzazione dei tumori, della risposta ai trattamenti oltreché per studi strutturali tissutali in vivo. Recentemente è stata approvata in Italia anche per l'uso clinico, oltre a quello scientifico.

Questa metodica di *imaging* consente di ottenere immagini nell'arco di pochi secondi (tanto da poterle acquisire in apnea respiratoria), con il vantaggio di ridurre il tempo complessivo di permanenza del paziente nella macchina.

Il principale beneficio legato all'impiego dell'MRI a 3T consiste in un significativo miglioramento del contenuto informativo delle immagini che passa dalla semplice rappresentazione morfologica ad un'analisi delle funzioni, normali o alterate, che può essere spinta fino al livello molecolare, indagando il substrato biochimico delle condizioni patologiche.

Il vantaggio più ovvio dello scanner a 3 Tesla è la più forte potenza del segnale prodotto dal magnete. Più è forte il campo magnetico, più sarà forte il segnale prodotto dai tessuti corporei. Segnali più potenti permettono di raggiungere una migliore risoluzione spaziale ed un miglior contrasto. Dunque, un aumento dell'intensità di campo da 1.5 Tesla a 3 Tesla corrisponde ad un campo magnetico doppio e ad un aumento di quattro volte della definizione delle immagini. Inoltre, l'MRI a 3T offre un aumento della capacità di risoluzione spaziale dell'immagine senza alterare il tempo di esecuzione dell'esame, consente di ottenere immagini cross sezionali del corpo umano ad elevatissimo dettaglio ed è in grado di analizzare strutture anatomiche anche della grandezza di un millimetro. Infatti,

l'utilizzo è pensato specificamente per lo studio di regioni di piccolo volume (ipofisi, ippocampo, nuclei propri del tronco, nervi cranici) e della corteccia cerebrale.

Benefici particolari sono previsti per la diagnosi di epilessia, neoplasie e sclerosi multipla. Nell'articolo *3 Tesla T2*-weighted brain MRI distinguishes multiple sclerosis from incidental white matter microangiopathic lesions* **[9]**, sono riportati i risultati ottenuti in uno studio in cui l'MRI a 3T è stata utilizzata per distinguere le lesioni celebrali della sclerosi multipla da quelle microangiopatiche nella sostanza bianca. Le lesioni della sostanza bianca sono frequentemente rilevate utilizzando tecniche di risonanza magnetica cerebrale. La maggior parte di queste lesioni sono microangiopatiche, ad esempio causate da malattia ischemica di piccoli vasi, ma la demielinizzazione, che avviene nella sclerosi multipla è una causa molto importante e le tecniche di risonanza magnetica convenzionali non sono in grado di distinguere i diversi tipi di lesioni. Gli autori hanno valutato se la tecnica di *imaging* a risonanza magnetica a 3 Tesla, che valuta le lesioni pesate in T2, permette di distinguere persone con sclerosi multipla da persone con lesioni cerebrali derivate da malattia ischemica dei piccoli vasi. Gli autori concludono che dall'analisi delle lesioni T2 con l'MRI a 3T si possono distinguere le lesioni cerebrali della sclerosi multipla della sostanza bianca dalle lesioni microangiopatiche e che l'applicazione clinica di questa tecnica potrebbe migliorare l'accuratezza della diagnosi **[9]**.

Un altro vantaggio non trascurabile sia sul piano diagnostico che economico, riguarda l'uso del mezzo di contrasto. Il potenziamento dell'impregnazione con il magnete a 3 Tesla è in stretta relazione con l'allungamento del tempo di rilassamento T1 piuttosto che con il miglioramento del rapporto segnale-rumore; tale caratteristica comporta una riduzione, anche di tre quarti della dose di mezzo di contrasto somministrata con beneficio per i pazienti e chiari riscontri economici.

2.3 Esempi di studi MRI condotti su gemelli

Tramite opportuni disegni di studio gemellare, ricorrendo all' MRI è possibile indagare l'ereditabilità di struttura e funzionalità cerebrale. Nell'articolo *What Twin Studies Tell Us About the Heritability of Brain Development, Morphology, and Function: A Review* **[10]** sono sintetizzati i risultati principali degli studi MRI effettuati su gemelli in età infantile,

adulta e anziana, volti a comprendere l'impatto di geni e ambiente su anatomia e funzionalità cerebrali nel corso della vita. In generale, dagli studi di MRI strutturale sui gemelli in età infantile si evince che vi è un'elevata ereditabilità dei parametri volumetrici globali del cervello, e delle densità locali di materia grigia e bianca. Vi è un'importante componente spazio-temporale, in quanto l'ereditabilità dipende sia dalla struttura che si sta considerando sia dall'età.

Inoltre, alcuni studi hanno analizzato la struttura del cervello in gemelli neonati con l'obiettivo di valutare l'impatto genetico e ambientale sulla morfologia celebrale nei soggetti appena nati. Questi studi fanno parte dell'*Early Brain Development Study* (Gilmore et al., 2010) dell'*Università del Nord Carolina*. Nel 2010 Gilmore insieme ai suoi colleghi presentò il primo studio sullo sviluppo celebrale di soggetti in età neonatale. I dati MRI strutturali furono acquisiti da 217 neonati alla quarantesima settimana gestazionale, 82 MZ, 100 DZ e da 35 gemelli disaccoppiati e furono analizzati i volumi corticali e sottocorticali. Questo studio ha dimostrato che non c'erano differenze statistiche significative nei volumi celebrali tra MZ e DZ, ma i volumi globali del tessuto, il volume intracranico totale e il volume del ventricolo erano altamente ereditabili, anche se con punteggi inferiori rispetto a quelli osservati in bambini più grandi e in adulti in altri studi (ad esempio, l'ereditabilità del volume globale di materia grigia era 56%, rispetto all'82% in bambini più grandi (Peper et al., 2008) e 82% negli adulti (Baare et al., 2001)) **[11]**.

Nel 2001 Thompson et al. hanno condotto uno studio sull'ereditabilità di alcune caratteristiche strutturali del cervello, quali la densità di materia grigia. In questo studio sono state trovate correlazioni significative tra gemelli MZ nell'area del linguaggio e nelle aree frontali e sensorimotorie. Inoltre, le correlazioni tra gemelli DZ nelle suddette aree sono risultate inferiori, eccetto che nella corteccia frontale.

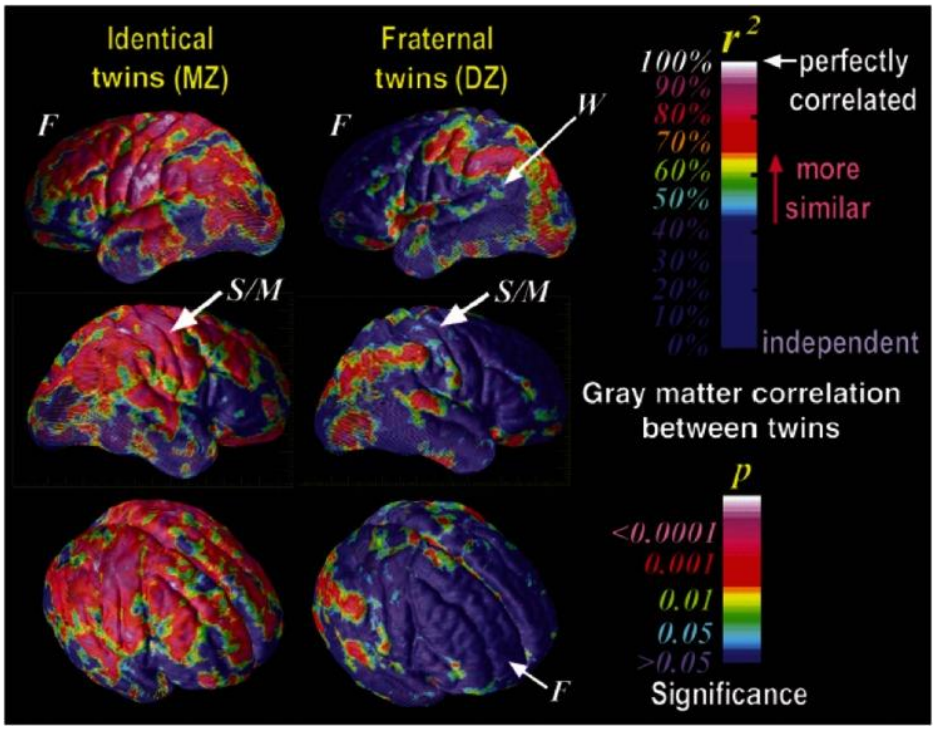

Figura 2.1: Genetic influence on Brain Structure, Thompson et al. 2001. Nota: F corteccia frontale, S/M sensorimotoria, W corteccia di Wernicke **[10]**.

Nella Figura 2.1 sono riportati i valori di significatività delle correlazioni intra-coppia tra gemelli MZ e DZ in termini di volume locale di materia grigia. I gemelli MZ risultano essere più correlati rispetto ai DZ nella maggior parte delle regioni cerebrali.

Capitolo 3

Materiali e metodi

1. Reclutamento dei dati dei soggetti e definizione del datset

La selezione dei soggetti è stata effettuata a partire dal campione complessivo sottoposto a indagini psicopatologiche e MRI a 3T presso il polo di Bosisio Parini, nell'ambito dei progetti SPES e "WHY ME?". A partire da tale campione è stato creato il dataset di analisi che inizialmente era composto da immagini T1-pesate di 152 soggetti di cui 128 gemelli, 54 MZ e 74 DZ, 72 femmine e 56 maschi, di età compresa tra 9 e 30 anni. Tra i 128 gemelli sono stati individuati 10 soggetti di cui non si disponeva dell'immagine MRI a 3T dell'altro gemello della coppia, che sono stati esclusi dalle analisi, e 59 coppie di cui erano disponibili le immagini T1-pesate di ambo i gemelli. Nel presente lavoro di tesi è stato assunto che la similarità neuroanatomica all'interno di ciascuna coppia fosse fortemente alterata dalla discordanza per sesso, pertanto sono state escluse dalle indagini 3 coppie DZ discordanti per sesso. Quindi, le analisi sono state condotte su un campione composto da 56 coppie, 25 MZ e 31 DZ, 50 maschi e 62 femmine di età compresa tra i 9 e i 30 anni. Dal momento che i partecipanti sono stati reclutati in qualità di volontari sani, in tutte le analisi riguardanti lo studio del diverso impatto dei fattori genetici e ambientali sulla morfologia celebrale, i soggetti sono stati considerati come un sottocampione della popolazione generale indipendentemente dal profilo psicopatologico. Tale considerazione ha permesso di non ridurre la numerosità del campione.

A fronte dell'interesse dei progetti SPES e Why Me? verso lo studio dei profili di rischio, in corso di analisi è stata ottenuta l'informazione riguardante il rischio psicotico di ciascun soggetto sulla base della scala ERIraos-CL **[4]**, che è stata utilizzata in un set di analisi preliminari riguardanti l'effetto del rischio psicotico sull'anatomia celebrale. Tali analisi sono state condotte sulle coppie di gemelli MZ e DZ discordanti per tale rischio. A partire dal campione complessivo di 59 coppie, 13 coppie sono risultate discordanti per rischio psicotico secondo le soglie imposte dalla scala ERIraos-CL, 45 sono risultate concordanti

per assenza di rischio psicotico e 1 coppia è concordante per la presenza di rischio. Delle 13 coppie discordanti per rischio psicotico 2 sono risultate discordanti anche per sesso ma in quest'altra fase di studio è stato scelto di non ridurre la numerosità del campione e di considerare anche queste coppie discordanti per sesso essendo queste discordanti anche per rischio psicotico.

2. Acquisizione dei dati sperimentali

Tutte le coppie di gemelli reclutate presso il polo di Bosisio Parini sono state invitate a svolgere una sessione di MRI multimodale a 3T. È stato impiegato uno scanner Philips Achieva 3T (Philips, the Netherlands) installato presso il CEntro Studi di Neuroimaging dell'età Evolutiva (CESNE) del polo di Bosisio Parini, diretto dal Dott. Filippo Arrigoni. I partecipanti sono stati sottoposti all'esame dopo aver letto e sottoscritto il consenso informato allo studio. Nel caso di soggetti minorenni, il consenso è stato visionato e firmato da entrambi i genitori.

La sessione di neuroimaging a 3T ha previsto l'acquisizione di una sequenza neuroanatomica, di una sequenza di diffusione (diffusion tensor imaging, DTI), di una sequenza spettroscopica presso la corteccia cingolata anteriore, e di due sequenze funzionali, sia a riposo (*resting state fMRI*) che durante un esercizio cognitivo con interferenze emotive (*task-based fMRI*) progettato specificamente per rispondere ai quesiti di progetto. Nella presente tesi sono state selezionate e analizzate le sole immagini neuroanatomiche ottenute utilizzando sequenze volumetriche MPRAGE (Magnetization Prepared RApid Gradient Echo) T1-pesate (voxel di 1 mm^3), al fine di estrarre parametri quantitativi di morfologia corticale e sottocorticale oggetto dei confronti gemellari condotti in questo lavoro.

La tecnica di acquisizione di immagini MPRAGE è descritta in dettaglio nel paragrafo successivo.

2.1 Sequenza MPRAGE

La sequenza PRAGE (*Magnetization-Prepared RApid Gradient-Echo*), è una delle sequenze di imaging più popolari. Essa consente di effettuare acquisizioni in 3D su tutto il campo di vista del cervello, permettendo di ottenere immagini ad alta risoluzione spaziale, con un ottimo contrasto tra i tessuti, in un tempo di acquisizione ridotto, sfruttando le potenzialità dell'*Imaging Magnetization Prepared* e gli impulsi rapidi *Gradient Echo*. In particolare, la sequenza comprende un impulso α di recupero di inversione (3D *inversion- recovery*) e N impulsi in radiofrequenza (di *flip angle* θ) di lettura equamente distribuiti nel tempo, con intervallo di tempo tra gli echi τ. Il tempo di ripetizione è definito come

$$T_R = T_I + N * \tau + T_D \qquad (3.1)$$

Dove T_I è l'intervallo di tempo tra l'impulso di inversione, e il primo impulso in RF di lettura, T_D è il tempo di ritardo **[12]**.

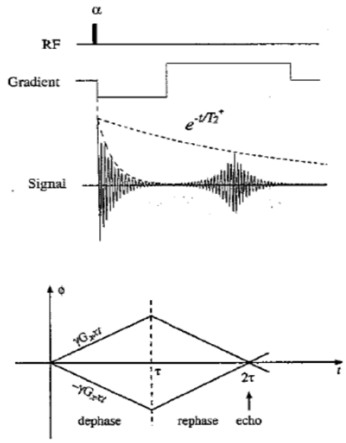

Figura 3.1: Sequenza Gradient echo **[1]**.

Nella Figura 3.1 è illustrata una tipica sequenza *Gradient Echo* in cui viene applicato un impulso in RF di un angolo α (solitamente in una sequenza *Gradient Echo* si ricorre a piccoli *flip angle* per velocizzare l'acquisizione). Dopo l'applicazione di tale impulso viene acceso un gradiente negativo lungo la direzione x. Di conseguenza, tutti gli spin che hanno posizioni diverse in direzione x acquisiranno fasi diverse, e quindi ci sarà una progressiva perdita di segnale. A questo punto, si applica un gradiente positivo della stessa intensità del precedente, che avrà l'effetto di stabilire una coerenza in fase tra gli spin, con la conseguente

rigenerazione del segnale. È possibile generare un treno di echi da un singolo segnale FID, cambiando ripetutamente il segno del gradiente applicato. Il numero di echi che si possono generare è limitato dalla costante di tempo T_2^*, e dalla velocità con cui il gradiente cambia segno. I tomografi moderni permettono di ottenere fino a 64 echi da un singolo segnale FID, consentendo di ricostruire un'immagine 2D da una singola eccitazione **[1]**.

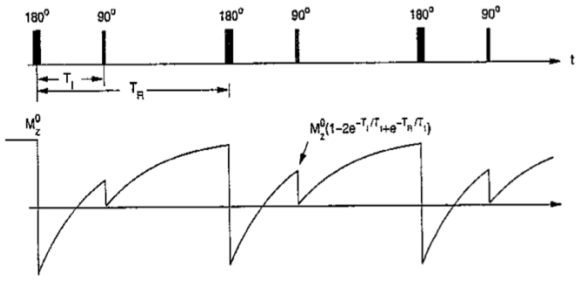

Figura 3.2: Sequenza Inversion-Recovery **[1]**.

La sequenza Inversion recovery è un'altra sequenza popolare per le immagini T1-pesate. Come mostrato nella Figura 3.2, questa sequenza inizia con un impulso in RF di 180° di preparazione (o impulso di inversione) e dopo un tempo T_I, è applicato un altro impulso in RF di 90°. La sequenza può esser scritta in forma abbreviata nel seguente modo

$$(180° - T_I - 90° - T_D)_N \qquad (3.2)$$

dove T_I è detto tempo di inversione e T_D è detto tempo di recupero (o ritardo) **[1]**.

Nella Figura 3.3 è stata rappresentata la sequenza MPRAGE in cui, per semplicità, sono stati omessi i gradienti.

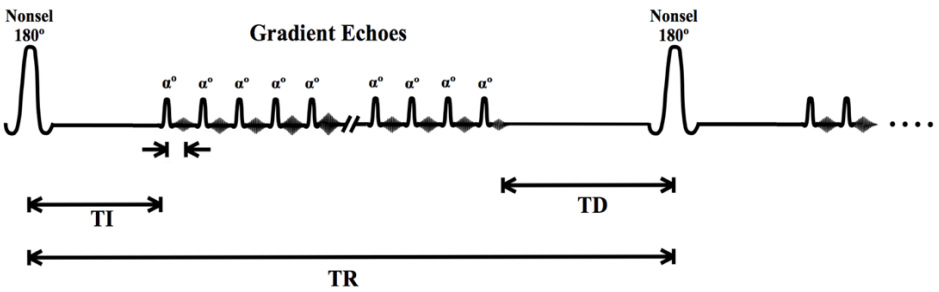

Figura 3.3: Sequenza MPRAGE.

3. Software per le analisi delle immagini neuroanatomiche

Per le analisi delle immagini di Risonanza Magnetica (RM) anatomiche sono stati adottati alcuni tra i principali software di elaborazione di neuroimmagini:

- **FreeSurfer**: è un software per elaborazioni avanzate di bioimmagini relative al cervello (sia anatomiche che funzionali), che opera sui sistemi operativi Linux e MacOS.
- **SPM 12** (*Statistical Parametric Mapping*): si tratta di un software implementato in Matlab che comprende un insieme di strumenti utili per analizzare immagini RM sia anatomiche che funzionali del cervello, oltre a segnali provenienti da tecniche quali spettroscopia nel vicino infrarosso (NIRS) ed elettroencefalografia (EEG).
- **NeuroMiner**: è un software che permette di applicare algoritmi di Machine Learning e analisi di regressione a dati clinici e/o di neuroimmagini.

I software sopra citati permettono di effettuare analisi morfologiche quantitative sulle immagini RM cerebrali. FreeSurfer e SPM 12 sono stati utilizzati in parallelo per effettuare analisi di tipo regionale (*region-based*) e locale (a livello del singolo voxel, *voxel-based*). FreeSurfer è stato utilizzato durante le fasi di *pre-processing* e di analisi ROI (*Region Of interest)-based*. Esso genera le superfici tridimensionali di separazione tra i tessuti principali del cervello, e fornisce in output una serie di parametri numerici regionali relativi a specifiche caratteristiche morfologiche dei tessuti (area superficiale, spessore corticale, curvatura media, volume). Le analisi *voxel-based* hanno previsto l'utilizzo di SPM 12 per il *pre-processing* delle immagini e l'estrazione delle mappe di densità di materia grigia e materia bianca, che in seguito sono state importate nel software di Machine Learning, NeuroMiner.

4. Analisi Region-based

4.1 Pre-processing delle immagini con FreeSurfer

Poiché le analisi di morfologia cerebrale di Freesurfer richiedono un onere computazionale notevole, i passaggi di pre-elaborazione e di ricostruzione delle superfici corticali sono stati svolti utilizzando una piattaforma computazionale *online*, il portale Virtual Imaging Platform (VIP) (http://vip.creatis.insa-lyon.fr). Le immagini sono state importate su VIP nel formato *NIFTI*, su questo portale viene automaticamente svolta una conversione al formato proprietario di *FreeSurfer*, ovvero il formato *.mgz*. I risultati delle analisi sono stati importati in un desktop PC con sistema operativo Linux Centos 7 della Fondazione Policlinico, che è stato utilizzato per la verifica di qualità e l'estrazione dei parametri regionali. In presenza di errori di elaborazione, sono state effettuate correzioni manuali ed è stata rilanciata l'analisi localmente. Nelle analisi locali e sul portale VIP, è stata utilizzata la versione *FreeSurfer 0.3.7*.

In fase di pre-elaborazione, *FreeSurfer* effettua la rimozione delle strutture non cerebrali, quali i tessuti ossei e le meningi (attraverso un processo ibrido che include trasformata watershed e deformazione superficiale), seguita dalla procedura di registrazione di Talairach, che permette di condurre le coordinate dell'immagine nel sistema di riferimento di Talairach. Viene quindi operata una segmentazione automatica della sostanza bianca sottocorticale e delle strutture profonde di sostanza grigia. Dopo la normalizzazione di intensità, che consente di correggere le disomogeneità indesiderate e rende uniforme l'intensità dei voxel di materia bianca, il software opera una classificazione dei voxel di materia bianca e materia grigia corticali. A seguito della correzione automatica dei difetti topologici per ogni emisfero, il software procede con la ricostruzione tridimensionale delle superfici corticali: le diverse superfici vengono evidenziate in relazione ai gradienti di intensità, identificando le aree che delimitano 1) sostanza bianca e sostanza grigia, 2) sostanza grigia e liquido cerebrospinale, nei punti in cui risulta evidente la differenza di intensità che definisce la transizione da un tessuto all'altro.

La registrazione delle superfici corticali individuali in un sistema di coordinate sferico permette l'individuazione delle diverse regioni corticali sulla base delle caratteristiche

strutturali di giri e solchi; utilizzando un approccio Bayesiano, *Freesurfer* parcellizza automaticamente la corteccia cerebrale in 34 regioni neuroanatomiche per ciascun emisfero (4 temporali mediali, 5 temporali laterali, 11 frontali, 5 parietali, 4 occipitali e 4 cingolate), secondo l'atlante di Desikan Killiany (Desikan et al., 2006), a livello delle quali vengono stimati i parametri neuroanatomici. La parcellizzazione avviene implementando una tecnica che assegna automaticamente un'etichetta neuroanatomica ad ogni regione della superficie corticale, basandosi sulle informazioni probabilistiche stimate a partire da un *training set* manualmente etichettato. Questa procedura incorpora sia informazioni geometriche, derivate da un modello corticale, che da una convenzione neuroanatomica. Il risultato è un *labeling* completo dei solchi e giri.

(https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation).

I parametri neuroanatomici di interesse sono generati al termine dell'intera pipeline di elaborazione (Figura 3.4) tramite il comando *recon-all*, che crea un insieme di file contenenti parametri numerici che descrivono le principali caratteristiche morfologiche celebrali. Tali parametri si trovano nella cartella *stats* generata automaticamente dal software e sono:

- *Area superficiale*: è una misura bidimensionale che rappresenta la dimensione della superficie esterna della corteccia. Questa misura viene fornita per tutte le regioni di interesse **[10]**. La misura è espressa in mm^2.

- *Spessore corticale*: è una misura monodimensionale che descrive lo spessore della corteccia che tipicamente è calcolato come la distanza tra i bordi di materia grigia/bianca e i confini di materia grigia/liquido celebrospinale. Questa misura è calcolata su un centinaio di punti lungo la superficie corticale e viene fornita una misura media per una specifica regione di interesse **[10]**. La misura è espressa in mm.

- *Curvatura media*: questo parametro si riferisce alla curvatura media dei solchi e giri della corteccia. *Freesurfer* fornisce in output i valori di curvatura media per ogni regione anatomica di parcellizzazione. La curvatura è misurata come 1/raggio, tale raggio si riferisce ad un cerchio inscritto. La curvatura media è la media di più curvature misurate su una determinata area, quindi l'unità di misura è 1/mm.

(https://surfer.nmr.mgh.harvard.edu/fswiki/MeanCurvature).

- *Volume*: è una misura tridimensionale della dimensione (volume) di una regione o struttura celebrale di interesse. Il volume è equivalente al prodotto tra lo spessore corticale e l'area superficiale **[10]**. La misura è espressa in mm^3.

Oltre ai parametri appena citati, il comando *recon-all* prevede il calcolo di ulteriori parametri volumetrici globali e di regioni sottocorticali, tra i quali:

- volume di materia grigia presente nei due emisferi

- volume di materia bianca presente nei due emisferi

- volume di segmentazione del cervello, cioè il volume di tutti i voxel non appartenenti né al background né al midollo spinale (brain segmentation volume)

- volume intracranico totale stimato

- volume di materia bianca corticale (cortical WM volume)

- volume di materia grigia corticale (Cortical GM volume)

- volume di materia grigia sottocorticale (subcortical GM volume)

- volume del chiasma ottico

-volume del cervelletto

-volume dell'ippocampo

-volume del plesso coroideo

-volume del caudato

-volume del putamen

-volume del pallidum

-volume del terzo e del quarto ventricolo

-volume dell'amigdala

-volume del tronco encefalico

I volumi sottocorticali appena citati vengono calcolati per ogni regione definita secondo un atlante contenente informazioni probabilistiche sulla localizzazione delle strutture anatomiche profonde.

(http://freesurfer.net/fswiki/SubcorticalSegmentation)

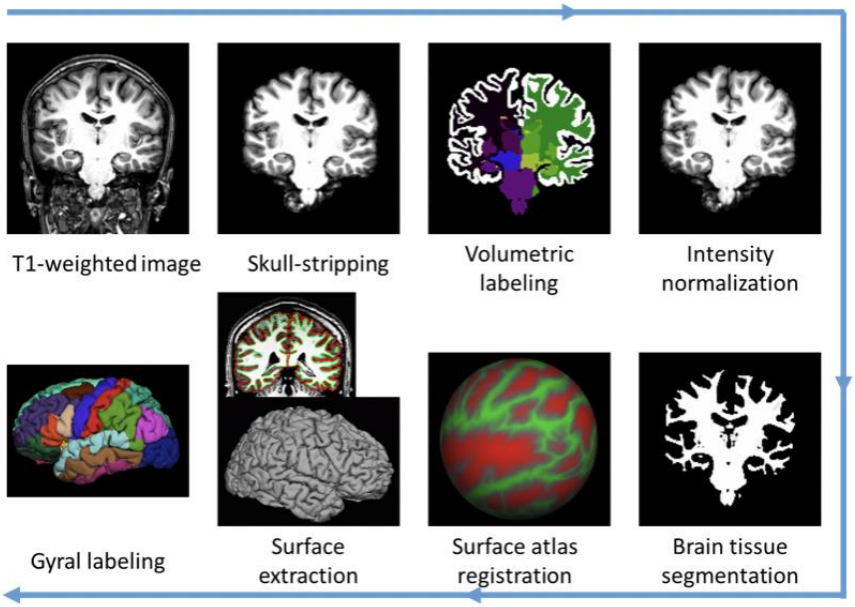

Figura 3.4: Pipeline di analisi delle immagini nel software *FreeSurfer*.

4.1.1 Analisi qualitativa dell'elaborazione delle immagini con FreeSurfer

I risultati delle elaborazioni appena descritte sono stati oggetto di controlli qualitativi per verificare che il software abbia correttamente identificato le superfici che delineano i contorni di materia grigia e materia bianca. La verifica di qualità (o *quality check)* è stata effettuata tramite ispezione visiva delle maschere tessutali e delle superfici ricostruite utilizzando il software *Freeview*. Affinché il software identifichi il percorso dell'immagine da analizzare, è stata impostata la giusta *directory* sul terminale mediante i seguenti comandi:

1)*tcsh*, comando che permette di impostare TENEX C Shell come interprete dei comandi

2)*setenv SUBJECTS_DIR /run/media/student/Tesi/subjects* definizione della directory in cui sono salvate le cartelle ottenute con FreeSurfer dei soggetti

3) *cd $SUBJECTS_DIR*, passaggio alla nuova directory dei soggetti

4)*Freeview*

Nei passaggi di elaborazione con *FreeSurfer*, anche nel caso in cui il comando *recon-all* termini senza interruzioni, potrebbero presentarsi imprecisioni o errori che influenzano la qualità delle superfici ricostruite e l'affidabilità dei parametri. Quindi, ogni immagine è stata sottoposta ad un'accurata ispezione visiva al fine di verificare la qualità di segmentazione tessutale e riconoscimento delle superfici *piale* (tra materia bianca e materia grigia, destra e

sinistra), e di *materia bianca* (destra e sinistra). Gli errori principali che possono verificarsi sono i seguenti:

1)Errori di rimozione dello scalpo

2)Errori di segmentazione

3)Errori di normalizzazione dell'intensità

4)Errori nella localizzazione della superfice piale

5)Difetti topologici

(https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/TroubleshootingData)

Durante la visualizzazione con *FreeView* si è riscontrata, in generale, una buona qualità dei risultati dell'elaborazione di *FreeSurfer*. Solo per il soggetto SBP0034 è stato riscontrato un errore nel riconoscimento della sostanza grigia e nella ricostruzione della superficie piale. Pertanto, si è ritenuto opportuno rilanciare l'analisi con *FreeSurfer* utilizzando come immagine originale l'immagine sottoposta a correzione del bias di intensità tramite SPM 12. Al termine di questa operazione, è stata effettuata una nuova ispezione visiva dei risultati, che hanno superato l'analisi qualitativa.

La buona qualità dei risultati dell'elaborazione è legata alla buona qualità delle immagini importate sul portale VIP e, quindi, in *FreeSurfer*. Infatti, è stato preventivamente verificato con il software *mricron* che tutte le immagini esportate dallo scanner RM in formato DICOM e convertite in formato NIFTI (tramite il software gratuito *dcm2nii*), fossero di buona qualità.

4.1.2 Estrazione dei parametri di interesse da FreeSurfer

Per tutte le immagini che hanno superato il *quality check* si è proceduto con l'estrazione dei parametri morfologici di interesse. Dopo aver impostato la directory sul terminale come descritto nel paragrafo 4.1.1, sono stati digitati i seguenti comandi al fine di estrarre i risultati in formato *.txt*

- aparcstats2table --subjects * --hemi rh --meas thickness --tablefile rh.thickness.txt, tale commando permette di integrare le misure di spessore generate dall'elaborazione e contenute nei file rh.aparc.stats per tutti i soggetti (*), generando un'unica tabella

denominata rh.thickness.txt, che contiene lo spessore di tutte le strutture anatomiche dell'emisfero destro definite secondo l'atlante di Desikan-Killiany ed etichettate nel file rh.aparc.annot.

- aparcstats2table --subjects * --hemi lh --meas thickness --tablefile lh.thickness.txt
- aparcstats2table --subjects * --hemi rh --meas meancurv --tablefile rh.meancurv.txt
- aparcstats2table --subjects * --hemi lh --meas meancurv --tablefile lh.meancurv.txt
- aparcstats2table --subjects * --hemi rh --meas area --tablefile rh.area.txt
- aparcstats2table --subjects * --hemi lh --meas area --tablefile lh.area.txt
- aparcstats2table --subjects * --hemi rh --meas volume --tablefile rh.volume.txt
- aparcstats2table --subjects * --hemi lh --meas volume --tablefile lh.volume.txt
- asegstats2table --subjects * --meas volume --tablefile subcortical.txt, tale commando combinerà tutti i file *volume* per tutti i soggetti (*) per generare un'unica tabella rinominata *volume.txt* che conterrà i globali e di tutte le strutture sottocorticali etichettate in aseg.mgz.

https://surfer.nmr.mgh.harvard.edu/fswiki/asegstats2table

Al termine di queste procedure di estrazione, tutti i file sono stati importati nel software Excel, dove le misure neuroanatomiche sono state aggiunte alle informazioni demografiche e psicopatologiche dei soggetti in vista delle analisi gemellari.

4.2 Analisi statistiche

Le analisi statistiche gemellari sono state incentrate, rispettivamente, sulla correlazione (*classical twin design*) e sulla differenza (*analisi di regressione*) intra-coppia del fenotipo considerato. La scelta di utilizzare approcci differenti è stata legata alla volontà di validare i risultati con due misure di somiglianza diverse. Inoltre, i metodi statistici sono stati implementati nell'ambito delle analisi dell'effetto del rischio psicotico sull'anatomia celebrale.

4.2.1 Analisi di correlazione fenotipica intra-coppia

La stima dell'entità delle influenze genetiche e ambientali su un fenotipo di interesse è condotta a partire da misure di somiglianza tra i fenotipi dei gemelli all'interno delle coppie MZ e DZ.

La correlazione, uno tra i principali metodi statistici utilizzati nell'ambito delle analisi gemellari, misura il grado di associazione tra i valori assunti dal parametro analizzato nei due gemelli che compongono le varie coppie. Il coefficiente di correlazione *r* è una misura dell'intensità dell'associazione tra le due variabili, mentre il segno indica il tipo di associazione:

- positivo, quando le due variabili aumentano o diminuiscono insieme,

- negativo, quando all'aumento dell'una corrisponde una diminuzione dell'altra o viceversa

L'analisi di correlazione è stata svolta in ambiente Matlab (R2018a) considerando misure neuroanatomiche e variabili demografiche di interesse.

Il punto di partenza è stato l'integrazione in un unico documento Excel delle informazioni di tipo demografico e clinico dei soggetti (codice, età, sesso, zigosità, codice del gemello corrispondente, punteggio all'ERIraos-CL, presenza/assenza di rischio psicotico sulla base di ERiraos-CL, presenza/assenza di diagnosi sulla base della *Mini International Neuropsychiatric Interview* (MINI), ecc…) e delle misure neuroanatomiche ottenute con FreeSurfer, suddivise in diversi fogli di calcolo in base alla tipologia. I parametri vengono calcolati per tutte le aree di parcellizzazione definite secondo l'atlante gyral-based di Desikan e Killiany quindi, in ciascun foglio sono stati riportati i valori del parametro nelle diverse regioni dell'atlante per tutti i soggetti che compongono il campione di studio.

In questa prima fase sono stati creati algoritmi in Matlab per il calcolo dei coefficienti di correlazione intra-coppia relativi a coppie MZ e DZ a partire dalle misure normalizzate. In particolare, l'*area superficiale* di ciascuna regione è stata normalizzata per l'area superficiale totale dell'emisfero corrispondente, lo *spessore corticale* di ciascuna regione è stato normalizzato per lo spessore corticale medio di ciascun emisfero e il *volume* è stato normalizzato per la stima del volume intracranico totale. Sono state svolte analisi di

correlazione separate per i diversi parametri morfologici, ottenuti con *FreeSurfer*, relativi alle diverse regioni celebrali. Per il calcolo dei coefficienti di correlazione è stata utilizzata la correlazione parziale. Questa funzione ha permesso la stima dei coefficienti di correlazione tra i valori assunti dal parametro nel gemello1 e nel gemello2 di ciascuna coppia al netto degli effetti dell'età e del sesso dei soggetti. Per ciascun parametro regionale (area, spessore corticale, curvatura media e volume) sono stati ottenuti 68 valori di correlazione intra-coppia per i MZ (r_{MZ}) e 68 valori di correlazione intra-coppia per i DZ (r_{DZ}), uno per ogni regione anatomica dell'atlante di Desikan e Killiany. Concordemente, sono stati estratti XX valori di correlazione relativi ai parametri volumetrici globali e sottocorticali. Per ciascun coefficiente di correlazione è stato estratto anche il *p-value*, il quale fornisce l'informazione sulla significatività statistica del coefficiente di correlazione. Sono stati considerati statisticamente significativi i valori di correlazione il cui *p-value* è minore di 0,05.

Confrontando i coefficienti di correlazione nelle coppie MZ rispetto ai coefficienti di correlazione nelle coppie DZ si possono fare considerazioni circa il diverso grado di similarità neuroanatomica nelle due classi di zigosità. Dalla letteratura MRI è noto che, globalmente i gemelli MZ hanno un grado di correlazione più alto dei gemelli DZ. Il confronto tra i coefficienti di correlazione è stato utilizzato per scegliere, nel modo opportuno, le successive analisi da implementare.

4.2.2 Stima dell'ereditarietà dei parametri morfologici nelle diverse regioni celebrali

La presente analisi si propone di valutare le influenze genetiche sui parametri morfologici cerebrali, ricorrendo ad un confronto tra i coefficienti di correlazione intra-coppia delle coppie di gemelli MZ (r_{MZ}) e delle coppie di gemelli DZ (r_{DZ}). Dalla letteratura è noto che quando il grado di correlazione delle coppie MZ è inferiore al grado di correlazione delle coppie DZ non è presente un contributo genetico al valore assunto dal parametro analizzato, mentre quando il grado di correlazione nelle coppie MZ è pari almeno al doppio al grado di correlazione nelle coppie DZ esiste un contributo genetico **[13]**. In quest'ultima situazione è possibile stimare l'ereditarietà (*h*) mediante la *formula di Falconer* (Eq. 3.3).

$$h = 2(r_{MZ} - r_{DZ}) \qquad (3.3)$$

Questa stima si basa su assunzioni ben precise, tra cui la già citata *equal environmental assumption* **[3]**. La differenza di correlazione (r_{MZ} - r_{DZ}) può essere usata per stimare in quale misura il patrimonio genetico contribuisce al valore assunto dal parametro in un soggetto. Per valutare la significatività statistica della differenza di correlazione è stata utilizzata la *trasformata di Fisher*. I valori trasformati in z di ciascun coefficiente di correlazione sono stati calcolati con le Eq. (3.4) e (3.5).

$$z_1 = 0{,}5 \times \ln\left(\frac{1+r_{MZ}}{1-r_{MZ}}\right) \qquad (3.4)$$

$$z_2 = 0{,}5 \times \ln\left(\frac{1+r_{DZ}}{1-r_{DZ}}\right) \qquad (3.5)$$

Il confronto tra due coefficienti di correlazione (r_{MZ} e r_{DZ}) calcolati su due campioni indipendenti (MZ e DZ), per rigettare l'ipotesi nulla che essi non siano significativamente differenti contro l'ipotesi alternativa che lo siano ($r_{MZ} > r_{DZ}$), pone il problema della forte asimmetria dei valori campionari alla quale si aggiunge quella della non omogeneità delle due varianze **[14]**. La simmetria e la omoschedasticità sono ricostruite mediante la trasformazione di Fisher, per cui il test diventa:

$$Z = \frac{z_1 - z_2}{\sigma_{z_{1-z_2}}} \qquad (3.6)$$

dove

- Z è il valore della distribuzione normale, unilaterale in rapporto all'ipotesi alternativa che i due campioni siano significativamente differenti
- $\sigma_{z_{1-z_2}}$ è l'errore standard della differenza precedentemente ottenuta con

$$\sigma_{z_{1-z_2}} = \sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}} \qquad (3.7)$$

- n_1 ed n_2 rappresentano rispettivamente il numero di coppie MZ e numero di coppie DZ

Mediante il comando Matlab *normcdf (Z,0,1)* è stata ottenuta la funzione di distribuzione cumultativa (o *cumulative distribution function*, *cdf*) dei valori di Z nella distribuzione normale con media 0 e deviazione standard 1. Il *p-value* di verifica dell'ipotesi nulla è stato calcolato come 1-*normcdf(Z,0,1)*; ciò equivale a dire che se la *cdf* di Z cade oltre il 95° percentile nella distribuzione normale con media 0 e deviazione standard 1 il test rigetta l'ipotesi nulla (vale l'ipotesi alternativa che $r_{MZ} > r_{DZ}$), con un *p-value* < 0,05. La stima dell'ereditarietà è stata considerata statisticamente significativa quando il *p-value* della differenza di correlazione è minore di 0,05.

Il calcolo di *h* mediante l'Eq. (3.3) è stato implementato per i parametri area, spessore corticale e curvatura media in tutte le regioni corticali definite secondo l'atlante Desikan e Killiany e per i parametri volumetrici delle regioni corticali e sottocorticali.

4.2.3 Modello gemellare univariato: ACE

Il modello ACE **[15]** è il metodo più frequentemente utilizzato per stimare l'influenza genetica e ambientale su un determinato carattere di studio (fenotipo). Si considerano come variabili osservabili, i valori fenotipici di un dato carattere rilevato sui gemelli, e si considerano variabili latenti (variabili non direttamente osservabili) quelle che modellano l'effetto di fattori genetici ed ambientali sul fenotipo in studio. Il caso più semplice è illustrato nella Figura 3.5 dove P è il fenotipo osservato sui due gemelli, mentre A, C ed E sono variabili latenti **[3]**.

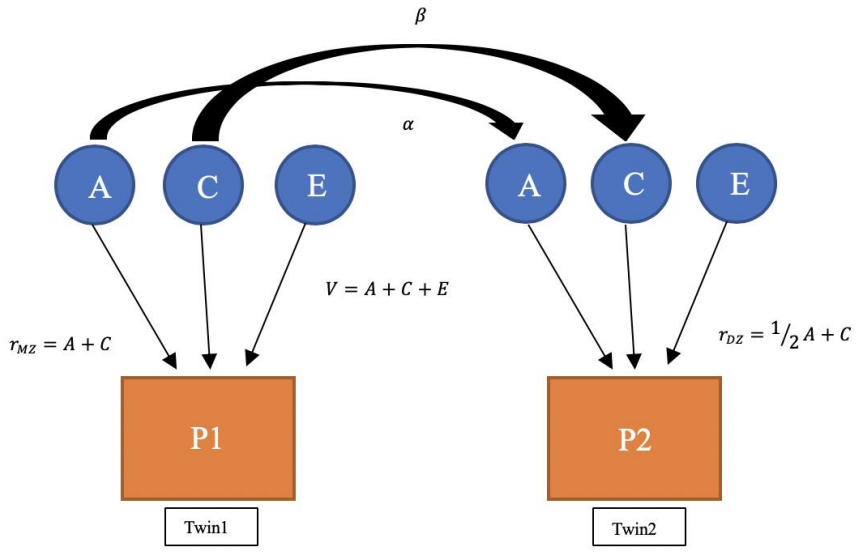

Figura 3.5: Modello ACE: P= fenotipo osservato su ogni gemello (i numeri 1 e 2 si riferiscono ai due gemelli di una stessa coppia); A=fattori genetici; C= fattori ambientali condivisi dai gemelli nella coppia; E= fattori ambientali non condivisi dai gemelli; V=varianza fenotipica totale; $\alpha = 1$ per i MZ oppure 0,5 per i DZ; $\beta = 1$ per i MZ e per i DZ.

In questo modello, la somma totale delle differenze (V) di un fenotipo, riscontrate tra i due individui, è suddivisa in tre fonti di variazione:

- A: rappresenta i fattori genetici che per le coppie MZ sono condivisi al 100% e per le coppie DZ sono condivisi al 50%.
- C: rappresenta i fattori ambientali comuni o condivisi ovvero quei fattori che rendono una coppia di gemelli più simili indipendentemente dal fatto che siano gemelli MZ o gemelli DZ. Corrispondono a quegli eventi che accadono nelle vite di entrambi i gemelli e li influenzano allo stesso modo.
- E: rappresenta i fattori ambientali unici o non condivisi che rendono i gemelli meno simili indipendentemente dal fatto che siano MZ o DZ. Corrispondono a quegli eventi che accadono durante il corso della vita di uno solo dei gemelli della coppia e che non influenzano (o influenzano in modo differente) l'altro gemello.

Il fatto che i gemelli MZ siano geneticamente identici, e che quelli DZ condividano, in media, il 50% del loro patrimonio genetico si traduce in una correlazione tra i fattori latenti genetici (α) per i due gemelli pari a 1 nel caso MZ e a 0,5 nel caso DZ. Assumendo la validità dell'*equal environmental assumption*, la correlazione tra i fattori ambientali condivisi (β) è posta uguale a 1 in entrambi i gruppi di zigosità. Quindi i gemelli, indipendentemente dalla

zigosità, sono ugualmente correlati per il loro grado di esposizione agli eventi ambientali comuni di rilievo per il tratto osservato. Inoltre, per definizione, non vi è correlazione intra-coppia nei fattori ambientali non condivisi **[3]**.

Osservando le correlazioni nei MZ e nei DZ è possibile formulare delle ipotesi riguardo all'entità dei contributi dei tre fattori causali (A, C, E). Dato che la correlazione descrive il grado di somiglianza tra gemelli, essa è determinata dai fattori A e C, mentre le differenze fra gemelli sono determinate dal fattore E. Queste considerazioni permettono di esprimere il coefficiente di correlazione delle coppie MZ mediante l'Eq. (3.8)

$$r_{MZ} = A + C \tag{3.8}$$

e il coefficiente di correlazione delle coppe DZ mediante l'Eq. (3.9) **[16]**.

$$r_{DZ} = ½A + C \tag{3.9}$$

Dall' Eq. (3.8) e dall' Eq. (3.9) si ottiene l'Eq. (3.10), nota come *formula di Falconer*.

$$A = 2(r_{MZ} - r_{DZ}) \tag{3.10}$$

dove A corrisponde all'ereditarietà (*h*). Inoltre, dall' Eq. (3.8) e dall' Eq. (3.9) si ottiene l'Eq. (3.11).

$$C = r_{MZ} - A \tag{3.11}$$

$$E = 1 - r_{MZ} \tag{3.12}$$

L'Eq. (3.12) deriva dal fatto che i gemelli MZ, supponendo che crescano nello stesso ambiente familiare, differiscono solo per i fattori ambientali unici **[17]**.

Inoltre, quando $r_{MZ} < r_{DZ}$, non è presente un contributo genetico alla varianza tra i cogemelli e l'analisi dell'ereditarietà è ininfluente. Quando $r_{MZ} > 2r_{DZ}$, è sicuramente presente un contributo genetico predominante. Quando $r_{DZ} < r_{MZ} < 2r_{DZ}$, probabilmente esiste un considerevole contributo dell'ambiente condiviso alla varianza e, al fine di quantificarlo, è possibile sviluppare il modello ACE **[13]**.

In questa fase sono stati stimati i coefficienti A, C ed E del modello ACE utilizzando le equazioni sopradescritte. La procedura ha previsto l'importazione nel software Excel dei valori dei coefficienti di correlazione delle coppie MZ e DZ, ottenuti mediante gli algoritmi sviluppati in Matlab, dei parametri analizzati in tutte le regioni celebrali di interesse e dei

rispettivi *p-value*. Il calcolo delle stime di A, C ed E è stato implementato per ciascun parametro, solo nelle aree celebrali in cui i coefficienti di correlazione risultavano statisticamente significativi rispetto al *p-value* (*p-value<0.05)*.

4.2.4 Regressione lineare multipla

In parallelo all'analisi gemellare classica (basata sui coefficienti di correlazione intra-coppia) sono state svolte analisi di regressione considerando come misura di similarità la differenza intra-coppia tra i valori del parametro in analisi. Questa fase di analisi ha previsto la creazione del modello di regressione lineare generalizzata multipla per valutare l'influenza relativa delle variabili zigosità, età e sesso (variabili indipendenti) sulle differenze intra-coppia in termini di neuroanatomia (variabili dipendenti).

Il modello di regressione lineare multipla ha la seguente forma

$$Y = w_1X_1 + w_2X_2 + \cdots + w_nX_n + b + \varepsilon \qquad (3.13)$$

con

- Y: variabile di risposta
- X_j: variabili predittive (o regressori)
- w_j: coefficienti che quantificano l'associazione tra X_j e la risposta Y
- ε: errore, ovvero la discrepanza tra la variabile di risposta e la predizione del modello
- b: intercetta

Considerando come variabile di risposta la differenza intra-coppia del fenotipo considerato (*delta*) e modellizzandola tramite le variabili indipendenti zigosità, età e sesso (uguali per entrambi i soggetti di ciascuna coppia avendo escluso le coppie di sesso opposto), il modello può essere riscritto nella forma:

$$delta = w_{zygosity}X_{zygosity} + w_{age}X_{age} + w_{sex}X_{sex} + b + \varepsilon \qquad (3.14)$$

con

- *delta:* variabile di risposta
- $w_{zygosity}$: coefficiente del regressore zigosità

- w_{age}: coefficiente del regressore età
- w_{sex}: coefficiente del regressore sesso

Il coefficiente di regressione esprime l'effetto marginale di ciascun predittore condizionato al valore corrente delle altre variabili predittive. Si tratta evidentemente di un'informazione che può rivelarsi utile per l'interpretazione del modello di regressione e che può fornire una misura di importanza relativa dei diversi predittori nei confronti della variabile di risposta **[18]**.

Per tutte le coppie è stata calcolata la differenza (*delta)* tra il valore assunto dal q-esimo parametro analizzato, nella ROI u-esima, nel gemello1 della coppia k-esima e il valore assunto dal q-esimo parametro analizzato, nella stessa ROI, nel gemello2 della coppia k-esima.

$$delta_{q,u,k} = |gemello1 - gemello2| \qquad (3.15)$$

con

- $delta$= differenza intra-coppia
- $q = 1, \dots, Q$ con Q=numero di parametri analizzati (area, spessore corticale, curvatura media, volume)
- $u = 1, \dots, U$ con U=numero di ROI definite secondo l'atlante di Desikan e Killiany
- $k = 1, \dots, K$ con K=numero di coppie di gemelli

La differenza è stata considerata in valore assoluto in quanto l'ordine con cui considero i due gemelli in ciascuna coppia è ininfluente (non è stato applicato alcun criterio di discordanza all'interno di ciascuna coppia).

Per costruire il modello sopra descritto, si è ricorso alla funzione disponibile in Matlab *fitglm*. È stato creato un modello di regressione lineare generalizzata avente la *matrice di design* sopradescritta. La stima dei coefficienti del modello, tramite il metodo dei minimi quadrati, è stata effettuata separatamente per ciascun parametro di ogni regione celebrale. In questa fase viene anche stimata la bontà del modello – di adattamento (*fitting)* del modello ai dati – tramite un confronto statistico χ^2 tra questo e il modello costante, senza regressori. Preliminarmente all'analisi dei coefficienti è stata operata una selezione dei modelli sulla base della loro bontà di *fitting*. In particolare, sono stati esclusi i modelli in cui il *p-value* del test χ^2 fosse inferiore a 0,05.

Nei modelli con comprovata bontà di *fitting* è stata esaminata l'influenza di ciascun regressore sulla variabile dipendente, ovvero la differenza intra-coppia nel parametro considerato. Tale influenza è stata valutata solo nei casi in cui il coefficiente del regressore fosse statisticamente significativo, ovvero il *p-value<0.05*.

Tramite analisi statistiche sui coefficienti è stato possibile identificare separatamente i parametri regionali in cui la differenza intra-coppia dei valori è influenzata dalle variabili zigosità, età e sesso. In presenza di un'associazione significativa, è stato approfondito il tipo di legame (proporzionalità diretta o inversa) presente tra ciascuna di queste variabili e la variabile di risposta. Per ciascun coefficiente w_j considerato significativo (*p-value<0.05*), si analizza il valore della statistica T *t-stat*. Si possono verificare due situazioni:

- *t-stat* è positivo: w_j aumenta con Y che aumenta. In questo caso la variabile indipendente X_j è direttamente proporzionale alla differenza intra-coppia.
- *t-stat* negativo, w_j diminuisce con Y che aumenta. In questo caso la variabile indipendente X_j è inversamente proporzionale alla differenza intra-coppia.

4.2.4.1 Effetto della zigosità sulla differenza fenotipica intra-coppia

L'analisi di significatività dei coefficienti del modello ha permesso di valutare l'effetto della zigosità sulle differenze intra-coppia nei parametri neuroanatomici, al netto degli effetti di variabili confondenti quali età e sesso.

La variabile indipendente zigosità è stata codificata come variabile binaria

$$\begin{cases} 1 \; se \; MZ \\ 0 \; se \; DZ \end{cases}$$

Quando il coefficiente $w_{zygosity}$ presenta un *t-stat* negativo, la variabile dipendente *delta* è inversamente legata alla variabile binaria *zigosità*. In altre parole, la differenza intra-coppia è maggiore per le coppie DZ, ovvero la similarità del valore del parametro è maggiore nelle coppie MZ rispetto alle coppie DZ. Viceversa, quando il valore della statistica T è positivo, la variabile risposta è legata secondo una proporzione diretta alla variabile binaria di cui sopra, di conseguenza, la similarità nelle coppie DZ è maggiore rispetto alle coppie MZ. Nella sezione risultati sono presentati i soli parametri neuroanatomici con valori di *t-stat* associati al regressore zigosità significativi (*p-value<0.05)*.

4.2.4.2 Effetto dell'età sulla differenza fenotipica intra-coppia

Poiché la morfologia cerebrale cambia continuamente durante lo sviluppo e nel corso della vita, è verosimile che anche le differenze cerebrali tra gemelli varino con l'età, in relazione a fattori genetici e ambientali. Oltre a rimuovere il contributo confondente di tale variabile nello studio degli effetti della zigosità, è stato valutato specificatamente l'effetto dell'*età* sulle differenze neuroanatomiche nelle coppie di gemelli (al netto degli effetti legati alla *zigosità* e al *sesso* delle coppie). A tale scopo in ciascun modello è stata valutata la significatività del coefficiente di regressione della variabile età. Quando in corrispondenza del coefficiente w_{age} il t-stat è negativo, la differenza intra-coppia *delta* è inversamente legata all'età, ovvero, la similarità nella coppia aumenta con la crescita. Quando in corrispondenza del coefficiente w_{age} il t-stat è positivo, la differenza intra-coppia è direttamente proporzionale all'età, quindi la similarità tra i due gemelli diminuisce con la crescita. Nella sezione risultati sono presentati i soli parametri neuroanatomici con valori di *t-stat* associati al regressore zigosità significativi (*p-value<0.05*).

4.2.4.3 Effetto del sesso sulla differenza fenotipica intra-coppia

In questa ulteriore analisi è stata valutata l'influenza del sesso sulla similarità neuroanatomica nelle coppie di gemelli a meno degli effetti dell'età e della zigosità.

La variabile sesso è stata codificata come variabile binaria

$$\begin{cases} 1\ se\ maschi \\ 0\ se\ femmine \end{cases}$$

Quando il *t-stat* del coefficiente w_{sex} è negativo, la similarità è maggiore per le coppie composte da gemelli di sesso maschile mentre, quando il *t-stat* è positivo, la similarità è maggiore nelle coppie di gemelli di sesso femminile. Nella sezione risultati sono presentati i soli parametri neuroanatomici con valori di *t-stat* associati al regressore zigosità significativi (*p-value<0.05*).

4.2.5 Analisi pilota dell'impatto del rischio psicotico sulla neuroanatomia

4.2.5.1 Selezione del campione di studio

L'analisi del rischio psicotico è stata condotta selezionando a partire dal campione gemellare complessivo (costituito da 59 coppie, 25 MZ e 34 DZ, di età compresa tra 9 e 30 anni) le coppie discordanti per rischio psicotico sulla base dei punteggi ottenuti alla scala ERIraos-CL. La soglia di punteggio utilizzata per la determinazione del rischio psicotico è pari a 12. Sono state identificate 13 coppie di gemelli discordanti per rischio psicotico (8 DZ e 5 MZ, di cui 14 femmine e 12 maschi) di cui 2 discordanti anche per sesso, 45 coppie concordanti per l'assenza di rischio psicotico (19 MZ e 26 DZ, di cui 41 maschi e 49 femmine) e 1 coppia MZ concordante per rischio, di sesso femminile.

Recenti studi in letteratura scientifica suggeriscono che l'insorgenza di disturbi psicotici sia scatenata da una complessa interazione tra fattori genetici e ambientali. Di conseguenza, sebbene nelle coppie discordanti i fratelli dei gemelli ad alto rischio siano risultati non a rischio per la scala ERIraos-CL, essi non sono stati considerati come soggetti sani ("gemelli di controllo") ma come soggetti con un rischio più alto rispetto alla popolazione normale media, ovvero come "gemelli a basso rischio".

4.2.5.2 Metodi di analisi dell'effetto del rischio psicotico sulla neuroanatomia

In questa fase lo scopo dell'analisi è di valutare le alterazioni dei parametri neuroanatomici di interesse legate al rischio psicotico. A tal fine sono state eseguite le seguenti analisi statistiche:

1) Confronto tra i valori del parametro nei gemelli ad alto rischio e nei gemelli a basso rischio delle coppie discordanti
2) Confronto tra i valori del parametro nei gemelli ad alto rischio e nei gemelli di controllo
3) Confronto tra i valori del parametro nei gemelli a basso rischio e nei gemelli di controllo

L'effetto del rischio psicotico è stato valutato sui seguenti parametri neuroanatomici: *area superficiale*, *spessore corticale*, *curvatura media* e *volume*. È stato creato, in ambiente

Matlab, un algoritmo in grado di prelevare dal file Excel contenente le informazioni dell'intero campione di 59 coppie, i dati delle coppie discordanti per il rischio. Per ogni coppia discordante l'algoritmo identifica il gemello a rischio e il gemello non a rischio e preleva dal dataset i valori assunti dal parametro considerato nei due gemelli. Per ogni parametro e regione celebrale dell'atlante di Desikan e Killiany, l'algoritmo crea due vettori di dati ognuno contenente in ciascuna cella il valore del parametro, rispettivamente nel gemello ad alto rischio psicotico e nel corrispondente gemello a basso rischio, come mostrato nella Figura 3.6.

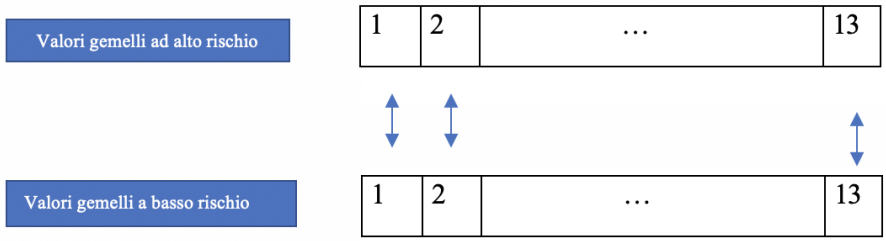

Figura 3.6: Sono rappresentati i due vettori di 13 celle, corrispondenti al numero di coppie discordanti per rischio psicotico, che rispettivamente contengono i valori del parametro analizzato nei soggetti ad alto rischio e nei soggetti a basso rischio.

I gemelli appartenenti alle coppie concordanti per l'assenza di rischio psicotico sono stati considerati come soggetti sani ("gemelli di controllo"). Nell'analisi è stato considerato uno solo dei gemelli della coppia concordante. Inoltre, è stata ispezionata l'età dei soggetti delle coppie discordanti in modo da prelevare dal campione dei gemelli di controllo 13 soggetti di età comparabile a quella delle coppie discordanti per il rischio.

Al fine di verificare se i dati nei vettori a confronto provenissero da una distribuzione normale, è stato applicato il *Jarque-Bera test.*

(*https://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/jarqbera.htm*)

Per la gran parte delle regioni celebrali i dati nei campioni a confronto si distribuiscono in modo normale, quindi è stato scelto di applicare test parametrici. In particolare, nel confronto tra i valori del parametro morfologico di interesse nei gemelli ad alto rischio e nei rispettivi gemelli a basso rischio è stato applicato il *paired-sample t-test*, mediante la funzione *ttest* disponibile in Matlab. La scelta del t-test per dati accoppiati è legata alla presenza di osservazioni appaiate, associate ai due gemelli di ciascuna coppia. La funzione, applicata a tutte le misure neuroanatomiche, fornisce in output un vettore contenente un valore logico

per ogni parametro celebrale, il quale indica la decisione del test riguardo l'ipotesi nulla che i dati nei due vettori a confronto (quindi i valori differenza punto a punto tra i due vettori) provengano da una distribuzione normale con media nulla con una soglia di significatività pari a $p=0.05$. Quindi, il test consente di verificare, per ogni regione celebrale, se vi sia una differenza significativa tra le medie dei valori nel gemello ad alto rischio e nel gemello a basso rischio:

- Quando il valore logico è pari a 1 l'ipotesi nulla è rigettata, ovvero la differenza tra le medie dei due campioni è significativamente diversa da zero e di conseguenza il parametro assume valori significativamente diversi nei due gruppi confrontati.
- Quando il valore logico è pari 0, l'ipotesi nulla è accettata pertanto, i valori assunti dal parametro nei due gruppi confrontati non sono significativamente diversi tra loro.

Per i confronti tra 1) i valori del parametro nei gemelli ad alto rischio e nei soggetti di controllo e 2) i valori del parametro nei gemelli a basso rischio e nei soggetti di controllo, è stato utilizzato il test parametrico *two-sample t-test*, il quale è implementabile in Matlab mediante la funzione *ttest2*. La scelta del t-test per dati non accoppiati è motivata dall'indipendenza dei soggetti nei due campioni in analisi. Il test restituisce la decisione riguardo l'ipotesi nulla che i dati nei due vettori a confronto, provenienti da due campioni indipendenti con distribuzione normale, abbiano stessa media.

L'interpretazione congiunta delle differenze emerse dai tre confronti per ciascuna misura morfologica cerebrale ha permesso di inferire sull'effetto del rischio psicotico sulla neuroanatomia e di fare supposizioni sui fattori (genetici e/o ambientali) che determinano tale effetto.

5. Analisi Voxel-based

In questa analisi parallela sono stati implementati algoritmi volti a valutare differenze nella somiglianza intra-coppia nell'anatomia del cervello locale (voxel per voxel) tra i due zigotismi. L'approccio analitico *voxel-based* si differenzia dalle tecniche *region-based* sopradescritte in quanto considera l'informazione neuroanatomica localmente, senza effettuare ipotesi a priori circa l'appartenenza di ciascun punto cerebrale a una determinata regione anatomica/funzionale. Nei confronti gemellari, tale approccio permette

l'identificazione di somiglianze neuroanatomiche intra-coppia in piccoli *cluster* cerebrali che non emergerebbero a livello regionale. Nella presente analisi sono state messe a confronto le densità locali di materia grigia e materia bianca all'interno di ciascuna coppia. Le mappe di similarità anatomica intra-coppia sono state oggetto di ulteriori confronti tra coppie MZ e DZ.

In particolare, dopo aver effettuato il pre-processing delle immagini RM anatomiche T1-pesate mediante il software *SPM 12,* le mappe "differenza" di densità di materia grigia/bianca relative a ciascuna coppia di gemelli sono state date in ingresso al software di Machine Learning (ML) *NeuroMiner* (version 0.999; https://www.pronia.eu/neurominer/). L'analisi ML ha avuto come scopo la creazione di un modello di classificazione che predica la zigosità delle coppie di gemelli a partire dall'immagine RM.

5.1 Pre-elaborazione delle immagini

Nell'ambito delle analisi *voxel-based* il pre-processing delle immagini è stato effettuato mediante il software SPM 12. Esso è implementato in ambiente Matlab e comprende un insieme di strumenti utili per analizzare immagini RM sia anatomiche che funzionali del cervello.

5.1.1 Visualizzazione e riallineamento delle immagini con SPM 12

L'importazione in SPM delle immagini T1-pesate avviene in formato NIFTI (Neuroimaging Informatics Technology Initiative). Le immagini DICOM in uscita dallo scanner RM sono state convertite in NIFTI tramite il software gratuito *dcm2nii* e importate nel software SPM 12. Le immagini originali sono state visualizzate per un'ulteriore verifica di qualità (già svolta nelle analisi *region-based* di *Freesurfer*) tramite la funzione *display* disponibile nell'interfaccia grafica di SPM 12.

Una prima fase di elaborazione ha previsto l'impostazione di un'origine e un orientamento comune a tutte le immagini che costituiscono il campione secondo i criteri seguenti. L'origine in ciascuna immagine è stata impostata in corrispondenza della commissura anteriore, ovvero quel piccolo fascio di sostanza bianca a livello dell'encefalo anteriore che svolge un'azione di connessione tra i due emisferi. Per agevolare il confronto delle immagini

dei soggetti con le mappe tessutali di riferimento utilizzate da SPM 12, la linea tra la commissura anteriore e posteriore deve essere parallela all'asse orizzontale e i due emisferi devono essere simmetrici. Nelle immagini che non rispettavano questi requisiti è stato effettuato un riallineamento modificando l'inclinazione del volume in radianti rispetto ai tre assi (pitch, roll, yaw) e reimpostando così l'inclinazione in ciascuna direzione.

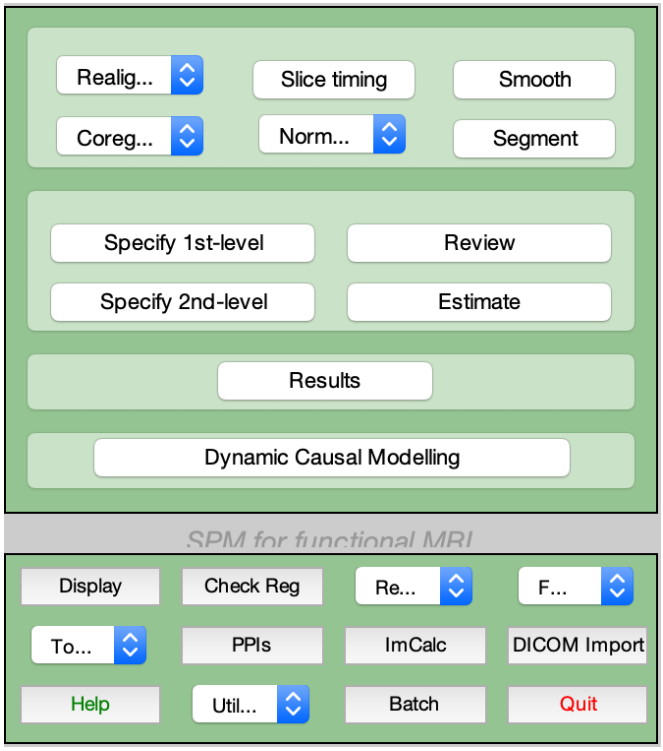

Figura 3.7: Interfaccia grafica SPM12.

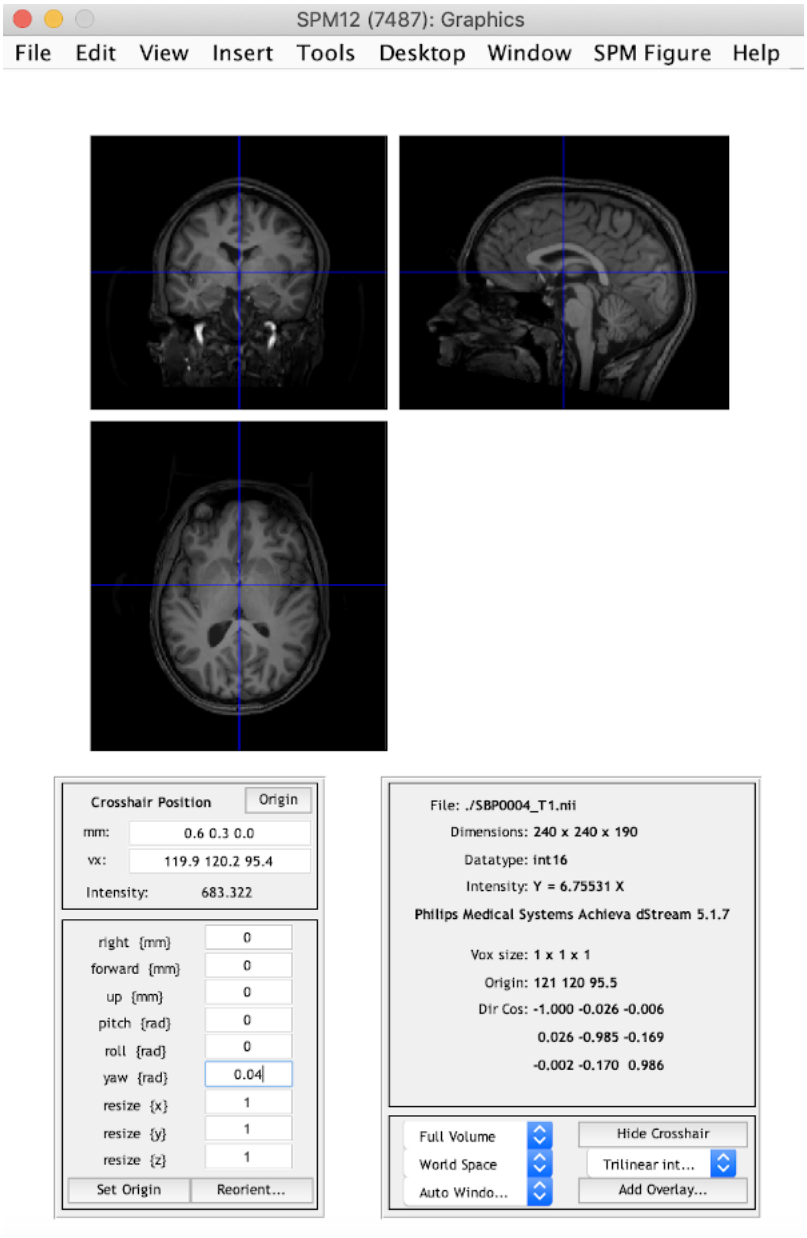

Figura 3.8: Interfaccia funzione Display.

5.1.2 Segmentazione dei tessuti celebrali con SPM 12

Per la segmentazione dei tessuti celebrali è stata utilizzata la funzione *segment* nell'interfaccia grafica di SPM 12. La procedura di segmentazione ha permesso di ottenere, a partire dall'immagine RM originale, una serie di mappe di densità dei diversi tessuti cerebrali. Il pannello di scelta dei parametri di elaborazione (*batch*) di SPM 12 per la segmentazione è rappresentato nella Figura 3.9.

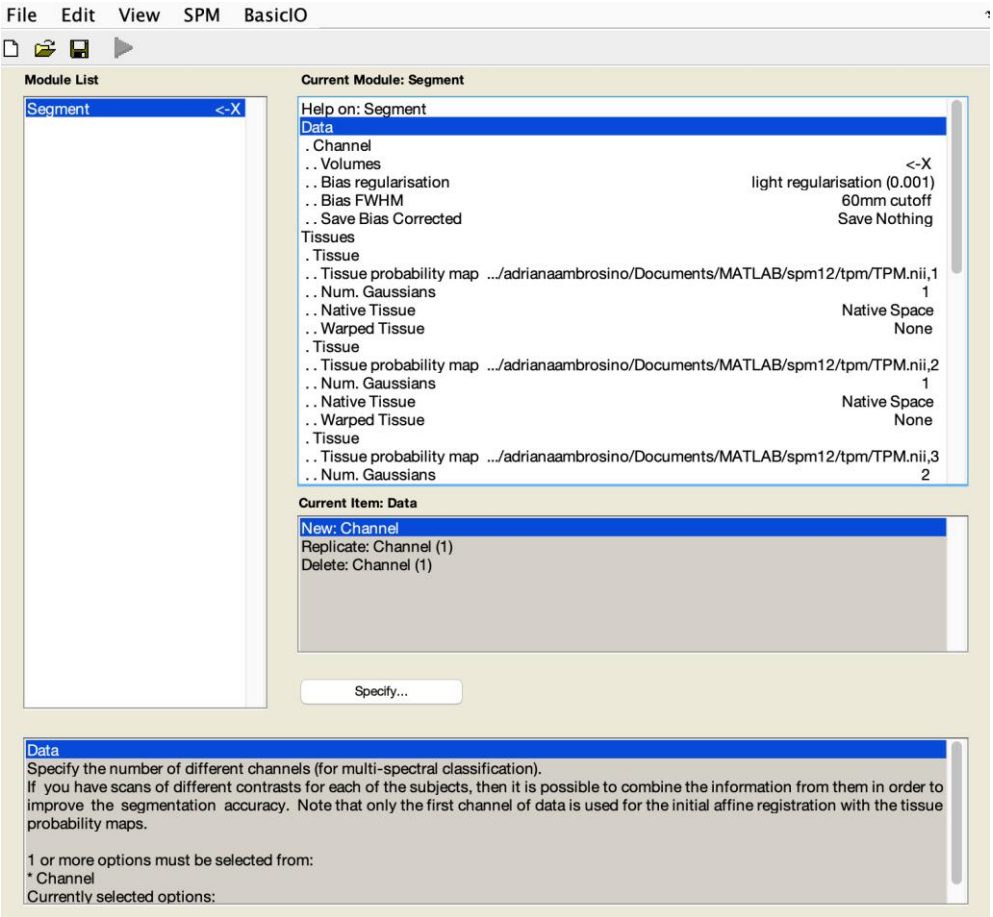

Figura 3.9: *Batch* di SPM 12 per la segmentazione.

Disponendo di un'unica modalità di immagini, ovvero le immagini T1-pesate, e disponendo di una sola immagine per ogni soggetto, è stata impostata la modalità di analisi a singolo canale. Inoltre, sono stati specificati i seguenti campi:

- In *Volumes*: sono state caricate le immagini da segmentare.
- *Bias regularisation*: è stata impostata una regolarizzazione del bias in quanto il *bias* del campo magnetico sui tessuti genera variazioni spaziali di intensità.
- *Save Bias Corrected*: in questo modo, si ha la possibilità di salvare le versioni delle immagini corrette rispetto alle inomogeneità legate all'intensità.

(https://www.fil.ion.ucl.ac.uk/spm/doc/spm12_manual.pdf)

In *Tissues* sono stati definiti i tessuti da identificare e le mappe tessutali a priori utilizzate dal software per l'identificazione dei diversi tessuti. In questa sezione è possibile richiedere l'esportazione delle mappe dei tessuti di interesse, ovvero materia grigia (GM) e materia

bianca (WM), in spazio nativo e/o standard. Nella presente analisi, il campo *Native Tissue* è stato impostato su *Native + Dartel imported* in modo che le immagini di materia grigia e materia bianca venissero fornite sia alla risoluzione delle scansioni originali (*Native*), sia ad una risoluzione più bassa che è utile per la registrazione non lineare con l'algoritmo DARTEL delle mappe tessutali di tutti i soggetti (*Dartel imported*) **[19]**.

Il riconoscimento dei tessuti si basa sull'utilizzo di mappe di probabilità tessutali (che il software conosce a priori) create a partire da una media di molti soggetti sani. L'algoritmo assegna automaticamente a ciascun voxel di ogni immagine un valore di probabilità di appartenenza a ciascun tessuto. Al termine della procedura di segmentazione, per ciascuna immagine di partenza, sono generati i seguenti file: i file *c1*.nii* sono le mappe tessutali di materia grigia, i file *c2*.nii* sono le mappe tessutali di materia bianca. I file *rc1*.nii e rc2*.nii* sono le versioni *Dartel Imported* delle immagini, cioè sono le stesse mappe tessutali, rispettivamente di materia grigia e di materia bianca, ma con una risoluzione più bassa.

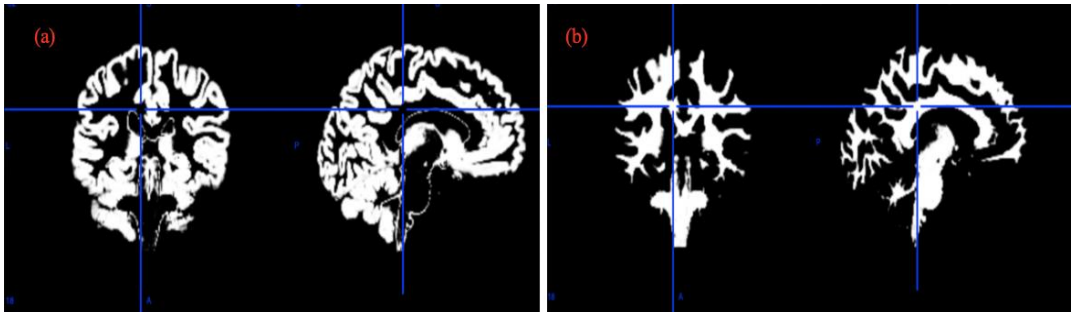

Figura 3.10: (a) Esempio di immagine c1*.nii. (b) Esempio di immagine c2*.nii. Scelta una *slice* di esempio, nell'immagine (a), sono presenti i voxel che con un'elevata probabilità appartengono alla materia grigia (mappa di densità di materia grigia) mentre, nell'immagine (b), sono rappresentati i voxel che con un'elevata probabilità appartengono alla materia bianca (mappa di densità di materia bianca).

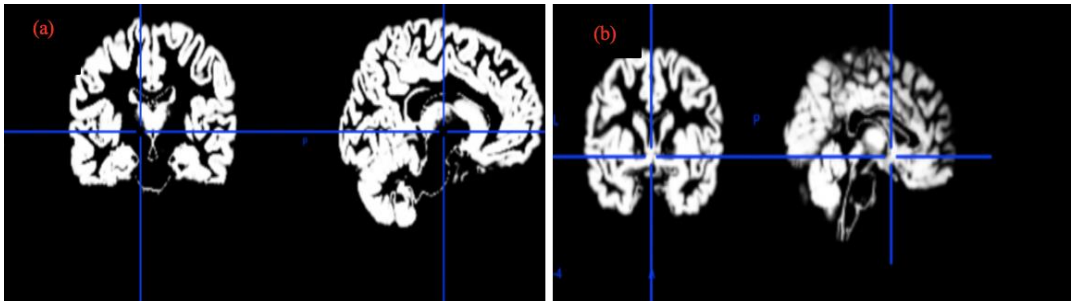

Figura 3.11: (a) Esempio di immagine c1*.nii. (b) Esempio di immagine rc1*.nii. Scelta una *slice* di esempio, nell'immagine (a), sono presenti i voxel che con un'elevata probabilità appartengono alla materia grigia alla risoluzione originale mentre, nell'immagine (b), sono rappresentati i voxel che con un'elevata probabilità appartengono alla materia grigia alla risoluzione *Dartel imported*.

Le immagini ottenute al termine della segmentazione sono state sottoposte ad un'ispezione visiva con il software *mricron* al fine di valutare la qualità della segmentazione. A ciascuna immagine originale sono state sovrapposte le mappe di densità di materia bianca e grigia per verificare che le aree di appartenenza ai rispettivi tessuti fossero state riconosciute correttamente.

5.1.3 Generazione del template anatomico con l'algoritmo DARTEL (Diffeomorphic Anatomical Registration using Exponentiated Lie algebra)

Un template, ottenuto a partire da un insieme di immagini del cervello, non è altro che una nuova immagine che accorpa, in qualche modo, le caratteristiche di tutte le immagini neuroanatomiche che compongono il campione di studio. Le mappe tessutali *Dartel imported* sono state utilizzate per la creazione del template basato sulle caratteristiche anatomiche del gruppo dei soggetti inclusi nello studio. Il template di gruppo è stato utilizzato come riferimento per la registrazione delle immagini tessutali dei diversi soggetti nello spazio template comune, in cui vi sia corrispondenza neuroanatomica puntuale tra soggetti.

Per la creazione del template sono state utilizzate le immagini T1-pesate dell'intero dataset di studio.

Il template è stato creato tramite l'algoritmo *DARTEL (Diffeomorphic Anatomical Registration using Exponentiated Lie algebra)*, disponibile in SPM 12, implementato per la

registrazione diffeomorfica (non lineare) di immagini tra più soggetti. La registrazione (o *image registration*) è un processo di sovrapposizione e allineamento di due o più immagini rappresentative di una stessa scena. Questo processo prevede che un'immagine sia considerata come immagine di riferimento, e che siano poi applicate trasformazioni geometriche o deformazioni locali alle altre immagini al fine di allineare queste ultime all'immagine di riferimento. Tale procedura permette di confrontare caratteristiche comuni tra immagini cerebrali di soggetti differenti.

L'implementazione di questo algoritmo è avvenuta usando la funzione *Run Dartel (create Templates)*. Vengono caricate tutte le immagini *rc1*.nii* e *rc2*.nii* ottenute al termine della segmentazione, e poi è avviato l'algoritmo.

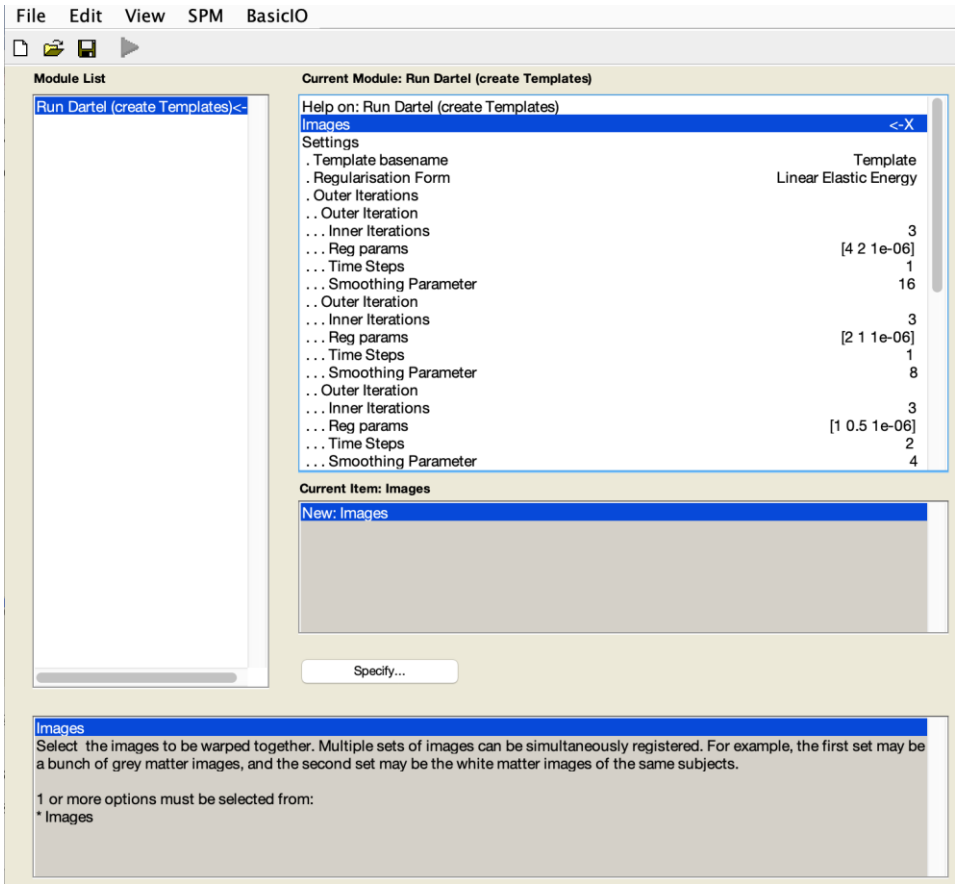

Figura 3.12: Interfaccia grafica della funzione *Run Dartel (create templates).*

L'algoritmo di registrazione è iterativo e consta dei seguenti passaggi. In un primo momento, a ciascuna immagine tessutale vengono applicate delle trasformazioni rigide in modo da allinearle il più possibile alle mappe di probabilità per ciascun tessuto. Successivamente

avviene la vera e propria registrazione: essa include l'allineamento delle immagini di materia grigia e simultaneamente l'allineamento delle immagini di materia bianca di tutti i soggetti. Tale allineamento inizia calcolando una media di tutte le immagini che viene utilizzata come template iniziale (immagine di riferimento). Ad ogni iterazione vengono calcolate (i) le deformazioni che si verificano per allineare le immagini al template (ovvero i campi di deformazione che vengono salvati nei file *u_rc1*.nii*) e (ii) il nuovo template calcolato come la media delle immagini ottenute applicando la trasformazione inversa (*Template*.nii*). Il ciclo viene ripetuto 18 volte e ad ogni iterazione il template diventa più accurato, migliorando la risoluzione spaziale. Al termine di questa procedura, alle immagini anatomiche iniziali di ogni soggetto viene applicata la deformazione finale stimata in tutti i voxel dell'immagine per la registrazione al template definitivo (*Template6.nii*).

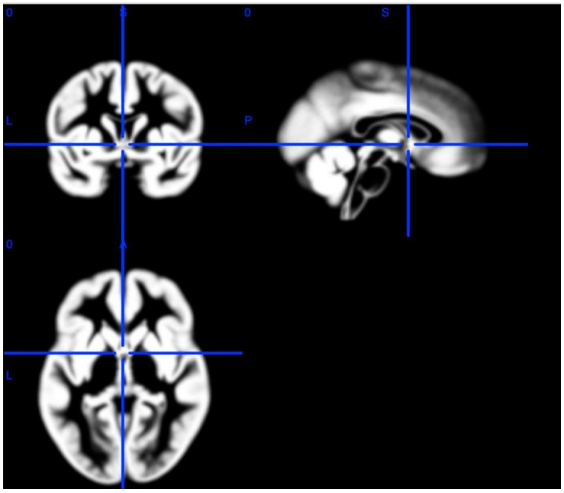

Figura 3.13: Template6.nii.

Dopo aver registrato tutte le immagini al template di gruppo, ogni voxel di ciascuna immagine corrisponde a un determinato punto neuroanatomico in tutte le immagini che compongono il dataset di studio. Inoltre, a partire dal template di gruppo è possibile ricondurre le immagini dallo spazio template allo spazio standard MNI in cui sono definiti i principali atlanti neuroanatomici.

5.1.4 Normalizzazione in spazio MNI (Montreal Neurological Istitute) del template DARTEL con SPM 12

Il template ottenuto con l'algoritmo DARTEL è definito nello spazio dei soggetti; è utile, a partire da questo template, entrare in uno spazio "standard" ovvero in uno spazio in cui ad una qualsiasi coordinata (x,y,z) dell'immagine, corrisponde una specifica regione neuroanatomica. Lo spazio MNI non è altro che un "cervello standard" definito dal *Montreal Neurological Istitute* a partire da un elevato numero di scansioni su soggetti di controllo. Durante l'operazione di normalizzazione le immagini di tutti i soggetti vengono portate allo spazio del template (infatti, si da in ingresso il *Template6.nii*) e poi sono riportate automaticamente allo spazio MNI. Al termine di questa procedura, le immagini dei diversi soggetti presentano una corrispondenza puntuale tra loro e con gli atlanti neuroanatomici in spazio standard MNI.

In questa fase sono stati dati in ingresso i campi di deformazione *u_rc1*.nii* ottenuti in uscita dall'algoritmo DARTEL ed è stata impostata la modalità *Preserve Amount* ("*modulation*") in modo che le immagini preservassero la quantità totale di segnale dall'intero cervello. Al fine di ridurre l'effetto di artefatti puntuali nelle immagini, è stato applicato un filtraggio spaziale passa-basso (*smoothing*) tramite un *Kernel Gaussiano* con una FWHM di 6 mm x 6 mm x 6 mm. Al termine della normalizzazione si ottengono le immagini *smwc1*.nii* e *smwc2*.nii* ovvero le immagini di materia grigia e materia bianca smussate, modulate e normalizzate in spazio MNI. A queste immagini è possibile sovrapporre un atlante, definito in spazio MNI, in modo da individuare le regioni neuroanatomiche corrispondenti ai diversi voxel dell'immagine.

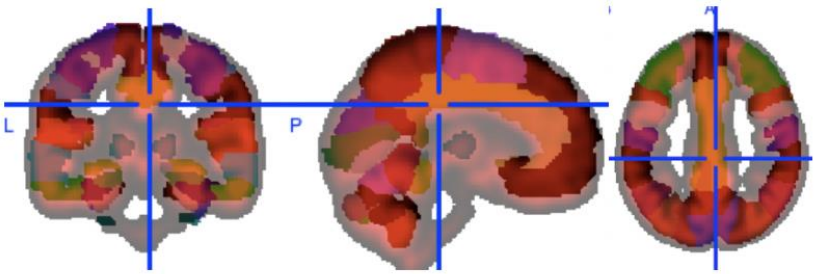

Figura 3.14: Esempio di immagine smwc1*.nii sovrapposta all'atlante AAL (*Automated Anatomical Labeling*), definito in spazio MNI. Rappresentazione di una *slice* nel piano coronale, sagittale e orizzontale.

Le immagini ottenute al termine del pre-processing con SPM 12 sono state utilizzate nelle analisi di Machine Learning. Eventuali differenze focali tra le immagini rifletteranno differenze focali nell'anatomia del cervello.

5.2 Calcolo delle mappe di differenza tessutale intra-coppia

A partire dalle mappe di materia grigia (*swmc1*.nii*) e dalle mappe di materia bianca (*swmc2*.nii*) ottenute al termine del pre-processing con SPM 12, sono state calcolate le mappe di differenza di tessutale intra-coppia che, nel presente lavoro, chiameremo *immagini delta*. Tali immagini sono rappresentative (in senso inverso) della similarità nei volumi locali dei due tessuti all'interno di ciascuna coppia. In particolare, è stata calcolata l'*immagine delta* di materia bianca e di materia grigia in ogni coppia MZ e in ogni coppia DZ.

Sono state utilizzate le funzioni *spm_vol* e *spm_read_vols* per caricare i volumi di ciascun gemello in Matlab e poiché si è assunto che i due gemelli fossero interscambiabili tra di loro (non è stato applicato alcun criterio di discordanza all'interno di ciascuna coppia), è stato calcolato il modulo della differenza. Successivamente è stata utilizzata la funzione *spm_write_vol* per salvare le nuove *immagini delta* che definiscono ciascuna coppia MZ e DZ.

5.2.1 Mappa di differenza tessutale intra-coppia (immagine *delta)*

L'immagine di differenza puntuale tra le mappe di densità tessutale dei due gemelli di ogni coppia è stata usata come indicatore di similarità neuroanatomica intra-coppia. Infatti, solo al termine di un confronto tra i due gemelli che compongono ciascuna coppia si può conoscere il livello di similarità presente. L'algoritmo di classificazione è stato utilizzato per valutare quanto le differenze neuroanatomiche *voxel-based* in coppie di gemelli possano predire la zigosità, e di conseguenza quanto siano influenzate da geni e ambiente. A tale scopo, l'algoritmo confronta a livello del singolo voxel le differenze neuroanatomiche nelle coppie di gemelli MZ e le differenze neuroanatomiche nelle coppie di gemelli DZ.

In corso di analisi è stato osservato che le *immagini delta* delle coppie di gemelli sia di materia grigia che di materia bianca mostravano una caratteristica particolare: ponevano in

evidenza una sorta di “gomitolo” di punti a differenza minima, che abbiamo ipotizzato essere riconducibile alla vascolarizzazione celebrale, Figura 3.15.

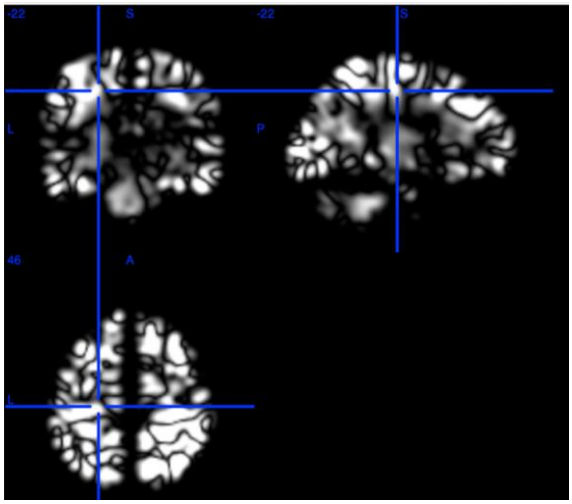

Figura 3.15: *Immagine delta* di materia bianca di una coppia di gemelli.

In Matlab la differenza tra due immagini è calcolata voxel per voxel in termini di differenza di livello di grigio; quindi l’immagine risultante che ha la dimensione dell’immagine originale, contiene in ciascun voxel la differenza tra i due voxel corrispondenti. Quando in un voxel la differenza tra i livelli di grigio è nulla, il voxel dell’immagine risulta rappresentato in nero (per convenzione rappresentativa). Nell’*immagine delta* dei gemelli è ricorrente la visualizzazione di questo “gomitolo” nero sia nelle immagini di materia bianca che nelle immagini di materia grigia. Inoltre, tale “gomitolo” era leggermente più evidente nelle mappe tessutali delle coppie MZ rispetto alle mappe tessutali delle coppie DZ. Questo pattern non emergeva così chiaramente nell’*immagine delta* tra due soggetti scelti in maniera casuale dal campione che, quindi, non costituiscono una coppia di gemelli (Figura 3.16). La mancata corrispondenza in soggetti non gemelli ha fatto presupporre che le linee a bassa intensità evidenziate siano corrispondenti a vasi sanguigni celebrali disposti in modo omogeneo nei gemelli di ciascuna coppia. Questa ipotesi preliminare richiede ulteriori approfondimenti.

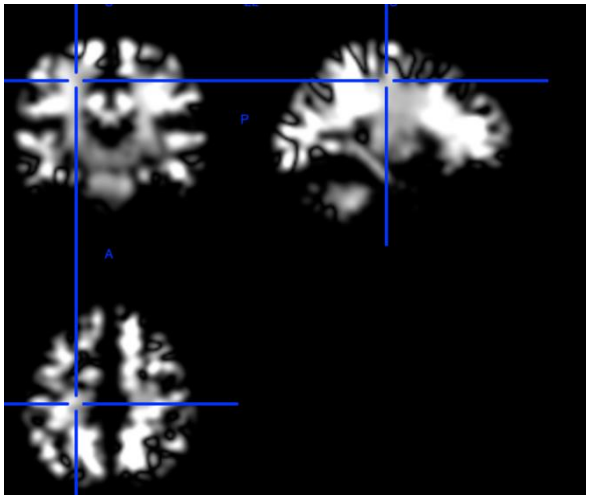

Figura 3.16: *Immagine delta* di materia bianca di due soggetti scelti in maniera casuale dal campione

Nell'immagine rappresentata nella Figura 3.16 sono visualizzate esattamente le stesse "fette" assiali, sagittali e coronali dell'immagine in Figura 3.15. La differenza voxel per voxel calcolata a partire dalle immagini di due soggetti non gemelli risulta essere quasi nulla in un numero ridotto di voxel, quindi emerge una corrispondenza puntuale ridotta. Sebbene siano osservabili alcuni tratti continui a bassa intensità, essi sono di gran lunga meno evidenti rispetto all'*immagine delta* dei soggetti gemelli. Dato che le immagini sono comparabili in quanto co-registrate nello spazio MNI, le differenze punto a punto tra le immagini delta dei gemelli e le immagini delta dei non gemelli sono corrispondenti tra le immagini.

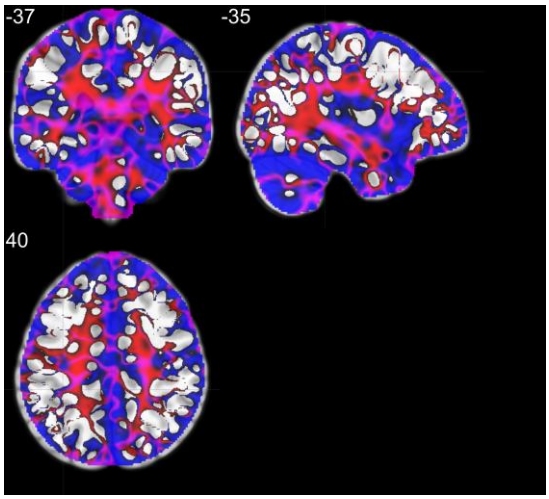

Figura 3.17: Sovrapposizione *immagine delta* di materia bianca con *immagine delta* di materia grigia di una coppia di gemelli

Nella Figura 3.17 sono state sovrapposte le mappe tessutali differenza di materia bianca con quelle di materia grigia. I punti evidenziati in fuxia corrispondono ai punti a bassa intensità.

Queste considerazioni pongono le basi per successive analisi di corrispondenza neuroanatomica dei punti che risultano a bassa intensità nei gemelli, associati a probabili vasi sanguigni celebrali, che potrebbe risultare molto più alta rispetto alla corrispondenza presente nella popolazione generale.

5.3 Analisi di Machine Learning

5.3.1 Applicazione del Machine Learning nel metodo gemellare

La teoria del Machine Learning è stata utilizzata per l'implementazione di un modello di classificazione che a partire dalle mappe di similarità neuroanatomica intra-coppia sia in grado di predire la zigosità delle coppie di gemelli che compongono il campione analizzato. I modelli di classificazione sono stati impiegati per individuare legami ricorrenti tra le variabili esplicative (voxel), dette *features*, che descrivono le immagini appartenenti ad una medesima classe. Si tratta di una classificazione binaria in quanto la variabile da predire può assumere due soli valori: MZ o dizigote DZ. L'algoritmo di classificazione utilizzato per la costruzione del modello è stato il Linear-SVM (descritto di seguito nel paragrafo 5.3.3) ed è stato implementato attraverso il software *NeuroMiner*.

```
*****************************************
****\                                 /****
*****\     ~~~
```

5.3.2 Inizializzazione in NeuroMiner

Analogamente al design di analisi *region-based*, poiché è stato assunto che la discordanza per sesso influenzi la similarità neuroanatomica nella coppia, sono state escluse dall'analisi le *immagini delta* appartenenti alle coppie discordanti per sesso. Inoltre, è stata esclusa dalle analisi di ciascun tessuto cerebrale l'*immagine delta* di una coppia scelta in maniera casuale dal campione al fine di utilizzarla per una validazione esterna (predizione) al termine della creazione del modello. Questo tipo di validazione è volta ad analizzare le effettive capacità di predizione del modello, utilizzandolo per predire la zigosità di un'immagine che l'algoritmo non ha analizzato in precedenza (esterna al campione di training).

Il campione è risultato composto dalle mappe di differenza tessutale intra-coppia (*immagini delta)* di materia bianca e di materia grigia di 55 coppie, 25 MZ e 30 DZ di cui 25 maschi e 30 femmine, di età compresa tra 9 e 30 anni.

Il dataset è costituito da *m* esempi, ovvero le *immagini delta*, che contengono intrinsicamente le informazioni di entrambi i gemelli di ciascuna coppia; *d* attributi esplicativi (o features) per ciascun esempio, ovvero il livello di grigio di ciascun voxel; una variabile target, ovvero la zigosità.

Sono stati necessari i seguenti passaggi per organizzare i dati prima dell'inizializzazione dell'analisi in NeuroMiner:

- Sono state suddivise le *immagini delta* delle coppie di gemelli in quattro cartelle, sulla base del tessuto rappresentato nell'immagine (materia bianca o materia grigia) e sulla base della zigosità, in modo da rendere più semplice la fase di caricamento dei dati nel software.
- Sono state create in Matlab una maschera di materia bianca e una maschera di materia grigia a partire dalle mappe di materia bianca e grigia del campione gemellare di studio. Tali maschere hanno permesso di escludere dall'analisi tutti i voxel che non appartengono alle regioni di interesse (voxel dello sfondo).
- È stata caricata nel workspace di Matlab una matrice che contiene le *variabili confondenti (*o *covariate)*; sono state considerate come variabili confondenti età e sesso. La matrice conteneva tante righe quante erano le *immagini delta* inserite in ingresso al software e tante colonne quante erano le variabili confondenti.

È stata posta particolare attenzione al fatto che le *immagini delta* e le variabili confondenti ad esse associate fossero inserite nei vari step di *NeuroMiner* nello stesso ordine.

A ciascuna classe (MZ, DZ) il software assegna automaticamente un'etichetta seguendo l'ordine con cui le immagini, appartenenti alle rispettive classi, sono inserite in NeuroMiner; durante le analisi delle immagini di materia bianca e di materia grigia, sono state inserite, in entrambi i casi, prima le immagini delle coppie MZ e poi delle coppie DZ quindi l'etichetta {1} indica la classe MZ e l'etichetta {-1} indica la classe DZ.

5.3.3 Linear Support Vector Machines

Uno degli algoritmi di Machine Learning maggiormente utilizzato in psichiatria è il Support Vector Machine (SVM) (Arbabshirani et al. 2017, Kambeitz et al. 2015, Orru et al. 2012), una tecnica multivariata di apprendimento automatico supervisionato che ottiene la massima efficacia nei problemi di classificazione binari ma viene utilizzato anche per problemi di classificazione multi-classe. Nello spazio bidimensionale SVM cerca di individuare un iperpiano che classifichi correttamente i dati in ingresso, per questo motivo l'algoritmo è rinominato Linear Support Vector Machine (Linear-SVM). Quando la superficie di separazione è una superficie complessa si parla di NonLinear-SVM.

L'idea che sta alla base dell'Linear-SVM è che date due classi di esempi multidimensionali linearmente separabili, tra tutti i possibili iperpiani di separazione, SVM determina quello in grado di separare le due classi con il maggior *margine* possibile.

Due insiemi di punti appartenenti a due classi binarie si dicono linearmente separabili se esiste un iperpiano, ovvero una retta per l'esempio bidimensionale, che li separa nello spazio $\mathbb{R}^d$, *spazio delle features*. Come illustrato nella Figura 3.19, esistono infinite possibili superfici di separazione lineare, apparentemente equivalenti tra loro. Tuttavia, se si considera la capacità di generalizzazione appare chiaro che la retta rossa e la retta verde si collocano a ridosso degli esempi delle due classi e quindi hanno una minore capacità di generalizzazione rispetto a una retta che risulta equidistante dai punti più vicini dei due insiemi, come la retta in nero.

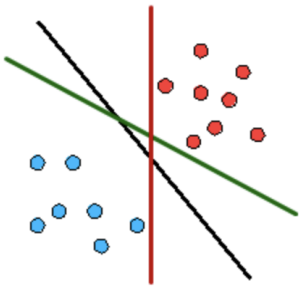

Figura 3.19: Le rette rappresentano possibili superfici di separazione lineare per i punti del diagramma.

Il *margine di separazione* è pari al doppio della minima distanza tra i punti delle rispettive classi e l'iperpiano separatore (δ). I punti delle due classi che si trovano alla minima distanza dall'iperpiano separatore, collocati quindi sugli iperpiani canonici, sono detti *vettori di supporto* e si rivelano più importanti delle altre osservazioni del dataset dal momento che, da soli, determinano la regola di classificazione. La soluzione del problema (equazione dell'iperpiano) può essere espressa come funzione di solo tali punti, indipendentemente dalla dimensionalità dello *spazio delle features* e dal numero di elementi del *training set*.

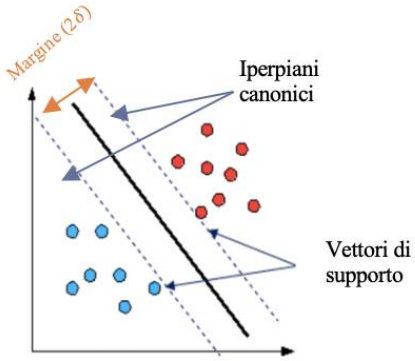

Figura 3.20: Superficie di separazione di massimo margine, iperpiani canonici e vettori di supporto.

Date due classi di esempi linearmente separabili e un *training set* contenente n campioni (x_1,y_1), …, (x_n,y_n), dove $x_i \in \mathbb{R}^d$ e $y_i \in \{+1.-1\}$, esistono diversi iperpiani in grado di eseguire la separazione voluta. Indicando con w i coefficienti angolari dell'iperpiano separatore e con b il suo termine noto, l'equazione dell'iperpiano è

$$w'x = b \tag{3.16}$$

mentre le equazioni degli iperpiani canonici sono

$$w'x - b - 1 = 0 \qquad (3.17)$$

$$w'x - b + 1 = 0 \qquad (3.18)$$

Il margine di separazione è pari a 2δ con

$$\delta = \frac{2}{||w||}$$

dove

$$||w|| = \sqrt{\sum_{j \in \mathrm{D}} w_j^2}$$

Per determinare i coefficienti w e b che identificano l'iperpiano di separazione ottimale si risolve il seguente problema di ottimizzazione quadratico con vincoli lineari

$$\min_{w,b} \quad \frac{1}{2}||w||^2 \qquad (3.19)$$

$$s.a \qquad y_i(w'x_i - b) \geq 1 \qquad i \in N \qquad (3.20)$$

La funzione obiettivo corrisponde alla massimizzazione del margine, espressa come minimizzazione del suo reciproco, mentre i vincoli impongono che ogni punto x_i cada nel semispazio corrispondente alla classe y_i.

Gli *n* punti di un training set risultano in generale non separabili linearmente, come illustrato nella Figura 3.21. In questi casi occorre attenuare i vincoli (3.20) sostituendoli con condizioni più deboli che prevedano la possibilità di errori di misclassificazione, e modificare la funzione obiettivo del problema di ottimizzazione. A questo scopo si definiscono le *variabili di scarto* $d_i, i \in N$, che per gli evenutali punti misclassificati x_i misurano la differenza positiva tra l'ordinata del punto e l'ordinata lungo l'iperpiano canonico che delimita la regione associata alla classe y_i **[18]**, come mostrato geometricamente nella Figura 3.21.

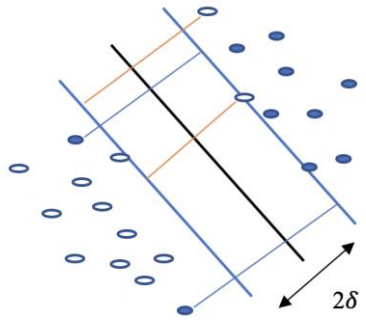

Figura 3.21: Superficie di separazione di massimo margine e iperpiani canonici per un training set non separabile linearmente.

È possibile formulare un secondo modello di ottimizzazione per il caso più generale di dataset non separabili linearmente

$$\min_{w,b,d} \frac{1}{2}||w||^2 + \lambda \sum_{i=1}^{m} d_i \qquad (3.21)$$

$$s.a \quad y_i(wx_i - b) \geq 1 - d_i \qquad i \in N \qquad (3.22)$$

$$d_i \geq 0 \qquad i \in N \qquad (3.23)$$

La funzione obiettivo contiene due termini, che esprimono rispettivamente il reciproco del margine di separazione e l'errore empirico. Il parametro λ svolge il ruolo di regolare il peso relativo della capacità di generalizzazione, rappresentata dal reciproco del margine, e dell'accuratezza nei confronti del training set, espressa dalla somma degli scarti. Il problema quadratico (3.21), può essere risolto passando ad una formulazione Lagrangiana e successivamente ad una formulazione duale. Oltre ad altri vantaggi, questo permette di identificare i *vettori di supporto* che corrispondono ai moltiplicatori di Lagrange diversi da zero **[18]**.

Nella formulazione Lagrangiana si indicano con $\alpha_i \geq 0$ i moltiplicatori di Lagrange relativi ai vincoli (3.22) e con $\mu_i \geq 0$ i moltiplicatori dei vincoli (3.23). È possibile formulare la funzione lagrangiana del problema (3.21) come

$$L(w, b, d, \alpha\ \mu) = \frac{1}{2}\left|\left|w\right|\right|^2 + \lambda \sum_{i=1}^{n} d_i - \sum_{i=1}^{n} \alpha_i [y_i(w' x_i - b) - 1 + d_i] - \sum_{i=1}^{n} \mu_i d_i \tag{3.24}$$

Per determinare la soluzione ottimale si annullano le derivate rispetto alle variabili w, b, d del problema (3.21), ottenendo le uguaglianze

$$\frac{\partial L(w,b,d,\alpha\ \mu)}{\partial w} = w - \sum_{i=1}^{m} \alpha_i y_i\, x_i = 0 \tag{3.25}$$

$$\frac{\partial L(w,b,d,\alpha\ \mu)}{\partial d} = \lambda - \alpha_i - \mu_i = 0 \tag{3.26}$$

$$\frac{\partial L(w,b,d,\alpha\ \mu)}{\partial b} = \sum_{i=1}^{m} \alpha_i\, y_i = 0 \tag{3.27}$$

che danno luogo alle condizioni

$$w = \sum_{i=1}^{m} \alpha_i\, y_i x_i \tag{3.28}$$

$$\lambda = \alpha_i + \mu_i \tag{3.29}$$

$$\sum_{i=1}^{n} \alpha_i\, y_i = 0 \tag{3.30}$$

Sostituendo queste ultime nell'espressione della funzione lagrangiana (3.24) si ottiene la funzione obiettivo del problema duale cioè quella formulazione in cui i parametri w e b sono espressi in funzione dei moltiplicatori α_i:

$$L(w, b, d, \alpha\ \mu) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} y_i\, y_k \alpha_i \alpha_k x_i{}' x_k \tag{3.31}$$

Con la condizione aggiuntiva $\alpha_i \leq \lambda, i \in N$. Le condizioni di complementarità di Karush-Kuhn-Tucker (KKT) applicate alla coppia di problemi primale e duale si traducono nelle disuguaglianze

$$\alpha_i[y_i(w' x_i - b) - 1 + d_i] = 0 \qquad i \in N \quad (3.32)$$

$$\mu_i(\alpha_i - \lambda) = 0 \qquad i \in N \quad (3.33)$$

In particolare, la relazione (3.33) permette di identificare i vettori di supporto, ovvero gli esempi più rappresentativi delle regole di classificazione estratte dal dataset. Le osservazioni x_i corrispondenti a moltiplicatori di Lagrange per cui vale la relazione

$0 < \alpha < \lambda$ si trovano infatti a una distanza di $1/||w||$ dall'iperpiano separatore e quindi si collocano su uno degli iperpiani canonici **[18]**.

5.3.4 Creazione dei modelli di classificazione delle immagini di materia bianca e grigia

Le analisi condotte sulle immagini di materia bianca e di materia grigia sono state implementate separatamente.

Il software NeuroMiner, per ogni immagine in ingresso, definisce in maniera automatica un numero di features pari al numero di voxel ($121 \times 121 \times 145$). Tuttavia, dando in ingresso una maschera, il software considera solo i voxel che ricadono nella maschera. Nell'analisi basata sulle immagini di materia grigia NeuroMiner ha considerato *40.2208 features* per ogni immagine *delta*, mentre nelle immagini di materia bianca ha considerato *179989 features*. Le *features* rappresentano le variabili esplicative di ciascun immagine sulla base delle quali l'immagine viene posizionata in un punto nello spazio $\mathbb{R}^d$, dove *d* è il numero di *features*. NeuroMiner, essendo un software di Machine Learning, è in grado di riconoscere dei pattern ricorrenti tra le *features* che descrivono ciascuna immagine e la classe target, ovvero la zigosità (nota a priori per queste osservazioni). Successivamente utilizzerà tali pattern per associare ad una nuova osservazione la classe di appartenenza (non nota a priori).

Generalmente, la scelta delle procedure di pre-processing da implementare nel software dipende dalla richiesta, dai dati disponibili e dal modello da applicare. Alcuni step sono comuni a tutte le tipologie di analisi statistiche, come ad esempio, la rimozione dell'effetto delle covariate (ad esempio, età e sesso). Inoltre, nell'ambito del Machine Learning con ampi set di *features* è pratica comune applicare procedure per la riduzione della dimensionalità. Tali procedure permettono di trasformare i dati descritti in uno spazio delle *features* con

un'elevata dimensionalità ad uno spazio delle *features* di dimensione inferiore (cioè con un numero di variabili esplicative ridotto) senza perdere le informazioni contenute nei dati (Van Der Maaten et al. 2009).

La *pipeline* del pre-processing è stata la seguente:

1. Scalatura dell'immagine per riportarla in un range da 0 a 1 (livelli di grigio normalizzati).
2. Rimozione degli effetti delle covariate (in questa fase NeuroMiner implementa la partial correlation).
3. Riduzione della dimensionalità applicando una PCA (Principal Component Analysis) preservando l'80% della varianza.
4. Nuova scalatura dell'immagine per riportarla in un range da 0 a 1 (livelli di grigio normalizzati).

È stato utilizzato l'algoritmo LISVM (LinearSupportVectors-Machines) il quale calcola l'equazione dell'iperpiano ottimo di separazione tra le due classi di zigosità. Come procedura di cross-validazione è stata utilizzata la *double nested CrossValidation (CV).* Essa prevede un ciclo di CV *interno (CV1)* all'interno di un ciclo di CV *esterno (CV2)*. Quest'ultimo è usato per accedere alla generalizzazione del modello (Filzmoser et al. 2009, Stone 1974). Entrambi i cicli *CV1* e *CV2* sono stati suddivisi in maniera randomica in 5 *folds* e rimescolati rispettivamente con 5 permutazioni.

In particolare, l'algoritmo suddivide il set di esempi che costituiscono il campione di partenza in 5 sottogruppi (*folds),* di cui uno viene tenuto fuori come *test set del CV2,* mentre l'unione degli altri 4 costituisce *il training set del CV2*. Alla prima iterazione l'insieme di training *del CV2* viene ulteriormente suddiviso in 5 sottogruppi di cui uno costituisce il *test set del CV1*, mentre l'unione degli altri 4 costituisce il *training set del CV1*. La procedura continua iterazione dopo iterazione fino a quando tutti e 5 i sottogruppi del *CV1* sono stati considerati almeno una volta come insieme di test. L'accuratezza è calcolata come media delle accuratezze ottenute in ogni iterazione. Al termine del *CV1*, l'algoritmo fa il training del modello sull'insieme dei soggetti che costituiscono il *training set del CV2* e poi applica la fase di test sul sottogruppo tenuto inizialmente fuori come *test set del CV2.* A questo punto passa alla iterazione successiva per il ciclo *esterno (CV2),* cioè viene definito il successivo *test set del CV2* e l'unione di tutti gli esempi rimanenti viene considerato come *training set*

del CV2. Nuovamente, l'insieme di training esterno viene splittato in 5 sottogruppi, e reinizia il ciclo interno fino a quando tutti i sottogruppi interni non sono stati utilizzati almeno una volta come insieme di test. Si passa alla successiva iterazione esterna fino a quando tutti i 5 sottogruppi esterni vengono considerati almeno una volta come insieme di test esterno; a quel punto, la procedura di cross-validazione si ferma.

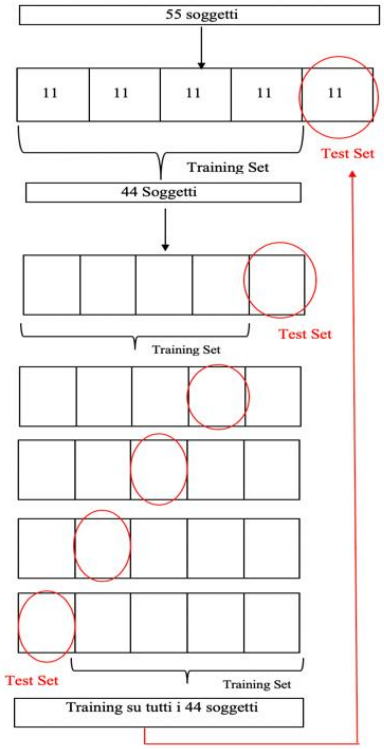

Figura 3.22: Rappresentazione schematica della *double nested CrossValidation*.

5.3.4.1 Calcolo dei pesi delle features nella procedura di classificazione

Il peso rappresenta, in termini assoluti, l'importanza di una specifica *feature* (voxel) nella procedura di classificazione, ovvero quanto un determinato voxel è determinante nella differenziazione dei due gruppi di zigosità. Dopo aver ottenuto il modello di classificazione ed averne estratto le misure di performance, è stato possibile, mediante il software NeuroMiner, implementare il calcolo del peso di ciascuna features.

L'output della procedura del calcolo dei pesi è un'immagine del cervello in cui il valore di ciascun voxel è stato calcolato come prodotto tra il peso del voxel e l'etichetta. Quindi, il segno in ciascun voxel può essere positivo o negativo a seconda che esso sia associato alla classe {1} o alla classe {-1}. Più nel dettaglio, il peso, in valore assoluto, porta

l'informazione sull'importanza del voxel nel modello di classificazione, mentre il segno porta l'informazione sulla classe in cui il valore numerico del voxel è maggiore. Ad un valore positivo del voxel corrisponde la situazione in cui il peso è stato moltiplicato con la classe etichettata come *classe {1}*, cioè MZ, ciò equivale a dire che il voxel in questa classe assume un valore più alto del valore dello stesso voxel nella *classe {-1}*, cioè DZ. Dato che il valore del voxel nelle immagini date in ingresso rappresenta il *delta* puntuale tra le coppie (intesa come differenza punto-punto), nei voxel associati a pesi positivi in uscita da NeuroMiner è presente una differenza maggiore nei MZ rispetto ai DZ o, alternativamente, la similarità nei DZ è maggiore che nei MZ. Viceversa, quando il segno è negativo, il peso è stato moltiplicato per la *classe {-1}*, cioè DZ, e quindi il valore del voxel dell'*immagine delta* è maggiore nei DZ e quindi la similarità di quel voxel sarà maggiore nelle coppie MZ.

5.3.4.2 Valutazione del modello di classificazione: Grafici ROC e Tabelle di contingenza

I grafici delle curve ROC (Receiving Operating Characteristics curve) consentono di valutare in modo visivo l'accuratezza di un classificatore ma anche di confrontare tra loro diversi modelli di classificazione. Un grafico ROC è un diagramma bidimensionale che rappresenta sull'asse delle ascisse la percentuale di false positive (FP), cioè il numero di predizioni errate per le osservazioni negative e sull'asse delle ordinate la percentuale di vere positive (VP), cioè il numero di predizioni esatte per le osservazioni positive **[18]**.

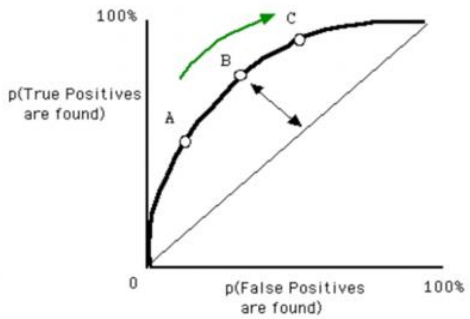

Figura 3.23: Diagramma della curva ROC per un classificatore.

Il punto di coordinate (0,1) rappresenta il classificatore ideale perché non commette nessun errore di predizione: la percentuale di FP = 0 e quella di VP = 1. Maggiore è l'area sotto la curva, migliore è la performance dell'algoritmo di classificazione.

La matrice di confusione, invece, è una tabella in cui ogni colonna rappresenta la classe predetta, mentre ogni riga della matrice rappresenta la classe reale, Figura 3.24.

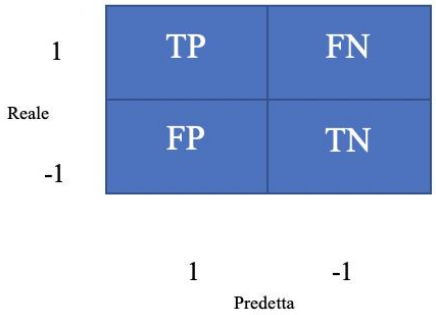

Figura 3.24: Tabella di contingenza.

Si definisce *sensibilità* il rapporto

$$\text{Sensibilità}=\frac{TP}{TP+FN}$$

ovvero la percentuale di immagini appartenenti alla classe {1}, predette come appartenenti alla classe {1}. Si definisce specificità il rapporto:

$$\text{Specificità}=\frac{TN}{TN+FP}$$

ovvero la percentuale di immagini appartenenti alla classe {-1}, predette come appartenenti alla classe {-1}.

Queste misure sono state calcolate per l'algoritmo di classificazione selezionato e hanno permesso di valutarne la sua capacità di discriminazione delle due classi.

4.3.5.3 Validazione esterna: predizione della classe target per una nuova immagine

In NeuroMiner è possibile caricare una o più *immagini delta* di cui il software non conosce la classe di appartenenza, sulle quali applica il modello di classificazione ottenuto attraverso gli step precedenti ed effettua la predizione di appartenenza a una determinata classe.

L’interfaccia grafica della funzione che permette il caricamento di nuovi dati è rappresentata nella Figura 3.25.

In questa specifica fase il software Neurominer è in grado di fornire l’informazione (con un’accuratezza che dipende dall’accuratezza del modello di classificazione) circa la zigosità della coppia di gemelli a partire dall’*immagine differenza tra le mappe tessutali di materia bianca* o *grigia* dei due soggetti che costituiscono la coppia.

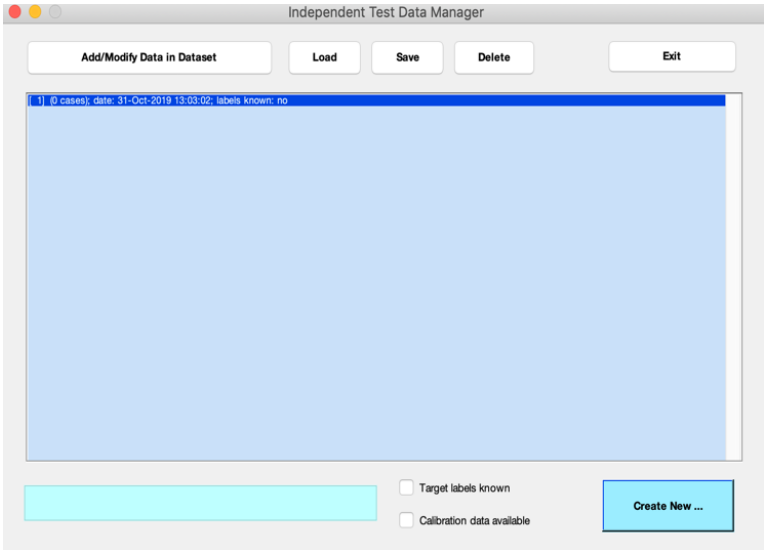

Figura 3.25: Interfaccia grafica *Indipendent Test Data Manager* di NeuroMiner.

Capitolo 4

Risultati e discussione

1. Risultati dell'analisi di correlazione

Per ogni parametro analizzato sono stati calcolati i coefficienti di correlazione parziale nei due gruppi di zigosità, in ogni regione dell'atlante di Desikan e Killiany. Le misure morfologiche risultate significativamente correlate in uno o entrambi i gruppi di gemelli sono riportate di seguito (*p-value<0.05)*.

Per il parametro *area superficiale,* sono emerse 30 regioni celebrali in cui le coppie MZ sono significativamente correlate (Tabella 4.1) e 19 regioni celebrali in cui le coppie DZ sono significativamente correlate (Tabella 4.2). Questi risultati sono in linea con ciò che è noto dalla letteratura ovvero che, globalmente, i gemelli MZ sono più simili tra di loro rispetto ai gemelli DZ. Inoltre, sia nei MZ che nei DZ sono emerse più regioni significativamente correlate nell'emisfero destro rispetto all'emisfero sinistro. Le regioni in cui è emersa una correlazione significativa nelle coppie MZ sono riportate nella Figura 4.1. Nella stessa figura sono riportati i coefficienti di correlazione delle coppie DZ nelle sole regioni (pari a 11) in cui anche questi sono risultati significativi (*p-value<0.05*). Dal confronto è possibile osservare che nelle coppie MZ il coefficiente di correlazione è maggiore rispetto al coefficiente di correlazione nelle coppie DZ in quasi tutte le regioni eccetto che nel *cuneo dell'emisfero destro* e nella porzione rostrale della *circonvoluzione frontale media dell'emisfero sinistro*. Quindi, in queste ultime regioni (oltre che nelle regioni con coefficienti di correlazione non significativi nei gemelli MZ), ci si aspetta che non vi sia un contributo dei fattori genetici sul valore assunto dal parametro.

ROI emisfero sinistro in cui r_{MZ} è significativo	rMZ	P-value	ROI emisfero destro in cui r_{MZ} è significativo	rMZ	P-value
lh_cuneus_area	0,525	0,010	rh_bankssts_area	0,544	0,007
lh_lateraloccipital_area	0,462	0,026	rh_cuneus_area	0,482	0,020
lh_lateralorbitofrontal_area	0,457	0,028	rh_inferiorparietal_area	0,489	0,018
lh_lingual_area	0,516	0,012	rh_inferiortemporal_area	0,573	0,004
lh_middletemporal_area	0,667	0,001	rh_lateralorbitofrontal_area	0,465	0,025
lh_parahippocampal_area	0,441	0,035	rh_lingual_area	0,567	0,005
lh_paracentral_area	0,663	0,001	rh_middletemporal_area	0,663	0,001
lh_parsopercularis_area	0,571	0,004	rh_paracentral_area	0,490	0,018
lh_pericalcarine_area	0,615	0,002	rh_parstriangularis_area	0,445	0,033
lh_rostralmiddlefrontal_area	0,519	0,011	rh_pericalcarine_area	0,719	0,000
lh_superiorparietal_area	0,471	0,023	rh_precuneus_area	0,647	0,001
lh_superiortemporal_area	0,415	0,049	rh_rostralanteriorcingulate_area	0,421	0,045
lh_transversetemporal_area	0,524	0,010	rh_superiorfrontal_area	0,676	0,000
lh_insula_area	0,611	0,002	rh_superiorparietal_area	0,563	0,005
			rh_transversetemporal_area	0,523	0,010
			rh_insula_area	0,425	0,043

Tabella 4.1: Regioni corticali dei due emisferi in cui il coefficiente di correlazione r_{MZ} è statisticamente significativo (*p-value<0.05)* per il parametro *area superficiale.*

ROI emisfero sinistro in cui r_{DZ} è significativo	rDZ	P-value	ROI emisfero destro in cui r_{DZ} è significativo	rDZ	P-value
lh_inferiortemporal_area	0,402	0,031	rh_cuneus_area	0,588	0,001
lh_isthmuscingulate_area	0,449	0,015	rh_inferiortemporal_area	0,395	0,034
lh_pericalcarine_area	0,437	0,018	rh_isthmuscingulate_area	0,490	0,007
lh_posteriorcingulate_area	0,508	0,005	rh_lingual_area	0,419	0,024
lh_rostralmiddlefrontal_area	0,570	0,001	rh_medialorbitofrontal_area	0,441	0,017
lh_superiorparietal_area	0,377	0,044	rh_middletemporal_area	0,515	0,515
			rh_parsopercularis_area	0,514	0,004
			rh_pericalcarine_area	0,692	0,000
			rh_precuneus_area	0,382	0,041
			rh_superiorparietal_area	0,500	0,006
			rh_supramarginal_area	0,410	0,027
			rh_frontalpole_area	0,437	0,018
			rh_insula_area	0,394	0,035

Tabella 4.2: Regioni corticali dei due emisferi in cui il coefficiente di correlazione r_{DZ} è statisticamente significativo (*p-value<0.05)* per il parametro *area superficiale.*

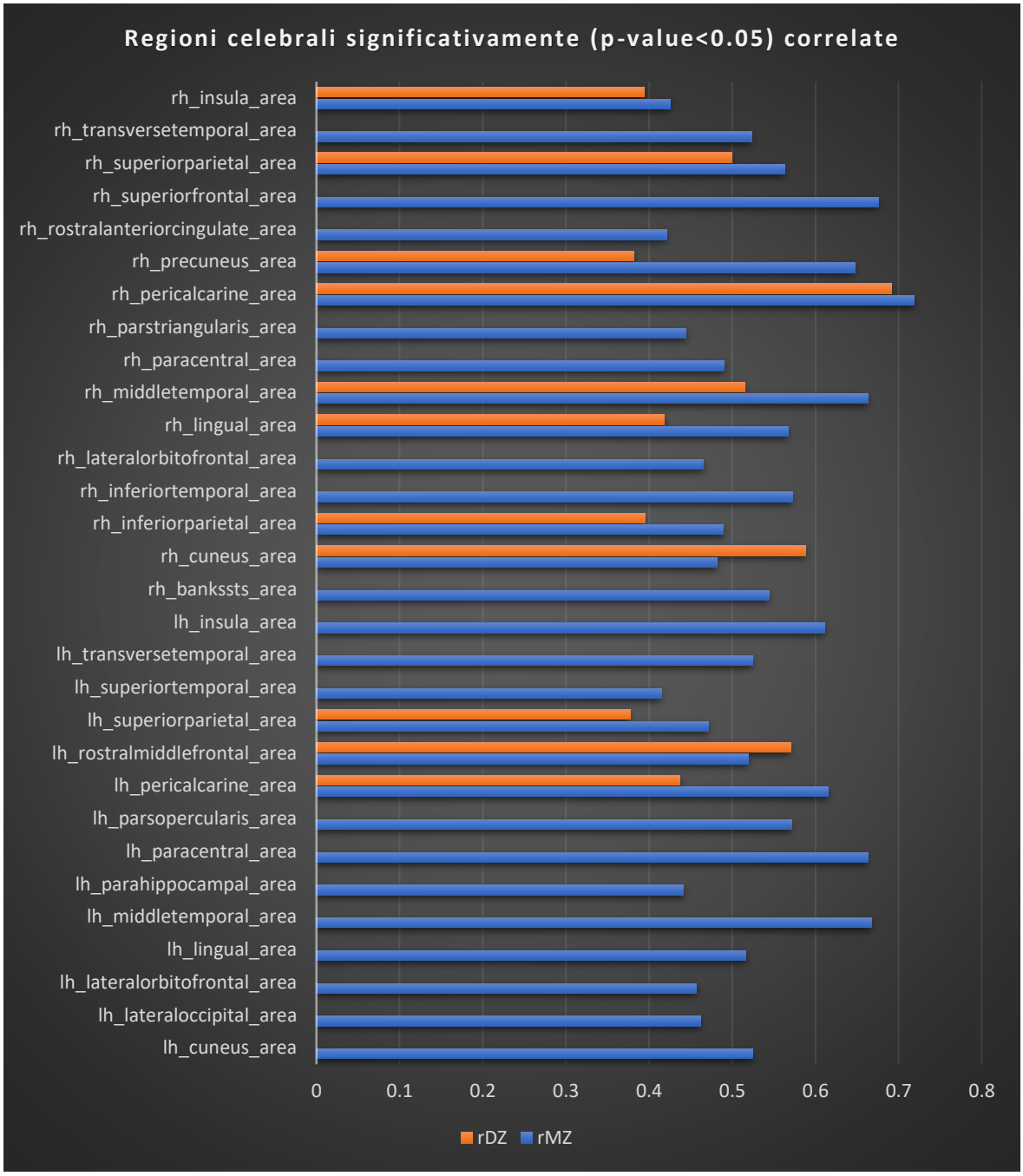

Figura 4.1: Regioni corticali in cui è emersa una correlazione significativa (*p-value<0.05)* del valore assunto dal parametro *area* nelle coppie MZ e confronto con il coefficiente di correlazione nelle coppie DZ quando anche questo è significativo (*p-value<0.05*) nelle stesse regioni.

Per lo *spessore corticale* sono emerse 19 regioni celebrali in cui le coppie MZ sono significativamente correlate e 13 regioni in cui le coppie DZ sono significativamente correlate (*p-value<0.05)*. Inoltre, nell'emisfero destro è emersa una correlazione significativa in un numero maggiore di regioni celebrali rispetto all'emisfero sinistro in

entrambi i gruppi di zigosità. Le regioni in cui è emersa una correlazione significativa nelle coppie MZ sono riportate nella Figura 4.2. Nella stessa figura sono riportati i coefficienti di correlazione delle coppie DZ quando questi sono significativi nelle stesse regioni dei MZ. Dall'analisi dei valori di correlazione di *spessore corticale* nei due gruppi, è emerso che il coefficiente di correlazione nei gemelli MZ è sempre maggiore del coefficiente di correlazione nei gemelli DZ. In altre parole, il grado di associazione tra i valori di spessore corticale nel gemello1 e nel gemello2 delle coppie MZ è sempre più elevato del grado di associazione presente nei due gemelli che costituiscono le coppie DZ.

ROI emisfero sinistro in cui r_{MZ} è significativo	**rMZ**	**P-value**	**ROI emisfero destro in cui r_{MZ} è significativo**	**rMZ**	**P-value**
lh_inferiortemporal_thickness	0,651	0,001	rh_cuneus_thickness	0,485	0,019
lh_isthmuscingulate_thickness	0,428	0,042	rh_fusiform_thickness	0,480	0,021
lh_lateraloccipital _thickness	0,491	0,017	rh_inferiortemporal_thickness	0,525	0,010
lh_middletemporal_thickness	0,433	0,039	rh_postcentral_thickness	0,488	0,018
lh_postcentral_thickness	0,745	0,000	rh_precentral_thickness	0,606	0,002
lh_posteriorcingulate_thickness	0,613	0,002	rh_rostralanteriorcingulate_thickness	0,655	0,001
lh_precentral_thickness	0,567	0,005	rh_superiorfrontal_thickness	0,644	0,001
lh_superiorparietal_thickness	0,558	0,006	rh_superiorparietal_thickness	0,566	0,005
lh_supramarginal_thickness	0,535	0,008	rh_superiortemporal_thickness	0,649	0,001
			rh_supramarginal_thickness	0,490	0,018

Tabella 4.3: Regioni corticali dei due emisferi in cui il coefficiente di correlazione r_{MZ} è statisticamente significativo (*p-value<0.05)* per il parametro *spessore corticale.*

ROI emisfero sinistro in cui r_{DZ} è significativo	rDZ	P-value	ROI emisfero destro in cui r_{DZ} è significativo	rDZ	P-value
lh_caudalanteriorcingulate_thickness	0,374	0,045	rh_isthmuscingulate_thickness	0,523	0,004
lh_fusiform_thickness	0,472	0,010	rh_lateraloccipital_thickness	0,391	0,036
lh_postcentral_thickness	0,590	0,001	rh_lateralorbitofrontal_thickness	0,408	0,028
lh_posteriorcingulate_thickness	0,553	0,002	rh_parahippocampal_thickness	0,431	0,020
lh_precentral_thickness	0,472	0,010	rh_postcentral_thickness	0,438	0,017
lh_insula_thickness	0,646	0,000	rh_superiortemporal_thickness	0,436	0,018
			rh_insula_thickness	0,737	0,000

Tabella 4.4: Regioni corticali dei due emisferi in cui il coefficiente di correlazione r_{DZ} è statisticamente significativo (*p-value<0.05)* per il parametro *spessore corticale.*

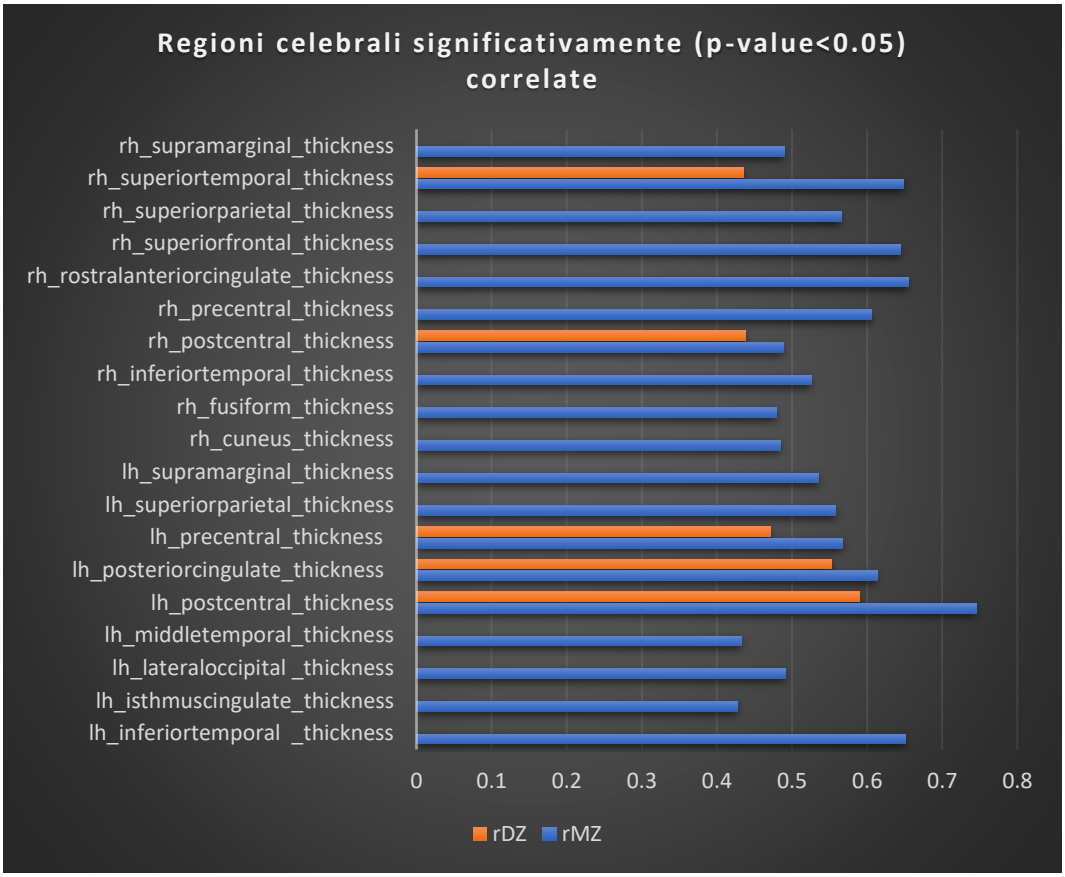

Figura 4.2: Regioni corticali in cui è emersa una correlazione significativa (*p-value<0.05)* del valore assunto dal parametro *spessore corticale* nelle coppie MZ e confronto con il coefficiente di correlazione nelle coppie DZ quando anche questo è significativo (*p-value<0.05*) nelle stesse regioni.

Per il parametro *curvatura media* sono emerse 9 regioni celebrali in cui le coppie MZ sono significativamente correlate e 15 regioni celebrali in cui le coppie DZ sono significativamente correlate (*p-value<0.05)*. Secondo le ipotesi del metodo gemellare, il fatto che vi sia un maggior numero di regioni con correlazione significativa nei gemelli DZ rispetto ai gemelli MZ suggerisce un minor effetto dei fattori genetici sui valori di curvatura media, rispetto ai fattori ambientali. Le regioni in cui è emersa una correlazione significativa nelle coppie MZ sono riportate nella Figura 4.3. Nella stessa figura sono riportati i valori dei coefficienti di correlazione delle coppie DZ quando essi risultano statisticamente significativi nelle regioni in cui sono significativi i coefficienti per le coppie MZ. Nella *circonvoluzione temporale media dell'emisfero sinistro* emerge un coefficiente di correlazione maggiore nelle coppie DZ rispetto alle coppie MZ. I valori di *curvatura media* in questa regione, oltre che nelle regioni con correlazione non significativa nelle coppie MZ, non sembrano fortemente influenzati dai fattori genetici.

ROI emisfero sinistro in cui r_{MZ} è significativo	**rMZ**	**P-value**	**ROI emisfero destro in cui r_{MZ} è significativo**	**rMZ**	**P-value**
lh_middletemporal_meancurvature	0,420	0,046	rh_cuneus_meancurvature	0,572	0,004
lh_superiorparietal_meancurvature	0,579	0,004	rh_middletemporal_meancurvature	0,416	0,048
lh_frontalpole_meancurvature	0,443	0,034	rh_parsopercularis_meancurvature	0,432	0,039
			rh_parstriangularis_meancurvature	0,452	0,030
			rh_supramarginal_meancurvature	0,562	0,005

Tabella 4.5: Regioni corticali dei due emisferi in cui il coefficiente di correlazione r_{MZ} è statisticamente significativo (*p-value<0.05)* per il parametro *curvatura media.*

ROI emisfero sinistro in cui r_{DZ} è significativo	rDZ	P-value	ROI emisfero destro in cui r_{DZ} è significativo	rDZ	P-value
lh_caudalanteriorcingulate_meancurvature	0,441	0,017	rh_caudalmiddlefrontal_meancurvature	0,467	0,011
lh_caudalmiddlefrontal_meancurvature	0,475	0,009	rh_lateraloccipital_meancurvature	0,428	0,020
lh_lateraloccipital_meancurvature	0,389	0,037	rh_pericalcarine_meancurvature	0,373	0,046
lh_middletemporal_meancurvature	0,441	0,017	rh_precuneus_meancurvature	0,373	0,046
lh_paracentral_meancurvature	0,454	0,013	rh_superiorfrontal_meancurvature	0,409	0,028
lh_rostralanteriorcingulate_meancurvature	0,420	0,023			
lh_superiorfrontal_meancurvature	0,577	0,001			
lh_superiorparietal_meancurvature	0,461	0,012			
lh_transversetemporal_meancurvature	0,377	0,044			
lh_insula_meancurvature	0,426	0,021			

Tabella 4.6: Regioni corticali dei due emisferi in cui il coefficiente di correlazione r_{DZ} è statisticamente significativo (*p-value<0.05)* per il parametro *curvatura media.*

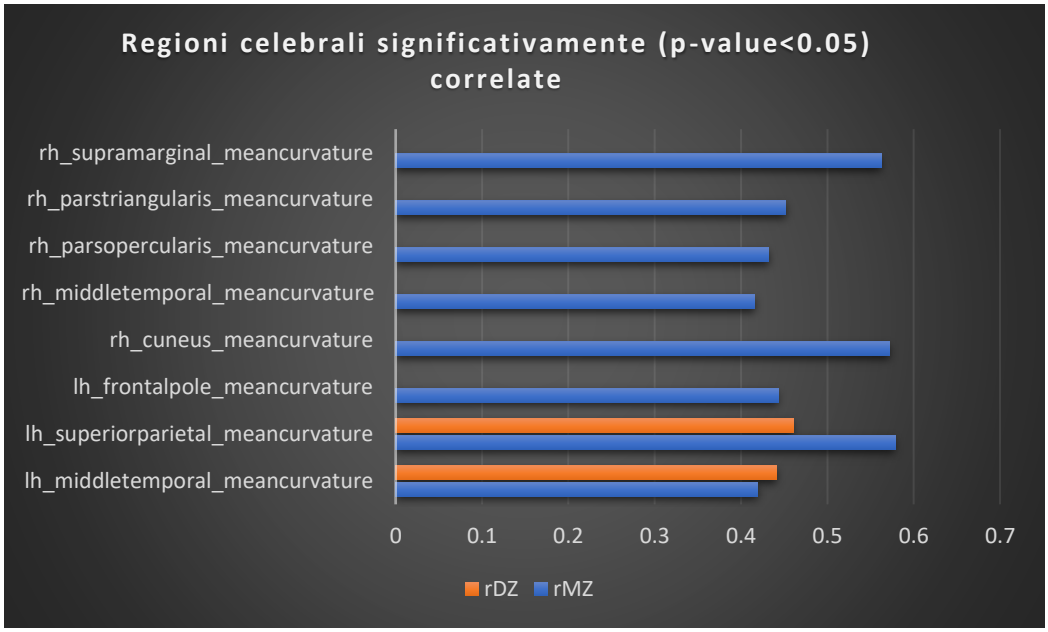

Figura 4.3: Regioni corticali in cui è emersa una correlazione significativa (*p-value<0.05)* del valore assunto dal parametro *curvatura media* nelle coppie MZ e confronto con il coefficiente di correlazione nelle coppie DZ quando anche questo è significativo (*p-value<0.05*) nelle stesse regioni.

Per il parametro *volume* sono emerse 38 regioni celebrali in cui le coppie MZ sono significativamente correlate e 11 regioni celebrali in cui le coppie DZ sono significativamente correlate (*p-value<0.05)*. Inoltre, sia nei MZ che nei DZ sono emerse più regioni significativamente correlate nell'emisfero destro rispetto all'emisfero sinistro. Le regioni in cui è emersa una correlazione significativa nelle coppie MZ sono riportate nella Figura 4.4. Nella stessa figura sono riportati i coefficienti di correlazione delle coppie DZ quando anche questi sono significativi. È emerso che il parametro *volume* è associato al numero massimo di regioni significativamente correlate nei gemelli MZ, rispetto al numero di regioni che risultano significativamente correlate per gli altri parametri. Quindi, il *volume corticale*, rispetto all'area superficiale, allo spessore e alla curvatura media corticali, potrebbe essere il parametro più influenzato da fattori genetici. Inoltre, nelle regioni in cui anche i coefficienti di correlazione relativi alle coppie DZ sono significativi, tali valori di correlazione sono minori dei coefficienti di correlazione relativi alle coppie MZ.

ROI emisfero sinistro in cui r_{MZ} è significativo	**rMZ**	**P-value**	**ROI emisfero destro in cui r_{MZ} è significativo**	**rMZ**	**P-value**
lh_cuneus_volume	0,654	0,001	rh_bankssts_volume	0,478	0,021
lh_entorhinal_volume	0,579	0,004	rh_cuneus_volume	0,522	0,011
lh_fusiform_volume	0,594	0,003	rh_inferiortemporal_volume	0,513	0,012
lh_lateraloccipital_volume	0,523	0,010	rh_isthmuscingulate_volume	0,501	0,015
lh_lingual_volume	0,654	0,001	rh_lingual_volume	0,497	0,016
lh_middletemporal_volume	0,678	0,000	rh_medialorbitofrontal_volume	0,586	0,003
lh_parsopercularis_volume	0,487	0,018	rh_middletemporal_volume	0,578	0,004
lh_pericalcarine_volume	0,561	0,005	rh_paracentral_volume	0,512	0,013
lh_postcentral_volume	0,625	0,001	rh_parsopercularis_volume	0,466	0,025
lh_precentral_volume	0,551	0,006	rh_parsorbitalis_volume	0,440	0,036
lh_rostralmiddlefrontal_volume	0,447	0,033	rh_parstriangularis_volume	0,658	0,001
lh_superiorparietal_volume	0,574	0,004	rh_pericalcarine_volume	0,479	0,021

lh_temporalpole_volume	0,448	0,032	rh_precentral_volume	0,428	0,042
lh_transversetemporal_volume	0,562	0,005	rh_rostralanteriorcingulate_volume	0,456	0,029
lh_insula_volume	0,681	0,000	rh_rostralmiddlefrontal_volume	0,635	0,001
			rh_superiorfrontal_volume	0,655	0,001
			rh_superiorparietal_volume	0,752	0,000
			rh_superiortemporal_volume	0,482	0,020
			rh_supramarginal_volume	0,477	0,021
			rh_frontalpole_volume	0,572	0,004
			rh_temporalpole_volume	0,415	0,049
			rh_transversetemporal_volume	0,510	0,013
			rh_insula_volume	0,577	0,004

Tabella 4.7: Regioni corticali dei due emisferi in cui il coefficiente di correlazione r_{MZ} è statisticamente significativo (*p-value<0.05)* per il parametro *volume.*

ROI emisfero sinistro in cui r_{DZ} è significativo	**rDZ**	**P-value**	**ROI emisfero destro in cui r_{DZ} è significativo**	**rDZ**	**P-value**
lh_paracentral_volume	0,428	0,021	rh_fusiform_volume	0,395	0,034
lh_pericalcarine_volume	0,399	0,032	rh_middletemporal_volume	0,434	0,019
lh_precuneus_volume	0,392	0,036	rh_precentral_volume	0,397	0,033
lh_supramarginal_volume	0,461	0,012	rh_supramarginal_volume	0,443	0,016
lh_frontalpole_volume	0,387	0,038	rh_frontalpole_volume	0,447	0,015
			rh_insula_volume	0,455	0,013

Tabella 4.8: Regioni corticali dei due emisferi in cui il coefficiente di correlazione r_{DZ} è statisticamente significativo (*p-value<0.05)* per il parametro *volume.*

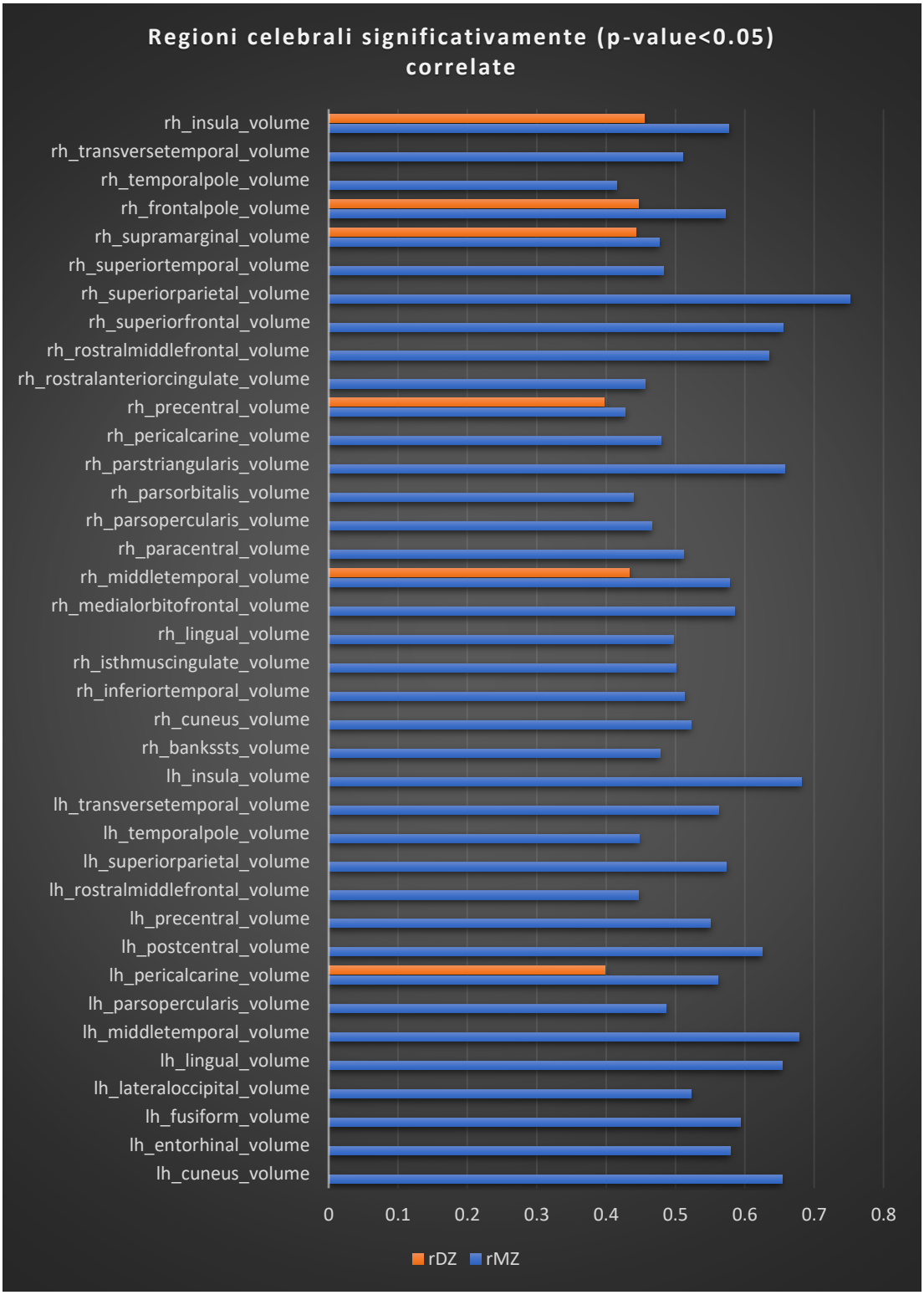

Figura 4.4: Regioni corticali in cui è emersa una correlazione significativa (*p-value<0.05)* del valore assunto dal parametro *volume* nelle coppie MZ e confronto con il coefficiente di correlazione nelle coppie DZ quando anche questo è significativo (*p-value<0.05*) nelle stesse regioni.

Tra le misure di *volume* globali e di regioni profonde, sono emerse 37 regioni sottocorticali aventi coefficienti di correlazione significativi nelle coppie MZ, 34 regioni sottocorticali in cui c'è correlazione significativa nelle coppie DZ. Nella Figura 4.5 sono riportate le regioni sottocorticali in cui è emersa significatività statistica dei coefficienti di correlazione delle coppie MZ. Inoltre, tali valori sono messi a confronto con i coefficienti di correlazione delle coppie DZ nelle regioni in cui anche questi risultano significativi. Tali risultati suggeriscono, ad esempio, un effetto genetico sui valori di *volume intracranico totale* , sui valori di *volume della corteccia cerebellare di entrambi gli emisferi*, sui valori di *volume di materia bianca della corteccia cerebellare di entrambi gli emisferi*, sul *volume del tronco encefalico*, sul v*olume del quarto ventricolo* oltre che nelle regioni più profonde dell'emisfero sinistro quali il *Putamen*, il *Caudato* e il *Pallidum* in quanto in tali regioni emerge una correlazione maggiore nei MZ rispetto ai DZ, mentre nelle regioni (pari a 4) in cui è emersa una maggiore correlazione nei DZ, il contributo genetico alla variabilità inter-individuale del parametro si può ritenere assente.

ROI	**rMZ**	**P-value**
Left-Cerebellum-White-Matter	0,702	0,000
Left-Cerebellum-Cortex	0,869	0,000
Left-Thalamus-Proper	0,539	0,008
Left-Caudate	0,699	0,000
Left-Putamen	0,600	0,002
Left-Pallidum	0,660	0,001
4th-Ventricle	0,811	0,000
Brain-Stem	0,898	0,000
Left-Hippocampus	0,598	0,003
Left-Amygdala	0,440	0,035
Left-VentralDC	0,763	0,000
Right-Cerebellum-White-Matter	0,802	0,000
Right-Cerebellum-Cortex	0,893	0,000
Right-Thalamus-Proper	0,737	0,000
Right-Caudate	0,661	0,001
Right-Putamen	0,815	0,000
Right-Pallidum	0,730	0,000

Right-Hippocampus	0,832	0,000
Right-Amygdala	0,474	0,022
Right-Accumbens-area	0,442	0,035
Right-VentralDC	0,682	0,000
Right-choroid-plexus	0,621	0,002
WM-hypointensities	0,435	0,038
non-WM-hypointensities	0,468	0,024
CC_Posterior	0,829	0,000
CC_Mid_Posterior	0,567	0,005
CC_Central	0,677	0,000
CC_Mid_Anterior	0,448	0,032
CC_Anterior	0,565	0,005
BrainSegVol	0,670	0,000
BrainSegVolNotVent	0,695	0,000
SupraTentorialVol	0,434	0,038
SupraTentorialVolNotVent	0,466	0,025
SupraTentorialVolNotVentVox	0,748	0,000
MaskVol	0,555	0,006
lhSurfaceHoles	0,485	0,019
EstimatedTotalIntraCranialVol	0,672	0,000

Tabella 4.9: Regioni in cui il coefficiente di correlazione r_{MZ} è statisticamente significativo (*p-value<0.05)* per il parametro *volume.*

ROI	**rDZ**	**P-value**
Left-Cerebellum-Cortex	0,575	0,001
Left-Thalamus-Proper	0,531	0,003
Left-Caudate	0,662	0,000
Left-Putamen	0,433	0,019
Left-Pallidum	0,543	0,002
4th-Ventricle	0,468	0,010
Brain-Stem	0,704	0,000
Left-Hippocampus	0,573	0,001
CSF	0,554	0,002
Left-VentralDC	0,580	0,001

Left-vessel	0,428	0,020
Right-Inf-Lat-Vent	0,423	0,022
Right-Cerebellum-White-Matter	0,417	0,024
Right-Cerebellum-Cortex	0,479	0,009
Right-Thalamus-Proper	0,753	0,000
Right-Caudate	0,438	0,018
Right-Putamen	0,604	0,001
Right-Pallidum	0,551	0,002
Right-VentralDC	0,440	0,017
5th-Ventricle	0,564	0,564
WM-hypointensities	0,609	0,000
Optic-Chiasm	0,460	0,012
CC_Posterior	0,517	0,004
BrainSegVol	0,657	0,000
BrainSegVolNotVent	0,670	0,000
CorticalWhiteMatterVol	0,421	0,023
SubCortGrayVol	0,427	0,021
SupraTentorialVolNotVent	0,590	0,001
SupraTentorialVolNotVentVox	0,426	0,021
MaskVol	0,701	0,000
BrainSegVol-to-eTIV	0,721	0,000
rhSurfaceHoles	0,577	0,001
SurfaceHoles	0,766	0,000
EstimatedTotalIntraCranialVol	0,386	0,039

Tabella 4.10: Regioni in cui il coefficiente di correlazione r_{DZ} è statisticamente significativo (*p-value<0.05)* per il parametro *volume.*

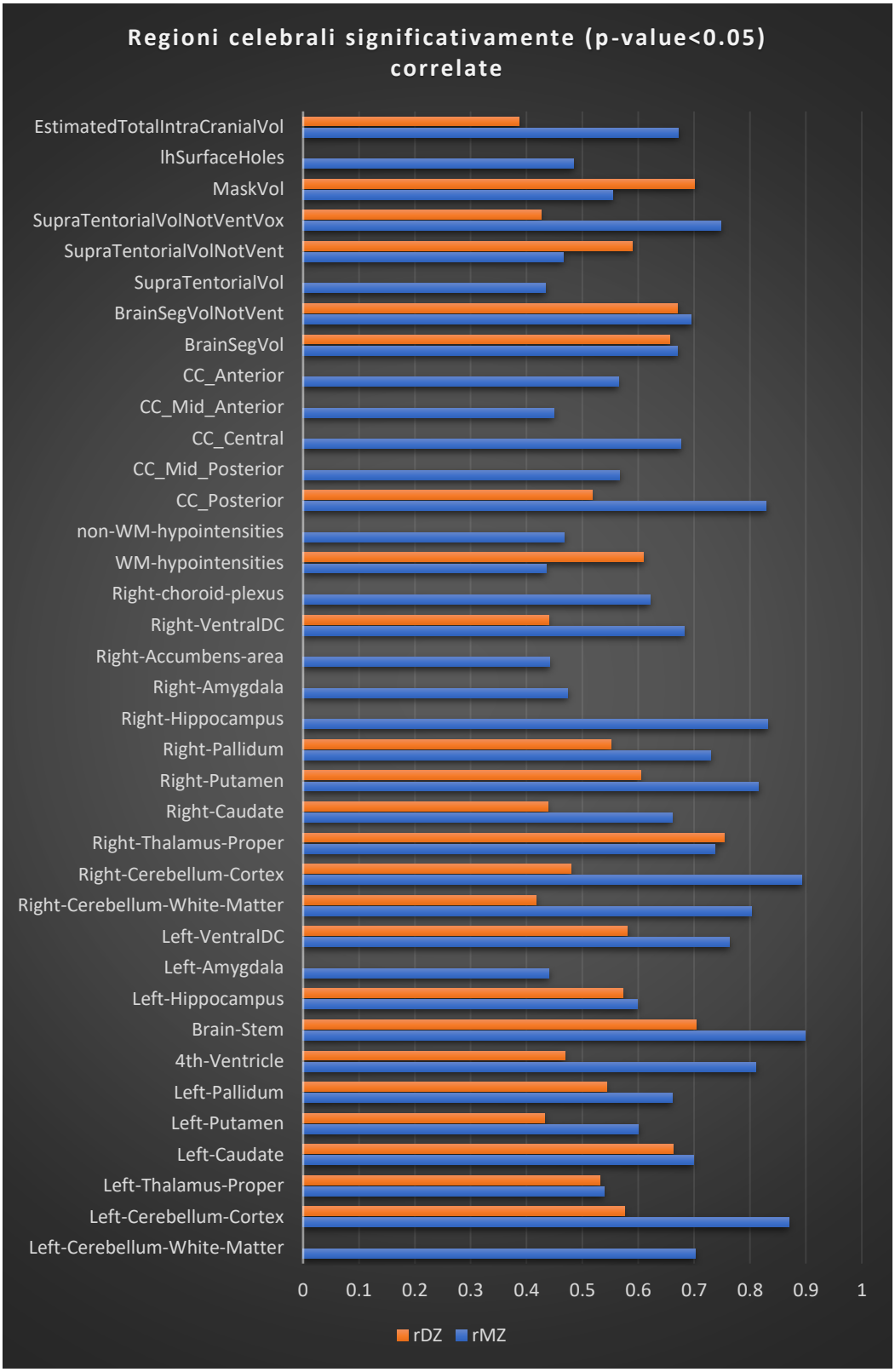

Figura 4.5: Regioni sottocorticali in cui è emersa una correlazione significativa (*p-value<0.05)* del valore assunto dal parametro *volume* nelle coppie MZ e confronto con il coefficiente di correlazione nelle coppie DZ quando anche questo è significativo (*p-value<0.05*) nelle stesse regioni

2. Analisi di ereditabilità dei parametri neuroanatomici

L'analisi di ereditabilità è stata condotta, per ciascun parametro, solo nelle regioni celebrali in cui $r_{MZ}>2r_{DZ}$. Le stime di ereditabilità ottenute sono state sottoposte ad un'analisi statistica (*p-value Fisher<0.05*) per valutarne la significatività.

Per il parametro *area superficiale*, i punteggi di ereditabilità nelle regioni corticali in cui essi sono significativi sono rappresentati nella Figura 4.6. Nella Figura 4.7 sono riportati i valori del *p-value* (*Fisher)* della differenza di correlazione tra MZ e DZ. Da un'ispezione visiva è possibile notare che la significatività statistica è alta, ovvero il *p-value* è basso, per stime maggiori di ereditabilità.

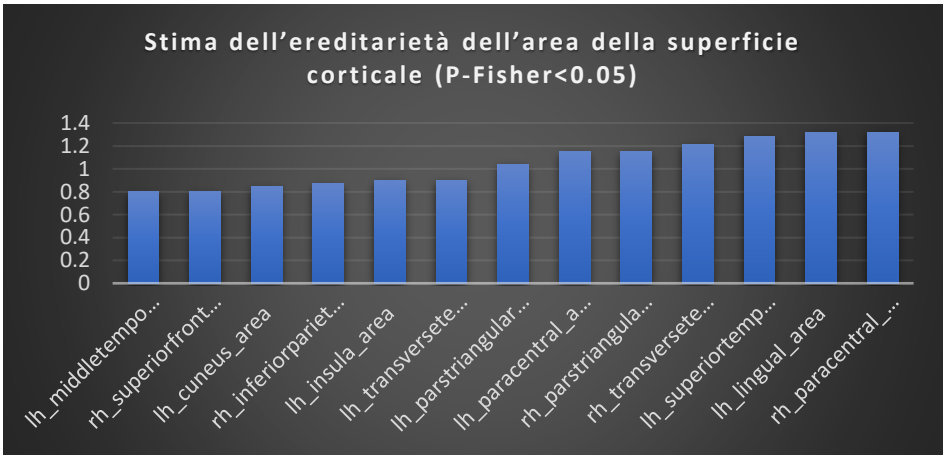

Figura 4.6: Stima dell'ereditarietà (ordine crescente di punteggio) del parametro *area superficiale* nelle regioni celebrali in cui è emersa una significatività statistica in accordo al Test di Fisher sulla differenza di correlazione.

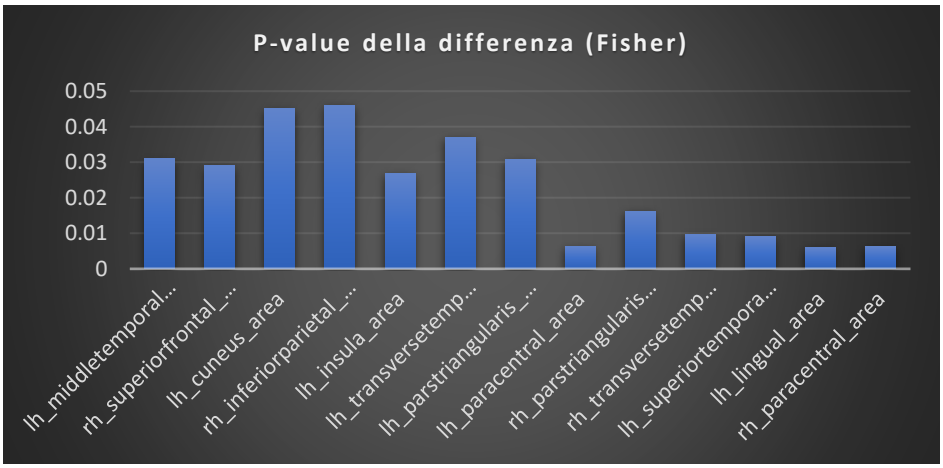

Figura 4.7: Valori del *p-value* della differenza di correlazione ottenuti mediante il Test di Fisher per il parametro *area superficiale.*

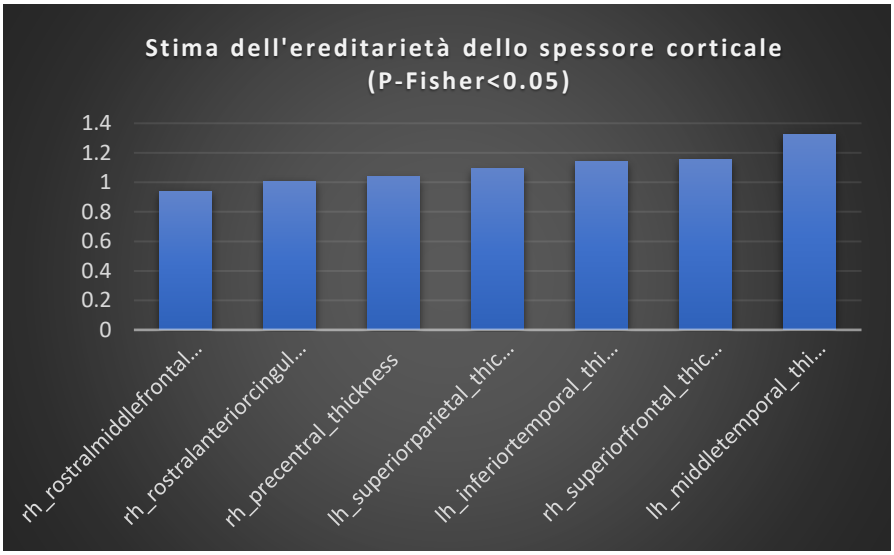

Figura 4.8: Stima dell'ereditarietà (ordine crescente di punteggio) del parametro *spessore corticale* nelle regioni celebrali in cui è emersa una significatività statistica in accordo al Test di Fisher sulla differenza di correlazione.

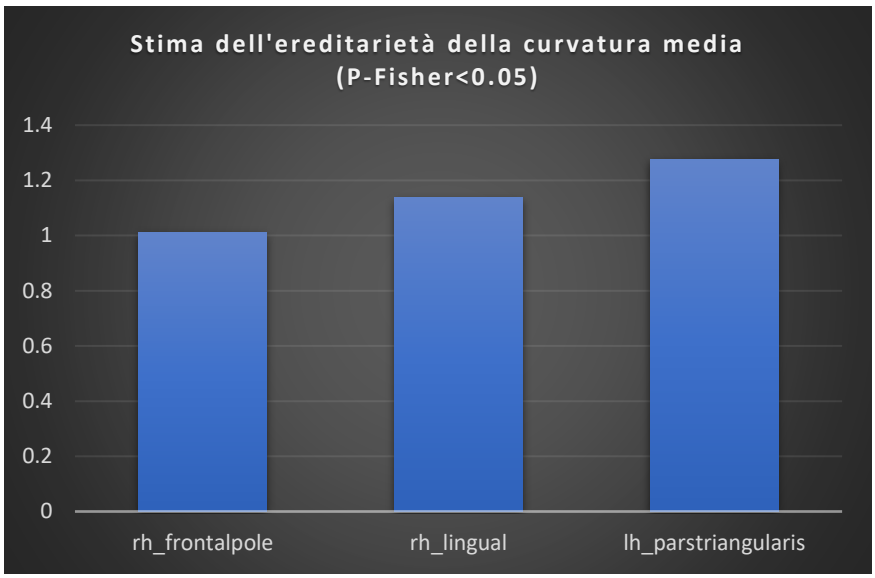

Figura 4.9: Stima dell'ereditarietà (ordine crescente di punteggio) del parametro *curvatura media* nelle regioni celebrali in cui è emersa una significatività statistica in accordo al Test di Fisher sulla differenza di correlazione.

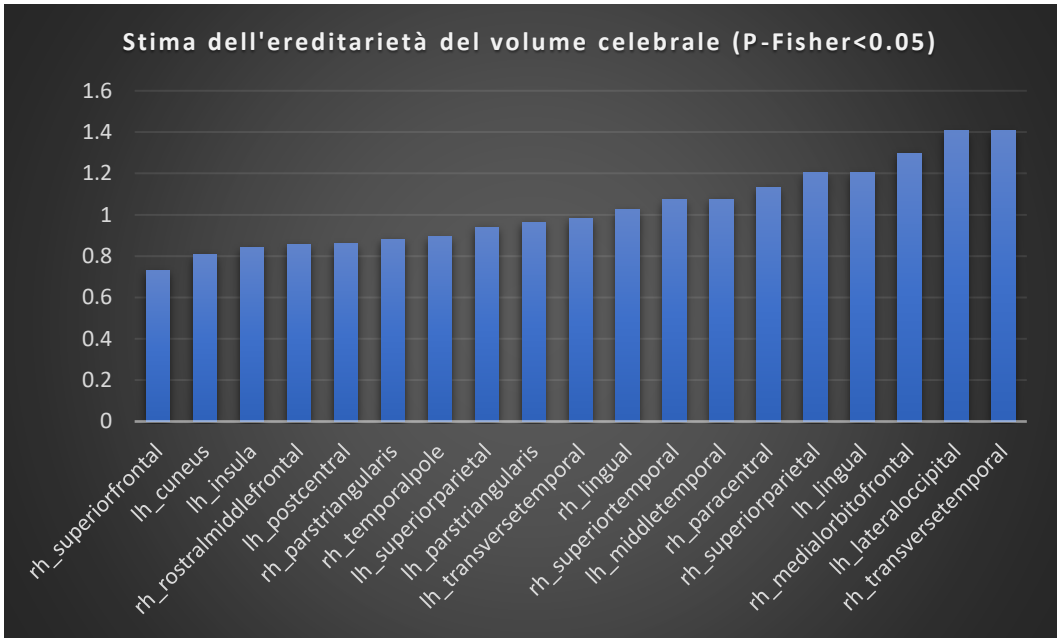

Figura 4.10: Stima dell'ereditarietà (ordine crescente di punteggio) del parametro *volume* nelle regioni celebrali in cui è emersa una significatività statistica in accordo al Test di Fisher sulla differenza di correlazione.

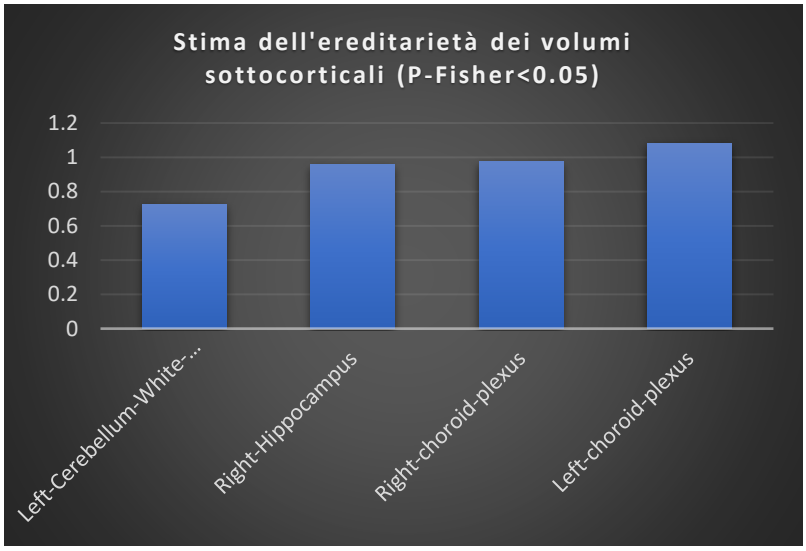

Figura 4.11: Stime significative (*p-value<0.05)* di ereditabilità del *volume* delle regioni sottocorticali.

Le stime di ereditabilità superano il valore unitario presumibilmente in due situazioni: quando l'*equal environmental assumption* è violata, ovvero quando i gemelli MZ condividono in misura maggiore l'ambiente rispetto ai gemelli DZ, oppure quando i gemelli DZ condividono meno del 50% del genoma. In entrambe queste situazioni è possibile ipotizzare che i fattori genetici risultano sovrastimati.

Sono emerse 13 regioni con ereditabilità significativa per il parametro *area superficiale*, mentre per il parametro *spessore corticale* sono emerse 7 regioni. Inoltre, sono emerse 19 regioni in cui il parametro *volume* è significativamente ereditabile e 3 sole regioni in cui il parametro *curvatura media* è significativamente ereditabile. In aggiunta, è emersa significatività statistica del punteggio di ereditabilità in 4 regioni sottocorticali.

Il *volume*, in linea con i risultati descritti nell'analisi di correlazione è risultato il parametro, con stime di ereditabilità significative in un numero maggiore di regioni celebrali rispetto agli altri parametri, in cui i fattori genetici giocano un ruolo determinante in tante regioni corticali.

L'*area superficiale* e lo *spessore* della *circonvoluzione temporale media* dell'emisfero sinistro mostrano punteggi di ereditabilità rispettivamente di 0.8 e 1.3, ovvero i geni influenzano maggiormente rispetto all'ambiente la variabilità inter-individuale dei valori di questi parametri. Lo *spessore* e il *volume* della *circonvoluzione parietale superiore* dell'emisfero sinistro, regione celebrale coinvolta nella processazione ed associazione degli stimoli somatosensoriali, sono ereditabili con punteggi rispettivamente di 1.1 e 0.9. Nella *circonvoluzione frontale superiore* dell'emisfero destro sono significativamente ereditabili i valori di *area superficiale* (0.8), di *spessore* (1.1) e di *volume* (0.7). Nella parte rostrale della *circonvoluzione frontale media* dell'emisfero destro è emersa un'ereditabilità significativa dello *spessore corticale* (0.9) e del *volume* (0.8). La circonvoluzione frontale superiore e la circonvoluzione frontale media sono regioni celebrali occupate dalla corteccia premotoria e dall'aria motoria supplementare, le quali svolgono un'azione di supporto nei confronti della corteccia motoria primaria (quindi sono implicate nei movimenti volontari). Nella *parte triangolare della circonvoluzione frontale inferiore* dell'emisfero sinistro è emersa un'ereditabilità significativa dei valori di *area superficiale* (1.1), di *volume* (0.9), di *curvatura media* (1.3) mentre nell'emisfero destro è emersa un'ereditabilità significativa dei valori di *area* (1.1) e di *volume* (0.9). La circonvoluzione frontale inferiore destra è stata tipicamente messa in relazione con compiti esegui/non esegui. In tale tipologia di compiti, il partecipante è chiamato ad inibire una risposta prepotente (ad esempio smettere di premere un pulsante quando appare un segnale rosso). La circonvoluzione frontale inferiore sinistra è estremamente importante per la produzione del linguaggio (qui è localizzata l'area di Broca). La variabilità tra i valori dei parametri morfologici di queste regioni tra gli individui è legata principalmente alla variabilità del patrimonio genetico. Questi risultati risultano in

linea con quanto noto dalla letteratura ovvero le regioni celebrali coinvolte nel linguaggio, nelle aree frontali e sensorimotorie sono subiscono principalmente l'influenza di fattori genetici. L'*area* e il *volume* del *cuneo sinistro* sono entrambi ereditabili con punteggio 0.8 mentre, nell'*insula* dell'emisfero sinistro sono ereditabili con punteggio di 0.9 e 0.8. La regione del cuneo è principalmente conosciuta per il suo coinvolgimento nel processo visivo di base. L'insula gioca un ruolo in diverse funzioni spesso legate all'emotività oppure alla regolazione dell'omeostasi corporea. Queste funzioni includono la percezione, il controllo motorio, l'auto-consapevolezza, le funzioni cognitive, e l'esperienza interpersonale. In relazione a queste alcuni ricercatori la sospettano coinvolta in aspetti anatomo-funzionali della psicopatologia. *L'area* e il *volume* della *circonvoluzione temporale trasversa* dell'emisfero sinistro sono ereditabili con punteggio di circa 0.9 mentre nell'emisfero destro con i punteggi 1.2 e 1.4. Tale circonvoluzione si trova nell'area della corteccia primaria uditiva. Nella *circonvoluzione linguale* dell'emisfero destro sono ereditabili i parametri di *volume* (1) e *curvatura media* (1.1) mentre sono ereditabili i valori di *area superficiale* (1.3) e di *volume* (1.2) nell'emisfero sinistro. Tale regione celebrale è principalmente legata ai processi visivi. L'*area* e il *volume* del *lobulo paracentrale* destro sono ereditabili con punteggi di 1.3 e 1.1, questa regione è la parte più centrale della circonvoluzione frontale superiore.

Nei *volumi sottocorticali* sono emerse significatività nel *plesso coroideo* di entrambi gli emisferi, nell'*ippocampo* dell'emisfero destro e nel v*olume di materia bianca presente nel cervelletto* dell'emisfero sinistro.

Negli studi di ricerca genetica in ambito psicologico e psichiatrico, l'ereditarietà non implica un determinismo genetico poiché i tratti comportamentali sono generalmente influenzati da geni multipli e fattori ambientali, il legame tra geni specifici e tratti/disturbi comportamentali è solo probabilistico: i geni rappresentano solo un fattore di rischio che aumenta la probabilità di sviluppare la malattia. Nella maggior parte dei casi, per i tratti complessi, l'influenza dell'ambiente è importante tanto quanto l'influenza genetica in quanto può modificare l'espressione genetica stessa.

3. Risultati modello ACE

Sebbene il modello ACE sia stato formulato per ogni parametro regionale in cui $r_{DZ}<r_{MZ}<2r_{DZ}$, esso è stato considerato significativo solo nelle regioni in cui i coefficienti di correlazione in coppie MZ e DZ sono significativi (*p-value<0.05).*

Nella Tabella 4.11 sono riportati i risultati ottenuti per il parametro *area superficiale.*

ROI	rMZ	rDZ	*p-value* MZ	*p-value* DZ	A	C	E
lh_pericalcarine_area	0,615	0,437	0,002	0,018	0,356	0,259	0,385
lh_superiorparietal_area	0,471	0,377	0,023	0,044	0,187	0,284	0,529
rh_inferiortemporal_area	0,573	0,395	0,004	0,034	0,355	0,218	0,427
rh_lingual_area	0,567	0,419	0,005	0,024	0,297	0,270	0,433
rh_middletemporal_area	0,663	0,515	0,001	0,004	0,297	0,367	0,337
rh_pericalcarine_area	0,719	0,692	0,000	0,000	0,054	0,664	0,281
rh_precuneus_area	0,647	0,382	0,001	0,041	0,532	0,116	0,353
rh_superiorparietal_area	0,563	0,500	0,005	0,006	0,127	0,436	0,437
rh_insula_area	0,425	0,394	0,043	0,035	0,063	0,362	0,575

Tabella 4.11: Modello ACE, per il parametro *area*, nelle regioni corticali in cui i coefficienti di correlazione sono statisticamente significativi (*p-value<0.05).* A: effetti genetici additivi; C: effetti ambientali condivisi; E: effetti ambientali non condivisi

Il modello ACE è stato sviluppato quando si è ipotizzato un contributo importante dei fattori ambientali condivisi alla varianza fenotipica. Nella regione *pericalcarina* dell'emisfero destro è emerso che le differenze del valore di area superficiale osservate tra i due gemelli sono attribuibili per il 5.4% a differenze di tipo genetico, mentre i restanti 66.4% e 28.1% della variabilità sono a carico, rispettivamente, di fattori ambientali condivisi e non condivisi. Si può quindi concludere che l'ambiente condiviso dai due gemelli sembri svolgere un ruolo importante nella determinazione del valore di area della superficie pericalcarina destra dei due individui. Un contributo elevato dell'ambiente è evidente anche nell'area della *circonvoluzione parietale superiore* dell'emisfero destro. Infatti, i fattori genetici contribuiscono alla variabilità solo per il 12.7% mentre i fattori ambientali condivisi per il 43.6% e quelli non condivisi per il 43.7%.

Nella Tabella 4.12 sono riportati i risultati ottenuti per il parametro *spessore corticale.*

ROI	rMZ	rDZ	*p-value* MZ	*p-value* DZ	A	C	E
lh_postcentral_thickness	0,745	0,590	0,000	0,001	0,311	0,434	0,255
lh_posteriorcingulate_thickness	0,613	0,553	0,002	0,002	0,121	0,492	0,387
lh_precentral_thickness	0,567	0,472	0,005	0,010	0,191	0,376	0,433
rh_cuneus_thickness	0,485	0,359	0,019	0,056	0,251	0,234	0,515
rh_postcentral_thickness	0,488	0,438	0,018	0,017	0,101	0,388	0,512
rh_superiortemporal_thickness	0,649	0,436	0,001	0,018	0,426	0,223	0,351

Tabella 4.12: Modello ACE, per il parametro *spessore corticale,* nelle regioni corticali in cui i coefficienti di correlazione sono statisticamente significativi (*p-value<0.05).* A: effetti genetici additive; C: effetti ambientali condivisi; E: effetti ambientali non condivisi.

Dall'analisi dei risultati del modello ACE, è emerso che le differenze tra i gemelli dei valori di spessore corticale nella *circonvoluzione postcentrale* e *nella circonvoluzione cingolata posteriore* dell'emisfero sinistro sono attribuibili principalmente a fattori ambientali condivisi dai gemelli. Infatti, nella circonvoluzione postcentrale sinistra il contributo della variabilità dell'espressione genetica al valore dello spessore è di 31.1%, il contributo della variabilità dell'ambiente condiviso, nelle diverse coppie, è del 43.4%, il contributo dell'ambiente non condiviso è del 25.5% mentre, nella circonvoluzione cingolata posteriore sinistra il contributo della variabilità genetica è di 12.1%, il contributo della variabilità dei fattori ambientali condivisi è del 49.2% e il contributo dei fattori ambientali non condivisi è del 38.7%.

Nella Tabella 4.13 sono riportati i risultati ottenuti per il parametro *curvatura media.*

ROI	rMZ	rDZ	*p-value* MZ	*p-value* DZ	A	C	E
lh_superiorparietal_meancurvature	0,579	0,461	0,004	0,012	0,236	0,342	0,421

Tabella 4.13: Modello ACE, per il parametro *curvatura media,* nelle regioni corticali in cui i coefficienti di correlazione sono statisticamente significativi (*p-value<0.05).* A: effetti genetici additive; C: effetti ambientali condivisi; E: effetti ambientali non condivisi.

Il modello ACE per il parametro curvatura media è risultato significativo solo nella *circonvoluzione parietale superiore* dell'emisfero sinistro. È emerso che la variabilità del parametro dipende per il 23.6% da fattori genetici, per il 34.2% da fattori ambientali

condivisi e per il 42.1% da fattori ambientali non condivisi. Dunque, secondo il modello, la curvatura media della circonvoluzione parietale superiore sinistra è determinata maggiormente da fattori ambientali rispetto che da fattori genetici.

Nella Tabella 4.14 sono riportati i risultati ottenuti per il parametro *volume.*

ROI	rMZ	rDZ	*p-value* MZ	*p-value* DZ	A	C	E
lh_pericalcarine_volume	0,561	0,399	0,005	0,032	0,325	0,237	0,439
rh_fusiform_volume	0,395	0,395	0,062	0,034	0,002	0,394	0,605
rh_middletemporal_volume	0,578	0,434	0,004	0,019	0,289	0,289	0,422
rh_precentral_volume	0,428	0,397	0,042	0,033	0,060	0,367	0,572
rh_supramarginal_volume	0,477	0,443	0,021	0,016	0,069	0,408	0,523
rh_frontalpole_volume	0,572	0,447	0,004	0,015	0,249	0,323	0,428
rh_insula_volume	0,577	0,455	0,004	0,013	0,244	0,333	0,423

Tabella 4.14: Modello ACE, per il parametro *volume,* nelle regioni corticali in cui i coefficienti di correlazione sono statisticamente significativi (*p-value<0.05).* A: effetti genetici additive; C: effetti ambientali condivisi; E: effetti ambientali non condivisi.

Dal modello ACE emerge un contributo importante dell'ambiente condiviso al valore di volume in ciascun gemello della coppia nella *circonvoluzione sopramarginale* dell'emisfero destro. Infatti, la variabilità del valore di volume di questa regione sembra dipendere per il 6.9% dalle differenze genetiche, per il 40.8% da fattori ambientali condivisi e per il 52.3% da fattori ambientali non condivisi.

Nella Tabella 4.15 sono riportati i risultati ottenuti per il parametro *volume* a livello globale e nelle regioni sottocorticali.

Regioni sottocorticali	rMZ	rDZ	*p_value* MZ	*p-value* DZ	A	C	E
Left-Cerebellum-Cortex	0,869	0,575	0,00000	0,00110	0,587	0,281	0,131
Left-Thalamus-Proper	0,539	0,531	0,00795	0,00303	0,016	0,523	0,461
Left-Caudate	0,699	0,662	0,00021	0,00009	0,073	0,625	0,301
Left-Putamen	0,600	0,433	0,00246	0,01903	0,335	0,265	0,400
Left-Pallidum	0,660	0,543	0,00060	0,00233	0,235	0,426	0,340
4th-Ventricle	0,811	0,468	0,00000	0,01040	0,685	0,126	0,189
Brain-Stem	0,898	0,704	0,00000	0,00002	0,389	0,509	0,102
Left-Hippocampus	0,598	0,573	0,00256	0,00116	0,051	0,548	0,402
Left-VentralDC	0,763	0,580	0,00002	0,00097	0,366	0,397	0,237
Right-Cerebellum-White-Matter	0,802	0,417	0,00000	0,02427	0,769	0,033	0,198
Right-Cerebellum-Cortex	0,893	0,479	0,00000	0,00856	0,827	0,066	0,107
Right-Caudate	0,661	0,438	0,00060	0,01759	0,446	0,214	0,339
Right-Putamen	0,815	0,604	0,00000	0,00052	0,423	0,392	0,185
Right-Pallidum	0,730	0,551	0,00008	0,00195	0,358	0,372	0,270
Right-VentralDC	0,682	0,440	0,00034	0,01694	0,484	0,198	0,318
CC_Posterior	0,829	0,517	0,00000	0,00408	0,624	0,205	0,171
BrainSegVol	0,670	0,657	0,00047	0,00011	0,027	0,644	0,330
BrainSegVolNotVent	0,695	0,670	0,00023	0,00007	0,050	0,645	0,305
SupraTentorialVolNotVentVox	0,748	0,426	0,00004	0,02129	0,645	0,103	0,252
EstimatedTotalIntraCranialVol'	0,672	0,386	0,00045	0,03861	0,571	0,100	0,328

Tabella 4.15: Modello ACE, per il parametro *volume,* nelle regioni sottocorticali in cui i coefficienti di correlazione sono statisticamente significativi (*p-value<0.05).* A: effetti genetici additive; C: effetti ambientali condivisi; E: effetti ambientali non condivisi.

Dall'analisi dei risultati del modello ACE nelle diverse regioni sottocorticali è emerso un contributo significativo dell'ambiente alle differenze del valore di volume tra i gemelli che compongono le coppie, nel *nucleo caudato* sinistro (C=62,5%), nel *pallidum* sinistro (C=42,6%) e destro (C=37,2%), nel *tronco encefalico* (C=50,9%) e *nella faccia ventrale* (o ipotalamica) *del diencefalo* (C=39,7%). Inoltre, è emersa una dipendenza significativa dall'ambiente condiviso dei valori di *BrainSegVol* (C=64,4%) e *BrainSegNotVent* (C=64,5%). Questi valori corrispondono, rispettivamente, al volume di tutti i voxel del cervello che non appartengono nè al background e nè al tronco encefalico e al volume di tutti i voxel del cervello che non appartengono al background, al tronco encefalico, ai ventricoli (laterale, laterale inferiore, terzo, quarto e quinto) e neanche al liquido celebrospinale e al plesso coroideo.

Questi risultati evidenziano che il solo fatto che il coefficiente di correlazione nei MZ sia maggiore del coefficiente di correlazione nei DZ non è sufficiente per valutare accuratamente l'impatto genetico sulle caratteristiche neuroanatomiche. Infatti, ad esempio, per il nucleo caudato, per il pallidum, per il tronco encefalico e per la faccia ipotalamica del diencefalo emerge una correlazione maggiore nei MZ che potrebbe far presuppore un'influenza genetica predominante. D'altra parte, il modello ACE ha rivelato un contributo maggiore dell'ambiente condiviso sui volumi di queste regioni. Questo risultato mette in luce la potenzialità del modello ACE nel fornire informazioni più dettagliate basandosi su un confronto tra i coefficienti di correlazione dei due gruppi e su una valutazione di *quanto* il coefficiente dei MZ sia maggiore rispetto al coefficiente dei DZ.

3.1 Discussione dei risultati

Il modello ACE ha permesso di scindere il contributo genetico e il contributo ambientale (condiviso e unico) sulla variabilità dei valori assunti dal parametro tra i gemelli che compongono le coppie in analisi.

Le somiglianze tra i gemelli non possono che derivare da ciò che hanno in comune ovvero i fattori A e C, mentre le differenze derivano da ciò che non hanno in comune, cioè E (che include fattori ambientali, errori di misurazione e dalla porzione di patrimonio genetico non condiviso). Quando i valori di correlazione sono alti e abbastanza diversi tra loro ($r_{MZ}>2r_{DZ}$) è possibile ipotizzare che il tratto in esame sia influenzato in larga parte da fattori di natura genetica. Quando i valori di correlazione dei due gruppi di zigosità sono comparabili è possibile ipotizzare che il tratto in esame sia influenzato in larga parte da fattori di natura ambientale condivisa (in questo caso si sviluppa il modello ACE per capire in che percentuale). Quando i valori di correlazione sono entrambi bassi e comparabili è possibile ipotizzare che il tratto in esame sia influenzato in larga parte da fattori di natura ambientale unica (e da errori di misurazione del parametro). Le domande sul *se* e sul *quanto* i fattori genetici influenzino le caratteristiche cerebrali, e di conseguenza i tratti caratteriali ed eventualmente l'insorgenza di disturbi psicopatologici, rappresentano solo i primi passi verso la comprensione delle basi delle differenze inter-individuali. Il passo successivo sarà quello di comprendere *quali geni* e *quali specifici fattori ambientali* sono coinvolti nell'espressione di un determinato fenotipo.

4. Risultati dell'analisi di regressione lineare

In questa fase, è stato ispezionato il valore della statistica T in corrispondenza dei coefficienti significativi per comprendere il tipo di legame presente tra il regressore e la variabile di risposta. I risultati dell'analisi di regressione sono stati interpretati alla luce delle stime di ereditabilità riportate nelle sezioni precedenti.

4.1 Effetto della zigosità sulla similarità intra-coppia

Dall'analisi dell'effetto della variabile zigosità sulla similarità neuroanatomica sono emerse significatività nelle regioni riportate nella Tabella 4.16.

ROI	$w_{zygosity}$	P-value	tStat	h	P-value del modello
lh_middletemporal_**area**	-0,0014	0,0224	-2,3526	0,799	0.0243
rh_entorhinal_**meancurvature**	0,0163	0,0152	2,5092	-1,13	0.0049
lh_parsorbitalis_**thickness**	0,0366	0,0478	2,0267	-0.667	0.0394
lh_insula_**thickness**	0,0301	0,0069	2,8157	-0.656	0.0193
rh_bankssts_**thickness**	0,0349	0,0374	2,1358	-0.232	0.0026
rh_insula_**thickness**	0,0304	0,0189	2,4229	-0.924	0.0284
Left-Cerebellum-Cortex	-1852,18	0,0070	-2,8069	0,587	0,0337
Right-Cerebellum-Cortex	-2411,78	0,0023	-3,1986	0,827	0,0098

Tabella 4.16: Regioni celebrali in cui sono emerse significatività statistiche del modello di regressione lineare multipla e dei coefficienti della variabile *zigosità*.

Nelle regioni celebrali riportate nella Tabella 4.16 la differenza intra-coppia dei valori del parametro risente della variabile zigosità della coppia. In particolare, per l'*area della superficie della circonvoluzione temporale media dell'emisfero sinistro* emerge una similarità maggiore nei gemelli MZ rispetto ai gemelli DZ. Inoltre, in questa regione, la stima di ereditarietà del parametro *area* è 79.9% e, come visto nel paragrafo 2 di questo capitolo, questa stima è significativa rispetto al *p-value di Fisher*. Dunque, in questa regione celebrale è ragionevole ipotizzare che la variabilità del valore di area tra gli individui sia in gran parte legata ad una variabilità genetica.

La similarità nei valori di *curvatura media nella corteccia entorinale destra* risulta maggiore nei gemelli DZ rispetto ai gemelli MZ. In linea con quanto detto in merito alle stime di

ereditabilità, una maggiore somiglianza nelle coppie DZ rende inutile lo studio dell'influenza dei fattori genetici sull'espressione della caratteristica in esame. Nei modelli sviluppati a partire dalle differenze intra-coppia dei valori di spessore corticale, è emersa una maggiore similarità nei DZ nella *parte orbitale della circonvoluzione frontale inferiore dell'emisfero sinistro*, *nell'insula di entrambi gli emisferi* e nel *solco temporale superiore destro*. Infatti, i punteggi di ereditabilità associati risultano non significativi. Nella *corteccia cerebellare di entrambi gli emisferi* emerge un consistente effetto della zigosità sulla similarità intra-coppia. Il volume di questa area celebrale è più simile nelle coppie MZ, quindi i fattori genetici sembrerebbero svolgere un ruolo determinante sui valori di volume della corteccia cerebellare.

4.2 Effetto dell'età sulla similarità intra-coppia

Dall'analisi dell'effetto della variabile età sulla similarità neuroanatomica, al netto degli effetti di zigosità e sesso, sono emerse significatività nelle regioni riportate nella Tabella 4.17

ROI	w_{age}	P-value	tStat	P-value del modello
lh_lateraloccipital_**area**	0,0002	0,0101	2,6707	0,0089
lh_precentral_**area**	-0,0002	0,0042	-2,9924	0,0010
rh_superiorparietal_**area**	-0,0002	0,0048	-0,2947	0,0333
rh_superiorparietal_**meancurvature**	-0,0003	0,0251	-2,3059	0.0418
rh_frontalpole_**meancurvature**	0,0008	0,01632	2,4822	0.0362
rh_temporalpole_**meancurvature**	-0,0010	0,0042	-2,9956	0.0165
lh_rostralanteriorcingulate_**meancurvature**	0,00045	0,01121	2,62972	0.0140
lh_pericalcarine_**thickness**	-0,0039	0,0069	-2,8144	0.0283
Optic-Chiasm	-2,671	0,0029	-3,1231	0,0074

Tabella 4.17: Regioni celebrali in cui sono emerse significatività statistiche del modello di regressione lineare multipla e dei coefficienti della variabile *età*.

Nelle regioni celebrali riportate nella Tabella 4.17 la variabile età influenza significativamente le differenze/similarità intra-coppia. La similarità dell'*area superficiale della circonvoluzione occipitale laterale dell'emisfero sinistro* si riduce con la crescita

mentre la similarità dell'*area superficiale della circonvoluzione precentrale sinistra* e del *lobulo parietale superiore destro* aumenta con la crescita. La similarità della *curvatura media nel lobulo parietale superiore* e nel *polo temporale destro* aumenta con la crescita mentre essa si riduce nel *polo frontale destro* e nella *regione rostrale della corteccia cingolata dell'emisfero sinistro*. La similarità dei valori di *spessore corticale della corteccia pericalcarina dell'emisfero sinistro* aumenta con la crescita così come anche la similarità del valore di volume del *chiasma ottico*.

L'interpretazione di questi risultati potrebbe essere la seguente: alla nascita i gemelli hanno un certo grado di similarità neuroanatomica influenzato da fattori genetici (condivisi al 100% nei MZ e circa al 50% nei DZ) e dall'ambiente condiviso (intra-uterino). Durante la crescita entrano in gioco esperienze di vita, che influenzano lo sviluppo dei due gemelli, sia comuni ai due soggetti, come ad esempio le esperienze condivise all'interno della famiglia (lavoro dei genitori, perdita di un genitore, aspetti normativi delle cure parentali, livello socio-culturale ed economico…) e al di fuori di essa (amicizie, ambiente scolastico) che esperienze non comuni ad entrambi, ovvero esperienze uniche nella vita di ciascun gemello. Una riduzione della similarità con la crescita potrebbe esser legata a esperienze che accadono nella vita di uno solo dei gemelli della coppia che condizionano il valore del parametro investigato solo nel gemello coinvolto, causandone una diversità rispetto all'altro gemello della coppia. In altre parole, il fattore "E" del modello ACE cambia durante la crescita, rendendo i due gemelli meno simili. Viceversa, un aumento della similarità neuroanatomica con la crescita potrebbe esser legato a quegli eventi che accadono nella vita di entrambi i gemelli e pertanto li influenzano allo stesso modo rendendoli più simili, indipendentemente dal fatto che siano monozigoti o dizigoti. In conclusione, nelle regioni in cui il valore del parametro con la crescita diventa più o meno simile tra fratelli gemelli, è presente un fattore che entra in gioco con la crescita, che potrebbe presumibilmente essere afferibile a influenze dell'ambiente rispettivamente condiviso e non condiviso oltre che ad una probabile evoluzione dell'espressione genetica con la crescita.

A supporto dell'ipotesi delle influenze ambientali, in corrispondenza delle regioni la cui similarità intra-coppia varia con la crescita è emersa un'ereditabilità, ovvero la porzione di varianza legata alla variabilità genetica, non significativa eccetto che per la curvatura media del polo frontale dell'emisfero destro.

4.3 Effetto della variabile sesso sulla similarità intra-coppia

Dall'analisi dell'effetto della variabile sesso sulla similarità neuroanatomica intra-coppia sono emerse significatività nelle regioni riportate nella Tabella 4.18.

ROI-AREA	w_{sex}	P-value	tStat	P-value del modello
lh_middletemporal**_area**	0,0013	0,026	2,2961	0.0243
lh_precentral**_area**	-0,0025	0,0024	-3,1909	0,0010
rh_supramarginal**_thickness**	0,0196	0,0220	2,3608	0.0222
rh_bankssts**_volume**	0,0001	0,0028	3,1381	0.00151

Tabella 4.18: Regioni celebrali in cui sono emerse significatività statistiche del modello di regressione lineare multipla e dei coefficienti della variabile *sesso*.

Le regioni corticali riportate nella Tabella 4.18 mostrano dei valori di differenza intra-coppia significativamente diversi a seconda che le coppie siano di sesso femminile o di sesso maschile. In particolare, *l'area della circonvoluzione temporale media dell'emisfero sinistro* è più simile nelle coppie di sesso femminile mentre *l'area della circonvoluzione precentrale dell'emisfero sinistro* è più simile nelle coppie di sesso maschile. Lo *spessore della corteccia sopramarginale dell'emisfero destro* è più simile nelle coppie di sesso femminile così come anche il *volume del solco temporale superiore destro*.

La maggiore similarità potrebbe essere associata a fattori ormonali genere-specifici o all'influenza di fattori ambientali condivisi in misura differente nelle coppie maschili e femminili, ipotizzando che tale ambiente condiviso influenzi il carattere investigato in misura maggiore rispetto all'ambiente non condiviso. Nonostante ciò, l'ipotesi secondo cui le coppie di sesso femminile e maschile condividano in misura differente l'ambiente, indipendentemente dal fatto che siano coppie MZ o DZ ed indipendentemente dall'età, se verificata, comporta una violazione dell'assunzione di uguaglianza ambientale.

Si osservi che il modello di regressione lineare, per il parametro area, applicato alla circonvoluzione temporale media dell'emisfero sinistro ha mostrato un coefficiente significativo sia per la variabile zigosità che per la variabile sesso ma considerando il valore assoluto del coefficiente emerge un maggiore effetto della zigosità sulla differenza intra-coppia rispetto all'effetto del sesso.

5. Impatto del rischio psicotico sulla neuroanatomia

In linea con le analisi soprariportate, l'analisi dell'impatto del rischio psicotico sulla neuroanatomia è stata implementata separatamente per ciascuna misura morfologica analizzata.

I risultati dei confronti tra gemelli di controllo, a basso rischio e ad alto rischio psicotico sono riportati di seguito.

Dal *ttest* tra i valori di **area** nei gemelli ad **alto rischio** e nei **gemelli a basso rischio**, sono emerse delle significatività riguardanti le regioni: *circonvoluzione temporale trasversa dell'emisfero destro*, *corteccia orbito-frontale mediale dell'emisfero sinistro*, *polo frontale destro*.

Dal *ttest2* tra i valori di **area** nei gemelli ad **alto rischio** e nei **gemelli di controllo**, sono emerse differenze significative per il *solco latero-occipitale dell'emisfero sinistro*.

Dal *ttest2* tra i valori di **area** nei gemelli a **basso rischio** e nei **gemelli di controllo** sono emerse differenze significative nelle seguenti regioni: *corteccia orbito frontale mediale dell'emisfero sinistro*, *parte triangolare della circonvoluzione frontale inferiore dell'emisfero sinistro*, *regione caudale della circonvoluzione frontale media dell'emisfero destro*.

Nella Figura 4.12 sono riportati i valori di area nei 3 gruppi nel solco occipitale laterale dell'emisfero sinistro. Per tale misura sono emerse significatività dal confronto dei valori nei gemelli ad alto rischio e nei gemelli di controllo ma non dagli altri due confronti. Si osservi che, anche se non sono state riscontrate differenze significative tra i gemelli ad alto rischio e i gemelli a basso rischio, i valori di area di questi ultimi sono leggermente maggiori rispetto ai valori di area nei gemelli di controllo.

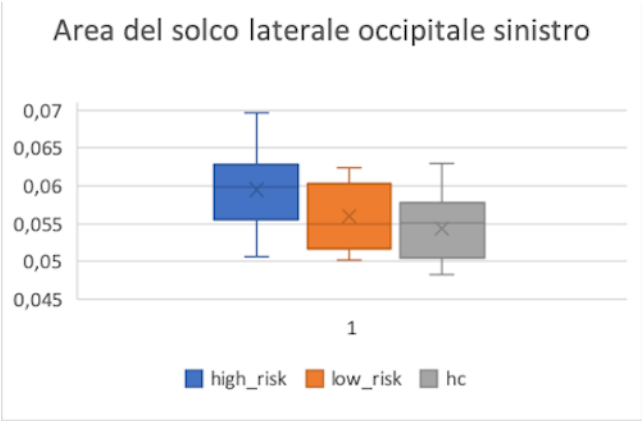

Figura 4.12: Valori di *area* nel solco laterale occipitale dell'emisfero sinistro nei gemelli ad alto rischio (in blue), nei gemelli a basso rischio (in arancione) e nei gemelli di controllo (in grigio).

Dal *ttest* tra i valori di **spessore corticale** nei gemelli **ad alto rischio** e nei **gemelli a basso** rischio emerge una differenza statisticamente significativa nella *parte rostrale della corteccia cingolata anteriore destra.*

Dal *ttest2* tra i valori di **spessore corticale** nei gemelli ad **alto rischio** e nei **gemelli di controllo** emergono differenze significative nella *parte rostrale della corteccia cingolata anteriore destra* e *nella parte triangolare della circonvoluzione frontale inferiore.*

Dal *ttest2* tra i valori di **spessore corticale** nei gemelli a **basso rischio** e nei **gemelli di controllo**, emerge una differenza statisticamente significativa nella *parte rostrale della corteccia cingolata anteriore destra.*

Nella Figura 4.13 sono messi a confronto i valori di spessore corticale nei tre gruppi nella regione rostrale della corteccia cingolata anteriore destra. Sia i gemelli ad alto rischio che i gemelli a basso rischio hanno valori di spessore corticale ridotti rispetto ai gemelli di controllo.

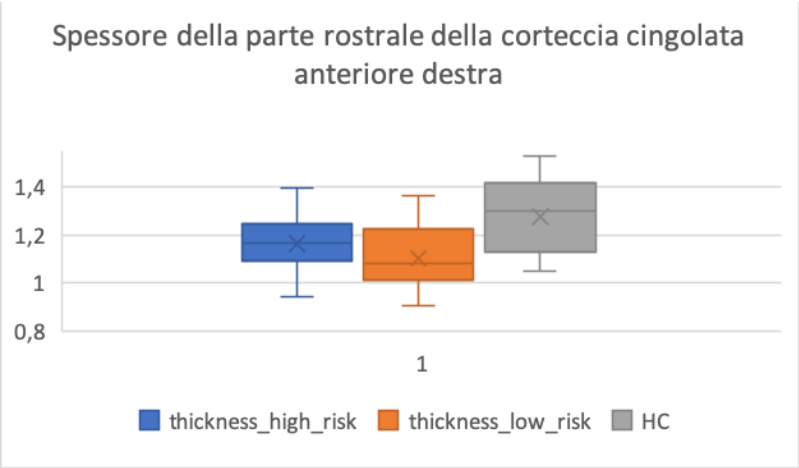

Figura 4.13: Valori di *spessore corticale* nella regione rostrale della corteccia cingolata anteriore destra nei gemelli ad alto rischio (in blue), nei gemelli a basso rischio (in arancione) e nei gemelli di controllo (in grigio).

Nella Figura 4.14 sono mostrati i valori di spessore corticale nella parte triangolare della circonvoluzione frontale inferiore. I gemelli ad alto rischio presentano valori di spessore corticale significativamente maggiori rispetto ai gemelli di controllo mentre, i gemelli a basso rischio non mostrano valori significativamente differenti dai controlli.

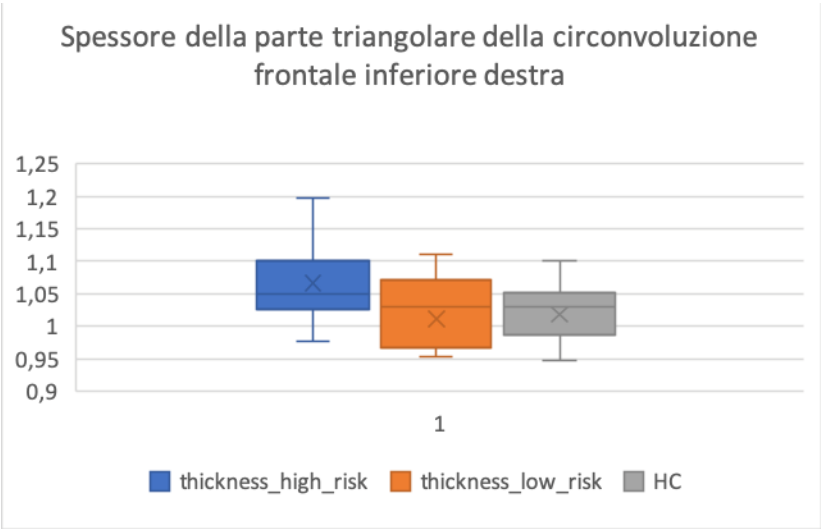

Figura 4.14: Valori di *spessore corticale* nella parte triangolare della circonvoluzione frontale inferiore destra nei gemelli ad alto rischio (in blue), nei gemelli a basso rischio (in arancione) e nei gemelli di controllo (in grigio).

Dal *ttest* tra i valori di **curvatura media** nei gemelli ad **alto rischio** e nei **gemelli a basso rischio** sono emerse delle significatività in corrispondenza delle regioni celebrali riportate nella Tabella 4.19.

ROI
Regione caudale della circonvoluzione frontale media sinistra
Corteccia entorinale destra
Circonvoluzione post-centrale di entrambi gli emisferi

Tabella 4.19: Regioni celebrali in cui sono emerse significatività statistiche dal *ttest* tra i gemelli ad alto rischio e i gemelli a basso rischio.

Dal *ttest2* tra i valori di **curvatura media** nei gemelli ad **alto rischio** e nei **gemelli di controllo** sono emerse delle significatività in corrispondenza delle regioni celebrali riportate nella Tabella 4.20.

ROI
Regione rostrale della circonvoluzione frontale media di entrambi gli emisferi
Circonvoluzione frontale superiore sinistra
Circonvoluzione temporale superiore sinistra
Circonvoluzione temporale trasversa sinistra
Regione caudale della circonvoluzione frontale media destra
Solco laterale occipitale destro
Corteccia cingolata posteriore destra
Circonvoluzione sopramarginale destra

Tabella 4.20: Regioni celebrali in cui sono emerse significatività statistiche dal *ttest2* tra i gemelli ad alto rischio e i gemelli di controllo.

Dal *ttest2* tra i valori di **curvatura media** nei **gemelli a basso rischio** e nei **gemelli di controllo** sono emerse delle significatività in corrispondenza delle regioni celebrali riportate nella Tabella 4.21.

ROI
Regione caudale della circonvoluzione frontale media di entrambi gli emisferi
Lobulo parietale inferiore sinistro
Circonvoluzione paracentrale di entrambi gli emisferi
Parte opercolare e parte orbitale della circonvoluzione frontale inferiore sinistra
Circonvoluzione postcentrale di entrambi gli emisferi
Precuneo di entrambi gli emisferi
Regione rostrale della circonvoluzione frontale media di entrambi gli emisferi
Cuneo destro
Giro fusiforme destro
Regione paraippocampale destra
Corteccia cingolata posteriore destra
Corteccia latero occipitale destra
Circonvoluzione sopramarginale destra

Tabella 4.21: Regioni celebrali in cui sono emerse significatività statistiche dal *ttest2* tra i gemelli a basso rischio e i gemelli di controllo.

Nella Figura 4.15 sono stati messi a confronto i valori di curvatura media nei tre gruppi nelle regioni in cui sono emerse delle significatività congiunte tra i gemelli ad alto rischio e i gemelli di controllo e i gemelli a basso rischio e gemelli di controllo. In queste regioni, sia i gemelli ad alto rischio che i gemelli a basso rischio hanno valori maggiori di curvatura media rispetto ai gemelli di controllo.

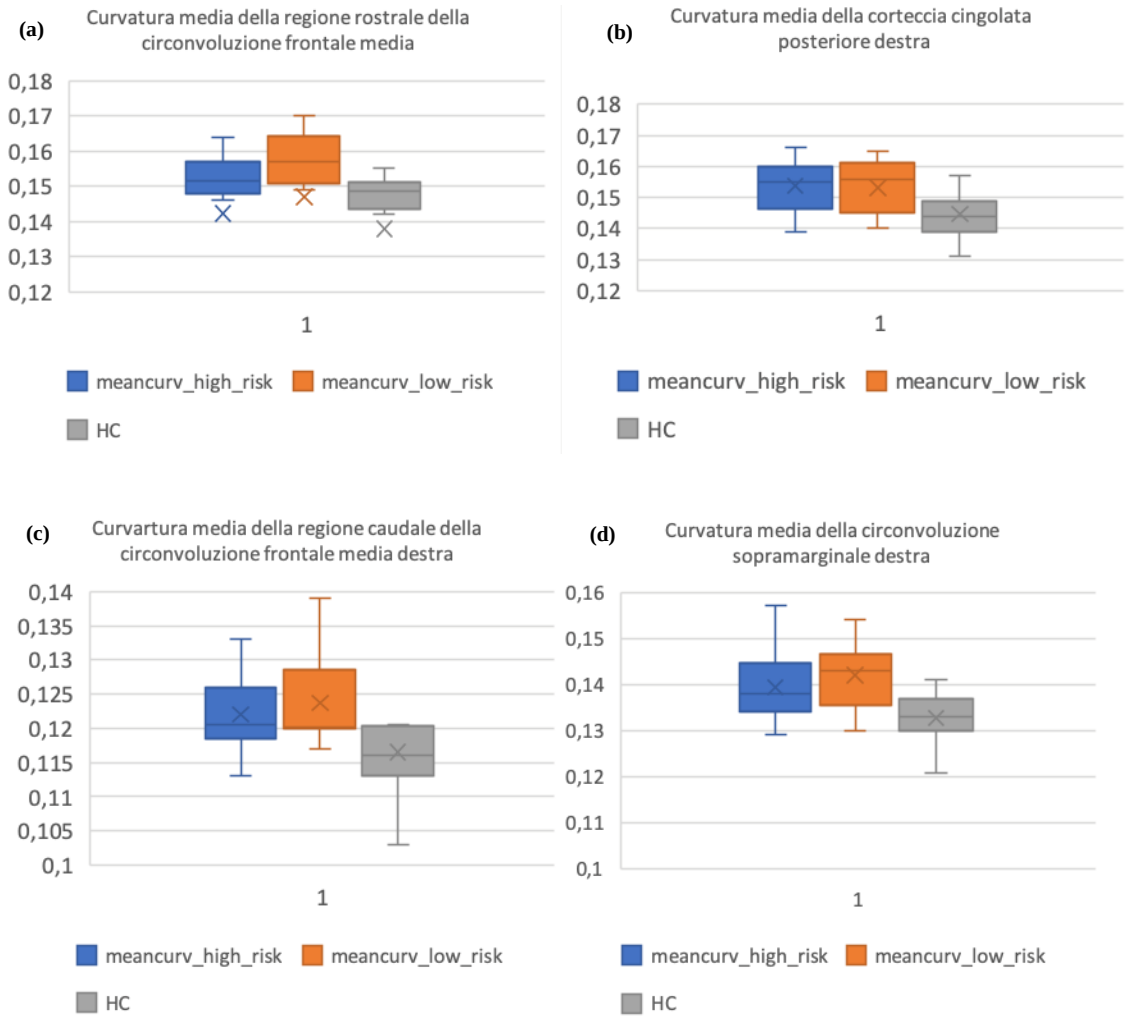

Figura 4.15: Confronto dei valori di *curvatura media* nella (a) regione rostrale della circonvoluzione frontale media, (b) nella corteccia cingolata posteriore destra, (c) nella regione caudale della circonvoluzione frontale media destra, (d) nella circonvoluzione sopramarginale destra, nei gemelli ad alto rischio (in blue), nei gemelli a basso rischio (in arancione) e nei gemelli di controllo (in grigio).

Dal *ttest* tra i valori di **volume** nei **gemelli ad alto rischio** e nei **gemelli basso rischio** sono emerse significatività nelle regioni riportate nella Tabella 4.22.

ROI
Solco superiore temporale sinistro
Circonvoluzione occipitale laterale sinistra
Corteccia orbitofrontale laterale sinistra
Corteccia orbitofrontale mediale sinistra
Precuneo entrambi gli emisferi
Circonvoluzione parietale inferiore destra
Istmo della circonvoluzione cingolata destra

Tabella 4.22: Regioni celebrali in cui sono emerse significatività statistiche dal ttest tra i gemelli ad alto rischio e i gemelli a basso rischio.

Dal *ttest2* tra i valori di **volume** nei **gemelli ad alto rischio** e nei **gemelli di controllo** sono emerse significatività nelle regioni riportate nella Tabella 4.23.

ROI
Solco superiore temporale sinistro
Regione caudale della circonvoluzione frontale mediale entrambi gli emisferi
Cuneo entrambi gli emisferi
Circonvoluzione parietale inferiore entrambi gli emisferi
Circonvoluzione occipitale laterale sinistra
Parte opercolare della circonvoluzione frontale inferiore entrambi gli emisferi
Parte triangolare della circonvoluzione frontale inferiore entrambi gli emisferi
Corteccia cingolata posteriore sinistra
Regione rostrale della circonvoluzione frontale mediale sinistra

Circonvoluzione frontale superiore sinistra
Circonvoluzione temporale superiore sinistra
Circonvoluzione fusiforme destra
Circonvoluzione paraippocampale destra
Circonvoluzione postcentrale destra
Circonvoluzione precentrale destra
Circonvoluzione frontale superiore destra
Circonvoluzione temporale superiore destra

Tabella 4.23: Regioni celebrali in cui sono emerse significatività statistiche dal ttest2 tra i gemelli ad alto rischio e i gemelli di controllo.

Dal *ttest2* tra i valori di **volume** nei **gemelli a basso rischio** e nei **gemelli di controllo** sono presenti differenze significative nelle regioni riportate nella Tabella 4.24.

ROI
Cuneo sinistro
Circonvoluzione occipitale laterale sinistra
Parte triangolare della circonvoluzione frontale inferiore sinistra
Circonvoluzione paraippocampale destra
Circonvoluzione temporale trasversa destra

Tabella 4.24: Regioni celebrali in cui sono emerse significatività statistiche dal ttest2 tra i gemelli a basso rischio e i gemelli di controllo.

Nella Figura 4.16 sono riportati i valori di volume nei tre gruppi nelle regioni in cui sia i gemelli ad alto rischio che i gemelli a basso rischio assumono valori significativamente diversi dai controlli. In queste regioni, i valori di volume sono più alti sia nei gemelli ad alto rischio che nei gemelli a basso rischio psicotico.

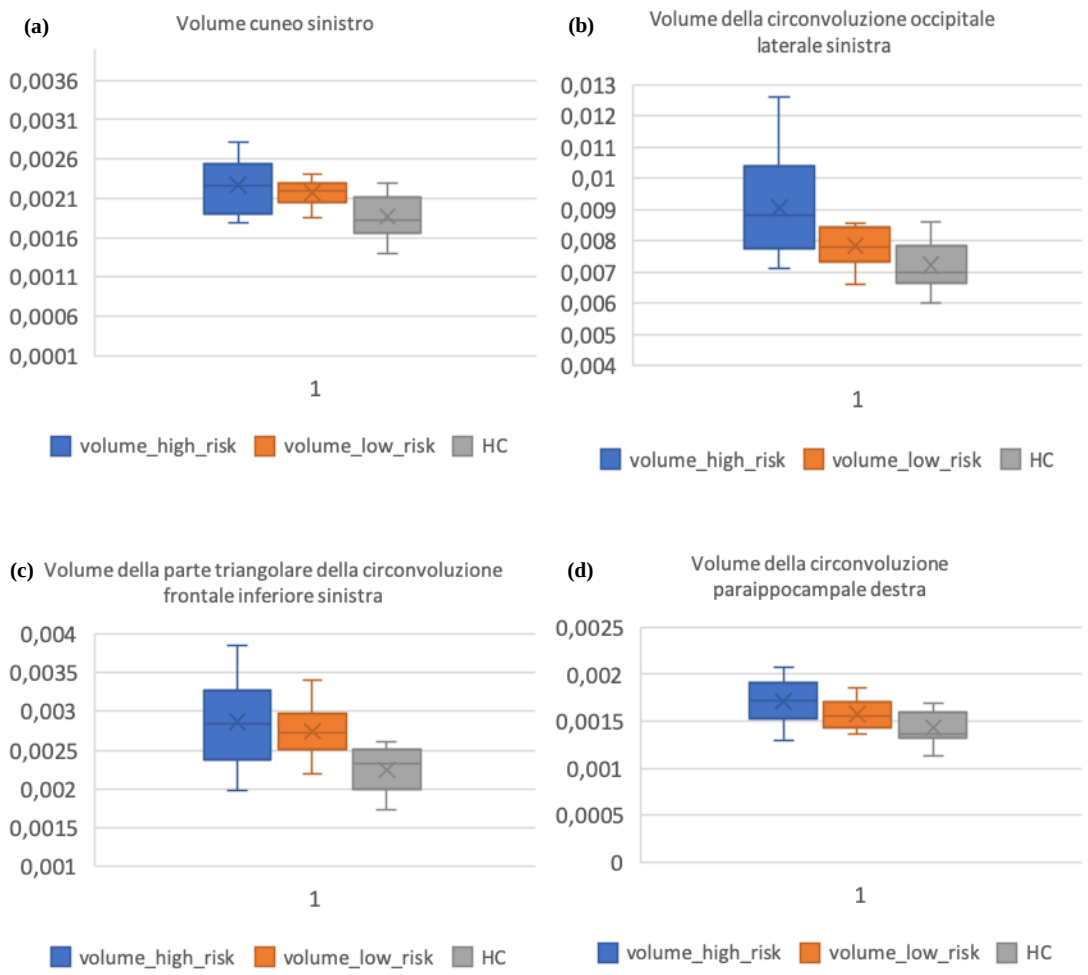

Figura 4.16: Confronto dei valori di *volume* nel (a) cuneo sinistro, (b) nella circonvoluzione occipitale laterale sinistra, (c) nella parte triangolare della circonvoluzione frontale inferiore sinistra, (d) nella circonvoluzione paraippocampale destra, nei gemelli ad alto rischio (in blue), nei gemelli a basso rischio (in arancione) e nei gemelli di controllo (in grigio).

Nella Figura 4.17 sono riportati i valori di volume nei tre gruppi nel solco superiore temporale sinistro. L'andamento dei valori mostra i gemelli ad alto rischio con valori di volume significativamente maggiori rispetto ai valori nei gemelli di controllo e significativamente maggiori rispetto ai valori nei gemelli a basso rischio, mentre i valori dei gemelli a basso rischio non sono significativamente diversi dai valori nei gemelli di controllo.

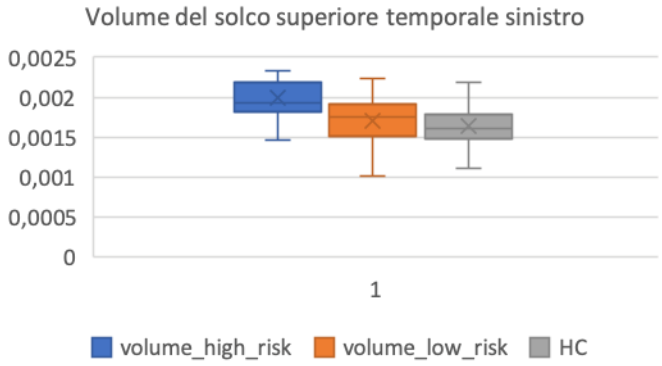

Figura 4.17: Valori di *volume* nel solco superiore temporale sinistro nei gemelli ad alto rischio (in blue), nei gemelli a basso rischio (in arancione) e nei gemelli di controllo (in grigio).

5.1 Discussione dei risultati

Secondo il razionale di studio, nei gemelli ad alto rischio eventuali alterazioni neuroanatomiche precoci dovrebbero far sì che le misure morfologiche associate differiscano dai valori assunti sia dai corrispettivi fratelli gemelli (non a rischio secondo la scala ERIraos-CL) che dai gemelli di controllo. Al contrario, si ipotizza che i valori dei parametri nei fratelli dei gemelli ad alto rischio, considerati a basso rischio, differiscano in misura minore o non differiscano dai valori nei gemelli di controllo.

Tuttavia, la psicopatologia è una condizione legata sia a fattori genetici che a fattori ambientali quindi è ragionevole ipotizzare che i fratelli dei gemelli, che risultano ad alto rischio per la scala ERI-raos-CL, abbiano una probabilità di sviluppare psicosi più alta rispetto alla popolazione normale media, e siano definiti perciò a basso rischio. Infatti, questi ultimi condividono con i rispettivi gemelli ad alto rischio sia il patrimonio genetico al 100% (MZ) o circa al 50% (DZ) che l'ambiente domestico, talvolta scolastico, ecc.

In linea con quanto ipotizzato, è stato riscontrato un numero consistente di regioni celebrali in cui la presenza di rischio psicotico in famiglia conduce ad un'alterazione dei valori assunti dai parametri analizzati anche nei fratelli dei gemelli ad alto rischio. L'effetto del rischio psicotico è risultato evidente soprattutto nei valori di *curvatura media* e di *volume* nelle diverse regioni celebrali e, in entrambi i casi, i valori nei gemelli a rischio sono maggiori rispetto ai valori nei gemelli di controllo. In particolare, nei grafici ottenuti per il parametro volume, è emerso che i valori nei gemelli a basso rischio sono intermedi tra i valori nei

gemelli ad alto rischio e i valori nei gemelli di controllo e in alcune regioni, i valori nei gemelli ad alto rischio si differenziano significativamente dai valori nei gemelli di controllo mentre i valori nei gemelli a basso rischio non si differenziano significativamente dai valori nei gemelli di controllo. Tale relazione permette di ipotizzare la presenza di influenze genetiche e ambientali condivise sulla misura morfologica e, indirettamente, sulla manifestazione del rischio. Infatti, il fatto che i valori dei gemelli a basso rischio non siano discostati significativamente dai valori nei gemelli di controllo ma si posizionino in scala verso i gemelli ad alto rischio suggerisce la condivisione di parte dei fattori predisponenti al rischio psicotico. In altre parole, il gemello ad alto rischio potrebbe essere all'inizio della fase prodromica della psicopatologia a causa di una complessa interazione di fattori, sia comuni che unici.

La presenza di stime di ereditabilità non significative nelle misure regionali di curvatura media permette di supporre che le alterazioni associate a basso e alto rischio siano legate in larga parte a fattori ambientali. Viceversa, la presenza di stime di ereditabilità significative nelle misure volumetriche permette di supporre che le alterazioni associate a basso e alto rischio siano legate in larga parte a fattori di natura genetica.

Trattandosi di un'analisi preliminare basata su un campione ridotto, è stato verificato che i risultati ottenuti siano in linea con le evidenze emerse in letteratura. Dalla ricerca bibliografica volta a identificare le regioni comunemente coinvolte nelle psicosi si evince che:

-la corteccia cingolata presenta disfunzioni e anomalie in presenza di disturbi psicotici, quali schizofrenia, depressione, disordine ossessivo compulsivo, autismo ecc **[20] [21].**

- le regioni mediali e laterali del lobo temporale presentano anormalità nei soggetti ad alto rischio psicotico. La disfunzione nel lobo temporale mediale, in particolare nell'ippocampo, sembra associata alla disregolazione del glutammato e della dopamina attraverso la rete ippocampale-striatale-mesencefalica che può portare ad alterazioni nell'attività neuronale che stanno alla base delle allucinazioni. Similmente, la disfunzione nel lobo temporale laterale può essere collegata al discorso disorganizzato e ad impedimenti nel linguaggio che sono osservabili nei soggetti ad alto rischio psicotico psicotico **[22]**.

- dalle immagini RM sono stati osservati cambiamenti nel volume di materia grigia e materia bianca del lobo frontale in soggetti affetti da rischio psicotico **[23]**.

- cambiamenti nel volume del lobo occipitale sono stati associati alla schizofrenia **[24]**.

6. Analisi dei modelli di classificazione ottenuti con NeuroMiner

L'applicazione dell'algoritmo Linear-SVM, implementato nel software *NeuroMiner*, alle mappe di differenza intra-coppia di materia bianca ha permesso di identificare l'equazione dell'iperpiano ottimo di separazione tra i due gruppi di zigosità con un'accuratezza del 78.2%. L'interfaccia grafica dei risultati ottenuti è riportata nella Figura 4.18.

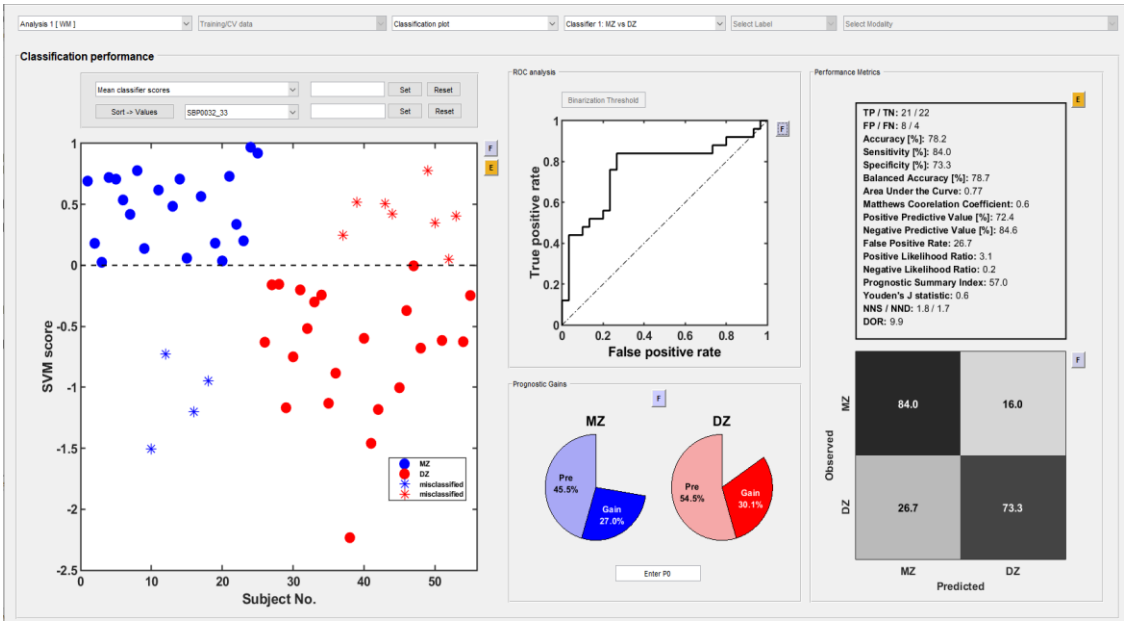

Figura 4.18: Interfaccia grafica dei risultati, ottenuti con le immagini di materia bianca, del software *NeuroMiner*.

Nel primo grafico a sinistra della figura 4.18, i punti in rosso rappresentano il posizionamento nello spazio delle *features* delle *immagini delta* di materia bianca delle coppie DZ predette correttamente dal classificatore, mentre gli asterischi in rosso indicano le *immagini delta* di materia bianca delle coppie DZ che sono state predette come appartenenti alla classe MZ. I punti in blu rappresentano le *immagini delta* di materia bianca delle coppie MZ classificate correttamente, mente gli asterischi in blu rappresentano le

immagini differenza di materia bianca delle coppie MZ classificate erroneamente come appartenenti alla classe DZ. La retta tratteggiata rappresenta l'iperpiano ottimo di separazione tra i due gruppi. Procedendo verso destra nella figura, è rappresentata la curva ROC la quale presenta sull'asse delle ascisse la percentuale di falsi positivi (ovvero le immagini delle coppie DZ attribuite erroneamente alla classe MZ) mentre sull'asse delle ordinate è rappresentata la percentuale di veri positivi (ovvero di immagini delle coppie MZ classificate correttamente). Come descritto nel paragrafo 5.3.4.2, maggiore è l'area sotto la curva ROC, migliore è la performance del classificatore: si osserva un'area pari a 0.77. Nella Figura 4.18, in alto a destra sono rappresentati gli indici che descrivono la metrica del modello di classificazione, mentre in basso a destra è raffigurata la tabella di contingenza. Inoltre, in basso nella parte centrale della schermata, sono rappresentati i grafici a torta della probabilità pre-test di ottenere una corretta classificazione basata sulla frequenza della classe target in ciascun gruppo.

Nella Figura 4.19 sono riportati i risultati del modello di classificazione ottenuti utilizzando le immagini di differenza intra-coppia di materia grigia.

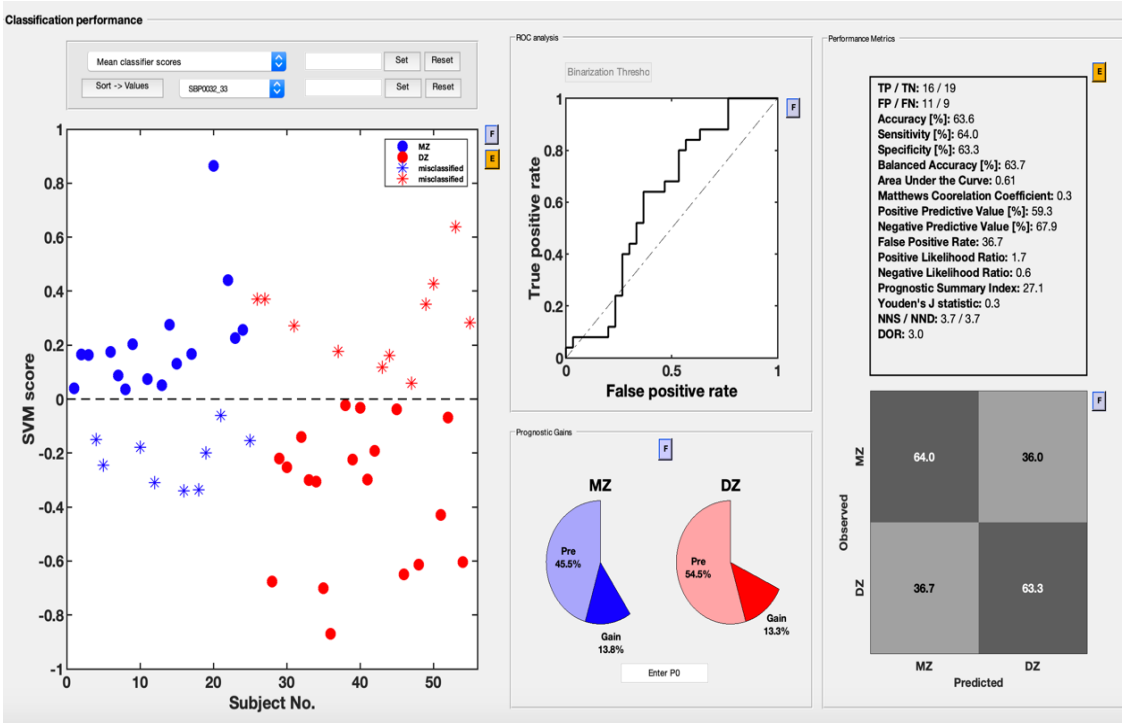

Figura 4.19: Interfaccia grafica dei risultati, ottenuti con le immagini di materia grigia, del software *NeuroMiner*.

Nella Figura 4.19 è evidente una performance del classificatore globalmente inferiore rispetto a quella ottenuta con le immagini di materia bianca: sono presenti molti errori di classificazione e l'andamento della curva ROC nelle prime osservazioni è addirittura al di sotto della bisettrice. L'accuratezza del classificatore è del 63.6%.

6.1 Pesi dei voxel nelle immagini di materia bianca e nelle immagini di materia grigia

Successivamente alla creazione del modello di classificazione, è stata implementata in *NeuroMiner* la procedura per il calcolo del peso avuto da ciascun voxel nella classificazione. L'output è un'immagine in cui il valore di ciascun voxel è stato ottenuto come prodotto tra il peso del voxel stesso e l'etichetta corrispondente alla classe in cui il voxel ha assunto il valore più alto (in cui la differenza punto a punto era maggiore). Nella Figura 4.20 è riportata l'immagine in uscita dal software in cui sono rappresentati i pesi di ciascun voxel delle immagini differenza di materia grigia e di materia bianca.

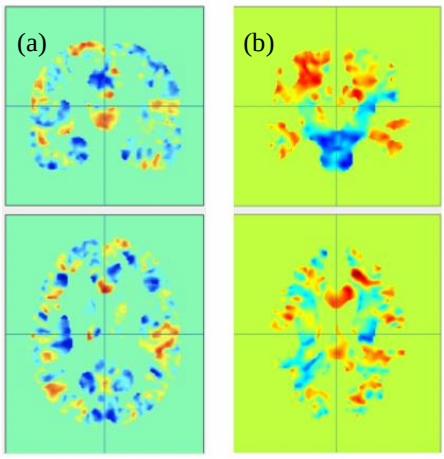

Figura 4.20: Immagine in uscita dal software *NeuroMiner* contenete il peso dei voxel nelle (a) immagini di materia grigia e (b) immagini di materia bianca.

Al termine di ognuna delle due analisi, al fine di interpretare al meglio i risultati, l'immagine contenente i pesi di tutti i voxel è stata estratta e sovrapposta ad un atlante neuroanatomico (*ch2better.nii*). Inoltre, sono stati impostati i range di visualizzazione [-3,-2] e [2,3], i quali permettono la visualizzazione delle sole aree celebrali in cui il peso è statisticamente significativo sulla base del *p-value* settato a 0.05. In linea con le diverse misure di

accuratezza dei modelli di materia bianca e grigia, non sono emersi voxel il cui peso è significativo nelle immagini di materia grigia (*p-value<0.05*), mentre nelle immagini di materia bianca sono emerse significatività nei voxel riportati nella Figura 4.21.

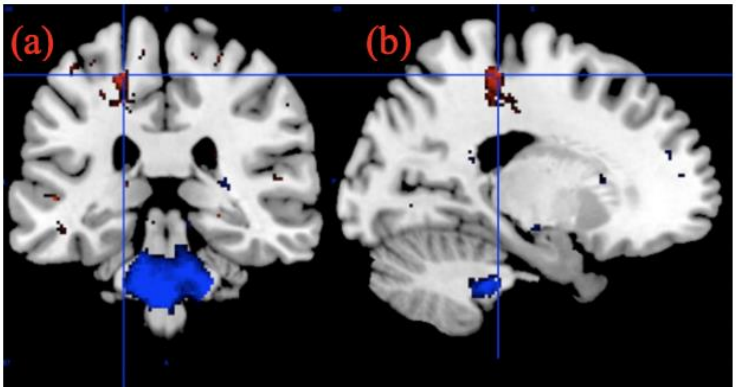

Figura 4.21: Rappresentazione nel piano coronale (a) e sagittale (b) dei pesi significativi (*p-value<0.05)* dei voxel dell'immagine di materia bianca sull'atlante *ch2better.nii.* In fuxia sono rappresentati i voxel in cui la quantità di materia bianca è più simile nelle coppie DZ. In blue sono rappresentati i voxel in cui la quantità di materia bianca è più simile nelle coppie MZ.

Successivamente, l'immagine delle regioni di interesse, associate a pesi significativi, è stata sovrapposta all'atlante AAL (*Automated Anatomical Labeling*) per la loro identificazione neuroanatomica.

È emerso che i gemelli DZ sono risultati più simili in termini di volume locale di materia bianca sottocorticale nelle regioni riportate nella Tabella 4.25.

ROI
Corteccia post-centrale di entrambi gli emisferi
Corteccia precentrale sinistra
Circonvoluzione temporale media di entrambi gli emisferi
Circonvoluzione temporale superiore di entrambi gli emisferi
Ippocampo

Tabella 4.25: Regioni celebrali in cui è emersa una similarità maggiore, in termini di materia bianca, nelle coppie DZ.

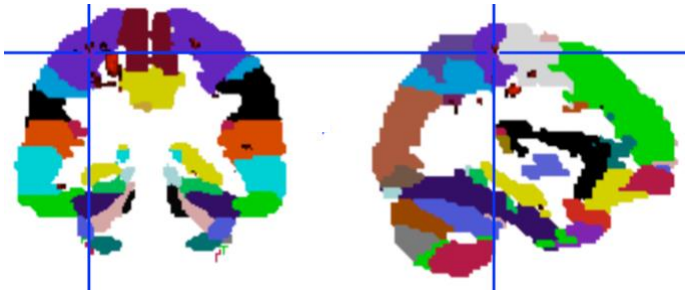

Figura 4.22: Sovrapposizione dei voxel che assumono valori significativamente diversi tra le coppie MZ e le coppie DZ, all'atlante AAL. In rosso sono rappresentati i voxel in cui la differenza intra-coppia, nel valore di materia bianca, è minore nelle coppie DZ rispetto alle coppie MZ. In viola è rappresentata la regione appartenente alla circonvoluzione postcentrale.

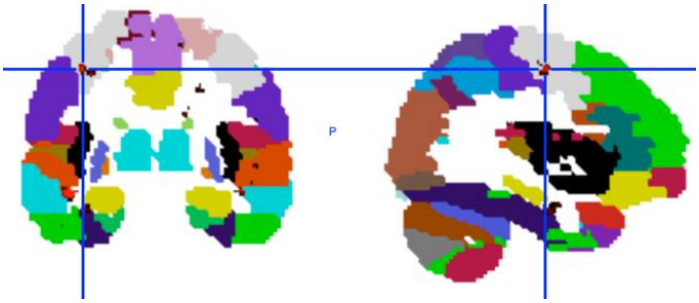

Figura 4.23: Sovrapposizione dei voxel che assumono valori significativamente diversi tra le coppie MZ e le coppie DZ, all'atlante AAL. In rosso sono rappresentati i voxel in cui la differenza intra-coppia, nel valore di materia bianca, è minore nelle coppie DZ rispetto alle coppie MZ. In grigio è rappresentata la regione appartenente alla circonvoluzione precentrale sinistra.

In linea con le aspettative, sono emersi più cluster cerebrali associati a una maggiore somiglianza in termini di materia bianca nei gemelli MZ rispetto ai DZ. In particolare, i gemelli MZ presentano una quantità più simile di materia bianca sottostante le regioni riportate nella Tabella 4.26.

ROI
Parte triangolare della circonvoluzione frontale inferiore destra
Corteccia linguale destra
Circonvoluzione occipitale media di entrambi gli emisferi
Circonvoluzione temporale superiore di entrambi gli emisferi
Solco temporale superiore sinistro
Circonvoluzione temporale inferiore sinistra
Parte orbitale della circonvoluzione frontale inferiore sinistra
Regione calcarina dell'emisfero sinistro
Precuneo destro
Corteccia cingolata anteriore destra
Corteccia sopramarginale emisfero destro
Putamen in entrambi gli emisferi
Pallidum
Regione paraippocampale destra
Cervelletto

Tabella 4.26: Regioni celebrali in cui è emersa una similarità maggiore, in termini di materia bianca, nelle coppie MZ.

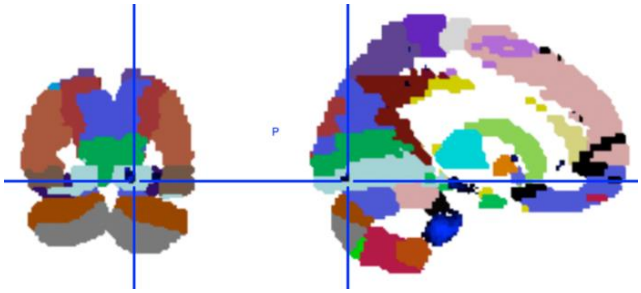

Figura 4.24: Sovrapposizione dei voxel che assumono valori significativamente diversi tra le coppie MZ e le coppie DZ, all'atlante AAL. In rosso sono rappresentati i voxel in cui la differenza intra-coppia, nel valore di materia bianca, è minore nelle coppie MZ rispetto alle coppie DZ. In azzurro è rappresentata la regione appartenente alla corteccia linguale destra.

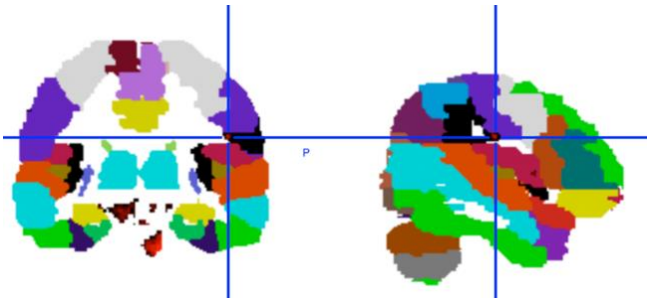

Figura 4.25: Sovrapposizione dei voxel che assumono valori significativamente diversi tra le coppie MZ e le coppie DZ, all'atlante AAL. In rosso sono rappresentati i voxel in cui la differenza intra-coppia, nel valore di materia bianca, è minore nelle coppie MZ rispetto alle coppie DZ. In nero è rappresentata la regione appartenente alla corteccia sopramarginale emisfero destro.

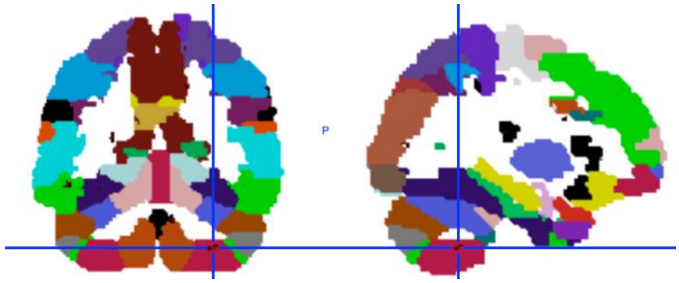

Figura 4.26: Sovrapposizione dei voxel che assumono valori significativamente diversi tra le coppie MZ e le coppie DZ, all'atlante AAL. In rosso sono rappresentati i voxel in cui la differenza intra-coppia, nel valore di materia bianca, è minore nelle coppie MZ rispetto alle coppie DZ. In fuxia è rappresentata la regione pozione di materia bianca appartenente al cervelletto nell'emisfero destro.

6.2 Predizione della zigosità di una nuova coppia

A titolo di esempio, è stata omessa dall'analisi di ciascun tessuto l'immagine differenza di una coppia scelta in maniera casuale dal campione. L'immagine apparteneva ad una coppia DZ ed è stata data in ingresso al modello in un secondo momento per effettuare una validazione esterna. L'immagine è risultata predetta in maniera corretta da entrambi i classificatori (materia bianca e materia grigia). Il risultato della predizione è stato visualizzato a partire dalla struttura *NM* creata dal software *NeuroMiner*.

Dopo aver caricato in Matlab la struttura NM, è stato inserito il comando '*NM.analysis{1}.OOCV{1}.BinResults{1}*'.

La schermata ottenuta è quella riportata nella Figura 4.27.

```
>> NM.analysis{1}.OOCV{1}.BinResults{1}

ans =

  struct with fields:

                 PerformanceMeasure: 'BAC'
   BinCV2Predictions_DecisionValues: {[1×625 double]}
          BinCV2Predictions_Targets: {[1×625 double]}
             MeanCV2PredictedValues: {[-0.5154]}
              StdCV2PredictedValues: {[0.3558]}
               CICV2PredictedValues: {[-1.3115 0.0767]}
                 BinProbPredictions: {[-1]}
            BinMajVoteProbabilities: {[0.0384 0.9616]}
```

Figura 4.27: Schermata ottenuta nella command window di Matlab dopo l'inserimento del comando per visualizzare i risultati della predizione della zigosità della nuova immagine. In corrispondenza di *BinProbPredictions* è possibile leggere la classe predetta. In questo caso si legge { -1} ovvero DZ.

Sebbene la validazione esterna condotta su una sola immagine non sia sufficiente a garantire l'efficacia del modello, questo risultato incoraggiante può essere il punto di partenza per successivi approfondimenti.

6.3 Discussione dei risultati

Dal confronto tra i risultati ottenuti utilizzando le immagini differenza di materia grigia e di materia bianca emerge un'accuratezza maggiore del classificatore che basa la predizione di zigosità sulle immagini differenza di materia bianca. Il classificatore è apparso in grado di individuare similarità/differenze nei valori di materia bianca e di attribuire a nuove immagini la corretta zigosità.

Dalla letteratura è noto che i gemelli MZ sono globalmente più simili dei gemelli DZ. Infatti, dalla sovrapposizione dei risultati ottenuti con l'atlante AAL è emerso un numero maggiore di regioni la cui quantità di materia bianca è più simile nelle coppie MZ, rispetto al numero di regioni in cui la quantità di materia bianca è più simile nelle coppie DZ. Inoltre, dall'analisi è emerso che il *volume di materia bianca presente nel cervelletto* è più simile nei gemelli MZ rispetto ai gemelli DZ. Questo risultato, era emerso anche dall'analisi di regressione lineare multipla effettuata a livello regionale.

Ripetendo l'analisi più volte in fase di ottimizzazione dei parametri del modello è stato osservato che, a causa della ridotta numerosità del campione, l'accuratezza del classificatore oscilla sia con le immagini di materia bianca che con le immagini di materia grigia. Questo rappresenta una limitazione nell'utilizzo degli algoritmi di classificazione implementati nel software *NeuroMiner* quando si ha a disposizione un campione poco numeroso.

Capitolo 5

Conclusioni e sviluppi futuri

Il lavoro sperimentale che è stato svolto nella presente tesi ha avuto tre scopi principali: (i) studiare e confrontare la morfologia celebrale di coppie di gemelli in età evolutiva al fine di individuare le caratteristiche strutturali ereditabili, (ii) individuare in che percentuali l'ambiente condiviso tra i gemelli e l'ambiente unico influenzino la varianza dei valori dei parametri morfologici nelle regioni celebrali, (iii) definire le regioni del cervello caratterizzate da alterazioni dei valori anatomici a causa del rischio di manifestare disturbi psicotici e fare ipotesi circa l'influenza di fattori genetici e ambientali. Il raggiungimento di (i) e (ii) ha previsto l'utilizzo di sofisticati strumenti di analisi di neuroimmagini e statistiche, di cui alcuni sviluppati specificatamente per lo studio di popolazioni gemellari.

Il *volume*, in linea con la letteratura, è risultato essere significativamente ereditabile in un numero elevato di regioni corticali (pari a 19), pertanto si può ipotizzare che sia il parametro morfologico più largamente influenzato da fattori genetici. Il punteggio più alto di ereditabilità è stato ottenuto nella *circonvoluzione temporale trasversa* in cui la totalità della varianza inter-individuale del volume è spiegata da differenze di tipo genetico. Si tratta della prima struttura corticale deputata all'elaborazione delle informazioni uditive che vengono percepite. Viceversa, l'ereditabilità della *curvatura media* è risultata significativa solo in tre regioni corticali, quindi è presumibile che la curvatura media delle diverse regioni del cervello sia meno legata a fattori genetici e più legata a fattori ambientali. Inoltre, nella *circonvoluzione temporale media dell'emisfero sinistro* dal confronto tra i coefficienti di correlazione nei MZ e nei DZ emerge una correlazione maggiore dei valori di curvatura media nelle coppie DZ che fa presupporre l'assenza di un impatto genetico sulla curvatura media della circonvoluzione e la presenza di un impatto ambientale predominante.

È stata riscontrata una maggiore correlazione dei parametri *area superficiale*, *spessore corticale* e *volume* nelle regioni corticali dell'emisfero destro in entrambi i gruppi di zigosità rispetto all'emisfero sinistro. Invece, per la *curvatura media* è emersa nelle coppie MZ una maggiore correlazione nelle aree celebrali dell'emisfero destro rispetto all'emisfero sinistro,

mentre nelle coppie DZ è emersa una maggiore correlazione nelle aree celebrali dell'emisfero sinistro rispetto all'emisfero destro.

Dal modello ACE è emerso che i valori di volume di diverse regioni sottocorticali risentono in larga parte dell'effetto dell'ambiente condiviso tra i cogemelli. In particolare, in ciascuna coppia il volume del cervello (ovvero il volume del cervello che rimane dopo aver rimosso il tronco encefalico, i ventricoli, il liquido celebrospinale e il plesso coroideo) sembra dipendere per una percentuale di circa il 64.5% dall'ambiente condiviso, per il 30.5% da fattori ambientali unici e solo per il 5% da fattori di natura genetica. Pertanto, tale caratteristica morfologica sembra esser spiegata in misura maggiore da fattori di natura ambientale condivisa. Quindi, poiché nei *volumi corticali* emerge un contributo rilevante dei fattori genetici, tale risultato fa supporre che ci sia anche un contributo rilevante dell'ambiente condiviso, che potrebbe essere concentrato nei *volumi sottocorticali*.

Dall'approccio regionale alternativo al metodo classico, basato su regressione lineare multipla, è emerso che il *volume regionale* della *corteccia cerebellare* è più simile nei gemelli MZ rispetto ai gemelli DZ.

Nell'ambito delle analisi *voxel-based* implementate nel software di Machine Learning NeuroMiner, dal confronto tra i risultati ottenuti utilizzando le immagini differenza di materia grigia e di materia bianca emerge un'accuratezza maggiore del classificatore che basa la predizione di zigosità sulle immagini differenza di materia bianca. Il classificatore è apparso in grado di individuare similarità / differenze nei valori di materia bianca e di attribuire a nuove immagini la corretta zigosità. Questa complementare analisi di classificazione di tipo *voxel-based* ha suggerito la presenza di influenze genetiche significative nei *volumi locali* di materia bianca, sulla base dei quali è stato possibile predire la zigosità di una nuova coppia. In particolare, il *volume* di *materia bianca* nella *corteccia cerebellare* è risultato più simile nei gemelli MZ rispetto ai gemelli DZ e determinante nella predizione. Tale risultato è supportato a livello regionale infatti, dall'analisi basata sui coefficienti di correlazione, è emersa un'ereditabilità del 72.5% del volume di materia bianca del cervelletto. Questi 2 approcci analitici, pur essendo parzialmente complementari, hanno condotto a risultati in linea, evidenziando come il volume di materia bianca nella corteccia cerebellare sia definito dal patrimonio genetico più che da fattori ambientali.

L’approccio con il modello di regressione lineare multipla ha anche evidenziato una maggiore similarità dei valori di *area superficiale* nella *circonvoluzione temporale media dell’emisfero sinistro* per le coppie MZ rispetto alle coppie DZ, oltre ad evidenziare una maggiore similarità dello stesso parametro per le coppie di sesso femminile. In aggiunta, mediante l’approccio basato sulla correlazione, in questa stessa regione è emersa un’ereditabilità dell’area superficiale del 79.9%. Anche in questo caso, i due approcci alternativi hanno condotto allo stesso risultato, secondo cui l’area superficiale della circonvoluzione temporale media dipende in misura maggiore da fattori genetici rispetto ai fattori ambientali. L'esatta funzione di questa struttura anatomica non è nota, ma si ritiene sia collegata a processi diversi come il contemplare a distanza, il riconoscimento dei volti noti, e il riconoscimento del significato delle parole durante la lettura.

Inoltre, l’analisi con modello lineare generalizzato ha messo in luce un effetto importante dell’età sulla similarità intra-coppia. In particolare, indipendentemente dalla zigosità della coppia, con la crescita si nota un aumento di similarità nei valori di

- *area superficiale* e di *curvatura media* del *lobulo parietale superiore dell’emisfero destro* (regione celebrale che contribuisce all'orientamento spaziale e all'integrazione delle informazioni visuo-spaziali)
- *area superficiale* della *circonvoluzione precentrale dell’emisfero sinistro* (la corteccia premotoria e l'area motoria supplementare occupano la circonvoluzione precentrale e svolgono un’azione di supporto nei confronti della corteccia motoria primaria, quindi sono implicate nei movimenti volontari)
- *curvatura media* del *polo temporale destro* (coinvolto nel linguaggio)
- *spessore* della *circonvoluzione pericalcarina dell’emisfero sinistro* (implicata nei processi visivi)
- *volume* del *chiasma ottico*

mentre si osserva una riduzione della similarità dei valori di

- *area superficiale* della *circonvoluzione latero-occipitale sinistra* (coinvolta nei processi visivi)
- *curvatura media* del *polo frontale destro* (coinvolto nei movimenti volontari)

- *curvatura media* della *corteccia cingolata anteriore sinistra* (regione in cui vengono elaborati, a livello inconscio, i pericoli ed i problemi cui un individuo è soggetto nel normale decorrere delle proprie esperienze).

Questi risultati potrebbero essere utili nell'ambito dello studio della morfologia celebrale in età evolutiva, in particolare nella comprensione dei fattori ambientali che possono influenzare l'espressione genetica nel corso della crescita.

Un altro ramo d'indagine preliminare di questo lavoro di tesi ha riguardato l'identificazione dei marker cerebrali di rischio psicotico. Grazie a un confronto preliminare tra soggetti a rischio, rispettivi gemelli e soggetti non a rischio è stato possibile individuare parametri morfologici in alcune regioni del cervello che risultano alterati nei soggetti a rischio di manifestare psicosi. Le analisi statistiche hanno permesso di individuare la *curvatura media* ed il *volume* come parametri neuroanatomici maggiormente affetti da alterazioni che anticipano la manifestazione dei sintomi: i risultati suggeriscono una variazione uniforme di questi parametri, che nei soggetti a rischio assumono valori maggiori rispetto ai relativi gemelli e ai gemelli di controllo. Sono stati riscontrati valori significativamente più elevati di *volume* e *curvatura* in presenza di sintomi pre-psicotici in diverse regioni (tra cui la corteccia cingolata, circonvoluzione occipitale e alcune regioni frontali) che sono precedentemente emerse in letteratura come coinvolte in disturbi di questo tipo. Le anomalie nei valori sono evidenti sia nei gemelli che sono definiti a rischio per la scala ERIraos-CL che nei fratelli di tali soggetti che costituiscono la coppia discordante. Tale risultato avvalora la tesi secondo cui lo sviluppo di una psicopatologia in età evolutiva è legato sia a fattori di natura genetica che a fattori di natura ambientale. Pertanto, dato che la curvatura media risulta essere poco ereditabile, è ragionevole ipotizzare che un'alterazione dei valori nei soggetti a rischio potrebbe essere legata a fattori di tipo ambientale, viceversa per il volume, che è risultato essere molto ereditabile, le anomalie nei valori dei soggetti a rischio potrebbero esser legate ad alterazioni a livello genetico.

L'individuazione, tramite analisi psicopatologiche e di neuroimaging in popolazioni gemellari, delle regioni e dei parametri morfologici modificati in presenza di rischio psicotico permetterebbe di individuare precocemente i soggetti a maggior rischio e pianificare strategie terapeutiche preventive. Inoltre, la quantificazione delle influenze

genetiche e ambientali sul valore di uno specifico parametro morfologico permette di approfondire, in caso di alterazione in presenza di patologia, i fattori genetico-ambientali specifici e di sviluppare misure preventive adeguate. Le suddette analisi sono state condotte su un campione poco numeroso (13 coppie di gemelli discordanti per il rischio psicotico), di conseguenza la potenza statistica dei risultati ottenuti è anch'essa ridotta. Uno sviluppo futuro prevede la ripetizione delle analisi sopradescritte su un campione più numeroso di dati.

Se si disponesse di un campione di gemelli discordanti per rischio psicotico abbastanza numeroso da poter applicare gli algoritmi di Machine Learning, un ulteriore sviluppo di questo lavoro di tesi potrebbe prevedere il calcolo dell'immagine differenza di ciascuna coppia di gemelli discordanti per il rischio psicotico e dell'immagine differenza di ciascuna coppia di gemelli concordanti per l'assenza di rischio psicotico. L'analisi condotta sulle immagini differenza dei due gruppi di zigosità potrebbe essere condotta in maniera analoga sulle immagini differenza dei due gruppi definiti rispettivamente in base alla concordanza e alla discordanza della presenza del rischio psicotico. Questo metodo permetterebbe di sviluppare un classificatore in grado di distinguere le immagini appartenenti a coppie discordanti per rischio psicotico dalle immagini di coppie concordanti per assenza/presenza di rischio psicotico, quindi capace di predire, a partire da un'immagine di risonanza magnetica, la presenza del rischio psicotico in nessuno, uno o entrambi i gemelli della coppia. Inoltre, un classificatore di questo tipo permetterebbe di individuare le regioni del cervello in cui le coppie concordanti per l'assenza di rischio sono più simili rispetto alle coppie discordanti e viceversa. Quindi, permetterebbe di capire in quali regioni celebrali e in che modo il rischio psicotico va ad inficiare sulla similarità intra-coppia. Se la coppia concordante risultasse più simile della coppia discordante in una determinata regione, questo potrebbe essere un indice dell'influenza del rischio psicotico di uno dei due soggetti sulla similarità intra-coppia, ovvero si potrebbero individuare le regioni in cui si evince una differenza nei valori tra il gemello a rischio e il gemello a basso rischio. Invece, se la coppia discordante fosse più simile della coppia di controllo in una determinata regione, questo potrebbe avvalorare la tesi secondo cui la presenza del rischio psicotico è in parte di natura genetica e si associa ad un'alterazione morfologica comune ai gemelli della coppia, indipendentemente dalla manifestazione dei sintomi.

In corso di analisi è stata notata una particolarità dell'immagine di differenza di volume tessutale tra coppie di gemelli, che mostrava una differenza puntuale quasi nulla in regioni che sembrerebbero rappresentare la vascolarizzazione celebrale. Se anche queste regioni non corrispondessero ai vasi sanguigni celebrali, sarebbe comunque interessante approfondire questo risultato preliminare per caratterizzare e localizzare maggiormente quelle regioni in cui è presente una corrispondenza così precisa nei due cervelli.

Il presente lavoro di tesi sperimentale è risultato essere innovativo oltre che unico nel suo genere in Italia. Inoltre, durante il percorso di ricerca sono state comprovate le potenzialità del *brain imaging* a risonanza magnetica, il quale si conferma uno strumento estremamente utile alla valutazione degli aspetti anatomici e strutturali del cervello, grazie alla scarsa invasività e all'elevata risoluzione di contrasto intrinseca. I software utilizzati per analisi di *brain imaging*, in particolare FreeSurfer e SPM 12, consentono un'analisi morfologica accurata e validata, oltre a rappresentare un metodo di analisi ampiamente utilizzato favorendo la comparabilità tra i diversi studi.

Bibliografia

[1] Zhi-Pei Liang, Paul C. Lauterbur. Principle of Magnetic Resonance Imaging: a signal processing perspective. 18 ott 1999.

[2] Lo studio dei gemelli- http://www.epiprev.it/materiali/2015/EP5-6/EP5-6_350_metodo.pdf

[3] Metodi statistici per lo studio dei gemelli. Corrado Fagnani, Sonia Brescianini, Emanuela Medda e Maria Antonietta Stazi Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute, Istituto Superiore di Sanità, Roma.

[4] Italian validation of the "Early Recognition Inventory for the retrospective assessment of the Onset of Schizophrenia Checklist": reliability, validity and instructions for use. A. Meneghelli, A. Alpi, M.T. Cascio, H. Häfner, K. Maurer, A. Preti1, A. Raballo, A. Cocchi.

[5] Intelligenza Artificiale e Big Data in ambito medico: prospettive, opportunità, criticità

Artificial Intelligence and Big Data in Medicine: scenarios, opportunities and critical issues. N. Musacchio, G. Guaita, A. Ozzello, M.A. Pellegrini, P. Ponzani, R. Zilich, A. De Micheli

[6] Maria Antonietta Stazi. I gemelli nella ricerca clinica: eredità ed ambiente, Giugno 2017.

[7] Breve Storia del Metodo Gemellare. F. Lorenzi, Facoltà di Scienze Matematiche Fisiche e Naturali, Università degli Studi di Roma Tor Vergata, Roma, Italia. Acta Genet Med Gemello l 46: 231-240 (1997)

[8] A. Carlo Altamura, Eleonora Maggioni, Taj Dhanoa, Valentina Ciappolino, Riccardo A. Paoli, Laura Cremaschi, Cecilia Prunas, Giulia Orsenigo, Elisabetta Caletti, Claudia M. Cinnante, Fabio M. Triulzi, Bernardo Dell'Osso, Lakshmi Yatham, Paolo Brambilla. The impact of psychosis on brain anatomy in bipolar disorder: a structural MRI study. *Journal of Affective Disorders*, Giugno 2018

[9] 3 Tesla T2*-weighted brain MRI distinguishes multiple sclerosis from incidental white matter microangiopathic lesions. N. Mistry, R. Abdel-Fahim, J. Dixon, C. Tench, P. Morris, N. Evangelou (Nottingham, GB)

[10] Jansen AG1, Mous SE, White T, Posthuma D, Polderman TJ. What twin studies tell us about the heritability of brain development, morphology, and function: a review. *Neuropsychol Rev*, 2015 Mar;25(1):27-46.

[11] Eleonora Maggioni, Letizia Squarcina, Nicola Dusi, Vaibhav A. Diwadkar, Paolo Brambilla. Twin MRI studies on genetic and environmental determinants of brain morphology and function in the early lifespan (under review)

[12] Jinghua Wang, Lili He, Hairong Zheng, Zhong-Lin Lu. Optimizing the Magnetization-Prepared Rapid Gradient-Echo (MP-RAGE) Sequence. Plos One, 30 May 2014.

[13] Twins methods quantitatively explore the genetic impact on children and adolescents brain gray matter volume. Xiao-Lu Chen1, Xiao-Wei Zhang3,8, Xiao Hou4, Xiao Li2, Xing-Shun Ma5, Xiao-Mei Hu6, Hua-Qing Meng2, Qian He7, Lian-Sheng Zhao8, Ying-Cheng Wang8, Yi-Xiao Fu2 & Tao Li8

[14] Soliani- Correlazione e covarianza- http://www.dsa.unipr.it/soliani/capu18.pdf

[15] Michael J. Zyphur, ZhenZhang, Adam P. Barsky, Wen-Dong Li. An ACE in the hole: Twin family models for applied behavioral genetics research

[16] A Twin Study of Early-Childhood Asthma in Puerto Ricans, Bunyavanich S, Silberg JL, Lasky-Su J, Gillespie NA, Lange NE, Canino G, et al. (2013) A Twin Study of Early-Childhood Asthma in Puerto Ricans. PLoS ONE 8(7): e68473. https://doi.org/10.1371/journal.pone.0068473

[17] The Oxford Handbook of Quantitative Methods, Vol. 2: Statistical Analysis. Todd D. Little. Chapter 10. Twin Studies and Behavior Genetics. Gabriella A. M. Blokland, Miriam A. Mosing, Karin J. H. Verweij, and Sarah E. Medland.

[18] Carlo Vercellis. Business intelligence. Modelli matematici e sistemi per le decisioni.

[19] John Ashburner VBM Tutorial, Marzo 2015. https://www.fil.ion.ucl.ac.uk/~john/misc/VBMclass15.pdf

[20] Bubb EJ, Metzler-Baddeley C, Aggleton JP. The cingulum bundle: Anatomy, function, and dysfunction. Neurosci Biobehav Rev., Sep 2018.

[21] uhammad Farid Abdul-Rahman, Anqi Qiu, Kang Sim. Regionally Specific White Matter Disruptions of Fornix and Cingulum in Schizophrenia. Plos One, April 14 2011.

[22] Emerging Temporal Lobe Dysfunction in People at Clinical High Risk for Psychosis

Paul Allen1,2,3*, Holly Moore4,5, Cheryl M. Corcoran2, James Gilleen1, Petya Kozhuharova1, Avi Reichenberg2 and Dolores Malaspina2

[23] 1Sindh Medical College, Dow University of Health Sciences, Karachi, Pakistan.

2Center for Mind & Brain, University of California, Davis, USA. Frontal lobe alterations in schizophrenia: a review

[24] Hassaan Tohid, MBBS, Muhammad Faizan, MBBS, and Uzma Faizan, MBBS. Alterations of the occipital lobe in schizophrenia