



*Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni*

Modelli e tecniche per l'analisi di dati radiomici

MOLIM - ONCOBRAIN LAB Metodi innovativi di imaging molecolare per lo studio di malattie oncologiche e neurodegenerative

Mara Sangiovanni, Nadia Brancati

Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio
Nazionale delle Ricerche (ICAR-CNR)

mara.sangiovanni@icar.cnr.it, nadia.brancati@icar.cnr.it

RT-ICAR-NA-2021-05

Dicembre 2021



Consiglio Nazionale delle Ricerche, Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR) – Sede di Cosenza, Via P. Bucci Cubo 8/9C, 87036 Rende, Italy, URL: www.icar.cnr.it – Sede di Napoli, Via P. Castellino 111, 80131 Napoli, URL: www.na.icar.cnr.it – Sede di Palermo, Via Ugo La Malfa 153, 90146 Palermo, URL: www.pa.icar.cnr.it



*Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni*

Modelli e tecniche per l'analisi di dati radiomici

MOLIM - ONCOBRAIN LAB Metodi innovativi di imaging molecolare per lo studio di malattie oncologiche e neurodegenerative

Mara Sangiovanni, Nadia Brancati

Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio
Nazionale delle Ricerche (ICAR-CNR)

mara.sangiovanni@icar.cnr.it, nadia.brancati@icar.cnr.it

RT-ICAR-NA-2021 -05

Dicembre 2021

I rapporti tecnici dell'ICAR-CNR sono pubblicati dall'Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche. Tali rapporti, approntati sotto l'esclusiva responsabilità degli autori, descrivono l'attività del personale e dei collaboratori dell'ICAR, in alcuni casi in un formato preliminare prima della pubblicazione definitiva in altra sede.

Modelli e tecniche per l'analisi di dati radiomici

MOLIM - ONCOBRAIN LAB Metodi innovativi di imaging molecolare per lo studio di malattie oncologiche e neurodegenerative

Mara Sangiovanni, Nadia Brancati

Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche
(ICAR-CNR)

Abstract

Il presente rapporto tecnico contiene la definizione dei modelli e delle tecniche per l'analisi di dati radiomici, utilizzati nell'ambito del progetto "MOLIM - ONCOBRAIN LAB: Metodi innovativi di imaging molecolare per lo studio di malattie oncologiche e neurodegenerative". A partire dai dati radiomici estratti dalle immagini mediche, sono stati definiti ed implementati metodi di Machine Learning allo scopo di classificare alcuni biomarker sia in ambito tumorale che in ambito neurodegenerativo, con particolare interesse rivolto al tumore al seno e al colon e ai morbi Alzheimer e Parkinson. Gli interessanti risultati ottenuti, nonostante l'esiguo numero di pazienti arruolati, aprono scenari a sviluppi futuri, confidando nella creazione di dataset di dati molto più ampi.

Keywords: radiomica, machine learning, classificazione, patologie tumorali, patologie neurodegenerative

Introduzione	6
Patologie tumorali	8
Tumore al seno	8
Tumore del colon retto	10
Modelli di apprendimento supervisionato	11
Tumore al seno	12
Tumore del colon retto	14
Metodologie di Machine Learning per patologie tumorali	16
Malattie Neurodegenerative	18
Morbo di Alzheimer	19
Morbo di Parkinson	21
Ulteriori tecniche di Feature selection e classificazione	22
Analisi dei dati del tumore al seno	23
Dati provenienti da dataset online	23
Dati di progetto	28
Marcatori molecolari	28
Definizione dei marcatori e loro valori	28
HUMAN EPIDERMAL GROWTH FACTOR RECEPTOR 2 - HER2	28
RECETTORI ESTROGENICI/PROGESTINICI - ER/PR	29
INDICE DI PROLIFERAZIONE - KI67	29
Classificazione molecolare	30
Decisioni su dati reali di progetto	30
HER2_score	30
ER_status_percent	31
PR_status_percent	31
KI67_level_percent	31
Classificazione molecolare	32
Casi particolari	32
Sottotipi istologici	33
Definizione dei sottotipi	33
Decisioni su dati reali di progetto	33
Classificazione dei tipi istologici	34
Casi particolari	35
Stadiazione del tumore	37
Definizione dei sottotipi	37
Decisioni su dati reali di progetto	37
Classificazione dei valori di stadiazione	37
Analisi dei dati delle malattie neurodegenerative	38
Morbo di Alzheimer	38
Dati clinici	39
DATI UNICZ	40
Dati da immagini mediche	40

Correlazione tra dati	40
Morbo di Parkinson	43
Dati clinici	45
DATI UNICZ	45
Dati da immagini mediche	45
Correlazione tra dati	46
Esperimenti preliminari sul dataset TCGA-BRCA	48
Marcatore PR	49
Marcatore ER	50
Istotipo	51
Stage	52
Discussione	54
Esperimenti di classificazione su dati reali di progetto	54
Approcci di selezione delle features e classificazione	54
Classificazione e validazione	56
Classificazione	56
Validazione	61
Implementazione e testing	61
Tumore al seno	61
Distribuzione dei dati	62
DATI SDN	62
DATI UNIME	63
Preprocessing	65
Risultati della classificazione e della validazione	65
Tumore al colon	67
Distribuzione dei dati	67
Preprocessing	68
Risultati della classificazione e della validazione	68
Morbo di Alzheimer	68
Distribuzione dei dati	69
DATI UNICZ FDG	69
DATI SDN AMY	70
Preprocessing	71
Risultati della classificazione e della validazione	71
Morbo di Parkinson	73
DATI UNICZ	73
Preprocessing	74
Risultati della classificazione e della validazione	74
Bibliografia	75

1. Introduzione

Il progetto “MOLIM - ONCOBRAIN LAB: Metodi innovativi di imaging molecolare per lo studio di malattie oncologiche e neurodegenerative” ha come obiettivo lo sviluppo di metodi innovativi per l’interpretazione delle immagini con il supporto dei sistemi bioinformatici. Viene proposto l’utilizzo di tecniche di radiomica che rappresenta un’innovativa modalità di analisi delle immagini diagnostiche, che garantisce una stima accurata della tipizzazione dei tessuti.

Per radiomica si intende l’analisi delle immagini mediche volta ad ottenere, tramite opportuni metodi matematici e l’uso dei computer, informazioni di tipo quantitativo non rilevabili direttamente tramite la semplice osservazione visiva da parte dell’operatore. Associata alla valutazione del radiologo risulta essere un potente strumento da impiegare in medicina insieme a dati epigenomici/genomici (radiogenomica), molecolari, clinici e psico-socio-ambientali, a scopo di supporto alla diagnosi e al monitoraggio dei pazienti durante il follow-up. La radiomica nasce quindi per sviluppare strumenti di supporto decisionale ed implica la combinazione di dati ricavati dall’imaging medicale con altre caratteristiche del paziente, quando disponibili, per aumentare la potenza dei modelli di supporto. I dati quantitativi estratti dalle immagini possono venire integrati in modelli predittivi multidisciplinari per la gestione del paziente, anche mediante l’utilizzo di metodiche di machine learning e di intelligenza artificiale, superando e arricchendo di fatto la quantità e la qualità dei dati associati ai singoli esami.

Il presente documento ha come obiettivo la descrizione di modelli e tecniche per l’estrazione di conoscenza da dati radiomici, mediante l’utilizzo di algoritmi di Machine Learning a partire dalle feature estratte dalle immagini mediche dei pazienti arruolati nell’ambito del progetto. Allo scopo di supportare il medico nelle sue diagnosi giornaliere, nel follow-up e nella scelta personalizzata delle terapie, viene effettuata l’integrazione dei descrittori numerici estratti dalle immagini con dati clinici/molecolari/morfologici, mediante l’utilizzo di algoritmi di Machine Learning. Il documento è così strutturato: dopo una breve introduzione sui modelli e le tecniche per il supporto alla diagnosi mediante dati radiomici, si entra nel dettaglio delle patologie oggetto di progetto. Viene successivamente effettuato uno studio sulle tecniche di apprendimento supervisionato, dopodichè inizia un’analisi dei dati, sia di quelli provenienti da dataset online sia dei dati reali di progetto. Infine, si passa a mostrare i risultati sugli esperimenti effettuati.

In Figura 1 è rappresentato uno schema generale di un tipico esempio di flusso di dati in ambito radiomico/radiogenomico. Il primo passo è quello della acquisizione delle immagini relative alla patologia oggetto di indagine, che nel contesto del progetto comprende due grandi tematiche: patologie tumorali e malattie neurodegenerative.

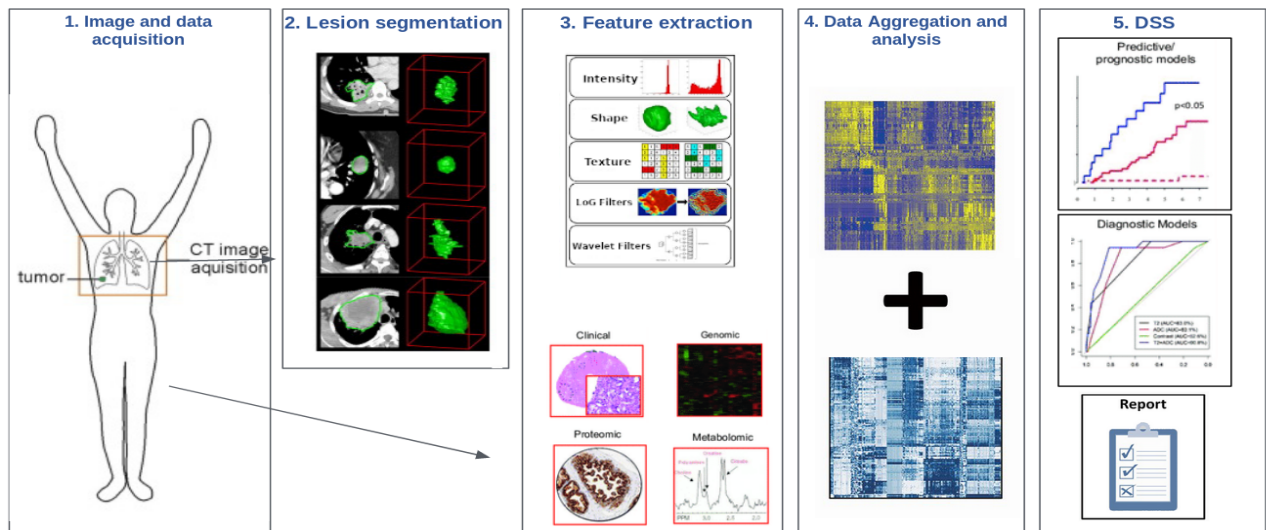


Figura 1: Flusso dati in ambito radiomico/radiogenomico

In ciascuna delle immagini acquisite si deve poi individuare la ROI (Region Of Interest), la regione di interesse più consona alla patologia trattata ed estrarla tramite tecniche di segmentazione (“*Lesion segmentation*” in Figura 1). Da queste regioni vengono poi estratti dei descrittori, detti *Features*, che catturano caratteristiche diverse delle immagini sia a livello di geometria, che a livello di distribuzione dei pixel, che a livello statistico (“*Feature extraction*” in Figura 1). Le feature radiomiche devono essere considerate, assieme a quelle provenienti dalla cartella clinica e da informazioni -omiche (ad esempio, per patologie tumorali: biomarkers di caratterizzazione dei tipi derivanti da espressione genica), come punto di partenza per la successiva estrazione di informazioni tramite metodologie di machine learning supervisionato. Obiettivo finale è la costruzione di modelli di apprendimento che consentano, a partire dai dati, la diagnosi di nuovi casi e/o la definizione di una eventuale stratificazione prognostica (punti 4 e 5 di Figura 1).

E’ stata effettuata una prima ricognizione della letteratura scientifica per individuare, per ciascuna delle aree di interesse, quali sono le modalità più affidabili per affrontare i vari aspetti su elencati ed iniziare a definire quali sono gli approcci possibili con i dati a disposizione. Si è considerata anche la possibilità di avvalersi di banche dati di libero accesso (ad esempio: TCIA) per iniziare a testare le metodologie individuate. Questi dati infatti potrebbero essere usati - laddove ci sia corrispondenza fra protocolli di acquisizione, metodologie di estrazione, feature estratte e dati clinici - per ampliare ulteriormente il dataset di input per la costruzione dei modelli di learning.

Le successive sezioni prenderanno in considerazione in maniera separata la letteratura individuata per ciascuna delle patologie di interesse. Si è inoltre deciso di separare le malattie oncologiche da quelle neurodegenerative.

2. Patologie tumorali

2.1. Tumore al seno

Le features in input saranno le seguenti:

- 14 features di forma (3D) – sono descrittori della dimensione e della forma tridimensionale e sono indipendenti dalla distribuzione dei livelli di grigio, e sono quindi calcolati solo sull'immagine originale;
- 18 features di primo ordine – sono descrittori della distribuzione dell'intensità dei voxel con metriche basilari e diffuse come ad esempio media e mediana;
- 68 features di secondo ordine o di texture – sono feature texturali calcolate su matrici che descrivono la distribuzione dei livelli di grigio e la loro rispettiva posizione. Sono estratte dalle matrici seguenti: Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Neighbouring Gray Tone Difference Matrix (NGTDM), Gray Level Dependence Matrix (GLDM).

Esse saranno estratte per tutte le tipologie di immagini a disposizione. Gli outcome attesi dal punto di vista medico comprendono: i) la predizione dell'istotipo tumorale (benigno vs maligno, se maligno di tipo Duttale, Lobulare...); ii) la predizione del fenotipo immunoistochimico (cioè ER, PgR, HER2, KI67); iii) la stratificazione prognostica (metastasi/sopravvivenza).

Esistono moltissimi approcci in letteratura che possono indicare la strada da perseguire come descritto ad esempio in [1], [2], e di seguito indichiamo quelli che sono sembrati i più interessanti. La maggior parte è basata su metodi di classificazione supervisionata delle features radiomiche/radiogenomiche e cliniche estratte. In alcuni casi, viene preventivamente estratto un sottoinsieme di features rilevanti da usare come discriminatori, come in Huang et al. [3], dove viene impiegato un passo di clustering non supervisionato per estrarre i descrittori più rilevanti dalle features radiomiche derivanti sia da PET che da RM, che vengono poi affiancate a quelle cliniche. Sul sottoinsieme di features estratte vengono poi testati diversi sistemi di classificazione con l'obiettivo di costruire un classificatore in grado di predire il tempo di sopravvivenza senza recidive, ed il grado tumorale. Diversi algoritmi sono stati testati, fra cui support vector machine (SVM), random forest (RF), e regressione logistica con termini di regolarizzazione in norma L1, L2 e di tipo ElasticNet. Tutte le features radiomiche sono state normalizzate tramite z-score prima del model training. Le capacità predittive vengono quantificate tramite una misura assai diffusa, e cioè l'AUC, Area Under receiver operator characteristic (ROC) Curve, ed i corrispondenti intervalli di confidenza, CI. I parametri del modello sono stati ottimizzati usando un approccio di cross validation su 10 ripetizioni. Il risultato migliore è ottenuto usando le features di texture derivanti da *Gray Level Co-occurrence Matrix* (GLCM) per il grado tumorale, ottenuto tramite regressione logistica con termine di regolarizzazione in norma L2 (AUC 0.76 (95% CI = [0.62, 0.83])). Per l'analisi di free survival invece, in cui il migliore classificatore è ottenuto tramite ElasticNet si ottiene invece una AUC di 0.75 (95% CI = [0.62, 0.88]) e di 0.68 (95% CI = [0.58, 0.81]) per le recidiva dopo 1 e 2 anni rispettivamente.

In [4], [5] Li et al. estraggono un totale di 38 descrittori matematici (chiamati CEIPs, Computer Extracted Phenotypes) dalle segmentazioni di immagini DCE-RM appartenenti al dataset online del Tumor Cancer Imaging Archive (TCIA [6]). Queste features includono dimensioni e forma della lesione tumorale, e caratteristiche di morfologia, texture, cinetica. Esse sono state descritte nel dettaglio del deliverable 2.1.1 “Analisi e definizione di modelli e tecniche per l'estrazione di dati radiomici attraverso l'utilizzo di imaging molecolare”.(Figura 2).

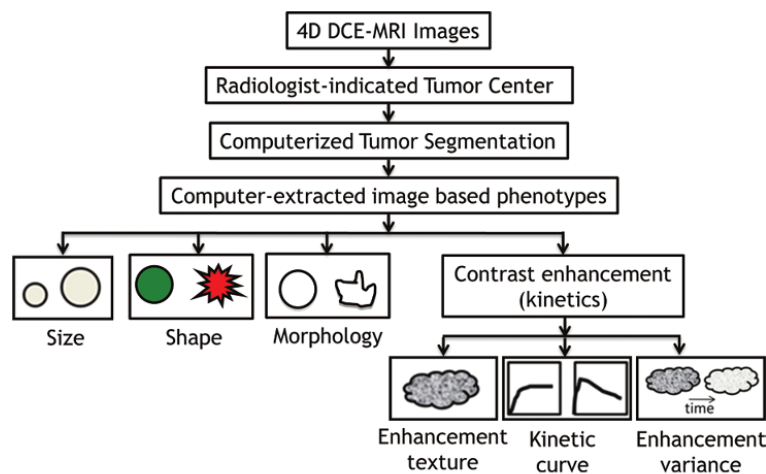


Figura 2: features estratte dal database TCIA

Le features estratte dalle immagini sono rese disponibili, inoltre, alla comunità scientifica e si candidano ad essere usate come test set per la scelta del modello. Alle features da immagine si affiancano inoltre le informazioni cliniche. Gli obiettivi della classificazione sono due: 1) la determinazione dello stato dei marker clinici dei recettori (ER/PR/HER2) e 2) la determinazione dei sottotipi di tumore a partire dallo stato dei marker molecolari immunohistochimici (normale, luminal A, luminal B, HER2-enriched, basal-like). La classificazione è stata fatta dividendo il problema generale in diversi problemi binari risolti tramite regressione, in cui la performance è stata determinata usando l'analisi della curva ROC tramite un modello binormale parametrico. L'approccio è stato di leave-one-out cross validation, con risultati che sono molto interessanti per la discriminazione del sottotipo molecolare, e migliorabili per la predizione della ricorrenza.

Un'altra direzione possibile è quella di estrarre dai dati tutte le feature radiomiche possibili e poi operare dei passi di feature selection prima e durante la classificazione come in [7], [8] dove viene fatta una prima selezione delle features che sono più rilevanti e poi si procede alla classificazione. In [7] la prima fase di selezione delle features è fatta tramite Feature ranking basato su Mann-Whitney U test, ed è seguita poi da un metodo di Embedded Feature Selection basato su Random Forest accoppiato alla classificazione tramite Support Vector Machines (SVMs). I risultati, valutati in termini di curve ROC, dipendono dal tipo di problema di classificazione e sono buoni per la distinzione Triple Negative vs NOT TN (NTN), e meno convincenti per la discriminazione degli altri tipi immunohistochimici. In [8] si

effettua prima un Feature Ranking ad ampio spettro tramite approccio ANOVA, e un passo successivo di raffinamento basato sul ranking ottenuto dalle features come risultato della classificazione tramite diversi metodi di regressione. Filtrate le features, si effettua poi la classificazione. Vengono presi in considerazione molti metodi diversi, fra cui Decision Trees, Linear Discriminant Analysis, SVM con differenti distanze, K-nearest neighbour classifiers (KNNs) in diversi setting di parametri, e cinque diversi metodi per l'Ensamble Learning fra cui AdaBoost. Analogamente al lavoro precedente, i risultati sono molto dipendenti dal tipo di problema di classificazione, e mentre per la discriminazione TN vs NTN si arriva ad accuratezze del 90%, non si ottengono risultati comparabili per gli altri tipi di classificazione immunistochemica.

In [9] dati genomici e radiogenomici provenienti dal dataset TCGA-BRCA sono stati utilizzati per la predizione di stadio tumorale, metastasi dei linfonodi, stato dei recettori degli estrogeni (ER), stato dei recettori del progesterone (PR) e lo stato del recettore del fattore di crescita epidermico umano 2 (HER2). Tutte le feature genomiche e radiomiche sono state standardizzate in modo da avere media zero e varianza 1, prima di passare alle analisi statistiche successive. Sono stati sperimentati due test statistici: il t-test e la regressione logistica mediante l'algoritmo LASSO. Per quanto riguarda il t-test, non sono state trovate feature genomiche che permettono di differenziare stadio e metastasi dei linfonodi, mentre molte feature genomiche permettono di predire stato ER e PR e infine una sola feature genomica, TP53 (ME) è significativamente associata allo stato HER2. Relativamente alla regressione logistica con LASSO, è stato dimostrato che le feature genomiche funzionano meglio delle radiomiche nella predizione di stato ER e PR. Per la predizione delle metastasi dei linfonodi e dello stato HER2, né le feature radiomiche, né quelle genomiche permettono di fare predizioni affidabili. Per quanto riguarda lo stadio tumorale, la feature radiomica più discriminante risulta essere il diametro effettivo, che è una misura della grandezza del tumore.

In [10] gli autori fanno utilizzo delle sole feature radiomiche per la predizione dello stadio tumorale e delle metastasi dei linfonodi. In particolare, per lo stadio tumorale 4 tipi di classificazione sono effettuati: a) stadio I versus stadio III, b) stadio II versus stadio III, c) stadio I versus stadio II, e d) stadio I+II versus stadio III. Nel caso della classificazione dei linfonodi, lo scopo è distinguere tra tumori con nodi negativi e quelli con uno o più nodi positivi. Per tutti i 5 tipi di classificazione è risultato che le feature più discriminanti sono la grandezza del tumore e l'area di superficie.

2.2. Tumore del colon retto

Le features di input sono estratte sulla scia di quelle del tumore al seno e sono accompagnate dalle informazioni cliniche. Gli obiettivi attesi includono: 1) determinare la stadiazione del tumore per supportare una eventuale terapia; 2) determinare la tipologia in termini di risposta/non risposta al trattamento farmacologico; 3) determinare la possibilità di successive metastasi ai linfonodi.

Anche nel caso di questa patologia esiste una letteratura piuttosto ampia, per una review si veda [11]. In [12] gli autori estraggono dalle immagini CT 150 features di texture e le usano per determinare la stadiazione del tumore. Usando un approccio di regressione LASSO viene fatta selezione delle features, e la predizione fatta tramite Mann-Whitney U-test. La performance viene valutata in termini di AUC. In [13] gli autori estendono il precedente lavoro usando le features estratte nel primo e l'informazione sulla stadiazione per predire la formazione di metastasi ai linfonodi. Gli approcci usati sono gli stessi del lavoro precedente. In [14] invece, allo scopo di predire la resistenza alla terapia farmacologica, vengono estratte dalle immagini di MRI multimodale più di 2400 features radiomiche. Il passo di riduzione del numero di features e loro selezione è basato su approccio di regressione LASSO, mentre il modello predittivo è stato costruito tramite regressione su analisi multivariabile. L'approccio è stato testato su una corte di pazienti completamente nuova rispetto ai dati di test. In [15] le features radiomiche estratte da immagini CT vengono usate per la predizione di markers genomici legati alla prognosi. Più di 300 features comprendenti anche informazioni da istogrammi e di texture vengono poi selezionate per la loro stabilità valutando gli intra-/inter-class correlation coefficients (ICCs), e per la loro rilevanza tramite analisi univariata. Il passo di feature selection viene fatto tramite algoritmo RELIEF e la parte di predizione tramite SVM. Le prestazioni sono valutate usando diverse metriche, tra cui AUC, come visibile in Figura 3.

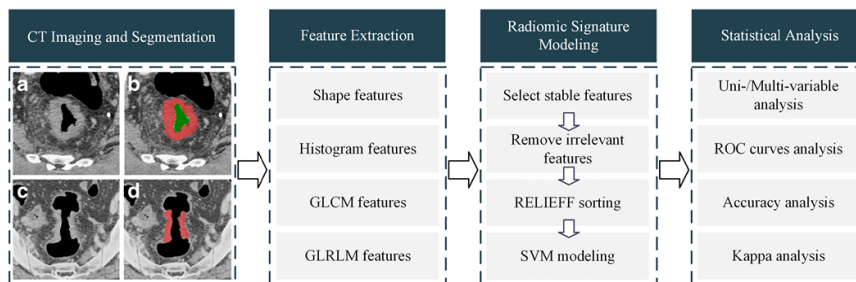


Figura 3: Estrazione di feature radiomiche da immagini CT e successiva elaborazione

L'ultimo lavoro qui considerato usa le feature radiomiche per predire l'instabilità dei microsattelliti (MSI) prima dell'intervento chirurgico. Dalle ROI del tumore annotate su immagini CT vengono estratte 254 feature che comprendono feature da grey-level co-occurrence matrices (GLCMs), run-length matrices (RLMs), local binary patterns (LBPs), fractal dimension (FD), intensity histogram (IH), angle co-occurrence matrices (ACMs). A queste vengono affiancate le features da informazioni cliniche. La fase di costruzione del modello viene effettuata tramite random forest (RF) e validata su una coorte distinta di pazienti. Le prestazioni sono misurate in termini di AUC, sensitivity e specificity.

3. Modelli di apprendimento supervisionato

Alla luce di quanto emerso, è possibile usare approcci molto differenti di apprendimento supervisionato per la creazione di modelli predittivi che sfruttino le features radiomiche. Non

emerge chiaramente una tecnica superiore alle altre, per cui si ritiene che questo punto vada affrontato tramite test su dati pubblici per poter iniziare a stabilire il sottoinsieme di metodologie che saranno poi effettivamente sviluppate. Emerge chiaramente però che l'applicazione di tecniche di Deep Learning è ancora molto insidiosa, in quanto per essere efficace necessita di enormi moli di dati [16]. A questo si aggiunge la difficoltà, intrinseca al metodo, di individuare eventuali markers che possano spiegare il fenotipo sottostante la predizione.

Dalla letteratura analizzata si evince che è molto importante affiancare ai dati radiomici anche i dati clinici o provenienti da biomarkers genomici. Andremo dunque a descrivere quali sono i dati non radiomici presenti in questo progetto ed utili agli obiettivi delineati.

3.1. Tumore al seno

Per il tumore al seno è prevista una ampia raccolta di dati non radiomici, che comprende sia dati clinici di base, sia markers di tipo istologico che biomarcatori genomici. In Tabella 1 sono riportate le informazioni per ognuno dei dati considerati. Per maggiori dettagli è possibile consultare il documento presente a [questo link](#) sul sito della Associazione Italiana di Oncologia Medica [17].

TIPO	NOME	DESCRIZIONE
DATI GENERALI	età	Età al momento della diagnosi
	genere	M/F
DATI ISTOLOGICI	presenza di biopsia	Si/No
	dimensioni del tumore	Dimensione della resezione
	presenza di linfonodi ascellari	Si/No
	grado istologico	Definisce la aggressività del tumore (G1/G2/G3)
	tipo istologico	Classificazione istologica del tumore
	presenza di invasione vascolare	Stabilisce se c'è infiltrazione nei vasi sanguigni (Si/No)
	multifocalità	Si/No
	presenza di linfociti tumorali	Si/No
	Livello di proliferazione tramite marker KI67	Misura quanto è aggressivo il tumore in base alla sua capacità di replicarsi.
	CLASSIFICAZIONE ISTOLOGICA	Classificazione istologica del tumore
CLASSIFICAZIONE MOLECOLARE	classificazione molecolare del tumore	Deriva dal risultato di specifici saggi sulla presenza di proteine espresse nel tumore. La classificazione

		comprende: Luminal A, LuminalB, Basal like, HER2-like.
RECETTORI ORMONALI	ER	Presenza o meno di recettori degli estrogeni. Può essere +/-
	PR	Presenza o meno di recettori dei progestinici Può essere +/-
	HER2	Presenza o meno di recettori del fattore di crescita dell'epidermide HER2.
	TN	Tripli negativi. Identifica la categoria di tumori che ha tutti I recettori precedenti negativi. (Si/No)
MARKERS DI ESPRESSIONE DI CITOCHERATINE	CK5/6	La presenza/assenza di citocheratine mioepiteliali è legata a diversi sottotipi tumorali. (Si/No).
	CK14	
	CK17	
MARKERS DI MUTAZIONI GENICHE	BRCA1	Mutazioni specifiche di questi geni sono associate ad una superiore probabilità di sviluppare tumori. Alcune mutazioni non sono patogeneiche, altre invece si. (Si/No)
	BRCA2	
	TP53	Mutazione associata ad alta instabilità genetica e superiore patogenicità. (Si/No)
INFORMAZIONI DIAGNOSTICHE / PROGNOSTICHE	CLASSIFICAZIONE TNM	T indica le dimensioni del tumore in una scala da 1 a 4; N indica se I linfonodi adiacenti presentano o meno infiltrazioni maligne (N0/N1),M indica se sono presenti metastasi o meno (M0/M1)
	stadio del tumore	Indica la situazione evolutiva del tumore. In ordine di gravità crescente si va dallo stadio 0 allo stadio IV, con sottostadi indicati dalle lettere A,B,C
	presenza di metastasi distanti	Si/No
	sito della metastasi	Descrizione del sito
	risposta al trattamento (followup)	Si/No

Tabella 1

I dati presentati nella Tabella 1 possono essere usati sia come input da affiancare alle features radiomiche che come output atteso in fase di classificazione. Si prenda ad esempio il problema della discriminazione fra pazienti sani e malati e, nel caso di malati, fra diversi sottotipi molecolari: in questo caso tutte le informazioni istologiche vengono usate come input; se invece il problema di classificazione è la predizione dei sottotipi

immunoistochimici, allora questi dati saranno usati come obiettivo atteso ed eliminati dagli input.

E' iniziata un'analisi di dataset di test poichè il TCIA mette a disposizione non solo le immagini, ma anche le features estratte da esse per un sottoinsieme curato di immagini. Oltre ad esse, ci sono i dati clinici che – pur non corrispondendo nella loro interezza – consentono di iniziare a definire le metodologie di costruzione dei classificatori per la diagnosi e/o predizione.

In particolare si considereranno i seguenti dati per cui si riporta la corrispondenza (a volte solo semantica) con i dati prodotti nel progetto:

- `age_at_initial_pathologic_diagnosis` → età
- `ajcc_tumor_stage_code` → stadio del tumore
- `breast_cancer_optical_measurement_histologic_type` → tipo istologico
- `breast_carcinoma_estrogen_receptor_status` → ER+/ER-
- `breast_carcinoma_progesterone_receptor_status` → PR+/PR-
- `her2_immunohistochemistry_level_result` → HER2+/HER2-
- `number_of_lymphnodes_positive_by_he` → presenza di linfonodi ascellari
- `PAM50_call` → classificazione molecolare

Si vuole qui sottolineare che: 1) pur non essendo ricchi come quelli prodotti dal progetto, questi dati consentono già di testare diversi approcci di classificazione; 2) che per quanto riguarda la classificazione molecolare, nel caso dei dati del TCIA non è ottenuta direttamente da saggi molecolari, ma ricavata a partire dai dati di espressione genica forniti dall'archivio del TCGA per i pazienti corrispondenti.

3.2. Tumore del colon retto

Anche per il tumore del colon retto è prevista una ampia raccolta di dati non radiomici, che comprende sia dati clinici di base, sia markers di tipo istologico che biomarcatori genomici. In Tabella 2 sono riportate le informazioni per ognuno dei dati considerati. Per maggiori informazioni si rimanda ancora una volta al già citato sito dell'AIOM [17] e al documento che si può trovare sullo stesso sito a [questo link](#).

TIPO	NOME	DESCRIZIONE
DATI GENERALI	età	Età al momento della diagnosi
	sex	M/F
DATI ISTOLOGICI	presenza di biopsia	Si/No
	dimensioni del tumore	Dimensione della resezione

	Sito del tumore	Ad esempio: colon, colon ascendente, colon sigma, retto...
	grado istologico	Definisce la aggressività del tumore (G1/G2/G3)
	tipo istologico	Classificazione istologica del tumore Es: Adenocarcinoma villosa, Adenoma tubulare...
	presenza di invasione vascolare	Stabilisce se c'è infiltrazione nei vasi sanguigni (Si/No)
	multifocalità	Si/No
	presenza di linfonodi tumorali	conteggio
CLASSIFICAZIONE E ISTOLOGICA	Classificazione istologica del tumore	Il nome della classe come da referto. Ad esempio:
CLASSIFICAZIONE E MOLECOLARE	classificazione molecolare del tumore	Deriva dal risultato di specifici saggi sulla presenza di proteine espresse nel tumore. La classificazione comprende: Luminal A, LuminalB, Basal like, HER2-like.
RECETTORI ORMONALI	ER	Presenza o meno di recettori degli estrogeni. Può essere +/-
	PR	Presenza o meno di recettori dei progestinici Può essere +/-
	HER2	Presenza o meno di recettori del fattore di crescita dell'epidermide HER2.
	TN	Tripli negativi. Identifica la categoria di tumori che ha tutti i recettori precedenti negativi. (Si/No)
MARKER GENOMICI DI PRESENZA DI MSI	Micro Satellite instability	La presenza/assenza determina la possibilità o meno di effettuare determinate terapie.
MARKERS DI MUTAZIONI GENICHE	KRAS/NRAS	Mutazioni specifiche di questi geni sono associate a sottotipi con prognosi diverse e alla resistenza ad alcune terapie con anticorpi monoclonali anti-EGFR
	BRAF	La mutazione in questo gene è associata a fenotipi diversi a seconda dei codoni coinvolti (V600/non-V600)
	TP53	Mutazione associata ad alta instabilità genetica e superiore patogenicità. (Si/No)
INFORMAZIONI DIAGNOSTICHE / PROGNOSTICHE	CLASSIFICAZIONE TNM-UICC 2009	Sono presi in considerazione caratteristiche diverse che includono le dimensioni del tumore T, la presenza di metastasi nei linfonodi N, e le metastasi a distanza M. Per ognuna delle caratteristiche esiste una suddivisione ulteriore in classi. Ad esempio: T0-T4 e successiva ulteriore classificazione usando le lettere a,b,c.

stadio del tumore	Indica la situazione evolutiva del tumore. In ordine di gravità crescente si va dallo stadio 0 allo stadio IV, con sottostadi indicati dalle lettere A,B,C.
presenza di metastasi distanti	Si/No
sito della metastasi	Descrizione del sito
risposta al trattamento (followup)	Si/No

Tabella 2

Non sarà possibile fare il test perché al momento, pur esistendo dataset che consentono di usare le immagini, non comprendono features radiomiche già estratte e disponibili. Per cui questa fase sarà rimandata alla effettiva disponibilità dei dati di progetto.

3.3. Metodologie di Machine Learning per patologie tumorali

Come è possibile dedurre dalle due tabelle, i dati in input per le due patologie sono estremamente simili, così come sono simili gli obiettivi di output. Si descriveranno brevemente qui quali sono gli algoritmi di machine learning che, alla luce di quanto estratto dall'indagine sulla letteratura esistente, si candidano ad essere i più adatti alla problematica in questione.

In particolare, dato l'alto numero delle features, si considereranno metodi che prevedono un passo di feature selection allo scopo di ridurre il numero di dati in input e prevenire così sia gli effetti sul learning derivanti da feature ridondanti o inutili, sia il cosiddetto "curse of dimensionality" che affligge gli algoritmi di learning in spazi ad alta dimensionalità.

I metodi di Feature Selection (FS) si possono suddividere dal punto di vista strutturale in tre grandi categorie: Filter, Wrapped, Embedded [18], mentre dal punto di vista funzionale possono essere suddivisi in due grandi categorie: i Ranker ed i metodi a Subset [19]. Per quanto riguarda l'aspetto funzionale, I metodi Ranker assegnano pesi ad ogni singola feature, in modo che l'ordinamento ricavato corrisponda alla loro rilevanza teorica. Sono anche detti metodi continui o di valutazione individuale. I metodi cosiddetti di valutazione dei sottoinsiemi (*subset evaluation*) producono possibili sottoinsiemi di feature usando strategie di ricerca. Dopodiché il sottoinsieme viene valutato tramite una funzione che determinerà il sottoinsieme finale di features scelte. I metodi possono essere o univariati o multivariati a seconda che considerino o meno ogni feature indipendentemente dalle altre.

Dal punto di vista strutturale, invece, i metodi Filter effettuano un preprocessing delle feature che è indipendente dal successivo passo di apprendimento e dipende da caratteristiche sottostanti ai dati. I metodi wrapper usano l'algoritmo di apprendimento scelto come una routine interna che consente di misurare quanto una feature sia utile al processo di apprendimento, testato su un set di validazione. Nei metodi Embedded, invece, il processo di FS è parte integrante del metodo di apprendimento, in modo che la ricerca sia guidata dal processo stesso di apprendimento [20-21]. In Figura 4 è riportata, in maniera schematica, la suddivisione dal punto di vista strutturale dei metodi di FS.

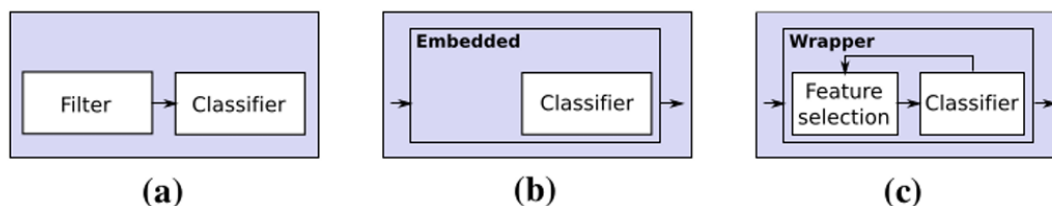


Figura 4: Suddivisione dei metodi di feature selection dal punto di vista strutturale: a) filter, b) embedded e c) wrapper

Ognuno di questi approcci ha i suoi pro ed i suoi contro in termini di velocità di calcolo e di accuratezza delle soluzioni trovate. I metodi Filters sono i più veloci e meno computazionalmente costosi, rischiano meno overfitting ma sono anche i meno accurati. I metodi wrappers sono i più accurati ma sono improponibili dal punto di vista computazionale su problemi grandi, e soffrono di overfitting. I metodi Embedded si collocano a metà fra i due. La Figura 5, tratta sempre da [20] riporta un sottoinsieme di metodi per cui si specificano sia le caratteristiche strutturali che funzionali.

	Uni/multivariate	Functional view	Structural view	Complexity
Chi-squared	Univariate	Ranker	Filter	nm
F score (Fisher score)	Univariate	Ranker	Filter	nm
Information gain	Univariate	Ranker	Filter	nm
ReliefF	Multivariate	Ranker	Filter	n^2m
mRMR	Multivariate	Ranker	Filter	nm^2
SVM-RFE	Multivariate	Ranker	Embedded	$\max(n, m)m^2$
CFS	Multivariate	Subset	Filter	nm^2
FCBF	Multivariate	Subset	Filter	$nm \log m$
INTERACT	Multivariate	Subset	Filter	nm^2
Consistency	Multivariate	Subset	Filter	nm^2

Figura 5: Tipologia di alcuni dei metodi più usati per la feature selection

In questo contesto terremo conto della letteratura analizzata e ci concentreremo sui seguenti metodi, di cui daremo una breve descrizione rimandando alla letteratura specifica per ulteriori approfondimenti.

Fra i metodi filter consideriamo: Fisher score [22], ANOVA f-value [23] e Logistic Regression [24]. Il primo filtra le feature basandosi su caratteristiche di similarità fra le features, mentre nel secondo si valutano, sotto l'ipotesi di distribuzione normale, le medie di diverse popolazioni; in ultimo la Logistic Regression è una tecnica statistica in cui vengono stimate le relazioni fra le variabili, valutando quanto il cambiamento in una variabile si riflette sulle altre.

I metodi filter vanno poi combinati con un classificatore, qui prendiamo in considerazione l'algoritmo di Support Vector Machine (SVM) [25]. SVM si basa sull'idea di trovare un iperpiano che divida al meglio un insieme di dati in diverse classi. L'algoritmo prende il nome dai cosiddetti "support vector", vettori di supporto, che corrispondono ai dati più vicini all'iperpiano. Tali punti dipendono dai dati che si sta analizzando e, se modificati, portano ad un cambiamento della posizione dell'iperpiano di separazione. Per questo motivo sono

considerati gli elementi critici di un set di dati. Affiancato al cosiddetto kernel trick le SVMs è possibile generare modelli non lineari in spazi ad alta dimensionalità.

Un classificatore Random Forest [26] è un algoritmo supervisionato che usa alberi di decisione e un approccio bagging per aggregare le decisioni prese singolarmente dagli alberi. Diversi alberi decisionali vengono infatti condensati in un unico modello. Le previsioni fatte individualmente dai singoli alberi potrebbero risultare non accurate ma, se combinate insieme sono mediamente più vicine al risultato “esatto”. L’output del Random Forest infatti è, nel caso di un problema di regressione, la media del risultato numerico restituito dai diversi alberi; oppure nel caso di un problema di classificazione, la classe individuata dal maggior numero di alberi. Random Forest restituisce naturalmente il ranking delle features, che è basata sul calcolo dell’out-of-bag error. RF ed è quindi un algoritmo di FS di tipo embedded. E’ un metodo estremamente veloce e poco pronò all’overfitting.

Support Vector Machines with Recursive Feature Elimination (SVM-RFE) [27] è un metodo di Embedded FS in cui si sfruttano i pesi assegnati alle variabili di input dal classificatore SVM. Tali pesi saranno usati per fare ranking delle features, eliminando progressivamente quelle meno rilevanti. E’ un buon compromesso fra tempi di esecuzione, accuratezza del risultato e rischio di overfitting.

4. Malattie Neurodegenerative

Per entrambe le malattie neurodegenerative le features in input saranno quelle estratte nella attività precedente e provenienti dall’OR1. Gli outcomes attesi dal punto di vista medico comprendono: 1) la estrazione delle features radiomiche derivate dalle immagini ibride MRI/PET che possono essere usate, assieme alle informazioni cliniche, per la predizione della progressione della patologia; 2) laddove possibile ricavare le segmentazioni delle aree cerebrali, si vuole testare l’uso combinato delle features radiomiche e delle features connettomiche, allo scopo di estrarre i descrittori migliori per la diagnosi *precoce* della malattia, punto importantissimo per entrambe le patologie.

Sia per il morbo di Alzheimer che per il Parkinson gli approcci di machine learning su dati radiomici ibridi sono meno diffusi, sia perché l’accesso alle immagini di test non è semplice come nel caso delle patologie tumorali sia perché l’approccio ibrido MRI/PET con traccianti specifici è stato definito e messo a punto molto recentemente. Ciò nonostante esistono già degli esempi in letteratura che possono essere utili come indicazione sulle tecniche più efficaci di machine learning per la classificazione di dati provenienti da immagini di queste patologie. Si può consultare, per una breve review sugli approcci di radiomica in entrambe le patologie, il lavoro [28]; mentre per l’uso di approcci di machine learning su dati di connettomica per diverse patologie, incluse quelle di interesse, si può considerare la estremamente dettagliata review in [29]; per un benchmarking dei diversi approcci di machine learning sui dati di connettomica provenienti da alcuni dei più importanti database di fMRI si può consultare [31].

4.1. Morbo di Alzheimer

In un lavoro di recentissima pubblicazione [32] dalle immagini di RM della regione ipocampale dei pazienti vengono estratte 495 features diverse. Tramite un approccio di Embedded Feature Selection, basato su Recursive Feature Elimination (RFE) per la selezione delle feature e su Support Vector Machine con un kernel RBF per la classificazione, gli autori dimostrano che le feature radiomiche provenienti dall'ipocampo possono essere usate per la diagnosi precoce di AD (Alzheimer's Disease). Viene usato un approccio di cross-validation in fase di apprendimento, e le prestazioni vengono espresse in termini di AUC.

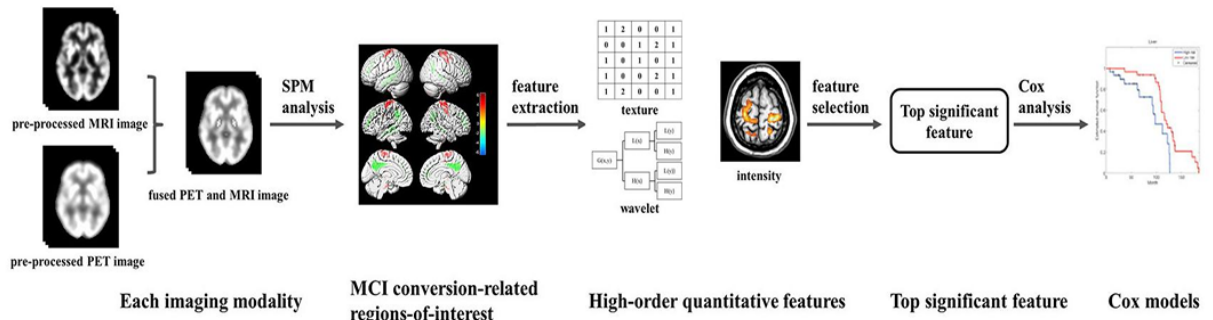


Figura 6: L'approccio metodologico usato in [34] per l'estrazione delle features derivanti da immagini ibride

Un approccio molto simile viene usato in [33] dove si usano dati da RM per distinguere AD dalla cosiddetta Demenza Vascolare (VD). Invece in [34] si usano, come in questo progetto, feature derivanti da immagini ibride PET/MRI (Figura 6). In questo caso vengono estratte 172 diverse features, fra cui features di primo e secondo livello; features di texture da Gray-Level Co-occurrence Matrix (GLCM), da Gray-Level Run-Length Matrix (GLRLM), da Gray-Level Size Zone Matrix (GLSZM), e da Neighborhood Gray-ToneDifference Matrix (NGTDM). Le features vengono usate per predire l'indice prognostico della malattia tramite un approccio di Feature ranking che sfrutta un modello di Cox minimizzato tramite LASSO (least absolute shrinkage and selection operator) con termine di regolarizzazione in norma L1. Viene usato un approccio di cross validation 70/30% e la performance viene analizzata tramite Harrell's Consistency index. Si dimostra che l'approccio combinato ha un potere predittivo superiore rispetto alle modalità singole di imaging.

Per quanto riguarda lavori in cui si usano invece features da connettoma si possono considerare i lavori [35-36] in cui si usano dati provenienti da risonanza magnetica funzionale (fMRI) in stato di riposo per costruire il connettoma dei pazienti.

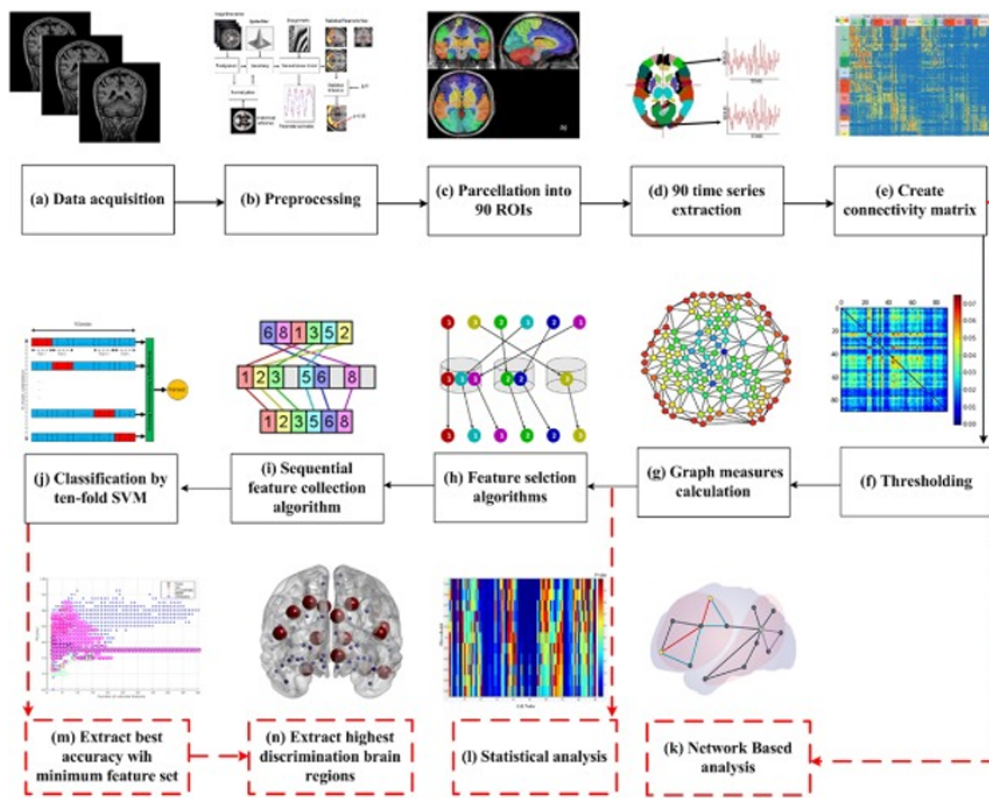


Figura 7: Approccio metodologico per sfruttare dati dei connettomi

I dati vengono usati nel primo per fare diagnosi e distinguere in maniera automatica pazienti sani da pazienti malati, nel secondo per definire la progressione della malattia e distinguere i pazienti che rimangono in stato di Mild Cognitive Impairment da quelli che poi svilupperanno la malattia di Alzheimer.

In Figura 7 è riportato l'approccio metodologico su cui è basato il secondo lavoro. A valle della costruzione del connettoma, quest'ultimo viene convertito in un grafo che viene poi analizzato tramite una serie di misure locali e globali, che costituiscono l'insieme 913 delle features su cui poi si basano le fasi successive. In particolare viene sviluppato un algoritmo di Feature Selection che si basa sul consenso di diversi approcci di feature ranking (Multivariate Minimal Redundancy Maximal Relevance (MRMR), Fisher score, Chi-square score, Gini score, Kruskal-Wallis test). Le feature estratte sono poi usate per addestrare un classificatore basato su Support Vector Machines (SVM) con un kernel lineare. Viene usato un approccio di Cross Validation e le prestazioni vengono analizzate tramite AUC.

In [37] invece si costruisce il connettoma metabolico sfruttando dati che vengono da PET con tracciante ^{18}F -fluorodeossiglucosio (FDG-PET) allo scopo di distinguere pazienti con MCI da pazienti AD. Anche in questo caso si calcola il connettoma e si estraggono le feature dalla matrice di similarità costruita tramite Kullback-Leibler divergence). Feature selection e classificazione seguono la metodologia SVM-RFE. Viene usato un approccio di leave-one-out cross validation e le prestazioni vengono misurate in termini di accuratezza precisione ed AUC.

Molti dei lavori qui citati hanno dati che vengono dal consorzio ADNI (Alzheimer's Disease Neuroimaging Initiative) [38] o dal progetto Humane Connectome (HUP) [39] i cui database sono accessibili sotto motivata richiesta e dopo autorizzazione, non risultando quindi disponibili per un uso che non sia ritenuto appropriato dal comitato di garanzia.

4.2. Morbo di Parkinson

Per il morbo di Parkinson la letteratura di riferimento è molto ridotta. Mancano al momento terapie, markers molecolari o da immagini che possano essere di supporto per la diagnosi la cura e la definizione della progressione della malattia.

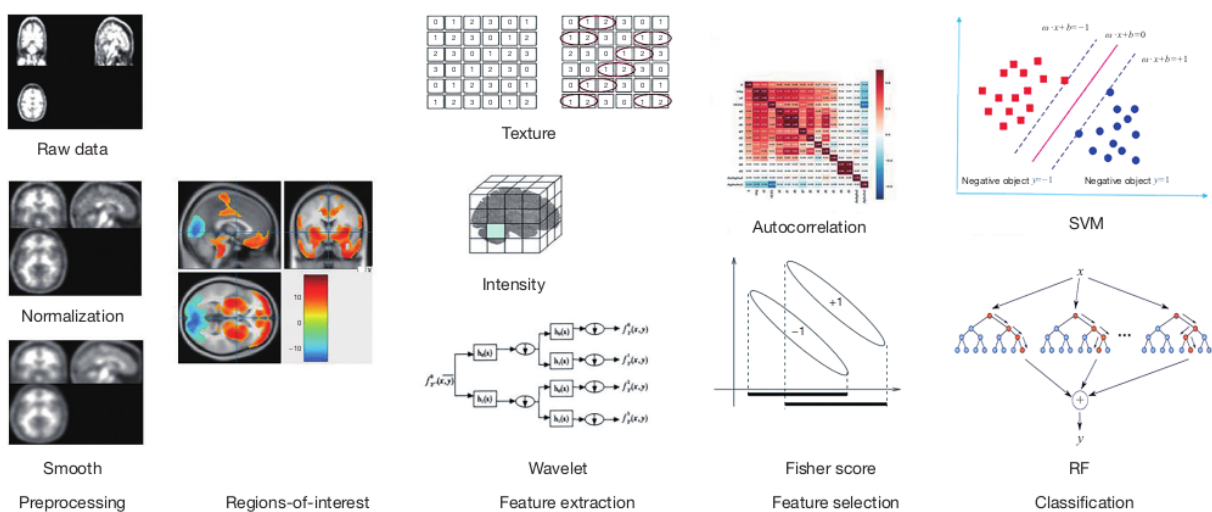


Figure 8: Metodologia usata in [40] per la classificazione di features da immagini.

In [40] un ampio insieme di features radiomiche vengono estratte dalle immagini PET ^{18}F -FDG, incluse feature di texture, wavelet, di istogramma e di intensità per un totale di più di 6000 per ogni paziente (figura 8). Tramite calcolo della autocorrelazione e l'uso del Fisher score, viene fatto feature selection e feature ranking per conservare solo le features informative. Per classificare i pazienti malati/sani si usano sia le SVMs che un approccio basato su Random Forests. Viene usato un approccio di crossfold validation e le prestazioni sono valutate in termini di AUC.

In [41] l'obiettivo è quello di predire la progressione della malattia combinando features cliniche e motorie con features "hand-crafted" classiche e radiomiche da immagini DAT-SPECT. La metodologia di classificazione è basata su Random Forest, l'approccio è di leave-one-out-cross validation e le prestazioni vengono calcolate in termini di errore assoluto.

In [42] si usano feature radiomiche estratte da immagini MRI di una particolare regione (Substantia Nigra) per discriminare pazienti sani da pazienti malati. In totale vengono usate 105 features radiomiche estratte da voxel di interesse che includono sia feature di forma che di texture. Tre diversi metodi di Feature Selection (ANOVA, RFE e Random Forest) vengono

usati per estrarre un ranking delle features. La classificazione viene fatta usando le SVMs con un approccio di cross-fold validation, e le prestazioni sono valutate in termini di AUC.

In [43] vengono usate assieme feature estratte da immagini single-photon emission computed tomography (SPECT) associate alle features cliniche di valutazione degli effetti della malattia sulle capacità motorie (presenza di rigidità, tremori, acinesia, disturbi del linguaggio,...). Vengono valutati molti diversi mezzi per effettuare la classificazione binaria in sani/malati: Support Vector Machines (SVMs), Random Forests (RFs), multilayer perceptron (MLP) Neural Network, Linear Regression, e K-NearestNeighbor. Le prestazioni sono valutate in termini di AUC con relativi intervalli di confidenza.

Per il morbo di Parkinson lavori di riferimento che facciano uso di dati di connettomica combinati con il machine learning praticamente non esistono.

Analogamente a quanto accade per Alzheimer, i dati per questa patologia possono essere ottenuti solo dopo motivata richiesta al consorzio Parkinson's Progression Markers Initiative (PPMI) [44].

4.3. Ulteriori tecniche di Feature selection e classificazione

Oltre ai metodi di feature selection e classificazione riportati nella sezione per le patologie tumorali, includiamo – tenendo conto della letteratura a supporto - anche una breve descrizione di altri due metodi che saranno usati per la classificazione a valle della feature selection: Multilayer Perceptron (MP) e K-nearest neighbor (K-NN).

Il perceptron è una rete formata da un certo numero di neuroni. Tramite uno strato singolo di neuroni si possono introdurre diverse funzioni di attivazione a soglia. Una rete del genere da sola non ha un particolare potere espressivo, e di norma quindi si introducono altri livelli di neuroni e si sostituisce la funzione di attivazione con la funzione logistica. In tal modo (poiché la funzione logistica è differenziabile), si possono usare algoritmi di gradient descent in fase di apprendimento, dando così vita a una rete chiamata *Multilayer Perceptron* [45]. Dunque la rete sarà formata da due livelli di unità di elaborazione, con d ingressi e con m uscite per il primo strato, a cui si aggiungono anche c uscite al secondo livello, dette unità nascoste perché le loro funzioni di attivazione non sono direttamente accessibili dall'esterno. In genere la funzione di attivazione delle unità nascoste sarà di tipo non lineare, così da garantire un maggiore potere espressivo.

Il K-NN [46] è considerato il più semplice algoritmo di machine learning, in quanto classifica un oggetto in base alla somiglianza con i suoi k vicini (neighbor), dove k è un intero positivo tipicamente non molto grande. La scelta del parametro k dipende dalle caratteristiche dei dati, in genere rappresentati come vettori in uno spazio. La distanza tra vicini è in genere calcolata usando la distanza euclidea oppure la distanza Manhattan, per dati numerici, ma consente l'uso di altre distanze definite in base alla natura dei dati: ad esempio la distanza di Hamming per stringhe. Lo spazio dei dati viene suddiviso in regioni in base alle posizioni reciproche dei

dati e alle loro caratteristiche di apprendimento. Un dato è assegnato ad una certa classe se questa è la più frequente fra i k esempi più vicini all'oggetto sotto esame.

5. Analisi dei dati del tumore al seno

5.1. Dati provenienti da dataset online

E' stata effettuata un'analisi preliminare di alcuni dati provenienti dal TCIA. Sono stati uniti i dati clinici pubblici della collezione TCGA-BRCA con le feature radiomiche estratte. Sono quindi state selezionate 173 variabili complete, ovvero con un numero di NA pari a 0. Definiti i pazienti e le variabili a disposizione e di interesse, si è passati ad una prima analisi descrittiva del dataset.

Questa fa uso degli strumenti statistici ed è di fondamentale importanza per comprendere le caratteristiche peculiari del caso che si sta studiando, oltre ad una panoramica delle incidenze di alcuni comportamenti, peculiarità o specifiche all'interno del sottogruppo dataset.

Sono stati prodotti degli istogrammi e delle tabelle di correlazione per valutare la distribuzione delle variabili oltre ad eventuali relazioni implicite. Occorre tenere in considerazione che è possibile effettuare delle descrizioni considerando in maniera condizionata una o più variabili, ovvero ad esempio l'età media al variare dei centri invianti o il numero di linfonodi al variare dello stadio tumorale. L'analisi è stata divisa e ripetuta per due diverse sequenze di risonanza magnetica, ovvero STIR e DCE (Figura 9).

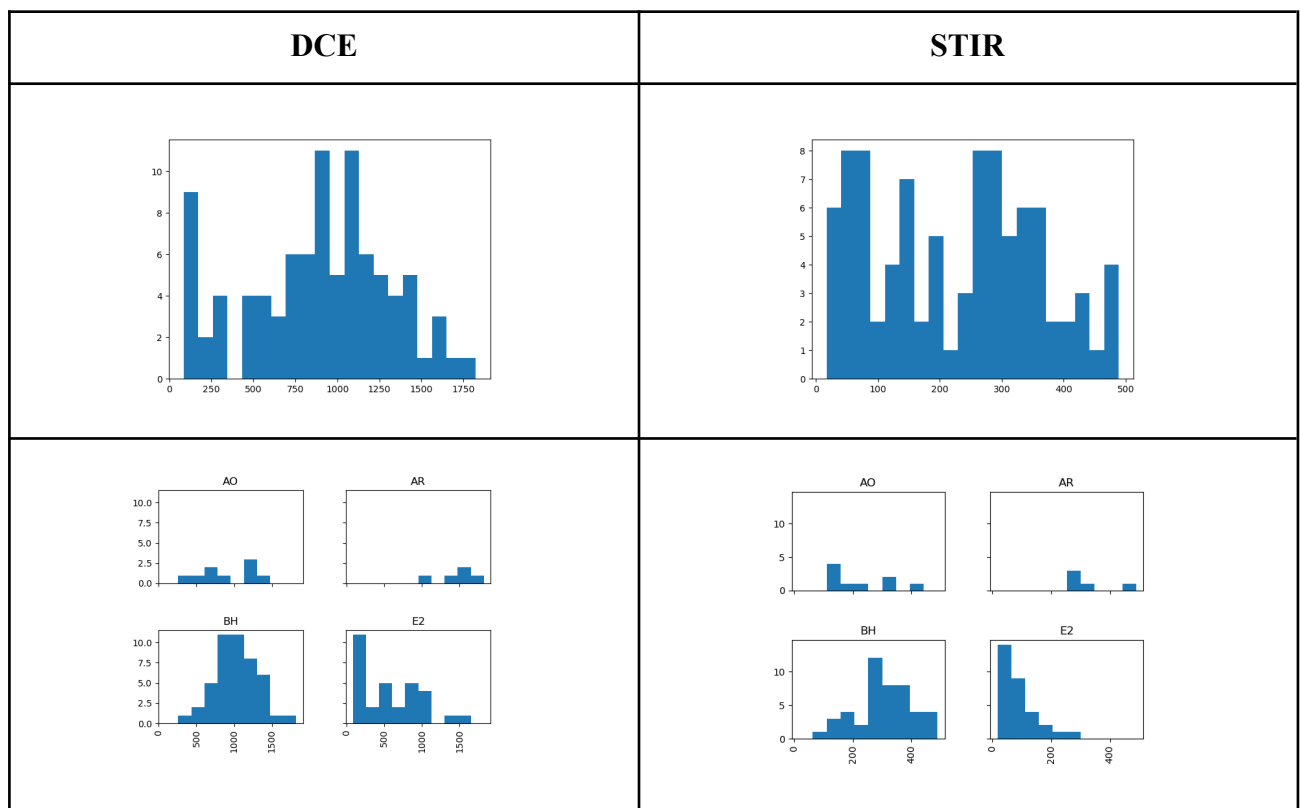


Figura 9: istogrammi di distribuzione delle variabili

Sono state effettuate ulteriori analisi per individuare le correlazioni tra le variabili estratte (figura 10). Ad esempio si può calcolare il coefficiente di correlazione di Pearson. A seconda del valore del coefficiente, è possibile avere un'indicazione su quanto due variabili seguano andamenti simili o siano legate tra loro.

In particolare, un modo utile per visualizzare il risultato di questa operazione è l'utilizzo di mezzi grafici, come ad esempio una "heatmap". Si può notare come le relazioni tra le variabili nei due casi siano simili, ad eccezione delle feature derivanti dall'analisi radiomica, che dipende anche dai livelli di grigio.

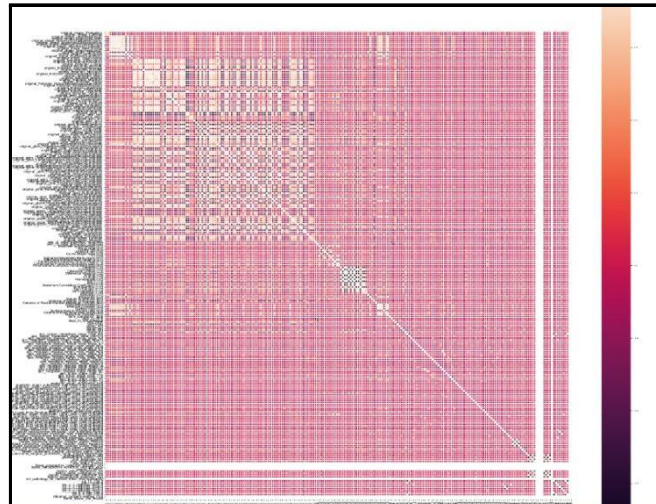


Figura 10: correlazione dei dati (heatmap)

In aggiunta, è evidente come le variabili "sesso" o "vital status" non influenzino in alcun modo lo scenario, essendo variabili costanti (i pazienti considerati sono tutti di sesso femminile e "alive" nel momento dell'esame).

Inoltre, si sono analizzate le caratteristiche cliniche dei 9 pazienti che hanno valori delle feature radiomiche legate al volume della lesione (S1, S3 e S4) al di sotto delle media. Rispetto allo stadio dei linfonodi (vedi Figura 11), la maggior parte dei pazienti risulta avere uno stadio dei linfonodi pari a N1.

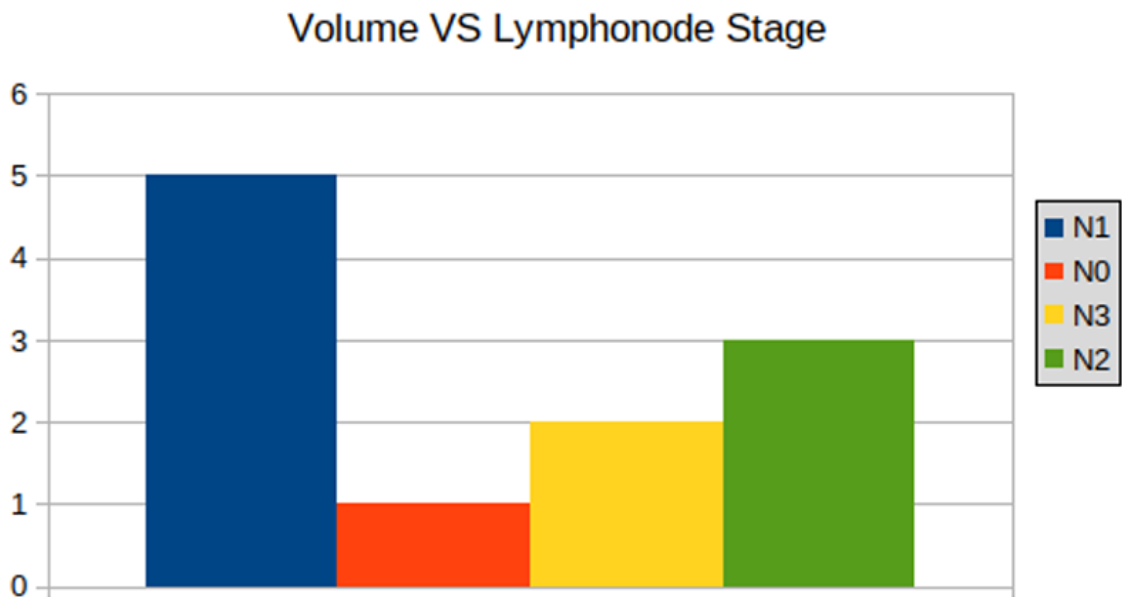


Figura 11: Distribuzione dello stadio dei linfonodi rispetto alle feature di Volume di 9 pazienti esaminati

In Figura 12 è mostrata la distribuzione dello stadio della neoplasia e in questo caso la maggior parte dei pazienti risulta avere uno stadio IIIB.

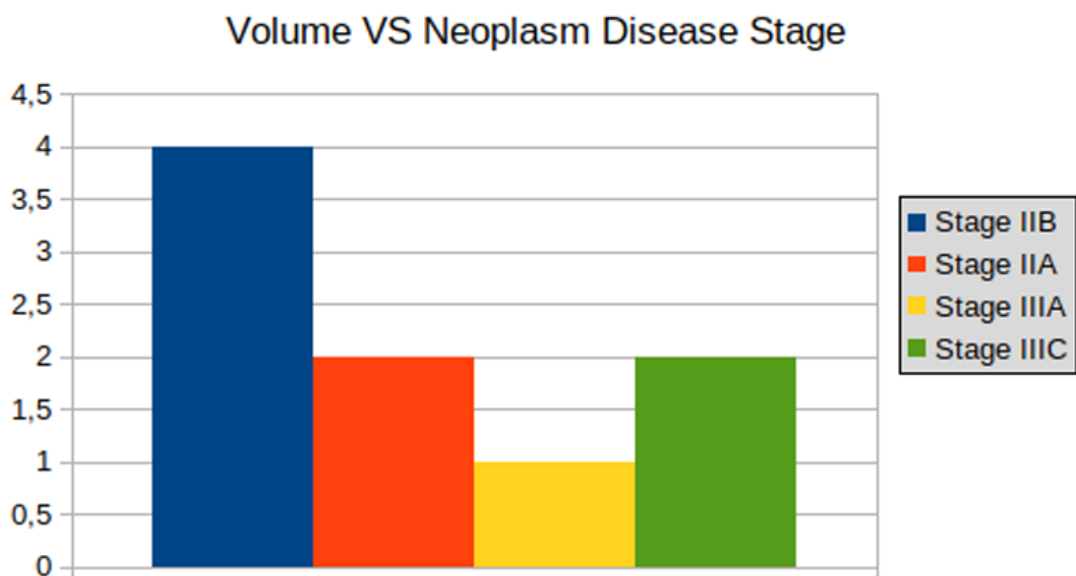


Figura 12: Distribuzione dello stadio della neoplasia rispetto alle feature di Volume di 9 pazienti esaminati

Rispetto ai recettori di estrogeno e progesterone, 7 dei 9 pazienti sono positivi ad entrambi i recettori (vedi Figure 13 e 14).

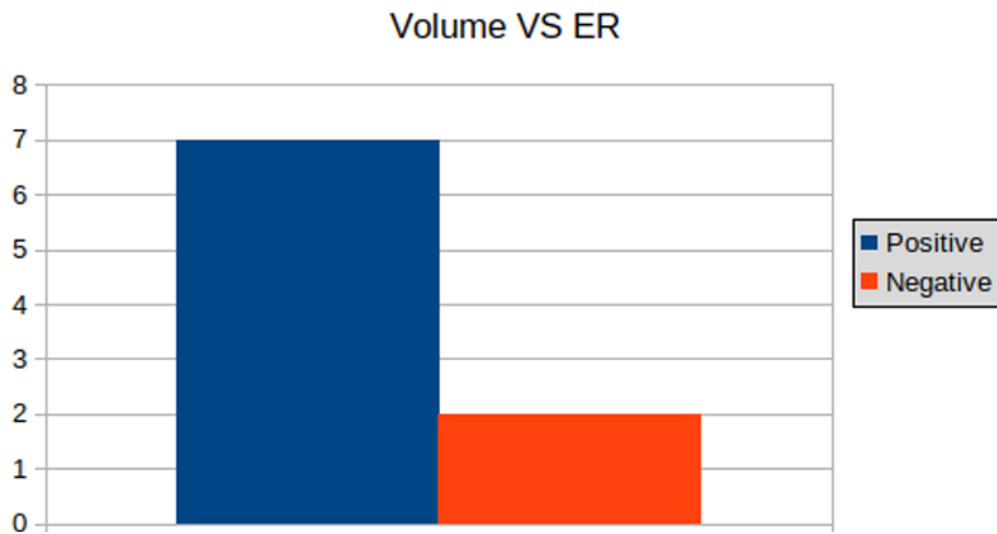


Figura 13: Distribuzione dello stato dei recettori dell'estrogeno rispetto alle feature di Volume di 9 pazienti esaminati

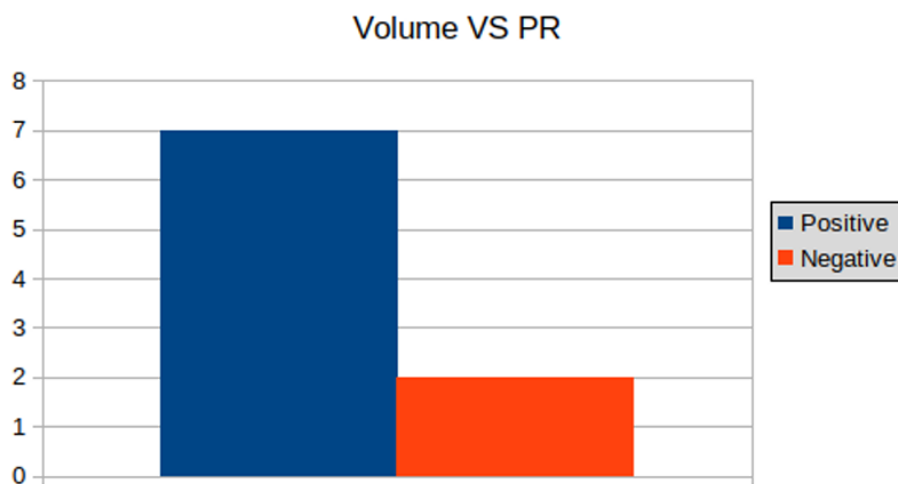


Figura 14: Distribuzione dello stato dei recettori del progesterone rispetto alle feature di volume di 9 pazienti esaminati

In figura 15 è illustrata la distribuzione dei pazienti rispetto all'istotipo. Per 7 di essi l'istotipo risulta essere Duttale.

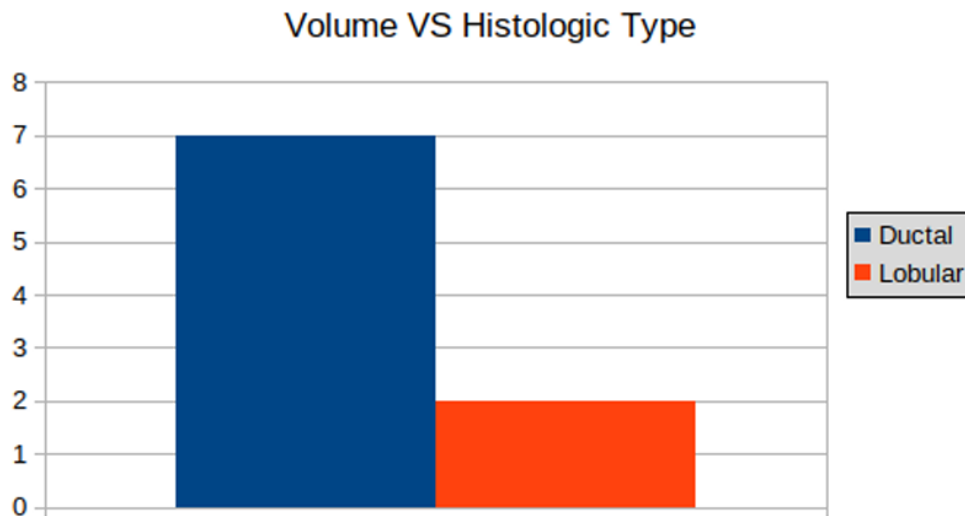


Figura 15: Distribuzione dell'istotipo rispetto alle feature di volume di 9 pazienti esaminati

Infine, relativamente allo stadio tumorale, soltanto 1 paziente su 9 ha stadio tumorale pari a T3 (vedi Figura 16).

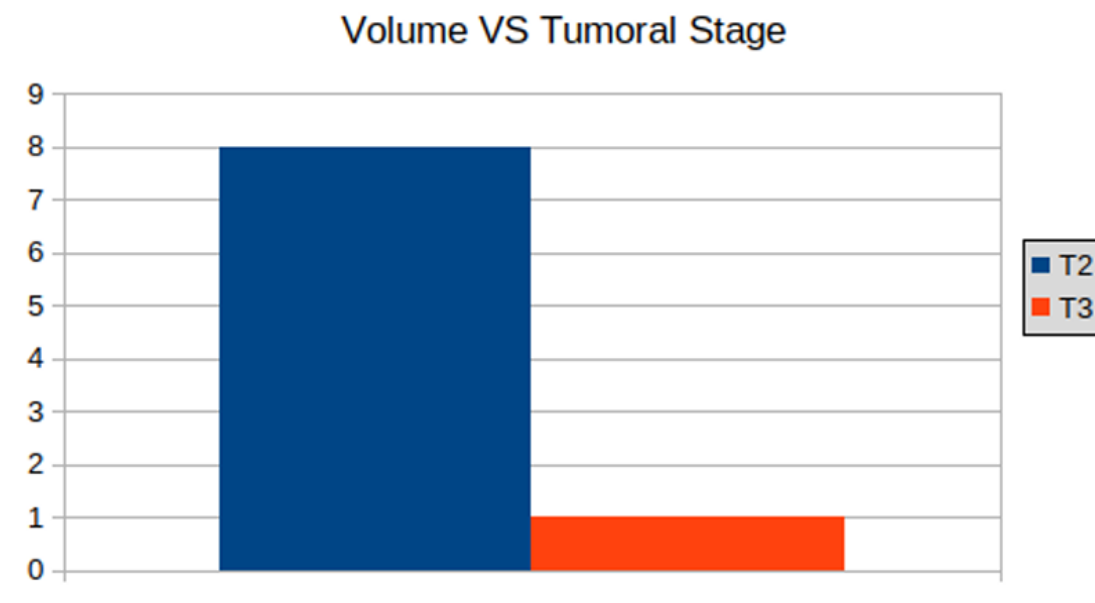


Figura 16: Distribuzione dello stadio tumorale rispetto alle feature di volume di 9 pazienti esaminati

Per quanto riguarda le feature radiomiche legate alla dinamiche ed analizzate nell'attività 2.1, era risultato che solo 3 pazienti avevano valori al di sotto della media, relativamente alle feature E1, E3 e E4. Analizzando le caratteristiche cliniche di questi 3 pazienti, risulta che 2 di essi hanno le stesse caratteristiche cliniche considerate precedentemente. In particolare, risultano avere uno stadio dei linfonodi pari a **N0**, sono **positivi** ai recettori di **estrogeno** e

progesterone, hanno istotipo **ductale**, lo stadio tumorale è pari a **T2** e infine, lo stadio della neoplasia è pari a **IIA**. Il terzo paziente invece ha in comune solo stadio dei linfonodi, la positività al recettore del progesterone e l'istotipo.

5.2. Dati di progetto

Di seguito si affronta il problema di valutare i dati clinici ricevuti dai partner al fine di usarli per le successive fasi di machine learning. Per rendere possibile la selezione delle features e la classificazione su un numero non elevato di campioni a disposizione rispetto al numero di features radiomiche estratte, si è deciso di accorpare e ridurre ad un numero inferiore di classi quei dati i cui valori sono usati come obiettivi di classificazione. In particolare quindi si analizzeranno i valori dei marcatori molecolari, andando ad estrarre anche i sottotipi individuabili mediante i loro valori. I sottotipi molecolari possono essere usati laddove nei dati manchino informazioni sul sottotipo istologico.

5.2.1. Marcatori molecolari

5.2.1.1. Definizione dei marcatori e loro valori

I valori dei marcatori HER2, ER, PR e IK67 sono sfruttati per dare la definizione del sottotipo molecolare, una classificazione utile in fase di diagnosi sia per la definizione della terapia, che per la successiva prognosi. Di seguito si analizzano i valori di ciascun marcatore alla luce della corrente letteratura, e allo scopo di definire i valori usati in fase di feature selection e classificazione.

HUMAN EPIDERMAL GROWTH FACTOR RECEPTOR 2 - HER2

Il valore di score di HER2 è estremamente rilevante perché prevede terapie profondamente differenti nel caso di score positivo/negativo.

Nei casi di equivocità (IHC=2+) si dovrebbe fare almeno il test FISH, o comunque prevedere una rivalutazione secondo quanto indicato dalla linee guida AIOM 2018 [17] (pag. 38, e Allegato 3 pag. 234) che a loro volta recepiscono le linee guida ASCO [47]. In realtà anche quando FISH è negativo ma il risultato non sembra coerente con quanto determinato dal patologo l'indagine andrebbe ripetuta. Risultano quindi definite le seguenti classificazioni ridotte:

SCORE HER2	TEST RESULT
Positivo	IHC > 3+
	FISH positivo
	IHC=2+ E FISH positivo
Negativo	IHC=0+, 1+
	FISH negativo
	IHC=2+, FISH negativo
Equivoco	IHC=2+

	FISH equivoco
--	---------------

RECETTORI ESTROGENICI/PROGESTINICI - ER/PR

Lo stato dei recettori, sia estrogenici che progestinici andrebbe riportato come percentuale e valutato come una variabile quantitativa continua come indicato dalle linee guida AIOM [17] (pag 38). Seguendo le linee dell'ASCO [47] si considerano positivi i tumori con almeno 1% di cellule positive. Esiste però una relazione tra i livelli di positività ed i benefici ottenuti con i trattamenti ormonali. Pertanto i tumori con elevati livelli di recettori sono quelli che hanno maggiori probabilità di beneficiare di una terapia ormonale. Per la determinazione del sottotipo Luminale A si usa il cutoff introdotto dalla Consensus di St. Gallen 2013 [48] che prevede la presenza di recettori progestinici positivi con valore di positività superiore al 20%. Risultano quindi definite le seguenti classificazioni ridotte:

STATUS ER/PR	TEST RESULT
Positivo	IHC > 1%
Negativo	IHC=0

INDICE DI PROLIFERAZIONE - KI67

L'attività proliferativa misurata con il *KI67 labeling index* è oggi un fattore prognostico riconosciuto. Purtroppo però l'alta variabilità del conteggio dell'IK67 fra vari laboratori non permette di fissare un valore di soglia unico al di sotto o al di sopra del quale il tumore possa essere definito a bassa o ad elevata attività proliferativa. L'AIOM sul KI67 si allinea alle posizioni definite nella Consensus di St. Gallen 2013 [48] secondo cui una percentuale del 20% è da considerarsi positiva [17]. La St. Gallen del 2015 [49] però afferma che gli score devono essere interpretati in base ai valori mediani del laboratorio in cui vengono fatti gli esami: *“Ki-67 scores should be interpreted in the light of local laboratory values: as an example, if a laboratory has a median Ki-67 score in receptor-positive disease of 20%, values of 30% or above could be considered clearly high; those of 10% or less clearly low.”* La successiva St. Gallen del 2017 [50] riprende e conferma quanto detto, ma suggerisce che il grading del tumore possa essere considerato al posto della proliferazione come indice prognostico. Risultano quindi definite le seguenti classificazioni ridotte:

STATUS KI67	TEST RESULT
Positivo	>= 20%
Negativo	< 20%

5.2.1.2. Classificazione molecolare

In base ai valori individuati per i marcatori molecolari è possibile definire una classificazione dei sottotipi tumorali che riflette in maniera molto accurata l'eventuale suddivisione su base genetica.

Tale classificazione è considerata di estrema importanza sia per la definizione della terapia personalizzata da proporre ai pazienti, sia per la definizione di una prognosi più attendibile [17], [50].

NOME	HER2 score	ER score	PR score	KI67 score
Luminal A-like	-	+	+*	-
Luminal B-like /HER2 negative	-	+	-	+/-
			+/-	+
Luminal B-like /HER2 positive	+	+	+/-	+/-
HER2-positive (non-luminal)	+	-	-	+/-
Triple Negative (ductal)	-	-	-	+/-

NOTE: * Per Luminal A-like deve essere PR $\geq 20\%$, come da indicazioni AIOM che recepiscono le indicazioni ASCO.

5.2.1.3. Decisioni su dati reali di progetto

Come già accennato, per rendere possibile sia la selezione delle features che la classificazione su un numero non elevato di campioni a disposizione rispetto al numero di features radiomiche estratte, si è deciso di accorpare e ridurre ad un numero inferiore di classi quei dati i cui valori sono usati come obiettivi di classificazione. Questo approccio si applica a tutti i marcatori molecolari.

HER2_score

Il campo è stato trasmesso con poca informazione sulle modalità di acquisizione. Poiché ha valori compresi fra 0 e 3, ed in alcuni casi c'è un * che indica che la FISH non è amplificata, se ne può dedurre 1) che si tratti dei valori di IHC, anche se non tutti hanno il simbolo '+' alla fine, e: 2) che, poiché in alcuni casi è stato segnalato che non c'è espressione del gene HER2 usando il simbolo: '*', sia stato eseguito per IHC ambiguo un'ulteriore indagine tramite FISH, come raccomandato.

Dunque, si è deciso di considerare tre classi, come rappresentato in tabella.

HER2 SCORE	VALORI	VALORE CLASSE
Positivo	3, 3+	2

Negativo	0, 1, 1+, 2*	0
Ambiguo	2, 2+	1

ER_status_percent

In questo caso i valori sono indicati in percentuali, e normalmente basta una percentuale diversa da zero perché sia considerato ER positivo. Le ultime risoluzioni dell'ASCO in materia [6] (non ancora recepite nelle linee guida AIOM) indicano che in caso di ER con percentuali basse in presenza di HER2 e PR negativi, il valore basso di ER vada segnalato come equivoco e da rivedere. Si è deciso in questo caso di non considerare questa raccomandazione e di tenere solo 2 classi.

ER STATUS	VALORI	VALORE CLASSE
Positivo	> 0%	1
Negativo	0	0

PR_status_percent

In questo caso i valori sono indicati in percentuali e, poiché esiste un limite di soglia usato per distinguere il sottotipo Luminal A-like dal B, si è deciso di considerare tre classi, di seguito descritte.

PR STATUS	VALORI	VALORE CLASSE
Positivo	$\geq 20\%$	2
Basso	> 0 and $< 20\%$	1
Negativo	0	0

KI67_level_percent

In questo caso i valori sono indicati in percentuali, e, come detto nel paragrafo precedente, sussistono diversi problemi nella classificazione in alto/basso di questo marcatore. Alla luce di quanto detto su, si è deciso di considerare soltanto due classi, alto e basso.

CLASSE HER2 SCORE	VALORI	VALORE CLASSE
Alto	$\geq 20\%$	1
Basso	$< 20\%$	0

Classificazione molecolare

Nella tabella seguente viene indicato come ottenere la classificazione molecolare usando i nuovi valori delle classi definite per i marcatori:

NOME	HER2 score	ER score	PR score	KI67 score
Luminal A-like	0	1	2	0
Luminal B-like /HER2 negative	0	1	0	0/1
			0/1/2	1
Luminal B-like /HER2 positive	2	1	0/1/2	0/1
HER2-positive (non-luminal)	2	0	0	0/1
Triple Negative (ductal)	0	0	0	0/1

Applicando questa regola, si ha la seguente distribuzione dei fenotipi molecolari per il set dei primi 29 pazienti (30 insiemi di features, dal momento che una paziente ha doppia segnalazione) forniti dal partner SDN:

CLASSE	SIGLA	NUMERO DI PAZIENTI
Luminal A-like	LA	2
Luminal B-like /HER2 negative	LBH2N	9
Luminal B-like /HER2 positive	LBH2P	5
HER2-positive (non-luminal)	NL	1
Triple Negative (ductal)	TN	7*
Sconosciuto	U	3*

NOTE: * si vedano i casi particolari descritti nel paragrafo successivo.

Casi particolari

PAZIENTE SDN_FDG_Bca_020: in questo caso la paziente ha una doppia segnalazione una per ciascuna mammella. Poichè i fenotipi molecolari descritti sono diversi, si è deciso di considerare ciascuna regione tumorale separatamente.

PAZIENTE SDN_FDG_Bca_011: in questo caso si ha ER=0 e PR=5, quindi fenotipo ER-/PR+. Questo è un evento molto raro e parecchio dibattuto, e mentre qualche autore lo considera come un nuovo sottotipo, le indicazioni dell'ASCO sono di ripetere gli esami,

perché probabilmente uno dei due valori è frutto di un artefatto [51], [52]. Poiché in questo caso il conteggio dei PR è davvero basso e, nello studio indicato [52], il 47% dei retesting dava ER-/PR- come fenotipo risultante, si è deciso di considerare PR=0, arrivando così ad una definizione di TN per questo paziente.

PAZIENTE SDN_FDG_Bca_007

In questi caso non è possibile una corretta assegnazione del fenotipo molecolare perché il livello di HER2 è equivoco. I due tipi possibili sono: LBH2N e LBH2P.

PAZIENTE SDN_FDG_Bca_026:

In questi caso non è possibile una corretta assegnazione del fenotipo molecolare perché il livello di HER2 è equivoco. I due tipi possibili sono: TN e NL.

5.2.2. Sottotipi istologici

Anche il sottotipo istologico è uno dei dati attesi come outcome nel processo di machine learning. Si andranno ad analizzare le classi presenti ed eventuali raggruppamenti, sia per il partner SDN che per il partner UNIME.

5.2.2.1. Definizione dei sottotipi

Per la definizione dei sottotipi di cancro alla mammella ci si riferisce alle già citate linee guida della Associazione Italiana di Oncologia Medica, che a loro volta recepiscono le indicazioni della World Health Organization 2019. Tale classificazione è riportata al paragrafo 3.1 del documento [17].

5.2.2.2. Decisioni su dati reali di progetto

In tabella è riportata la suddivisione dei dati rispetto alla annotazione fornita dal partner SDN.

SOTTOTIPO (IT)		SOTTOTIPO (EN)		SIGLA	NUMERO DI PAZIENTI
Sarcoma Balstico	Mieloide	Blastic Sarcoma	Myeloide	BMS	1
Carcinoma invasivo	duttale	Invasive Carcinoma	Ductal	DC	42*
Carcinoma lobulare		Lobular Carcinoma		LC	5*
Carcinoma infiltrante di tipo non speciale		Non Specific Carcinoma	Type	NST	32*
Adenosi sclerosante		Sclerosing adenosys		SA	1*
Resezione di tumore primario		Surgery Primary Tumor		SPT	1**

Sconosciuto	-	-	11
TOTALE	TOTALE	-	93

Note: *: si vedano i casi particolari per ulteriori informazioni; **: in questo caso c'è un errore, in quanto la resezione di un tumore primario non è un tipo istologico.

Analogamente si riporta la suddivisione dei dati per il partner UNIME.

SOTTOTIPO (IT)		SOTTOTIPO (EN)		SIGLA UNIME	NUMERO DI PAZIENTI
Carcinoma duttale in situ		Ductal Carcinoma in Situ		DCIS	1
Carcinoma invasivo	duttale	Invasive Carcinoma	Ductal	IDC	17*
Carcinoma lobulare		Lobular Carcinoma		LC	2
Carcinoma infiltrante		Infiltrating Carcinoma		IC	1
Carcinoma infiltrante	tubulare	Tubular carcinoma	infiltrating	CA Tubulare Infiltrante	1
Sconosciuto		-		-	2
TOTALE	TOTALE	-		-	24

Note: *: si vedano i casi particolari per ulteriori informazioni;

Classificazione dei tipi istologici

La tabella su riportata può essere rivista alla luce delle indicazioni che si trovano nelle già citate linee guida. In esse si considerano il tipo DC ed il tipo NST come una unica classe che sarà quindi chiamata DC/NST. Vengono accorpati in una classe tutti i pazienti che hanno un sottotipo sconosciuto o non utilizzabile. Si ha dunque che il numero di classi si riduce molto, ma purtroppo si ottiene praticamente una sola classe per il sottotipo, con un profondo sbilanciamento rispetto alle altre. Nella tabella che segue sono riportati i risultati ottenuti.

SOTTOTIPO (IT)		SOTTOTIPO (EN)		SIGLA	NUMERO DI PAZIENTI
BMS		Blastic Sarcoma	Myeloide	BMS	1
Carcinoma invasivo/infiltrante speciale	duttale di tipo non	Invasive Carcinoma/ Non Specific Type Carcinoma	Ductal	DC/NST	74

Carcinoma lobulare	Lobular Carcinoma	LC	5
Sconosciuto o non utilizzabile	-	-	13
TOTALE	TOTALE	-	93

Analogamente si riportano i risultati ottenuti per il partner UNIME. Anche in questo caso sono stati raggruppati i tipi IDC con il tipo IC.

SOTTOTIPO (IT)	SOTTOTIPO (EN)	SIGLA	NUMERO DI PAZIENTI UNIME
Carcinoma duttale in situ	Ductal Carcinoma in Situ	DCIS	1
Carcinoma duttale invasivo/Carcinoma infiltrante di tipo non speciale	Invasive Ductal Carcinoma/ Non Specific Type Carcinoma	DC/NST	14
Carcinoma lobulare	Lobular Carcinoma	LC	2
Carcinoma tubulare infiltrante	Tubular infiltrating carcinoma	CA Tubulare Infiltrante	1
Sconosciuto o non utilizzabile	-	-	6
TOTALE	TOTALE	-	24

Casi particolari

PAZIENTE SDN_FDG_Bca_041: in questo caso la paziente ha annotato istotipo DC con la annotazione “(NOS)”, che sta per Not Otherwise Specified. Nelle linee guida questo sottotipo viene chiamato NAS (Non Altrimenti Specificato) e secondo le indicazioni WHO 2019, riprese dalle linee guida AIOM [1], questo tipo va raggruppati con i carcinomi di tipo NST.

PAZIENTE SDN_FDG_Bca_043: in questo caso la paziente ha annotato istotipo DC+MC, cioè carcinoma duttale e carcinoma mucinoso. Poichè la classificazione è incerta, ed il carcinoma MC molto raro, è stato contato come DC.

PAZIENTE SDN_FDG_Bca_044: in questo caso la paziente ha annotato istotipo DC con la annotazione “(cytoincluded)”, a indicare che l’assegnazione è avvenuta non tramite saggio immunoistochimico ma tramite ago aspirato. E’ stato contato come DC.

PAZIENTE SDN_FDG_Bca_081: in questo caso la paziente ha indicato istotipo SA con annotazione (atypical). La adenosi sclerosante è considerata una patologia benigna, che in qualche caso può essere prodromo di carcinomi mammari. In ogni caso, mancando tutte le informazioni che servono per la classificazione, è sostanzialmente inutilizzabile in fase di learning e test.

PAZIENTI UNIME_FDG_BCa_001 UNIME_FDG_BCa_022 UNIME_FDG_BCa_023 UNIME_FDG_BCa_024: in tutti questi casi non ci sono dati di Risonanza Magnetica, per cui non è possibile includerli nelle fasi di classificazione che sono basate, in prima istanza, su questa tipologia di esame diagnostico.

5.2.3. Stadiazione del tumore

La stadiazione del tumore è effettuata a partire dalla rilevazione di tre caratteristiche diverse che coinvolgono le dimensioni del tumore primitivo (T), lo stato dei linfonodi regionali (N) e la presenza di metastasi distanti (M). Le diverse combinazioni possibili dei tre parametri vengono poi raggruppate in stadi, utili a descrivere lo stato della patologia e fortemente associati alla prognosi del paziente.

5.2.3.1. Definizione dei sottotipi

Per la classificazione degli stadi in base ai valori di T,N,M si può consultare la Tabella 3.2 delle linee guide AIOM [17].

5.2.3.2. Decisioni su dati reali di progetto

In tabella è riportata la suddivisione dei dati in base ai possibili valori di stadiazione per i pazienti SDN.

STADIO	NUMERO DI PAZIENTI
IIA	15
IIB	12
IIIA	28
IIIB	14
IIIC	2
IV	11
Non specificato	11
TOTALE	93

Per i pazienti UNIME invece non ci sono informazioni di stadiazione esplicitamente riportati.

Classificazione dei valori di stadiazione

La tabella su riportata può essere semplificata allo scopo di rendere possibile il machine learning con un numero così limitato di dati. Si raggrupperanno dunque gli stadi indipendentemente dalla sottostadiazione (data dalla lettere a fianco al numero romano).

Si ottiene così la seguente tabella degli stadi

STADIO	NUMERO DI PAZIENTI
II	27
III	44

IV	11
Non specificato	11
TOTALE	93

6. Analisi dei dati delle malattie neurodegenerative

Descriviamo qui, il set di dati clinici generati nell'ambito del progetto e che saranno usati nella fase di generazione dei modelli di apprendimento.

6.1. Morbo di Alzheimer

TIPO	NOME	DESCRIZIONE
DATI GENERALI	età	Età al momento della diagnosi
	sexo	M/F
	educazione	In anni, da 4 a 17.
DIAGNOSI	Diagnosi	Definizione della patologia
	MMSE	
TOKEN TEST (test dei gettoni)	TOKEN	Test di valutazione della comprensione orale. Valore fra 0 e 36.
COWAT (CONTROLLED ORAL WORD ASSOCIATION TEST)	COWAT	Misura la produzione di parole orali, il risultato è un numero intero.
RAVLT (Rey Auditory Verbal Learning Test)	R.I	Valuta le capacità di rievocazione immediata e differita. In genere è un numero con almeno un decimale.
	R.D.	
FCSRT (Free And Cued Selective Reminding Test)	IFR	Sono indici che vengono usati per misurare le capacità di rievocazione della memoria. Ogni test comprende in genere dodici figure per cui il punteggio totale varia fra 0 e 12. I test sono: Rievocazione Libera Immediata (IFR), Rievocazione Immediata Totale (ITR), Rievocazione Differita Libera (DFR), Rievocazione Differita Totale (DTR). L'Indice di Sensibilità al Suggerimento Semantico (ISC) si ricava a partire da IFR e ITR.
	ITR	
	DFR	
	DTR	
	ISC	
DIGIT	FW	Il digit test forward (FW) e backward (BW) serve per misurare il funzionamento della memoria di lavoro. Valore

	BW	numerico, è il numero massimo di numeri in sequenza ricordati dal paziente in 3 test su 4.
FAB		Frontal Assessment Battery. Vengono esplorate le funzioni del lobo frontale in 6 test, il cui valore varia da 0 e 3 per un totale quindi di 18 punti.
STROOP	W	Serve per valutare l'attenzione visiva. Si articola in tre test W (word), C (color), CW (COLOR WORD) prova di interferenza). Gli output dei test sono due, un tempo (la differenza fra il tempo nella prova CW e la media nelle altre due) ed il numero degli errori (terza prova rispetto alla media delle prime due).
	C	
	CW	
WEIGL		Test per valutare la capacità di cogliere somiglianze. I punteggi grezzi sono numeri interi, ma quando si effettuano le correzioni sono numeri con almeno 2 cifre decimali.
JLO V		Judgment of Line Orientation, è un test di elaborazione spaziale. In genere è una stima di quanti task, su un totale di 30, sono stati correttamente risolti.
BECK II		Test dello stato di depressione di un paziente. In genere contiene una serie di test con risposte correlate ad intensità crescenti. Il punteggio totale individua delle fasce di stadiazione (ad esempio per 9 domande con 4 risposte: 0-9, prestazione normale; 10-15 lieve; 16-19 da lieve a moderata; 20-29 = da moderata a grave; 30-36 = depressione grave).
HAMA		Test Hamilton Rating Scale for Anxiety. Misura lo stato d'ansia. E' composta da 14 domande, per ognuna delle quali ci possono essere 4 risposte con punteggi di gravità crescente da 0 a 4. Ci sono delle fasce di stadiazione associate (< 17 lieve entità, 18-24 da lieve a moderata; 25-30 da moderata a grave)

6.1.1. Dati clinici

Di seguito si affronta il problema di valutare i dati clinici ricevuti dai partner al fine di usarli per le successive fasi di machine learning. Per rendere possibile la selezione delle features e la classificazione su un numero non elevato di campioni a disposizione rispetto al numero di features radiomiche estratte, si è deciso di accorpare e ridurre ad un numero inferiore di classi quei dati i cui valori sono usati come obiettivi di classificazione.

I dati ricevuti dal partner UNICZ comprendono dati clinici, test cognitivi e features radiomiche ma purtroppo non comprendono i dati sulla presenza di beta-amiloide.

DATI UNICZ

I dati clinici includono, oltre a sesso ed età, anche il valore dell'indicatore di educazione, che è fondamentale per il corretto bilanciamento dei risultati ottenuti dai test cognitivi. Questi ultimi comprendono tutti i test descritti nei precedenti documenti di progetto.

Sono inoltre disponibili i dati relativi all'uptake del mezzo di contrasto FDG per le regioni di interesse, e cioè: Lobo Parietale, Lobo Occipitale, Lobo Temporale, Lobo Frontale, Corteccia limbica, Ippocampo ed Amigdala.

6.1.2. Dati da immagini mediche

I dati, estratti tramite il tool Freesurfer [68] da immagini cliniche comprendono:

- area, curvatura media, volume, spessore di tutte le strutture **corticali** sia per l'emisfero destro che per il sinistro (dati freesurfer *aparc*)
- volume delle aree **subcorticali** sia per **emisfero destro** che per il **sinistro** (dati freesurfer *aseg*)
- volume delle **sottoregioni** (nuclei) del **Talamo**
- volume delle **sottoregioni** della **Amigdala** e dell'**Ippocampo**
- volume delle quattro **sottoregioni** del **Tronco encefalico**

Le colonne 'Left-WM-hypointensities', 'Right-WM-hypointensities', 'Left-non-WM-hypointensities', 'Right-non-WM-hypointensities' sono state rimosse perché tutte valorizzate a 0.

6.1.3. Correlazione tra dati

In collaborazione con i partner clinici è stata effettuata un'analisi di correlazione tra un sottoinsieme di dati, particolarmente significativi ai fini diagnostici. In particolare, lo scopo è stato identificare correlazioni specifiche tra 1) dati relativi ai **test cognitivi**, presenti nella tabella del paragrafo 7.1, 2) dati di **uptake** del mezzo di contrasto FDG e 3) dati provenienti da immagini, in particolare relativi ai **volumi**. I sottoinsiemi di dati per cui analizzare eventuali correlazioni sono stati identificati dai partner clinici successivamente ai primi esperimenti effettuati per la classificazione diagnostica.

Sono stati scelti gli indici di correlazione di Pearson [69] e di Spearman [70]. I risultati sono disponibili come immagini negli allegati al presente documento, da 1 a 3. Per semplicità riportiamo nelle successive figure solo qualche esempio rappresentativo.

In figura 1, riportiamo il grafico delle correlazioni di Pearson tra i dati dei test e di uptake.

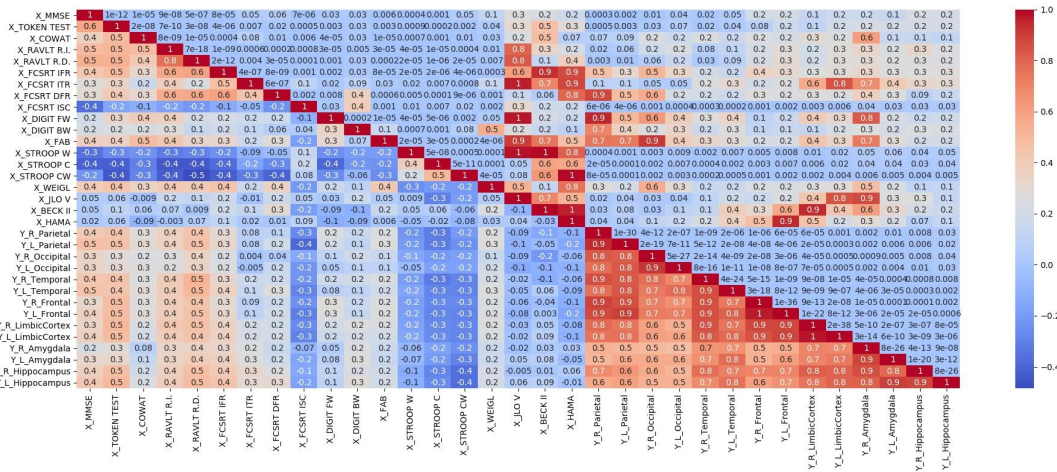


Fig. 1: Correlazione di Pearson tra dati dei test e dati di uptake

In particolare, nella diagonale inferiore sono riportati gli indici di correlazione, mentre nella diagonale superiore sono indicati i corrispondenti *p-value*.

Per semplicità di lettura, tale grafico viene accompagnato da quello presente in figura 2.

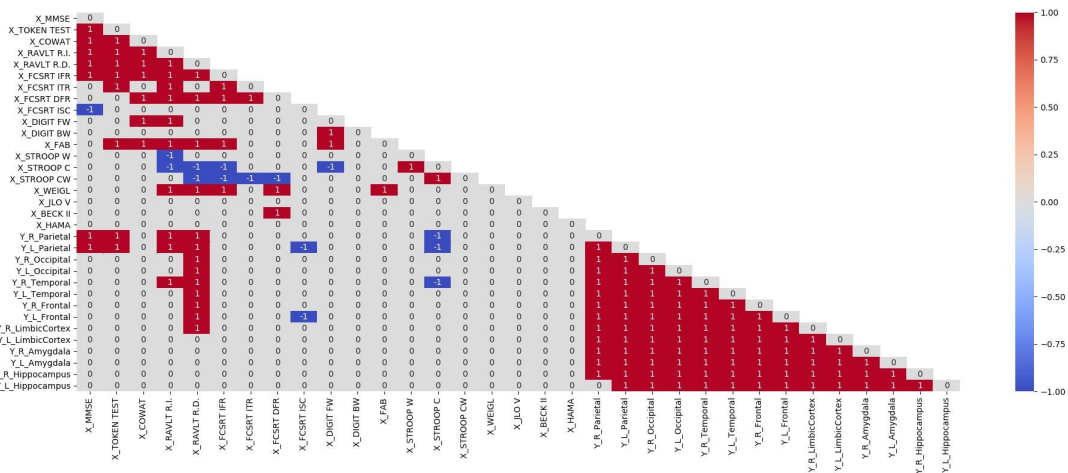


Fig. 2: Correlazione di Pearson rilevanti tra dati dei test e dati di uptake

In tale grafico, il valore 1 (-1) si trova in corrispondenza di indici di correlazione maggiori uguali di 0.3 (minori uguali di -0.3) e con *p-value* minore 0 uguale a 0.05, valore solitamente scelto come soglia di significanza. In questo modo, i valori 1 (-1) si troveranno in corrispondenza di alta correlazione (anti-correlazione) tra i dati. Ad esempio, dal grafico di figura 2, si evince che R_Parietal e L_Parietal sono altamente correlati con MMSE e TOKEN_TEST.

In statistica è particolarmente noto il problema dei test multipli, che si verifica quando si considera un insieme di inferenze statistiche contemporaneamente oppure si deduce un sottoinsieme di parametri selezionati in base ai valori osservati. Più variabili vengono incluse, più è probabile che si verifichino inferenze errate. Diverse tecniche statistiche sono state sviluppate per evitare che ciò accada, consentendo di confrontare direttamente i livelli di significatività per confronti singoli e multipli. Queste tecniche di correzione generalmente richiedono una soglia di significatività più rigorosa per i confronti individuali, in modo da

compensare il numero di inferenze fatte. La correzione che solitamente viene utilizzata è quella di Bonferroni [71]. Tale correzione controlla il Family-wise error rate (FWER), che è definito come la probabilità di ottenere almeno un errore di Tipo I (Falso positivo) all'interno della correzione, cioè la probabilità che almeno un p -value porti a rigettare l'ipotesi nulla.

La correzione di Benjamini-Hochberg [72] permette di fare un migliore controllo dei falsi positivi utilizzando il False discovery rate (FDR), che è definito come la proporzione di falsi positivi (errori di Tipo I) tra le ipotesi rigettate dal test e quelle che pensiamo siano falsi positivi.

Abbiamo effettuato esperimenti utilizzando entrambi questi tipi di correzioni, dal momento che, nonostante la correzione di Bonferroni sia quella più utilizzata, l'abbiamo ritenuta piuttosto "conservativa" nel caso dei dati analizzati.

Inoltre, abbiamo applicato la correzione per "sotto-famiglie" di dati. Così ad esempio, nel caso dell'analisi di correlazione tra dati di uptake e test cognitivi, i test sono stati suddivisi in "sotto-famiglie" e la correzione è stata applicata ad ogni sotto-famiglia VS i dati di uptake.

Nelle figure 3 e 4 sono riportate ad esempio le correzioni di Bonferroni e di Benjamini-Hochberg rispettivamente, per l'MMSE.

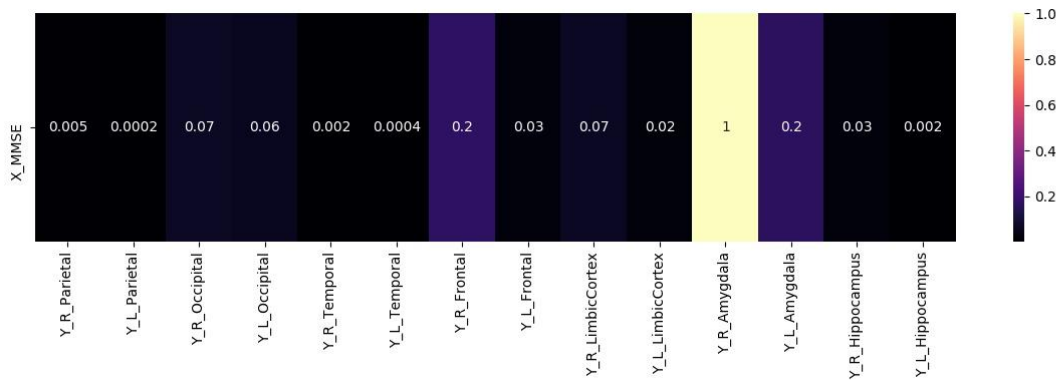


Fig. 3 Correzione di Bonferroni MMSE VS Dati di Uptake

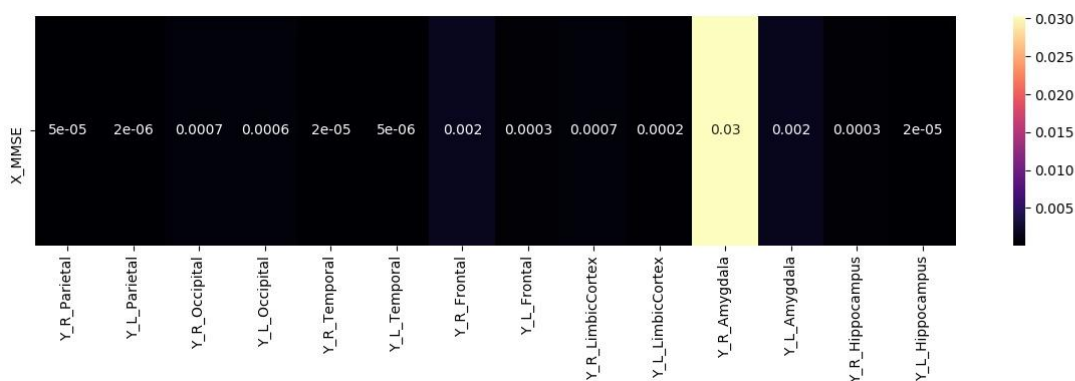


Fig. 4 Correzione di Benjamini-Hochberg MMSE VS Dati di Uptake

Come si può osservare, in entrambi i casi c'è la conferma che R_Parietal e L_Parietal sono altamente correlati. Inoltre, nel caso della correzione di Benjamini-Hochberg solo la R_Amygdala sembra non essere correlata a MMSE.

Negli allegati indicati, sono riportati tutti i valori calcolati con gli indici di correzione sia di Bonferroni che di Benjamini-Hochberg. In ogni caso, nel seguito indicheremo i dati che sono

altamente correlati facendo riferimento alla correzione di Benjamini-Hochberg. In particolare, facendo riferimento al p-value corretto indichiamo i dati di test che hanno un p-value corretto minore di 5^{-4} . In questo caso, quindi avremo che:

- MMSE è correlato con: R_Parietal, L_Parietal, R_Temporal, R_Temporal, L_Frontal, L_LimbicCortex, R_Hippocampus e L_Hippocampus;
- TOKEN_TEST è correlato con tutti tranne Occipital e Amygdala, sia destro che sinistro;
- RAVLT R.I. e RAVLT R.D. sono correlati con tutti tranne Occipital e Amygdala, sia destro che sinistro;
- Tra gli FCSRT, l'ISC è correlato con L_Parietal, mentre l'IFR con R_LimbicCortex, L_Amygdala e con Hippocampus sia destro che sinistro;

Relativamente ai dati provenienti da immagini e in particolare dati di volume non risulta nessuna particolare correlazione nè con i dati dei test, nè con quelli di uptake, prendendo una soglia pari a 5^{-4} . Scegliendo invece una soglia meno conservativa, come ad esempio un classico 0.005, troviamo che relativamente ai dati di uptake, l'lh_parsopercularis risulta correlato con R_Temporal, l'lh_posteriorcingulate risulta correlato con R_Amygdala, l'lh_superiortemporal risulta correlato con R_Hippocampus, e infine lh_temporalpole è correlato con R_parietal.

Considerando questa soglia anche per i dati di test, ritroviamo qualche correlazione anche in questo caso:

- lh_lingual è correlato sia con COWAT che con TOKEN_TEST;
- lh_lateraloccipital è correlato con RAVLT R.I.;
- lh_caudalmiddlefrontal e lh_cuneus sono entrambi correlati con FCSRT ITR;
- lh_lateralorbitofrontal è correlato con BECK II.

6.2. Morbo di Parkinson

Il morbo di Parkinson comprende tutti i test su descritti ed in più una serie di test specifici di questa patologia perché strettamente legati ai disturbi motori tipici della patologia (H&Y test, e UPDRS test).

TIPO	NOME	DESCRIZIONE
DATI GENERALI	età	Età al momento della diagnosi
	sexo	M/F
	educazione	In anni, da 4 a 17.
DIAGNOSI	Diagnosi	Definizione della patologia
	MMSE	
TOKEN TEST (test dei gettoni)	TOKEN	Test di valutazione della comprensione orale. Valore fra 0 e 36.
COWAT (CONTROLLED ORAL WORD)	COWAT	Misura la produzione di parole orali, il risultato è un numero intero.

ASSOCIATION TEST)		
RAVLT (Rey Auditory Verbal Learning Test)	R.I	Valuta le capacità di rievocazione immediata e differita. In genere è un numero con almeno un decimale.
	R.D.	
FCSRT (Free And Cued Selective Reminding Test)	IFR	Sono indici che vengono usati per misurare le capacità di rievocazione della memoria. Ogni test comprende in genere dodici figure per cui il punteggio totale varia fra 0 e 12. I test sono: Rievocazione Libera Immediata (IFR), Rievocazione Immediata Totale (ITR), Rievocazione Differita Libera (DFR), Rievocazione Differita Totale (DTR). L'Indice di Sensibilità al Suggerimento Semantico (ISC) si ricava a partire da IFR e ITR.
	ITR	
	DFR	
	DTR	
	ISC	
DIGIT	FW	Il digit test forward (FW) e backward (BW) serve per misurare il funzionamento della memoria di lavoro. Valore numerico, è il numero massimo di numeri in sequenza ricordati dal paziente in 3 test su 4.
	BW	
FAB		Frontal Assessment Battery. Vengono esplorate le funzioni del lobo frontale in 6 test, il cui valore varia da 0 e 3 per un totale quindi di 18 punti.
STROOP	W	Serve per valutare l'attenzione visiva. Si articola in tre test W (word), C (color), CW (COLOR WORD) prova di interferenza). Gli output dei test sono due, un tempo (la differenza fra il tempo nella prova CW e la media nelle altre due) ed il numero degli errori (terza prova rispetto alla media delle prime due).
	C	
	CW	
WEIGL		Test per valutare la capacità di cogliere somiglianze. I punteggi grezzi sono numeri interi, ma quando si effettuano le correzioni sono numeri con almeno 2 cifre decimali.
JLO V		Judgment of Line Orientation, è un test di elaborazione spaziale. In genere è una stima di quanti task, su un totale di 30, sono stati correttamente risolti.
BECK II		Test dello stato di depressione di un paziente. In genere contiene una serie di test con risposte correlate ad intensità crescenti. Il punteggio totale individua delle fasce di stadiazione (ad esempio per 9 domande con 4 risposte: 0-9, prestazione normale; 10-15 lieve; 16-19 da lieve a moderata; 20-29 = da moderata a grave; 30-36 = depressione grave).
HAMA		Test Hamilton Rating Scale for Anxiety. Misura lo stato d'ansia. E' composta da 14 domande, per ognuna delle quali ci possono essere 4 risposte con punteggi di gravità

		crescente da 0 a 4. Ci sono delle fasce di stadiazione associate (< 17 lieve entità, 18-24 da lieve a moderata; 25-30 da moderata a grave)
H&Y		Valutazione di Hoehn and Yahr dei sintomi della malattia. In ordine crescente da 0 a 5, con valori intermedi 1.5 e 2.5
UPDRS	I	Unified Parkinson's Disease Rating Scale. Misura vari aspetti: I (Attività mentale, comportamento, umore), II (attività della vita giornaliera), III (esame motorio) per cui I valori vanno da 0 a 4 in ordine di gravità crescente; IV (complicazioni della terapia nell'ultima settimana) valori si/no. Il numero di domande in ogni test può variare.
	II	
	III	
	IV	

6.2.1. Dati clinici

Di seguito si affronta il problema di valutare i dati clinici ricevuti dai partner al fine di usarli per le successive fasi di machine learning. Per rendere possibile la selezione delle features e la classificazione su un numero non elevato di campioni a disposizione rispetto al numero di features radiomiche estratte, si è deciso di accorpate e ridurre ad un numero inferiore di classi quei dati i cui valori sono usati come obiettivi di classificazione.

I dati ricevuti dal partner UNICZ comprendono dati clinici, test cognitivi e features radiomiche ma purtroppo non comprendono i dati sulla presenza di beta-amiloide.

DATI UNICZ

I dati clinici includono, oltre a sesso ed età, anche il valore dell'indicatore di educazione, che è fondamentale per il corretto bilanciamento dei risultati ottenuti dai test cognitivi. Questi ultimi comprendono tutti i test descritti nei precedenti documenti di progetto.

Sono inoltre disponibili i dati relativi all'uptake del mezzo di contrasto FDG per le regioni di interesse, e cioè: Lobo Parietale, Lobo Occipitale, Lobo Temporale, Lobo Frontale, Corteccia limbica, Ippocampo ed Amigdala.

6.2.2. Dati da immagini mediche

I dati, estratti tramite il tool Freesurfer [68] da immagini cliniche comprendono:

- area, curvatura media, volume, spessore di tutte le strutture **corticali** sia per l'emisfero destro che per il sinistro (dati freesurfer *apar*)
- volume delle aree **subcorticali** sia per **emisfero destro** che per il **sinistro** (dati freesurfer *aseg*)
- volume delle **sottoregioni** (nuclei) del **Talamo**
- volume delle **sottoregioni** della **Amigdala** e dell'**Ippocampo**
- volume delle quattro **sottoregioni** del **Tronco encefalico**

Le colonne 'Left-WM-hypointensities', 'Right-WM-hypointensities', 'Left-non-WM-hypointensities', 'Right-non-WM-hypointensities' sono state rimosse perché tutte valorizzate a 0.

6.2.3. Correlazione tra dati

Così come per il morbo di Alzheimer, anche nel caso del Parkinson in collaborazione con i partner clinici è stata effettuata un'analisi di correlazione tra un sottoinsieme di dati, particolarmente significativi ai fini diagnostici. In particolare, lo scopo è stato identificare correlazioni specifiche tra 1) dati relativi ai **test cognitivi e motori**, presenti nella tabella del paragrafo 7.2, 2) dati di **uptake** del mezzo di contrasto FDOPA e 3) dati provenienti da **immagini**, in particolare un sottoinsieme identificato dai modelli di classificazione. I sottoinsiemi di dati per cui analizzare eventuali correlazioni sono stati identificati dai partner clinici successivamente ai primi esperimenti effettuati per la classificazione diagnostica. Anche in questo caso, sono stati scelti gli indici di correlazione di Pearson e di Spearman. I risultati sono disponibili come immagini negli allegati al presente documento, da 4 a 6. Per semplicità riportiamo nelle successive figure solo qualche esempio rappresentativo. In figura 5, riportiamo il grafico delle correlazioni di Pearson tra dati test e di uptake.

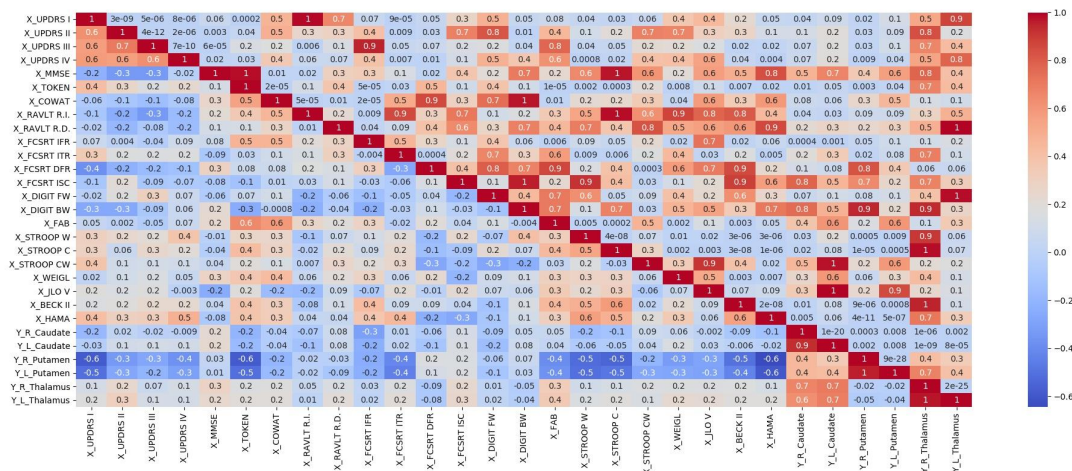


Fig. 5 Correlazione di Pearson tra dati dei test e dati di uptake

In particolare, nella diagonale inferiore sono riportati gli indici di correlazione, mentre nella diagonale superiore sono indicati i corrispondenti *p-value*.

Per semplicità di lettura, tale grafico viene accompagnato da quello presente in figura 6.

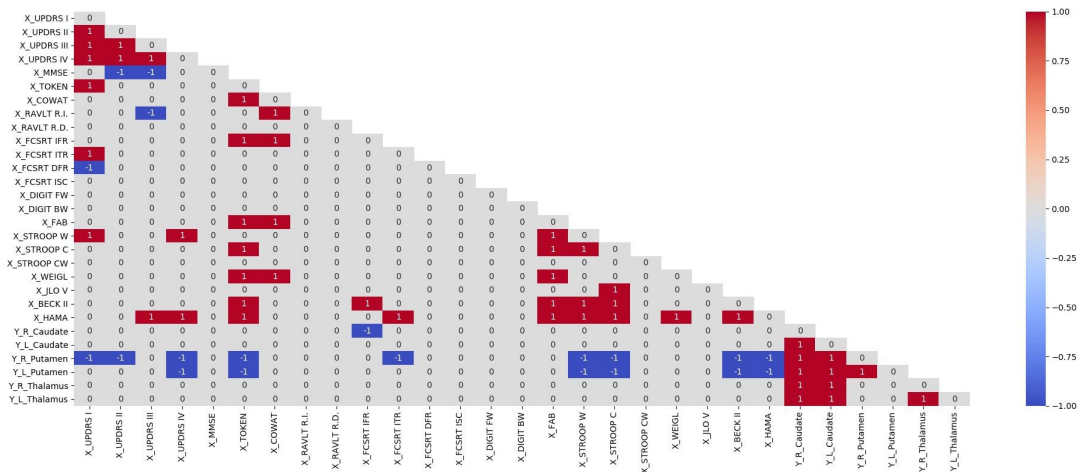


Fig. 6: Correlazione di Pearson tra dati dei test e dati di uptake: valori rilevanti

In tale grafico, il valore 1 (-1) si trova in corrispondenza di indici di correlazione maggiori uguali di 0.3 (minori uguali di -0.3) e con *p-value* minore o uguale a 0.05, valore solitamente scelto come soglia di significanza. In questo modo, i valori 1 (-1) si troveranno in corrispondenza di alta correlazione (anti-correlazione) tra i dati. Ad esempio, dal grafico di figura 6, si evince che R_Putamen è altamente anti-correlato con UPDRS I, UPDRS II e UPDRS IV.

Relativamente alle correzioni di Bonferroni e di Benjamini-Hochberg, riportiamo due grafici di esempio nelle figure successive.

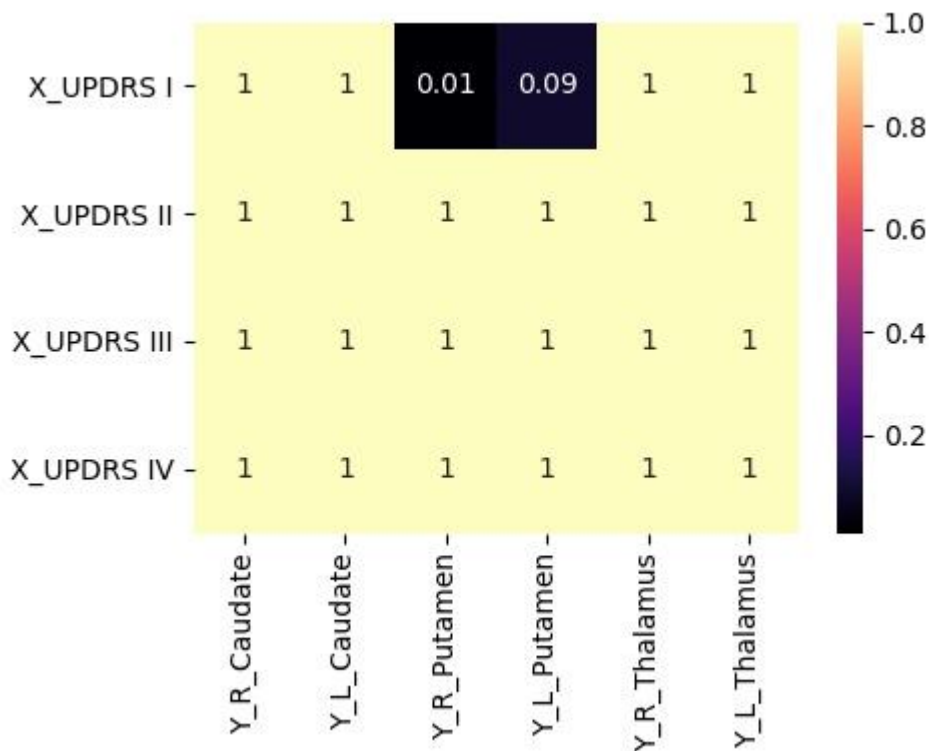


Fig. 7: Correzione di Bonferroni UPDRS VS Dati di Uptake

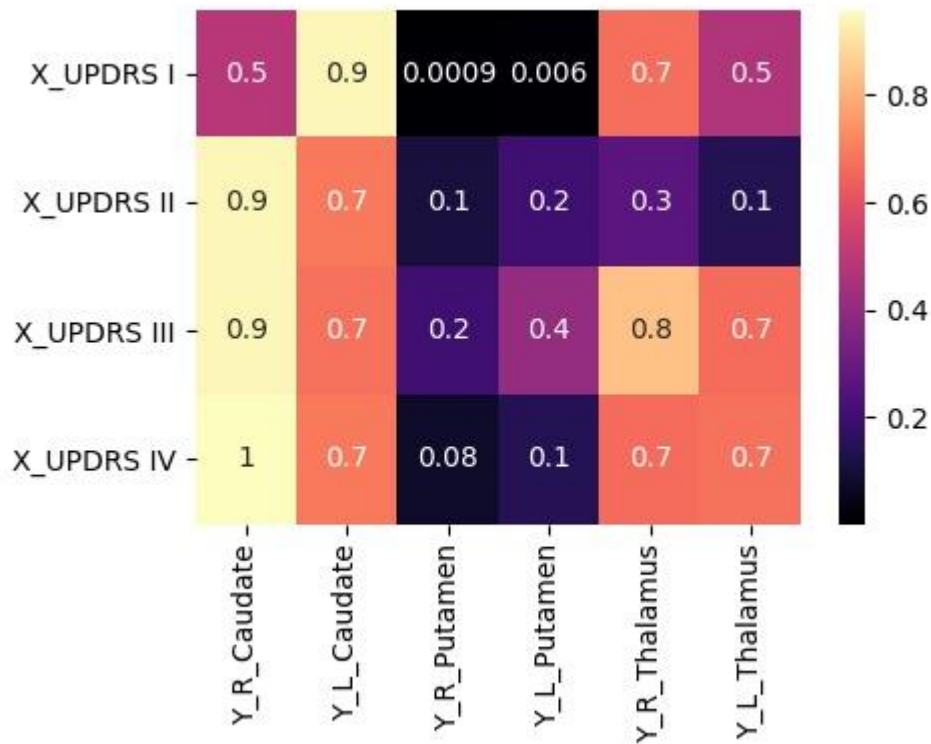


Fig. 8: Correzione di Benjamini-Hochberg UPDRS VS Dati di Uptake

L'ipotesi precedentemente formulata relativamente al R_Putamen correlato con UPDRS I e UPDRS II viene confermata in parte dalla correzione di Benjamini-Hochberg, dove risulta che soltanto UPDRS I ha una forte correlazione con R_Putamen. Inoltre, avendo suddiviso i dati in "sotto-famiglie", viene evidenziata anche una correlazione abbastanza forte tra UPDRS I e L_Putamen.

Sempre facendo riferimento all'indice di correzione di Benjamini-Hochberg, nel caso del Parkinson le correlazioni che vengono fuori coinvolgono soltanto i Putamen e i particolare essi risultano correlati con i test: BECK, UPDRS I, FAB, FCSRT ITR, HAMA, STROOP C, STROOP W e TOKEN.

Relativamente ai test cognitivi e alle loro possibili correlazioni con i dati provenienti da immagini, FCSRT ITR risulta particolarmente correlato con rh_postcentral_thickness, mentre STROOP C risulta correlato con rh_pericalcarine_thickness. Nessuna correlazione è invece emersa tra dati di uptake e dati estratti dalle immagini.

7. Esperimenti preliminari sul dataset TCGA-BRCA

In attesa di ricevere le feature radiomiche e i dati clinici per tutti i pazienti arruolati per il progetto MOLIM, relativamente al cancro al seno, è stato analizzato il dataset di immagini presente nel portale "The Cancer Imaging Archive" (TCIA), in particolare la collezione

Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA) [53,54], contenente 91 campioni. Sono stati effettuati i primi esperimenti in termini di classificazione, correlando feature radiomiche e cliniche.

In particolare, in questa prima fase, tutte le feature radiomiche presenti nel dataset pubblico, insieme con l'età delle pazienti sono state utilizzate per la classificazione.

Come outcome degli esperimenti sono stati utilizzati gli stessi outcome previsti per il progetto in esame e cioè:

- Recettore del progesterone (PR)
- Recettore dell'estrogeno (ER)
- Istotipo
- Stage

Non vengono invece utilizzati né il marcatore Ki67, perché non presente nel dataset pubblico né il marcatore HER2, dal momento che per 39 campioni su 91 tale dato non è disponibile.

Relativamente agli esperimenti sono stati utilizzati i seguenti algoritmi:

- Random Forest (RF) [55]
- K-nearest neighbor (KNN) [56]
- Naive Bayes (NB) [57]
- Support vector machine con kernel non lineare (SVM) [58]
- Multi-layer perceptron (MLP) [59]
- Decision Tree (DT) [60]

La classificazione è avvenuta sia in versione semplice che utilizzando l'approccio cosiddetto "ensemble". In particolare per la versione ensemble è stato utilizzato Adaboost [61].

Dal momento che i campioni sono in numero molto limitato (è stato utilizzato il metodo Synthetic Minority Over-sampling Technique for Nominal and Continuous (SMOTE-NC), che diversamente dal classico SMOTE elabora dati sia numerici che categorici [62, 63]. SMOTE genera osservazioni "sintetiche" a partire dalla classe di minoranza e li aggiunge al set di dati esistenti. I record artificiali della classe di minoranza vengono generati basandosi sulla similarità nello spazio dei predittori. Per ciascun record x_i appartenente alla classe di minoranza vengono create k osservazioni e solo quelle più vicine sono prese come campioni.

Per la strategia di testing è stata utilizzata una 10 Fold Cross validation (10-Fold-CV) e una Leave One Out Cross Validation (LOOCV) [64, 65], essendo tutti pazienti differenti. Vengono quindi di seguito riportati i risultati per gli algoritmi elencati e relativamente alle due strategie di testing utilizzate.

7.1. Marcatore PR

Nel caso del recettore del progesterone, è stata effettuata una classificazione binaria, essendo infatti presenti campioni annotati con la sola informazione sulla presenza/assenza del recettore: PR+ e PR-.

Nella tabella 1 sono riportati i risultati per la 10 Fold Cross validation. Sia nel caso della classificazione di base, che utilizzando un metodo ensemble, l'algoritmo Random Forest (rf) fornisce il miglior risultato in termini di accuratezza, di AUC e di F1 score.

10-Fold-CV					
Basic Algorithm					
algorithm	accuracy	precision	recall	f1score	auc
dt	0,73	0,57	0,57	0,56	0,57
knn	0,48	0,48	0,48	0,41	0,47
mlp	0,59	0,39	0,47	0,39	0,47
nb	0,37	0,54	0,53	0,36	0,65
rf	0,78	0,57	0,60	0,58	0,66
svm	0,47	0,52	0,54	0,43	0,55
Adaboost					
algorithm	accuracy	precision	recall	f1score	auc
nb	0,54	0,43	0,49	0,39	0,58
rf	0,79	0,62	0,63	0,61	0,62
svm	0,47	0,55	0,56	0,44	0,56

Tabella 1: 10-Fold Cross Validation per PR

Nella tabella 2 sono riportati i risultati per la Leave One Out Cross validation. Anche in questo caso, sia relativamente alla classificazione di base, che al metodo ensemble, l'algoritmo Random Forest (rf) fornisce il miglior risultato in termini di accuratezza, di AUC e di F1 score. Naturalmente, nel caso della LOOCV, le misure di performance forniscono lo stesso risultato, dal momento che sono calcolate su un singolo paziente.

LOOCV				
Basic algorithm				
algorithm	accuracy	precision	recall	f1score
dt	0,73	0,73	0,73	0,73
knn	0,53	0,53	0,53	0,53
mlp	0,53	0,53	0,53	0,53
nb	0,42	0,42	0,42	0,42
rf	0,76	0,76	0,76	0,76
svm	0,49	0,49	0,49	0,49
Adaboost				
algorithm	accuracy	precision	recall	f1score
nb	0,71	0,71	0,71	0,71
rf	0,76	0,76	0,76	0,76
svm	0,48	0,48	0,48	0,48

Tabella 2: LOOCV per PR

7.2. Marcatore ER

Anche nel caso del recettore dell'estrogeno, è stata effettuata una classificazione binaria, essendo infatti presenti campioni annotati con la sola informazione sulla presenza/assenza del recettore: ER+ e ER-.

Nella tabella 3 sono riportati i risultati per la 10 Fold Cross validation. In questo caso, l'algoritmo Random Forest (rf) utilizzato come algoritmo di base per la classificazione, fornisce il miglior risultato in termini di accuratezza, di AUC e di F1 score.

10-Fold-CV					
Basic algorithm					
algorithm	accuracy	precision	recall	f1score	auc
dt	0,71	0,48	0,48	0,47	0,48
knn	0,53	0,51	0,54	0,44	0,57
mlp	0,69	0,45	0,61	0,48	0,60
nb	0,37	0,44	0,55	0,34	0,70
rf	0,83	0,64	0,66	0,64	0,70
svm	0,48	0,56	0,59	0,44	0,66
Adaboost					
algorithm	accuracy	precision	recall	f1score	auc
nb	0,58	0,46	0,52	0,40	0,68
rf	0,83	0,58	0,61	0,59	0,71
svm	0,44	0,56	0,59	0,41	0,59

Tabella 3: 10-Fold Cross Validation per ER

Anche nel caso della Leave One Out Cross validation, il Random Forest (rf) è l'algorithmo che fornisce il miglior risultato (vedi tabella 4).

LOOCV				
Basic algorithm				
algorithm	accuracy	precision	recall	f1score
dt	0,70	0,70	0,70	0,70
knn	0,65	0,65	0,65	0,65
mlp	0,71	0,71	0,71	0,71
nb	0,29	0,29	0,29	0,29
rf	0,82	0,82	0,82	0,82
svm	0,54	0,54	0,54	0,54
Adaboost				
algorithm	accuracy	precision	recall	f1score
nb	0,82	0,82	0,82	0,82
rf	0,81	0,81	0,81	0,81
svm	0,47	0,47	0,47	0,47

Tabella 4: LOOCV per ER

7.3. Istotipo

Nel caso dell'istotipo, le classi sono principalmente due, *Infiltrating Ductal* e *Infiltrating Lobular*. Solo per 2 delle 91 pazienti è fornito un istotipo identificato come *Other* e un altro identificato come *Mixed Histology*. Essendo solo 2 i campioni che si differenziano dagli altri, essi sono stati esclusi dalla classificazione, in modo da avere anche in questo caso la possibilità di effettuare una classificazione binaria (*Infiltrating Ductal* vs *Infiltrating Lobular*).

La tabella 5 mostra i risultati nel caso della 10 Fold Cross validation, mentre la 6 nel caso della LOOCV. In questo caso, per entrambe le strategie, il Random Forest (rf) fornisce il

miglior risultato quando viene utilizzato all'interno di un ensemble (nel caso della 10 Fold CV, viene presa la AUC maggiore).

10-Fold-CV					
Basic algorithm					
algorithm	accuracy	precision	recall	f1score	auc
dt	0,75	0,46	0,44	0,45	0,44
knn	0,51	0,47	0,41	0,38	0,43
mlp	0,44	0,26	0,49	0,28	0,66
nb	0,38	0,51	0,52	0,35	0,66
rf	0,84	0,51	0,61	0,54	0,76
svm	0,19	0,34	0,49	0,18	0,51
Adaboost					
algorithm	accuracy	precision	recall	f1score	auc
nb	0,66	0,49	0,50	0,42	0,57
rf	0,84	0,52	0,61	0,54	0,79
svm	0,19	0,34	0,49	0,18	0,49

Tabella 5: 10-Fold Cross Validation per Istotipo

LOOCV				
Basic algorithm				
algorithm	accuracy	precision	recall	f1score
dt	0,74	0,74	0,74	0,74
knn	0,48	0,48	0,48	0,48
mlp	0,69	0,69	0,69	0,69
nb	0,38	0,38	0,38	0,38
rf	0,81	0,81	0,81	0,81
svm	0,20	0,20	0,20	0,20
Adaboost				
algorithm	accuracy	precision	recall	f1score
nb	0,27	0,27	0,27	0,27
rf	0,84	0,84	0,84	0,84
svm	0,20	0,20	0,20	0,20

Tabella 6: LOOCV per Istotipo

7.4. Stage

Nel caso dello stage, i dati di partenza mostrano 7 differenti valori per lo stage:

1. Stage I
2. Stage IA
3. Stage II
4. Stage IIA

5. Stage IIB
6. Stage IIIA
7. Stage IIIC

Dal momento che i campioni sono solo 91, per poter ottenere una classificazione quantomeno accettabile, è stato necessario “accorpare” le classi. In particolare, le classi sono state identificate in questo modo:

1. **Stage I**, che comprende Stage I e Stage IA
2. **Stage II**, che comprende Stage II, Stage IIA e Stage IIB
3. **Stage III**, che comprende Stage IIIA e Stage IIIC.

In questo caso, è stata quindi effettuata una classificazione a 3 classi.

10-Fold-CV				
Basic algorithm				
algorithm	accuracy	precision	recall	f1score
dt	0,44	0,38	0,35	0,35
knn	0,30	0,28	0,30	0,25
mlp	0,38	0,33	0,37	0,30
nb	0,27	0,29	0,39	0,26
rf	0,46	0,28	0,32	0,29
svm	0,28	0,27	0,22	0,21
Adaboost				
algorithm	accuracy	precision	recall	f1score
nb	0,35	0,30	0,37	0,27
rf	0,48	0,28	0,33	0,29
svm	0,27	0,27	0,22	0,22

Tabella 7: 10-Fold Cross Validation per Stage

In tabella 7 sono mostrati i risultati relativi alla 10 Fold Cross validation. Come si può osservare, in questo caso non viene raggiunto neanche il 50% in termini di accuratezza per nessuno degli algoritmi utilizzati. Il motivo è dovuto al fatto che la 10-Fold Cross validation utilizzata con 91 campioni effettua il training su 81 pazienti e il test su 10 pazienti e probabilmente 81 pazienti, nonostante l'utilizzo di un algoritmo di generazione di osservazioni "sintetiche" è un numero troppo limitato per poter effettuare un training a 3 classi.

I risultati si fanno invece più interessanti nel caso della Leave One Out Cross validation (vedi tabella 8). In questo caso una SVM a più classi permette di ottenere delle performance intorno al 60% di accuratezza.

LOOCV				
Basic algorithm				
algorithm	accuracy	precision	recall	f1score
dt	0,43	0,43	0,43	0,43
knn	0,29	0,29	0,29	0,29
mlp	0,42	0,42	0,42	0,42
nb	0,30	0,30	0,30	0,30
rf	0,46	0,46	0,46	0,46
svm	0,60	0,60	0,60	0,60
Adaboost				
algorithm	accuracy	precision	recall	f1score
nb	0,46	0,46	0,46	0,46
rf	0,48	0,48	0,48	0,48
svm	0,60	0,60	0,60	0,60

Tabella 8: LOOCV per Stage

7.5. Discussione

Dai risultati ottenuti fino a questo momento, sembra chiaro che 91 campioni sono un numero troppo limitato per poter ottenere delle accuratze interessanti in termini di classificazione dei biomarcatori molecolari. Comunque i risultati più interessanti si ottengono nel caso della classificazione binaria e in particolare è l'istotipo a dare i risultati più promettenti.

In termini di algoritmi invece, il Naive Bayes e l'SVM forniscono quasi sempre i risultati peggiori, esclusa la classificazione a tre classi per lo Stage, dove è proprio SVM a mostrare i risultati migliori.

A questo punto si intende procedere ad uno studio più approfondito della tipologia delle feature radiomiche da utilizzare per la classificazione, dal momento che alcune delle feature usate a questo scopo potrebbero introdurre rumore, e quindi contribuire in maniera negativa al risultato della classificazione. Infatti, anche in letteratura sono presenti una serie di studi che approfondiscono l'analisi delle feature più significative in termini di classificazione [66, 67].

8. Esperimenti di classificazione su dati reali di progetto

8.1. Approcci di selezione delle features e classificazione

I dati raccolti nel progetto (e analizzati nel dettaglio, per ogni patologia, nei paragrafi a seguire) possiedono caratteristiche tali da necessitare di un attento approccio metodologico alla fase di apprendimento supervisionato.

Bilanciamento delle classi e feature selection

Per ovviare al problema dello sbilanciamento delle classi si è usato un approccio di letteratura chiamato *SMOTE* (Synthetic Minority Over-sampling Technique) [62,63], tramite cui vengono generati, in fase di apprendimento, un certo numero di campioni appartenenti alle classi sottorappresentate in modo da ottenere un dataset perfettamente bilanciato in termini di numero di campioni per ogni classe.

Inoltre, il numero di *features* per ogni paziente è molto elevato rispetto al numero di campioni, in tutte le patologie considerate. Si è scelto dunque di effettuare un passo di *feature selection* (FS) prima della classificazione, in modo da individuare soltanto le features più rilevanti per l'apprendimento ed evitare gli effetti del cosiddetto “curse of dimensionality”.

E' ormai noto, dalla letteratura, che non esiste un singolo algoritmo di feature selection che sia il migliore in assoluto, ma che è necessario sperimentare e trovare quale è l'algoritmo che produce i risultati migliori per uno specifico dataset. A questo scopo si è scelto di considerare, fra i diversi algoritmi esistenti, il sottoinsieme descritto in Tabella 9.

Algoritmo	Acronimo di progetto	Tipo (Ranker/Subset)	Approccio (Filter/Embedded)
Chi squared	CHI	Ranker	Filter
Fisher score	FS	Ranker	Filter
Gini index	GI	Ranker	Filter
Mutual Information	MI	Ranker	Filter
RelieFF	RF	Ranker	Filter
LR-RFE	LR	Ranker	Embedded
CFS	CFS	Subset	Filter

Tabella 9. Elenco degli algoritmi di Feature Selection usati.

Si è scelto di considerare diversi approcci di feature selection basati su metodi di tipo *filter*, che restituiscono un ranking (in genere completo, se si esclude il metodo Mutual Information) delle features perché sono un ottimo compromesso fra bontà delle soluzioni trovate e tempi di calcolo. Sono poi stati considerati due algoritmi più pesanti in termini di tempi di calcolo ma che forniscono soluzioni in genere più accurate e cioè Recursive Feature Elimination con Lasso Regression (LR-RFE) che è un metodo embedded, e CFS (Correlation-Based feature selection) che è un metodo basato sulla correlazione e valuta sottoinsiemi di features (risultando per questo motivo il più lento di tutti gli algoritmi considerati).

8.2. Classificazione e validazione

Anche in questo caso vale la stessa considerazione fatta per gli algoritmi di feature selection: non esiste un algoritmo di classificazione universalmente efficace, ma bisogna cercare sperimentalmente quale fornisce i risultati migliori per il dataset considerato. Si deve inoltre tenere conto che il sottoinsieme di features estratte al passo precedente influisce sui risultati ottenuti con uno specifico algoritmo. Sono stati presi in considerazione diversi approcci, elencati in tabella 10, fra quelli più noti ed usati in letteratura per la loro efficacia.

Algoritmo	Acronimo di progetto
Decision Tree	DT
K-nearest-neighbor	KNN
Multi Layer Perceptron	MLP
Naive Bayes	NB
Random Forest	RF
Support Vector Machines	SVM

Tabella 10. Elenco degli algoritmi di classificazione usati.

8.2.1. Classificazione

Come già detto, non esiste un metodo univoco ed affermato che sia possibile usare come riferimento per trattare i dati radiomici usati in questo progetto. Il primo obiettivo da perseguire è quello della riduzione del numero di features usate per la classificazione. A questo scopo si è deciso per un approccio a cascata in cui una prima selezione delle features viene effettuata usando i metodi che restituiscono un ranking completo (Chi, Fisher, Gini index, ReliefF). Si estraggono da questi ranking un sottoinsieme n predefinito di features, con $n=[5,10,15,20,25,40,45,50]$, che vengono poi date in input a tutti gli algoritmi di classificazione. In base al migliore risultato ottenuto dai classificatori (valutato in termini di f1score) si individuano le features più rilevanti: le features vengono aggregate sui diversi fold (si veda il paragrafo sulla Validazione dei dati) usando come misura la minima posizione ottenuta dalla singola feature nei vari fold ed il numero totale di fold in cui tale posizione è stata ricoperta. Le features vengono poi ordinate per rilevanza in base a questo criterio, e vengono estratte tutte quelle che almeno una volta hanno avuto minima posizione

n . Su questo sottoinsieme si ripete il procedimento di classificazione, usando però gli altri metodi di selezione delle features (Mutual Information, CFS, LR-RFE) con un insieme m più ristretto di soglie per il numero di features da selezionare, con $m=[1, \dots, 10]$. Anche in questo caso si sceglie il risultato migliore in termini di f1score e si individuano, applicando lo stesso criterio su descritto, la combinazione di features più promettenti e algoritmo di classificazione. Queste informazioni vengono usate per addestrare e generare un modello di predizione che sarà poi validato (laddove possibile) usando dati provenienti da un diverso dataset. Il modello di predizione così ottenuto, se sufficientemente efficiente, sarà usato nella costruzione della piattaforma. Il processo a due fasi è illustrato in figura 1.

Questo tipo di approccio è stato usato per la gran parte dei dati di progetto, escluso i dati di tumore al colon, su cui si è sperimentato un approccio lievemente differente, anche se impostato sullo stesso principio di un affinamento successivo delle features selezionate.

Per i dati di tumore al colon, infatti, si è voluto indagare un approccio basato su Ensemble Feature Selection. La ricerca di features tramite metodi ensemble è basata sulla supposizione che features molto rilevanti per il dominio in esame dovrebbero essere scelte da più algoritmi di features selection, e quindi considerare i risultati di più selettori dovrebbe portare alla individuazione di sottoinsiemi più stabili di features. A questo scopo si considerano i ranking ottenuti dai quattro metodi che restituiscono un ordinamento completo delle features in input (Chi, Fisher, Gini index, ReliefF). Anche in questo caso si estraggono da questi ranking un sottoinsieme n predefinito di features, con $n=[5, 10, 15, 20, 25, 40, 45, 50]$, che vengono poi date in input a tutti gli algoritmi di classificazione. In base al migliore risultato ottenuto dai classificatori (valutato in termini di f1score) si individuano le features più rilevanti: le features vengono aggregate sui diversi fold (si

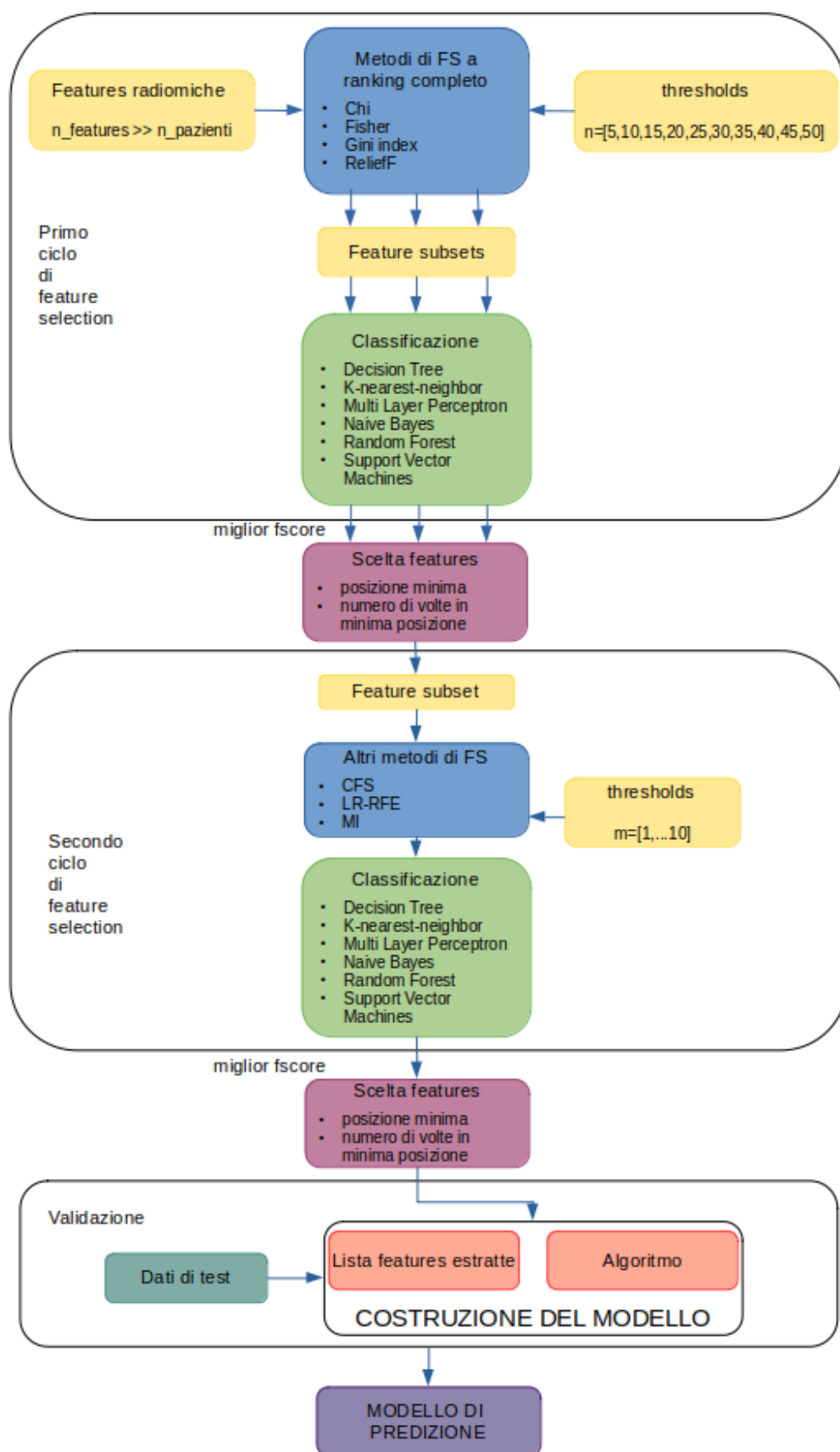


Figura 1: Processo di selezione e di classificazione per Cancro alla Mammella, Morbo di Parkinson e Morbo di Alzheimer

veda il paragrafo sulla Validazione dei dati) usando come misura la minima posizione ottenuta dalla singola feature nei vari fold ed il numero totale di fold in cui tale posizione è stata ricoperta. Per ognuno dei metodi ranker suindicati sono state selezionate tutte le features con minima posizione $p=\{1,2\}$ per un totale di 20 features, alcune delle quali comuni ai vari metodi. Una controindicazione dell'approccio di Ensemble Feature Selection è che nella fase di aggregazione delle features è possibile che diversi metodi scelgano diverse features fra loro fortemente correlate. A questo scopo viene calcolata la correlazione di Pearson tra le diverse features. Tutte le features che hanno correlazione reciproca $c > 0.9$ vengono eliminate, tranne una (selezionata a caso) per ogni gruppo. Per ridurre ulteriormente la presenza di features ridondanti si effettua un passo ulteriore di feature selection usando il metodo CFS. Le features così individuate vengono poi usate per un ultimo ciclo di feature selection tramite metodi a ranking non completo (LR-RFE, MI). Anche in questo caso si sceglie il risultato migliore in termini di f1score e si individua quindi il sottoinsieme di features più promettenti e l'algoritmo di classificazione. Queste informazioni vengono usate per addestrare e generare un modello di predizione che, se sufficientemente efficiente, sarà usato nella costruzione della piattaforma. Il processo di filtraggio delle features è rappresentato in figura 2.

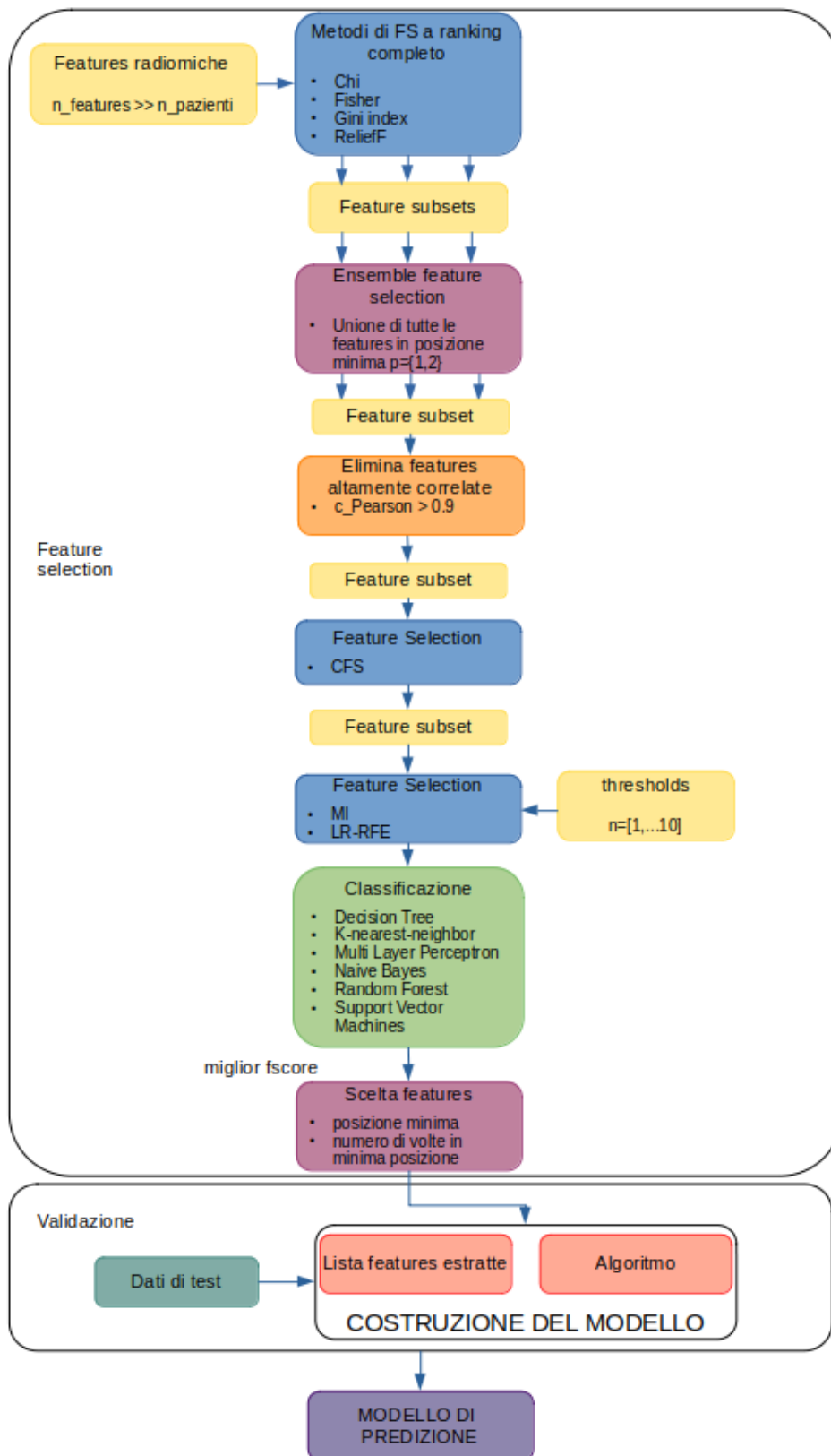


Figura 2: Processo di selezione e classificazione per Cancro al Colon

8.2.2. Validazione

Per poter effettuare la validazione dei risultati ottenuti si è scelto di procedere come segue, data l'esiguità del numero di campioni in input: si è utilizzato un approccio di Ten-Fold Cross Validation nelle fasi di feature selection in cui determinare quali features usare. Una volta determinato il sottoinsieme di features più rappresentative si è usato un approccio di Leave One Out Validation per determinare l'efficacia del sottoinsieme di features estratte e l'algoritmo usato per addestrare il modello finale.

8.3. Implementazione e testing

Si è scelto Python come linguaggio in cui implementare il codice per il machine learning. Tale scelta è stata pressoché immediata in quanto il linguaggio è: 1) non proprietario; 2) supportato da innumerevoli librerie per il calcolo scientifico e per il machine learning (quali *numpy*, *SciPy*, o *scikit-learn*), che sono uno *standard de facto* nella comunità scientifica. La gran parte degli algoritmi e degli approcci descritti nei paragrafi precedenti sono supportati dalle già citate librerie di Python, mentre per alcuni degli algoritmi di feature selection si è usata la libreria *scikit-feature* [10] sviluppata dalla Arizona State University.

Il testing degli algoritmi è stato eseguito in diverse condizioni:

1. con tutte le features disponibili, cioè senza effettuare il passo di *Feature Selection* allo scopo di stabilire un riferimento generico sui risultati che si otterrebbero senza selezione.
2. effettuando la Feature Selection e variando il numero di features estratte per individuare il sottoinsieme più rilevante;
3. usando tutte le features da immagini assieme, oppure tenendo separate le diverse modalità (ADC, SUB, PC, T2);
4. usando tutti, una parte o nessun dato clinico assieme ai dati radiomici, con l'obiettivo di esplorare la possibilità di individuare insiemi ridotti di features, con il fine ultimo di semplificare le pratiche cliniche.

8.4. Tumore al seno

L'obiettivo finale di questa porzione di progetto è la costruzione di modelli di machine learning che, opportunamente addestrati, siano di supporto alla diagnosi. Al termine dello studio effettuato si è determinato che definire dei modelli che abbiano come outcome il tipo istologico è purtroppo impossibile con i dati in possesso, poiché praticamente tutti i pazienti risultano essere di uno stesso sottotipo. Si è determinato che è invece possibile usare come outcome il valore dei markers

molecolari. L'obiettivo in questo caso sarebbe di evitare l'analisi istologica deducendo le informazioni direttamente dalle immagini RMI.

Le features cliniche utilizzabili (cioè con valori definiti per tutti gli outcome) sono:

- età
- dimensioni tumore (considerato come area ottenuta moltiplicando le due dimensioni fornite)

Il numero totale di pazienti per cui i valori dei markers sono presenti è: 80 per i dati del partner SDN, 19 per i dati del partner UNIME.

8.4.1. Distribuzione dei dati

DATI SDN

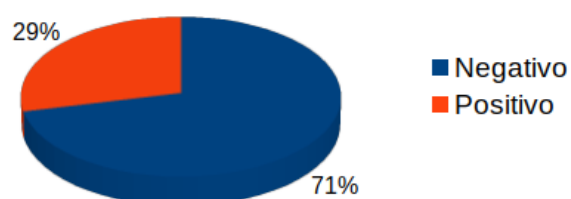
Nell'ambito della classificazione di marker molecolari la suddivisione dei valori è lievemente differente rispetto alle soglie usate nella clinica. I valori usati per distinguere le varie classi sono riportate nella seguente tabella.

Marker	Marker Negativo	Marker Positivo
ER	< 10%	>=10%
HER2	= 0	>1
KI67	<14%	>=14%
PR	< 10%	>=10%

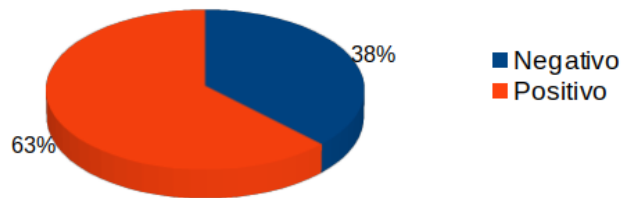
Quindi per i quattro marker molecolari la distribuzione delle classi è riportata nei grafici a seguire.

In generale le classi sono sbilanciate, in alcuni casi in maniera molto forte, come ad esempio nel caso di HER2.

Marker HER2 - distribuzione delle classi

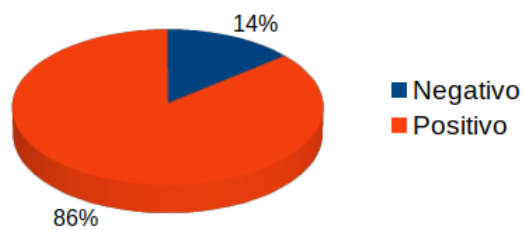


Marker ER - distribuzione delle classi



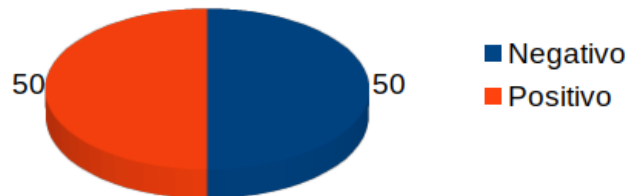
Nel caso del marker KI67 questa condizione è particolarmente evidente.

Marker KI67 - distribuzione delle classi



Invece nel caso del marker PR la distribuzione è perfettamente bilanciata.

Marker PR - distribuzione delle classi

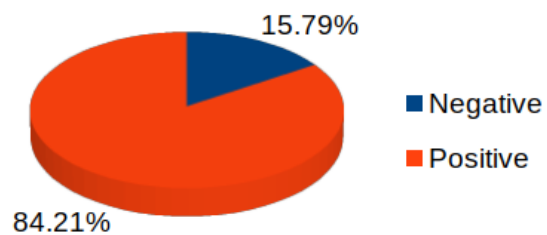


DATI UNIME

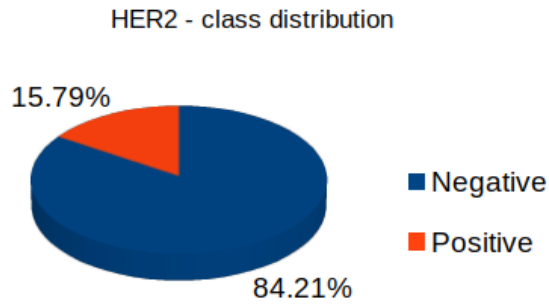
Anche per i dati del partner UNIME si riporta, per i quattro marker molecolari, la distribuzione delle classi.

Nel caso di ER i dati sono molto sbilanciati a favore della classe dei positivi.

ER - class distribution

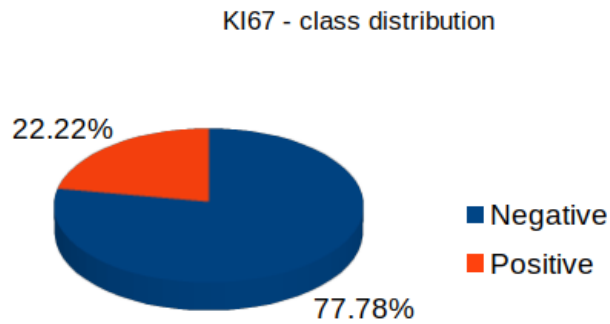


Discorso simile per HER2, in cui a prevalere sono invece i negativi.



Anche per
sbilanciamento verso la classe dei negativi.

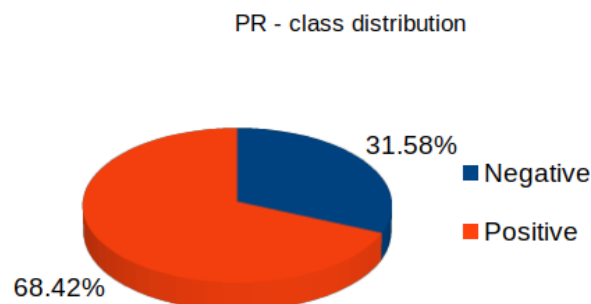
KI67 si ha un forte



Per PR la situazione

è appena migliore.

Per la fase di apprendimento sono stati usati solo i dati di SDN, in modo da partire da un insieme omogeneo (rispetto alla macchina usata per l'acquisizione delle immagini). Per i dati di UNIME non è stato possibile farlo perchè sono in numero ridotto, e mal distribuiti nelle varie classi. I dati di UNIME sono stati usati per effettuare i test sui modelli di predizione generati.



8.4.2. Preprocessing

In questo paragrafo si descrivono tutti i passaggi che vengono effettuati prima della classificazione. I dati in input sono costituiti dalle *features* radiomiche, descritte nei precedenti documenti, estratte dalle immagini di risonanza magnetica (RM) dei pazienti. Per ogni RM di un singolo paziente sono disponibili quattro diverse tipologie di immagini, da ognuna delle quali viene estratta una diversa informazione riguardante la lesione. Le tipologie sono: *T2-weighted* (T2), *apparent diffusion coefficient* (ADC), *post contrast* (PC), e *subtraction* (SUB). Nel caso delle immagini SUB e PC sono state estratte immagini a diversi tempi dopo la somministrazione del mezzo di contrasto. Per scegliere quale delle immagini usare, per le PC è stato scelto il tempo t in cui il valore di assorbimento medio del mezzo di contrasto è massimo (valore della colonna *original_firstorder_Mean*). Di conseguenza, nelle immagini SUB è stato preso il corrispondente tempo $t-1$. Per ogni tipologia di immagine sono state estratte 1197 features che sono state usate prima separatamente e poi considerate tutte insieme, avendo quindi per ogni paziente un totale di 4788 features. Ad esse si aggiungono i dati clinici descritti precedentemente. Prima di essere elaborati i dati sono stati riscaldati nell'intervallo [0,1].

8.4.3. Risultati della classificazione e della validazione

I modelli predittivi costruiti sono stati valutati mediante analisi delle performance di classificazione dei pazienti sulla base degli outcome clinici considerati. In particolare, il potere di classificazione di tali modelli è stato valutato considerando i seguenti parametri:

- Sensibilità, la proporzione dei veri positivi;
- Specificità, la proporzione dei veri negativi;
- Accuratezza, la proporzione dei pazienti correttamente classificati.

Nel contempo, i risultati dell'analisi sulle possibili fonti di variabilità che potessero influire sulla misura delle features radiomiche utilizzate per la costruzione dei modelli predittivi sono stati oggetto di un articolo scientifico di recente pubblicazione [73].

Per la definizione del sottotipo molecolare si è scelto di considerare separatamente i quattro marker coinvolti, in quanto il numero di pazienti in input non consente la classificazione diretta sulle classi esistenti. La combinazione dei valori ottenuti per i singoli marker ER, HER2, KI67, PR potrà dare indicazioni sul sottotipo molecolare così come definito nella tabella del paragrafo 1.2.

Come descritto nei paragrafi precedenti, il testing relativo alla prima fase di feature selection è effettuato prendendo in considerazione tutte le possibili combinazioni di dati. Tutti i risultati sono disponibili come allegati al presente documento (allegato 7).

Di seguito vengono riportati i risultati più significativi (in termini di f1score) per il passo finale di classificazione con validazione Leave-One-Out (tabella 1). Le features individuate per i diversi markers sono disponibili come allegato al presente documento (allegato 8). Il modello di predizione, costruito a partire dalle features e dall' algoritmo individuati, è stato poi testato sui dati del partner UNIME, non usati per la fase di addestramento. I risultati dei test sono riportati in tabella 2.

Risultati della classificazione LOO per I markers molecolari.								
Marker	Tipo features	Tipo immagine	algoritmo di classificazione	numero di features	accuracy	precision	recall	f1score
ER	radiomiche da singola immagine	T2	SVM	11	0.8125	0.872	0.82	0.845
HER2	radiomiche	T2/ADC	RF	5	0.8	0.667	0.609	0.636
KI67	radiomiche da singola immagine	PC	MLP	2	0.846	0.966	0.851	0.905
PR	radiomiche da singola immagine	PC	MLP	3	0.725	0.725	0.725	0.725

Tabella 1

Risultati dei test sui dati UNIME						
Marker	numero pazienti	per classe (N/P)	accuracy	precision	recall	f1score
ER	19	P=16	0.7895	0.8333	0.9375	0.8823
		N=3	0.7895	0	0	0
HER2	19	N=16	0.8421	0.8421	1	0.91
		P=3	0.8421	0	0	0
KI67	18	N=14	0.778	0.75	0.92	0.85
		P=4	0.778	0.5	0.75	0.6

PR	19	N=6	0.6842	0	0	0
		P=13	0.6842	0.684	1	0.812

Tabella 2

8.5. Tumore al colon

Si è determinato che è possibile usare come outcome per questi dati solo la possibile presenza di linfonodi, in quanto i dati acquisiti sono solo di pazienti con patologia e per altri obiettivi possibili (ad esempio grading e staging) non tutti i pazienti hanno informazione. L'obiettivo in questo caso sarebbe di predire il coinvolgimento dei linfonodi direttamente dalle immagini (N0/N+).

L'unica features clinica utilizzabile (cioè con valori definiti per tutti gli outcome) è:

- età

Il numero totale di pazienti forniti dal partner UNIMIB è 52.

8.5.1. Distribuzione dei dati

La distribuzione delle classi è riportata nel grafico che segue.



Purtroppo le classi sono molto sbilanciate.

8.5.2. Preprocessing

I dati in input sono costituiti dalle *features* radiomiche, descritte nei precedenti documenti, estratte dalle immagini dei pazienti. Per ogni immagine sono state estratte 168 features. Prima di essere elaborati i dati sono stati riscaldati nell'intervallo [0,1].

8.5.3. Risultati della classificazione e della validazione

Come descritto nei paragrafi precedenti, il testing relativo alla prima fase di feature selection è effettuato prendendo in considerazione tutte le possibili combinazioni di dati. Tutti i risultati sono disponibili come allegati al presente documento (allegato 9).

Di seguito vengono riportati i risultati più significativi (in termini di f1score) per il passo finale di classificazione con validazione Leave-One-Out (tabella 3). Le features individuate sono disponibili come allegato al presente documento (allegato 10). Il modello di predizione, costruito a partire dalle features e dall' algoritmo individuati, è stato poi testato su di un singolo paziente non inserito nel dataset di creazione del modello. Su questo paziente il modello ha dato il risultato corretto.

Risultati della classificazione LOO per la diagnosi di coinvolgimento dei linfonodi in pazienti affetti da tumore al colon							
Tipo	Tipo features	algoritmo di classificazione	numero di features	accurac y	precision	recall	f1score
N0/N+	radiomiche	MLP	7	0.804	0.897	0.854	0.875

Tabella 3

8.6. Morbo di Alzheimer

Al termine dello studio effettuato sui dati si sono individuate due possibili linee di indagine. Sui dati di PET-FDG del partner UNICZ si è determinato che è possibile usare come obiettivo della classificazione la presenza/assenza di malattia e, in quest'ultimo caso, diagnosticarne la progressione determinando se è monolaterale/bilaterale. Sui dati PET-AMY del partner SDN si è deciso, nonostante l'esiguità dei campioni, di provare a determinare se i pazienti sono positivi/negativi al marker della beta amiloide. Il partner UNIME ha messo inoltre a disposizione un piccolo dataset PET-FDG e PET-AMY su cui si è stabilito di effettuare i test dei modelli.

Si è deciso in tutti i casi di NON usare i dati provenienti dai test cognitivi, in modo da affidarsi esclusivamente alle informazioni provenienti da immagini e dai dati clinici di base, con l'obiettivo di rendere più rapida ed agevole la fase di supporto alla diagnosi. Infatti, in questa configurazione,

si potrebbe prescindere dalla somministrazione dei test cognitivi, operazione in genere estremamente lunga, ed affidarsi unicamente al responso dell'imaging.

Le features cliniche utilizzabili (cioè con valori definiti per tutti gli outcome) sono:

- età
- sesso
- grado di istruzione

Inoltre sono disponibili, per i dati UNICZ e UNIME, anche i valori di uptake del mezzo di contrasto in diverse regioni cerebrali dell'emisfero destro e sinistro (Lobo Parietale, Occipitale, Temporale, Frontale, Corteccia Limbica, Ippocampo, Amigdala).

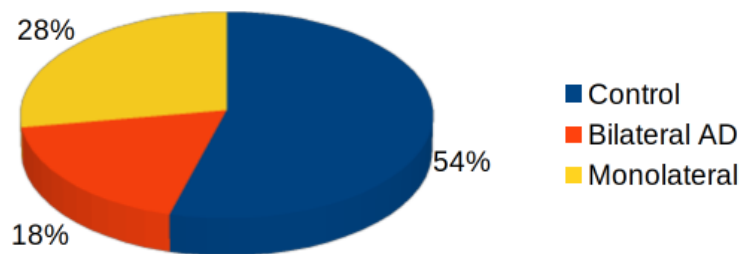
Il numero totale di pazienti FDG di UNICZ è 94, per SDN ne sono disponibili 27 per AMY e 19 per FDG, mentre UNIME ne ha 4 AMY e 4 FDG.

8.6.1. Distribuzione dei dati

DATI UNICZ FDG

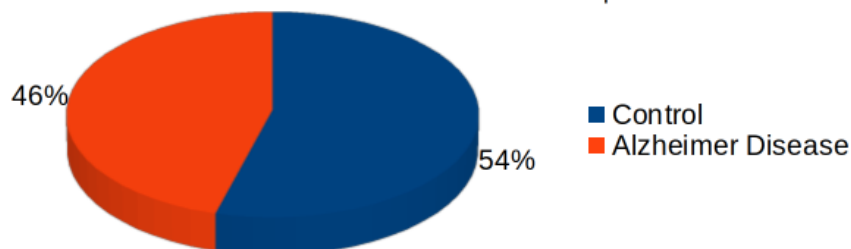
I dati presenti sono suddivisi in casi di controllo/malato, e per i malati è presente anche l'indicazione sulla tipologia della malattia, cioè se monolaterale o bilaterale.

Alzheimer Disease - distribuzione delle classi per sano/AD Mono/AD Bilaterale



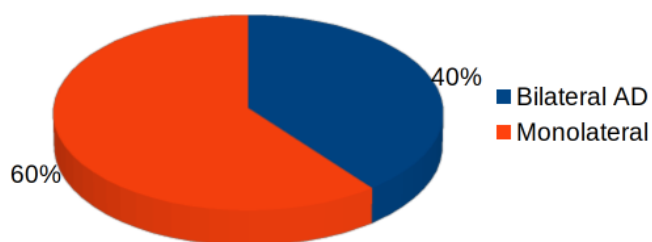
Dal grafico in figura si vede che i dati sono sbilanciati, anche se in modo lieve, con una predominanza dei casi di controllo rispetto ai malati.

Alzheimer Disease - distribuzione delle classi per sano/malato



Per quanto riguarda la distribuzione nelle due classi monolaterale/bilaterale si ha che la prima classe è più presente della seconda, e dunque la mancanza di bilanciamento è più sensibile (Figura 2).

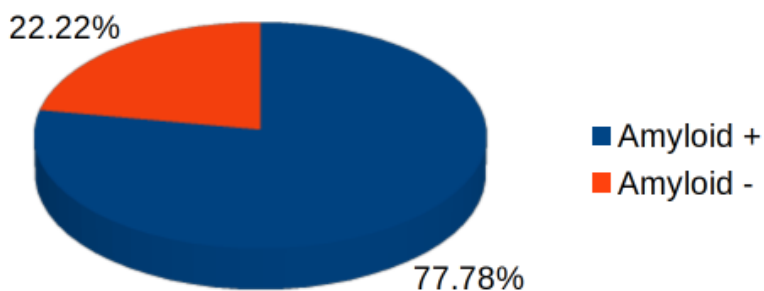
Alzheimer Disease - distribuzione delle classi per AD Mono/Bilaterale



DATI SDN AMY

I dati SDN sono fortemente sbilanciati, con la classe degli amiloide positivo che è preponderante.

Morbo di Alzheimer - Distribuzione delle classi per test amiloide



8.6.2. Preprocessing

I dati in input sono costituiti dalle *features* di volume, spessore, etc. come descritte nei precedenti paragrafi, estratte dalle immagini multimodali PET/RMI dei pazienti. Si ribadisce la scelta di non usare le features cliniche da test cognitivi, al fine di permettere una diagnosi legata il quanto più

possibile ai soli dati clinici essenziali e alle immagini mediche. Le features sono in totale 461 per ogni paziente, sia per i dati UNICZ che per i dati SDN ed UNIME.

Prima di essere elaborati i dati sono stati riscaldati nell'intervallo [0,1].

8.6.3. Risultati della classificazione e della validazione

DISCRIMINAZIONE CONTROLLO/MALATO E MONOLATERALE/BILATERALE (DATI UNICZ FDG)

Dato il numero ristretto di pazienti, si è deciso di separare il problema a tre classi, e si effettua in due passi la distinzione fra sano e malato, e quella fra AD monolaterale e bilaterale.

Di seguito vengono riportati i risultati più significativi e descritto l'approccio che ha portato alla selezione dei modelli per ognuno dei problemi di classificazione considerati. Come descritto nei paragrafi precedenti, il testing relativo alla prima fase di feature selection è effettuato prendendo in considerazione tutte le possibili combinazioni di dati. Tutti i risultati sono disponibili come allegati al presente documento (allegato 11).

In tabella 4 vengono riportati i risultati più significativi (in termini di f1score) per il passo finale di classificazione con validazione Leave-One-Out. Le features individuate per i diversi markers sono disponibili come allegato al presente documento (allegato 12).

Risultati della classificazione LOO per la diagnosi di Alzheimer e valutazione mono/bilaterale							
Tipo	Tipo features	algoritmo di classificazione	numero di features	accuracy	precision	recall	f1score
Controllo/Malato	radiomiche /uptake	DT	1	0.894	0.902	0.86	0.881
Controllo/Malato	radiomiche	MLP	6	0.851	0.854	0.814	0.834
Monolaterale/Bilaterale	radiomiche /uptake	RF	23	0.791	0.815	0.846	0.830
Monolaterale/Bilaterale	radiomiche	SVM	12	0.674	0.75	0.692	0.72

Tabella 4

Per entrambi gli obiettivi di classificazione, sono riportati i risultati che si ottengono usando come features di addestramento solo quelle provenienti da immagini o la combinazione di queste con le features di uptake. Nella distinzione fra controlli e malati, l'uso delle features di uptake migliora le prestazioni e riduce il numero delle features. Accade lo stesso nel caso della distinzione fra AD monolaterale e bilaterale, con la differenza che il numero di features usate è più alto.

DISCRIMINAZIONE PAZIENTI AMYLOIDE NEGATIVO/POSITIVO (DATI SDN e UNIME)

Si precisa qui che, a differenza di tutti gli altri dataset di tutto il progetto, per questo dataset si è dovuto effettuare la classificazione tramite cross fold validation usando tre fold, invece che i dieci solitamente considerati, perché il numero di campioni a disposizione è troppo ridotto per poter fare altrimenti. Inoltre, per la stessa ragione, si è scelto di non usare le features cliniche a disposizione. I risultati completi sono disponibili come allegati al presente documento (allegato 11).

In tabella 5 vengono riportati i risultati più significativi (in termini di f1 score) per il passo finale di classificazione con validazione Leave-One-Out. Le features individuate per i diversi markers sono disponibili come allegato al presente documento (allegato 12).

Risultati della classificazione LOO per la diagnosi differenziale amiloide negativo/positivo							
Tipo	Tipo features	algoritmo di classificazione	numero di features	accuracy	precision	recall	f1score
Controllo/ Malato	radiomiche	SVM	6	0.714	0.882	0.789	0.834

Tabella 5

Per quanto riguarda il test sui dati UNIME, per tutti e quattro i pazienti disponibili il predittore ha correttamente assegnato la classe amiloide positiva. Purtroppo non sono disponibili dati di test appartenenti alla classe dei pazienti amiloide negativi.

8.7. Morbo di Parkinson

Sui dati PET-FDOPA messi a disposizione dal partner UNICZ è naturale costruire un classificatore per distinguere i casi di controllo dai casi di malattia. Vale la pena precisare che i pazienti segnalati

come controllo sono affetti comunque da patologie motorie e cognitive, ma non diagnosticati come Parkinson's. Per completezza si riporta che il partner SDN ha messo a disposizione dei dataset di pazienti le cui PET sono ottenute con traccianti differenti (AMY e FDG). Purtroppo questi dati non possono essere usati come test, e sono troppo pochi per poter considerare altre tipologie di classificazione.

Si è deciso anche in questo caso di NON usare i dati provenienti dai test cognitivi, in modo da affidarsi esclusivamente alle informazioni provenienti da immagini e dai dati clinici di base, con l'obiettivo di rendere più rapida ed agevole la fase di supporto alla diagnosi. Infatti, in questa configurazione, si potrebbe prescindere dalla somministrazione dei test cognitivi, operazione in genere estremamente lunga, ed affidarsi unicamente al responso dell'imaging. Le features cliniche utilizzabili (cioè con valori definiti per tutti gli outcome) sono:

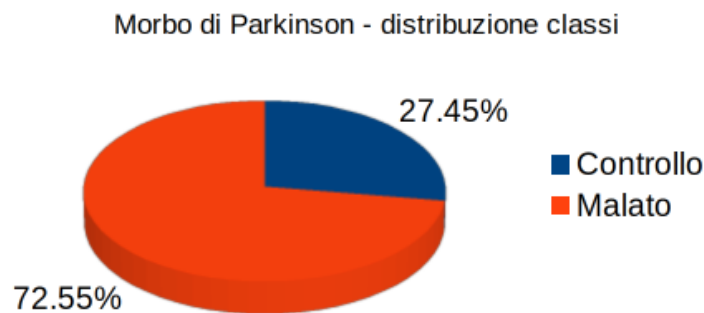
- età
- sesso

Inoltre sono disponibili anche i valori di uptake del mezzo di contrasto in diverse regioni cerebrali dell'emisfero destro e sinistro (Caudato, Putamen, Talamo).

Il numero totale di pazienti di UNICZ è 51.

DATI UNICZ

Dal grafico in figura 1 si vede che i dati sono sbilanciati con una predominanza delle diagnosi di Parkinson's rispetto ai casi di controllo.



8.7.1. Preprocessing

In questo paragrafo si descrivono tutti i passaggi che vengono effettuati prima della classificazione. I dati in input sono costituiti dalle *features* di volume, spessore, etc. come descritte nei precedenti

paragrafi, estratte dalle immagini multimodali PET/RMI dei pazienti. Si ribadisce la scelta di non usare le features cliniche da test cognitivi, al fine di permettere una diagnosi legata il quanto più possibile ai soli dati clinici essenziali e alle immagini mediche. Le features sono in totale 461 per ogni paziente.

Prima di essere elaborati i dati sono stati riscaldati nell'intervallo [0,1].

8.7.2. Risultati della classificazione e della validazione

Di seguito vengono riportati i risultati più significativi e descritto l'approccio che ha portato alla selezione dei modelli per ognuno dei problemi di classificazione considerati. Come descritto nei paragrafi precedenti, il testing relativo alla prima fase di feature selection è effettuato prendendo in considerazione tutte le possibili combinazioni di dati. Tutti i risultati sono disponibili come allegati al presente documento (allegato 13).

In tabella 6 vengono riportati i risultati più significativi (in termini di f1score) per il passo finale di classificazione con validazione Leave-One-Out. Le features individuate per i diversi markers sono disponibili come allegato al presente documento (allegato 14). In mancanza di dati da altri partner, il modello di predizione, costruito a partire dalle features e dall'algorithmo individuato, è stato validato su di un unico paziente, eliminato dal dataset al momento della costruzione del modello. Su questo paziente il modello ha dato il risultato corretto.

Risultati della classificazione LOO per la diagnosi differenziale controllo/ malato di Parkinson							
Tipo	Tipo features	algoritmo di classificazione	numero di features	accurac y	precision	recall	f1score
Controllo/ Malato	radiomiche /uptake	KNN/SVM	2	0.980	0.974	1	0.987
Controllo/ Malato	radiomiche	SVM	27	0.922	0.946	0.946	0.946

Tabella 6

Sono qui riportati i risultati che si ottengono sia usando come features di addestramento solo quelle provenienti da immagini sia la combinazione di queste con le features di uptake. L'uso delle features di uptake migliora le prestazioni e riduce notevolmente il numero delle features.

9. Bibliografia

- [1] B. Reig, L. Heacock, K. J. Geras, and L. Moy, "Machine learning in breast MRI," *J. Magn. Reson. Imaging*, vol. 52, no. 4, pp. 998–1018, 2020, doi: 10.1002/jmri.26852.
- [2] L. J. Grimm and M. A. Mazurowski, "Breast Cancer Radiogenomics: Current Status and Future Directions," *Acad. Radiol.*, vol. 27, no. 1, pp. 39–46, Jan. 2020, doi: 10.1016/j.acra.2019.09.012.
- [3] S. Huang *et al.*, "Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis," *Npj Breast Cancer*, vol. 4, no. 1, Art. no. 1, Aug. 2018, doi: 10.1038/s41523-018-0078-2.
- [4] H. Li *et al.*, "Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set," *Npj Breast Cancer*, vol. 2, no. 1, Art. no. 1, May 2016, doi: 10.1038/npjbcancer.2016.12.
- [5] H. Li *et al.*, "MR Imaging Radiomics Signatures for Predicting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays," *Radiology*, vol. 281, no. 2, pp. 382–391, May 2016, doi: 10.1148/radiol.2016152110.
- [6] "Welcome to The Cancer Imaging Archive," *The Cancer Imaging Archive (TCIA)*. <https://www.cancerimagingarchive.net/> (accessed Nov. 09, 2020).
- [7] Z. Liu *et al.*, "Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study," *Clin. Cancer Res.*, vol. 25, no. 12, pp. 3538–3547, Jun. 2019, doi: 10.1158/1078-0432.CCR-18-3190.
- [8] T. Xie *et al.*, "Machine Learning-Based Analysis of MR Multiparametric Radiomics for the Subtype Classification of Breast Cancer," *Front. Oncol.*, vol. 9, 2019, doi: 10.3389/fonc.2019.00505.
- [9] W. Guo *et al.*, "Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data," *J. Med. Imaging*, vol. 2, no. 4, p. 041007, Sep. 2015, doi: 10.1117/1.JMI.2.4.041007.
- [10] E. S. Burnside *et al.*, "Using computer-extracted image phenotypes from tumors on breast magnetic resonance imaging to predict breast cancer pathologic stage," *Cancer*, vol. 122, no. 5, pp. 748–757, 2016, doi: <https://doi.org/10.1002/cncr.29791>.
- [11] N. Horvat, D. D. B. Bates, and I. Petkovska, "Novel imaging techniques of rectal cancer: what do radiomics and radiogenomics have to offer? A literature review," *Abdom. Radiol.*, vol. 44, no. 11, pp. 3764–3774, Nov. 2019, doi: 10.1007/s00261-019-02042-y.
- [12] C. Liang *et al.*, "The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer," *Oncotarget*, vol. 7, no. 21, pp. 31401–31412, Apr. 2016, doi: 10.18632/oncotarget.8919.

- [13] Y. Huang *et al.*, “Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer,” *J. Clin. Oncol.*, vol. 34, no. 18, pp. 2157–2164, May 2016, doi: 10.1200/JCO.2015.65.9128.
- [14] X. Zhou *et al.*, “Radiomics-Based Pretherapeutic Prediction of Non-response to Neoadjuvant Therapy in Locally Advanced Rectal Cancer,” *Ann. Surg. Oncol.*, vol. 26, no. 6, pp. 1676–1684, Jun. 2019, doi: 10.1245/s10434-019-07300-3.
- [15] L. Yang *et al.*, “Can CT-based radiomics signature predict KRAS/NRAS/BRAF mutations in colorectal cancer?,” *Eur. Radiol.*, vol. 28, no. 5, pp. 2058–2067, May 2018, doi: 10.1007/s00330-017-5146-8.
- [16] Z. Zhu, E. Albadawy, A. Saha, J. Zhang, M. R. Harowicz, and M. A. Mazurowski, “Deep learning for identifying radiogenomic associations in breast cancer,” *Comput. Biol. Med.*, vol. 109, pp. 85–90, Jun. 2019, doi: 10.1016/j.compbiomed.2019.04.018.
- [17] “AIOM | Associazione Italiana di Oncologia Medica – Official Website,” *AIOM*. <https://www.aiom.it/> (accessed Nov. 12, 2020).
- [18] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction: Foundations and Applications*. Springer, 2008.
- [19] L. Yu and H. Liu, “Efficient feature selection via analysis of relevance and redundancy,” *J. Mach. Learn. Res.*, vol. 5, pp. 1205–1224, Oct. 2004.
- [20] V. Bolón-Canedo, N. Sánchez-Marroño, and A. Alonso-Betanzos, “Feature selection for high-dimensional data,” *Prog. Artif. Intell.*, vol. 5, no. 2, pp. 65–75, May 2016, doi: 10.1007/s13748-015-0080-y.
- [21] Y. Saeys, I. Inza, and P. Larrañaga, “A review of feature selection techniques in bioinformatics,” *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, Oct. 2007, doi: 10.1093/bioinformatics/btm344.
- [22] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.
- [23] D. A. Freedman, *Statistical models: theory and practice*. Cambridge University Press, 2009.
- [24] D. W. Hosmer and S. Lemeshow, *Applied logistic regression*. Wiley New York, 2000.
- [25] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, “Support vector regression machines,” in *Advances in neural information processing systems*, 1997, pp. 155–161.
- [26] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [27] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene Selection for Cancer Classification using Support Vector Machines,” *Mach. Learn.*, vol. 46, no. 1, pp. 389–422, Jan. 2002, doi: 10.1023/A:1012487302797.
- [28] C. Salvatore, I. Castiglioni, and A. Cerasa, “Radiomics approach in the neurodegenerative brain,” *Aging Clin. Exp. Res.*, pp. 1–3, Aug. 2019, doi: 10.1007/s40520-019-01299-z.
- [29] C. J. Brown and G. Hamarneh, “Machine Learning on Human Connectome Data from MRI,” *ArXiv161108699 Cs Q-Bio Stat*, Nov. 2016, Accessed: Nov. 11, 2020. [Online]. Available: <http://arxiv.org/abs/1611.08699>.

- [30] “Benchmarking functional connectome-based predictive models for resting-state fMRI,” *NeuroImage*, vol. 192, pp. 115–134, May 2019, doi: 10.1016/j.neuroimage.2019.02.062.
- [31] K. Dadi *et al.*, “Benchmarking functional connectome-based predictive models for resting-state fMRI,” *NeuroImage*, vol. 192, pp. 115–134, May 2019, doi: 10.1016/j.neuroimage.2019.02.062.
- [32] K. Zhao *et al.*, “Independent and reproducible hippocampal radiomic biomarkers for multisite Alzheimer’s disease: diagnosis, longitudinal progress and biological basis,” *Sci. Bull.*, vol. 65, no. 13, pp. 1103–1113, Jul. 2020, doi: 10.1016/j.scib.2020.04.003.
- [33] T. Xie *et al.*, “Machine Learning-Based Analysis of MR Multiparametric Radiomics for the Subtype Classification of Breast Cancer,” *Front. Oncol.*, vol. 9, 2019, doi: 10.3389/fonc.2019.00505.
- [34] H. Zhou *et al.*, “Dual-Model Radiomic Biomarkers Predict Development of Mild Cognitive Impairment Progression to Alzheimer’s Disease,” *Front. Neurosci.*, vol. 12, 2019, doi: 10.3389/fnins.2018.01045.
- [35] A. Khazaei, A. Ebrahimzadeh, and A. Babajani-Feremi, “Identifying patients with Alzheimer’s disease using resting-state fMRI and graph theory,” *Clin. Neurophysiol.*, vol. 126, no. 11, pp. 2132–2141, Nov. 2015, doi: 10.1016/j.clinph.2015.02.060.
- [36] S. H. Hojjati, A. Ebrahimzadeh, and A. Babajani-Feremi, “Identification of the early stage of alzheimer’s disease using structural mri and resting-state fmri,” *Front. Neurol.*, vol. 10, no. AUG, 2019, doi: 10.3389/fneur.2019.00904.
- [37] M. Wang, Z. Yan, and J. Jiang, “Brain metabolic connectome classify mild cognitive impairment into Alzheimer’s dementia *,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, Jul. 2019, pp. 32–35, doi: 10.1109/EMBC.2019.8857104.
- [38] “ADNI Alzheimer’s Disease Neuroimaging Initiative.” <http://adni.loni.usc.edu/> (accessed Nov. 11, 2020).
- [39] “Human Connectome Project | Data.” <http://www.humanconnectomeproject.org/data/> (accessed Nov. 11, 2020).
- [40] Y. Wu *et al.*, “Use of radiomic features and support vector machine to distinguish Parkinson’s disease cases from normal controls,” *Ann. Transl. Med.*, vol. 7, no. 23, Dec. 2019, doi: 10.21037/atm.2019.11.26.
- [41] A. Rahmim *et al.*, “Improved prediction of outcome in Parkinson’s disease using radiomics analysis of longitudinal DAT SPECT images,” *NeuroImage Clin.*, vol. 16, pp. 539–544, Jan. 2017, doi: 10.1016/j.nicl.2017.08.021.
- [42] Z. Cheng *et al.*, “Radiomic Features of the Nigrosome-1 Region of the Substantia Nigra: Using Quantitative Susceptibility Mapping to Assist the Diagnosis of Idiopathic Parkinson’s Disease,” *Front. Aging Neurosci.*, vol. 11, 2019, doi: 10.3389/fnagi.2019.00167.

- [43] R. Mabrouk, B. Chikhaoui, and L. Bentabet, "Machine Learning Based Classification Using Clinical and DaTSCAN SPECT Imaging Features: A Study on Parkinson's Disease and SWEDD," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 170–177, Mar. 2019, doi: 10.1109/TRPMS.2018.2877754.
- [44] "Parkinson's Progression Markers Initiative ." <https://www.ppmi-info.org/> (accessed Nov. 11, 2020).
- [45] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.
- [46] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Am. Stat.*, vol. 46, no. 3, pp. 175–185, 1992.
- [47] A. C. Wolff *et al.*, "Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Update," *J. Clin. Oncol.*, vol. 31, no. 31, pp. 3997–4013, Oct. 2013, doi: 10.1200/JCO.2013.50.9984.
- [48] A. Goldhirsch *et al.*, "Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013," *Ann. Oncol.*, vol. 24, no. 9, pp. 2206–2223, Sep. 2013, doi: 10.1093/annonc/mdt303.
- [49] A. S. Coates *et al.*, "Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015," *Ann. Oncol.*, vol. 26, no. 8, pp. 1533–1546, Aug. 2015, doi: 10.1093/annonc/mdv221.
- [50] G. Curigliano *et al.*, "De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017," *Ann. Oncol.*, vol. 28, no. 8, pp. 1700–1712, Aug. 2017, doi: 10.1093/annonc/mdx308.
- [51] K. H. Allison *et al.*, "Estrogen and Progesterone Receptor Testing in Breast Cancer: ASCO/CAP Guideline Update," *J. Clin. Oncol.*, vol. 38, no. 12, pp. 1346–1366, Jan. 2020, doi: 10.1200/JCO.19.02309.
- [52] N. M. Foley *et al.*, "Re-Appraisal of Estrogen Receptor Negative/Progesterone Receptor Positive (ER–/PR+) Breast Cancer Phenotype: True Subtype or Technical Artefact?," *Pathol. Oncol. Res.*, vol. 24, no. 4, pp. 881–884, Oct. 2018, doi: 10.1007/s12253-017-0304-5.
- [53] wiki.cancerimagingarchive.net/display/DOI/TCGA+Breast+Phenotype+Research+Group+Data+sets#19039112036220c66a5a436f90e4a0b54367bfae
- [54] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, *Journal of Digital Imaging*, Volume 26, Number 6, December, 2013, pp 1045-1057.
- [55] Ho, Tin Kam. "Random decision forests." *Proceedings of 3rd international conference on document analysis and recognition*. Vol. 1. IEEE, 1995.

- [56] Dasarathy, Belur V. "Nearest neighbor (NN) norms: NN pattern classification techniques." IEEE Computer Society Tutorial (1991).
- [57] Langley, Pat, Wayne Iba, and Kevin Thompson. "An analysis of Bayesian classifiers." *Aai*. Vol. 90. 1992.
- [58] Vapnik, Vladimir. *The nature of statistical learning theory*. Springer science & business media, 2013.
- [59] Rumelhart, D. E. "Learning internal representations by error propagation." *Parallel distributed processing 1* (1986): 318-362.
- [60] Magee, J. F. (1964). *Decision trees for decision making* (pp. 35-48). *Harvard Business Review*.
- [61] Zhang, Cha, and Yunqian Ma, eds. *Ensemble machine learning: methods and applications*. Springer Science & Business Media, 2012.
- [62] Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." *Journal of artificial intelligence research* 16 (2002): 321-357.
- [63] <https://imbalanced-learn.readthedocs.io/en/stable/index.html>
- [64] McLachlan, Geoffrey J., Do, Kim-Anh; Ambroise, Christophe (2004). *Analyzing microarray gene expression data*. Wiley.
- [65] Molinaro, A. M.; Simon, R.; Pfeiffer, R. M. (2005-08-01). "Prediction error estimation: a comparison of resampling methods". *Bioinformatics*. 21 (15): 3301–3307.
- [66] Chen, C. H., Chang, C. K., Tu, C. Y., Liao, W. C., Wu, B. R., Chou, K. T., ... & Huang, T. C. (2018). Radiomic features analysis in computed tomography images of lung nodule classification. *PloS one*, 13(2), e0192002.
- [67] Filograna, L., Lenkowicz, J., Cellini, F., Dinapoli, N., Manfrida, S., Magarelli, N., ... & Valentini, V. (2019). Identification of the most significant magnetic resonance imaging (MRI) radiomic features in oncological patients with vertebral bone marrow metastatic disease: a feasibility study. *La radiologia medica*, 124(1), 50-57.
- [68] "FreeSurferWiki - Free Surfer Wiki." <https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferWiki> (accessed May 10, 2021).
- [69] Weisstein, Eric W. "[Statistical Correlation](https://mathworld.wolfram.com/StatisticalCorrelation.html)". *mathworld.wolfram.com*.
- [70] Corder, G. W. & Foreman, D. I. (2014). *Nonparametric Statistics: A Step-by-Step Approach*, Wiley.
- [71] Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8, 3-62.
- [72] Benjamini Y, Hochberg Y (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing" (PDF). *Journal of the Royal Statistical Society, Series B*. 57 (1): 289–300.
- [73] Romeo, V., Cavaliere, C., Imbriaco, M., Verde, F., Petretta, M., Franzese, M., Stanzione, A., Cuocolo, R., Aiello, M., Basso, L., Amitrano, M., Lauria, R., Accurso, A., Brunetti, A., &

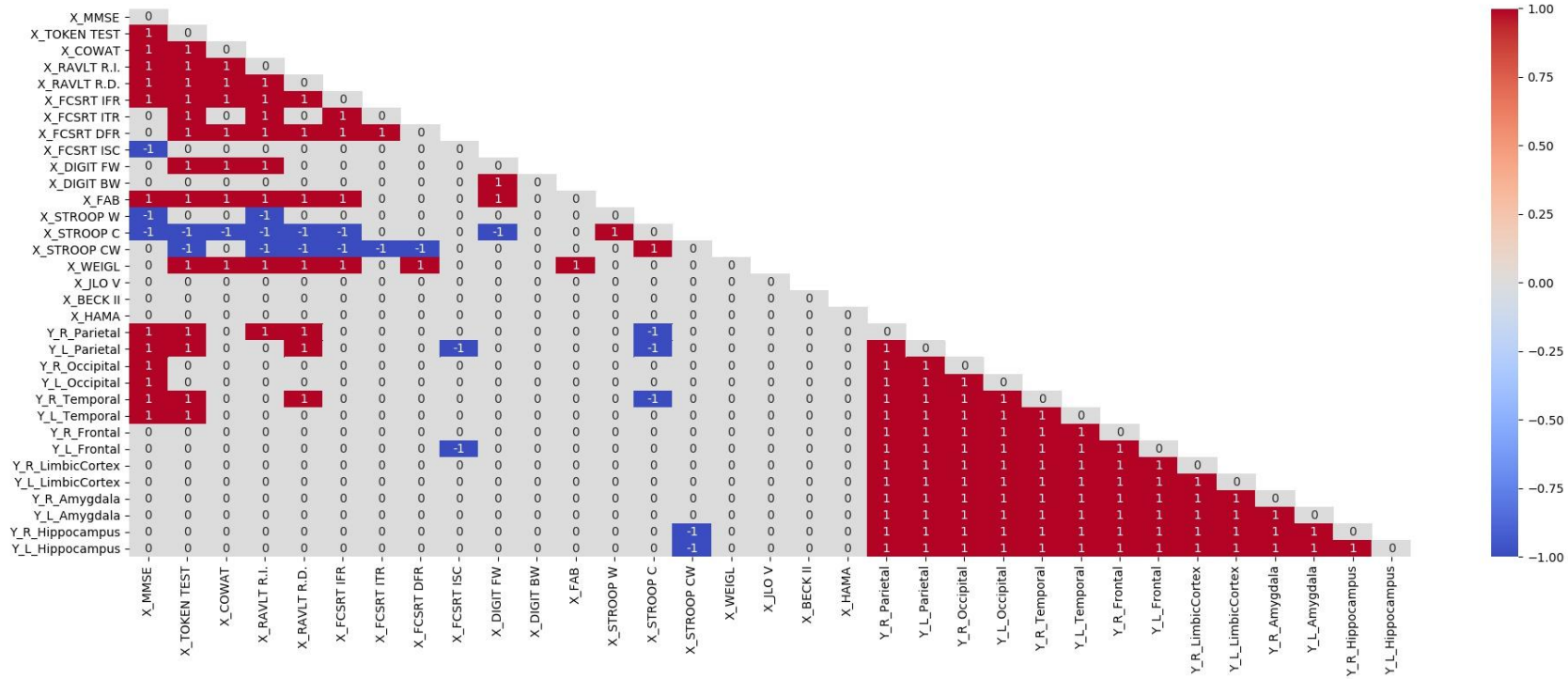
Salvatore, M. (2020). Tumor segmentation analysis at different post-contrast time points: A possible source of variability of quantitative DCE-MRI parameters in locally advanced breast cancer. *European journal of radiology*, 126, 108907.

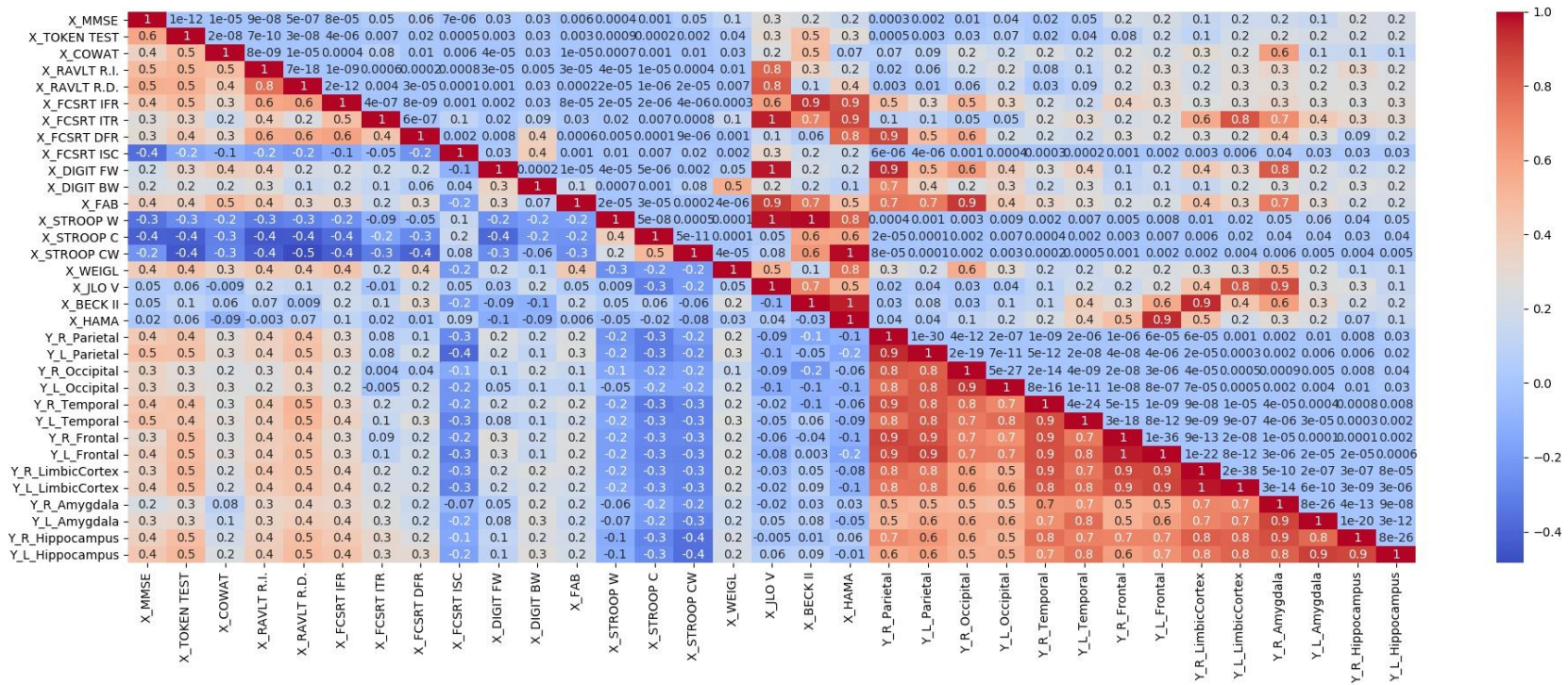
Allegato 1

Morbo di Alzheimer

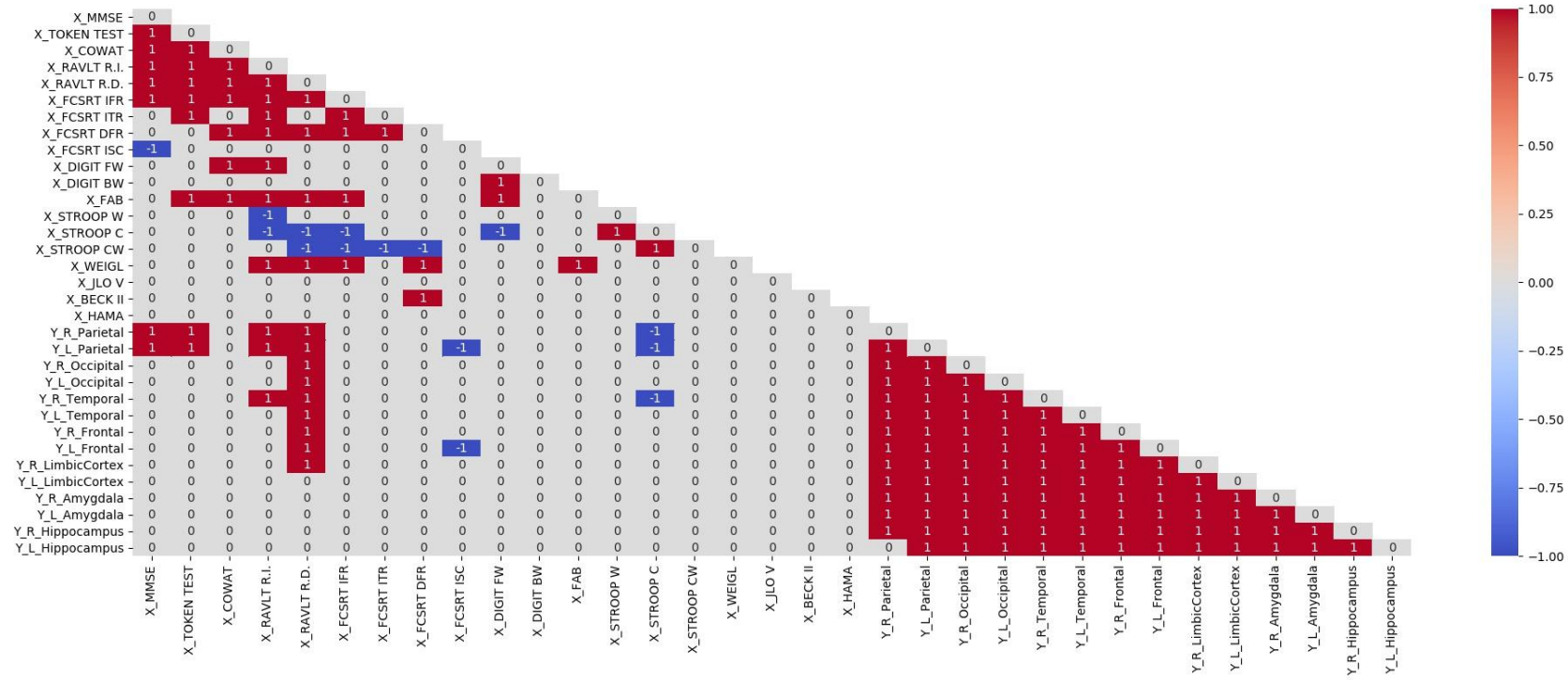
Test Cognitivi vs Dati di Uptake

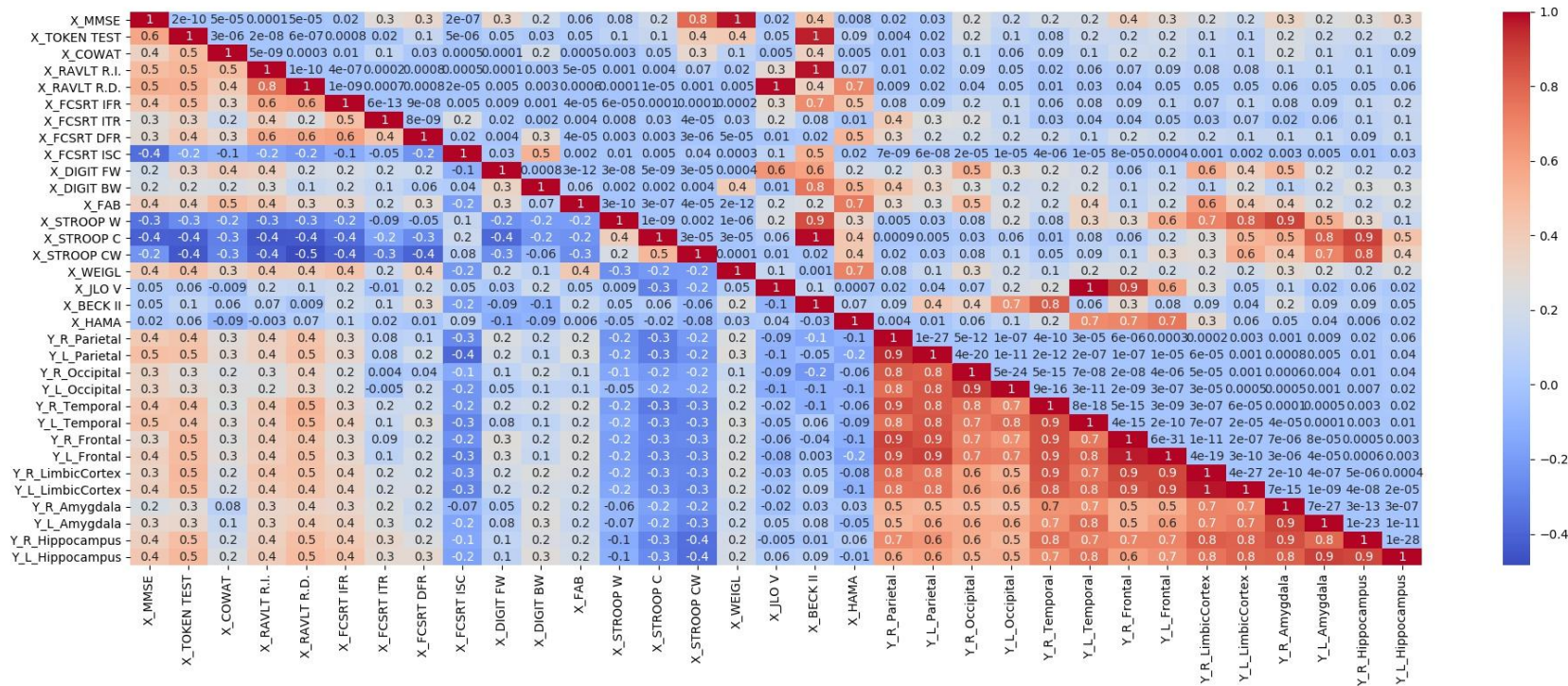
Correlazione di Pearson



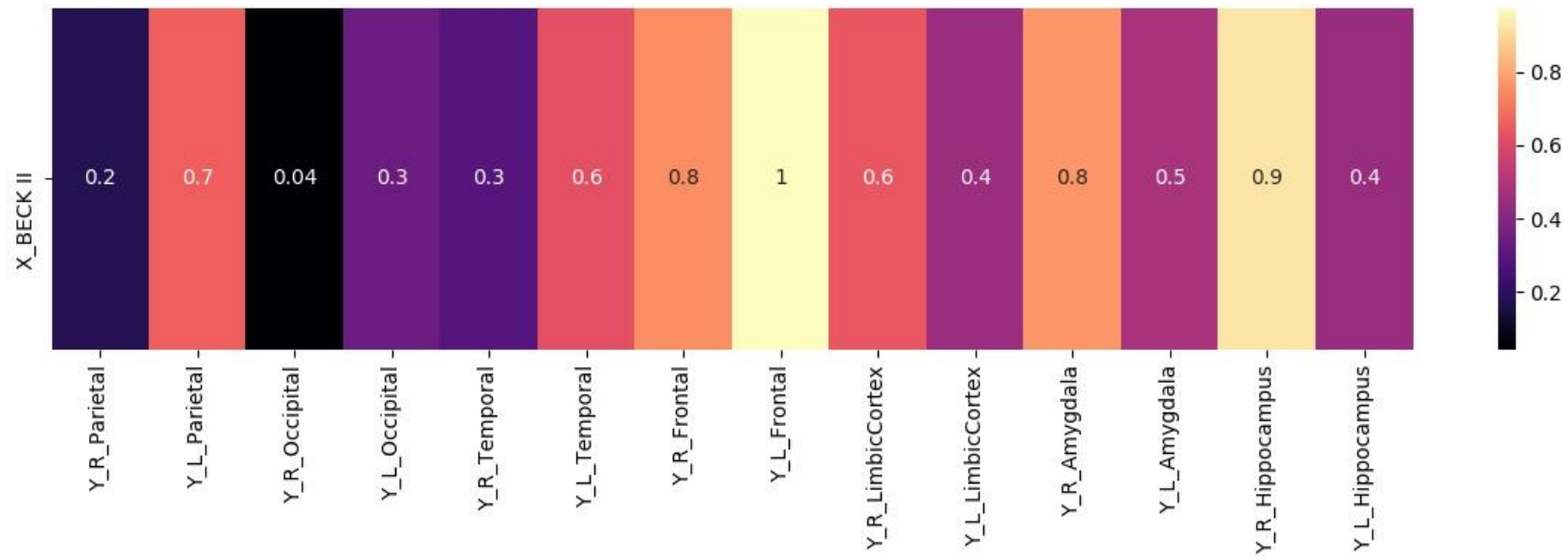


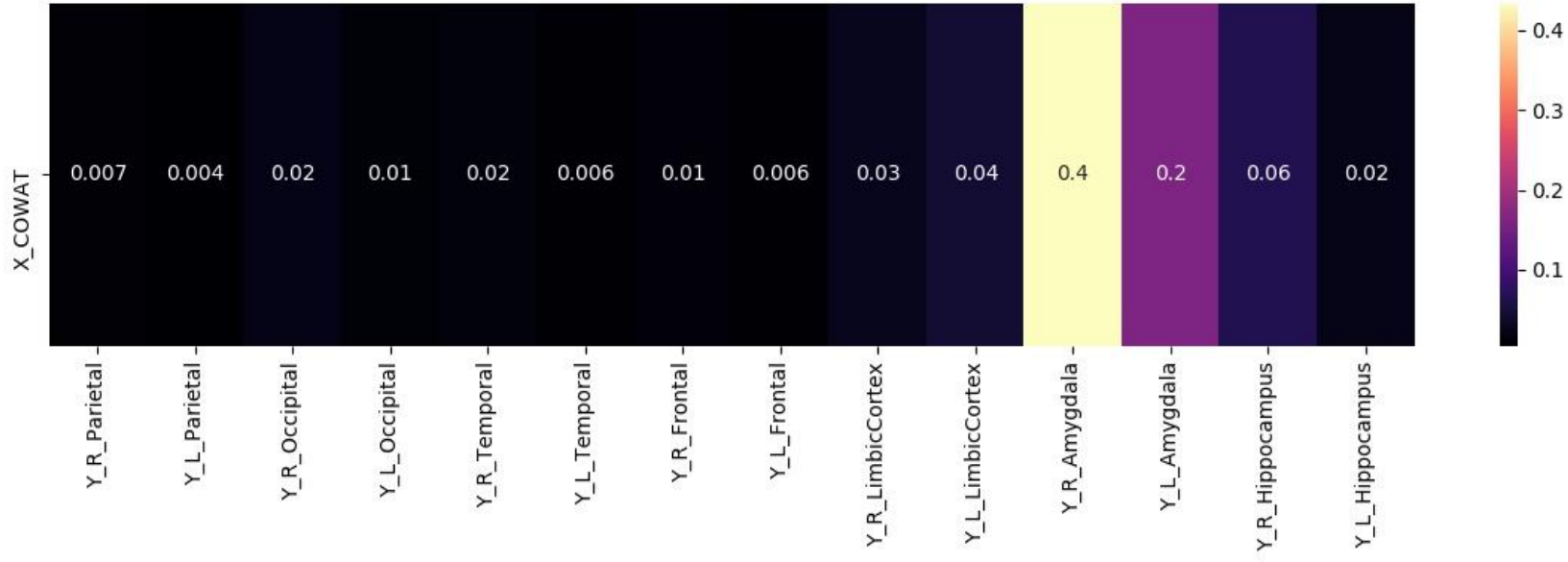
Correlazione di Spearman

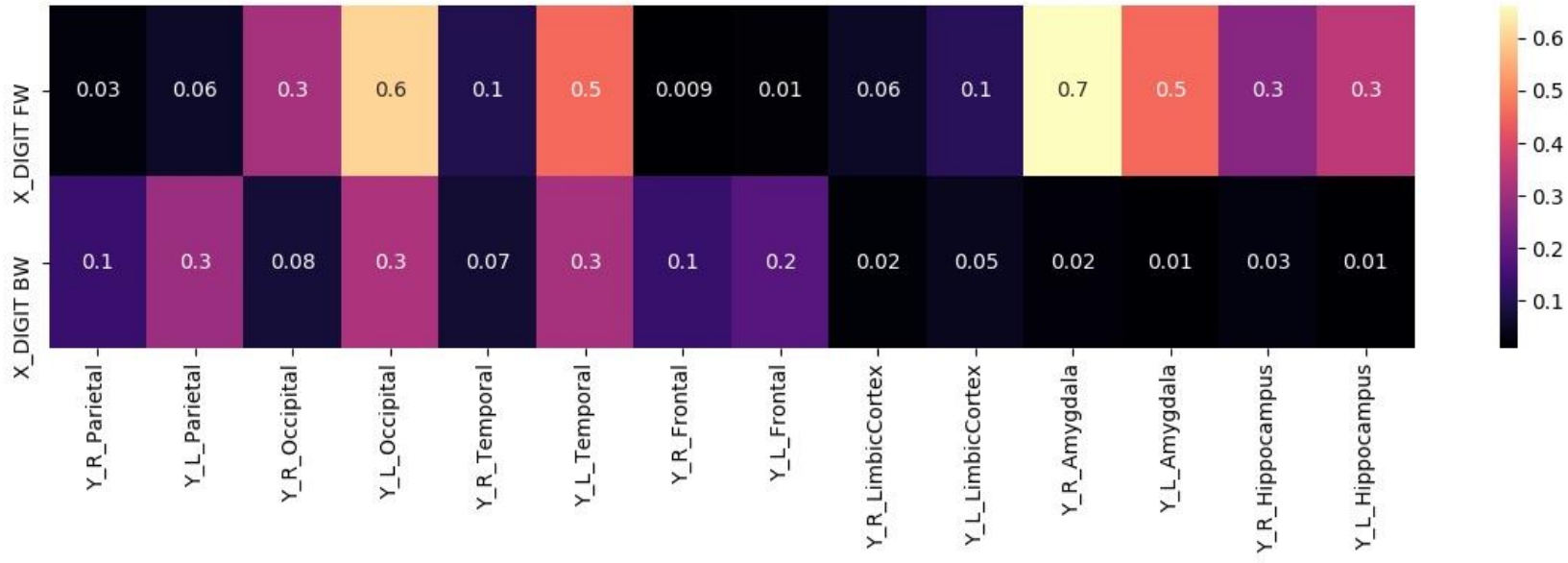


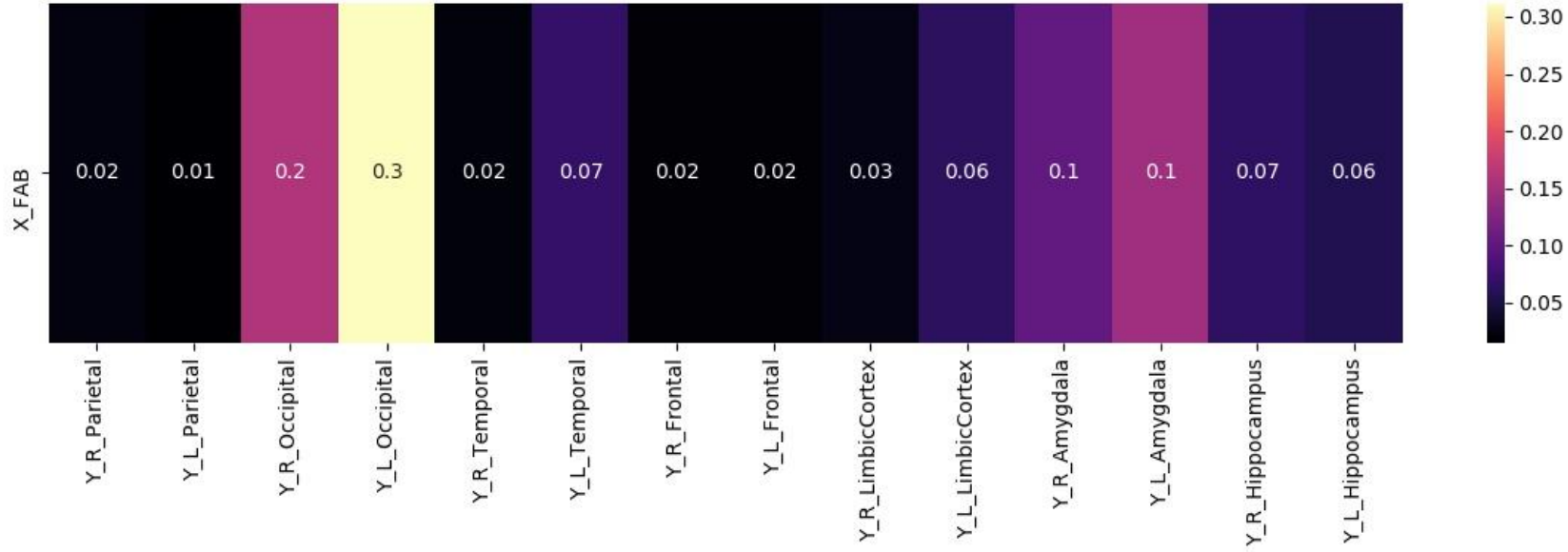


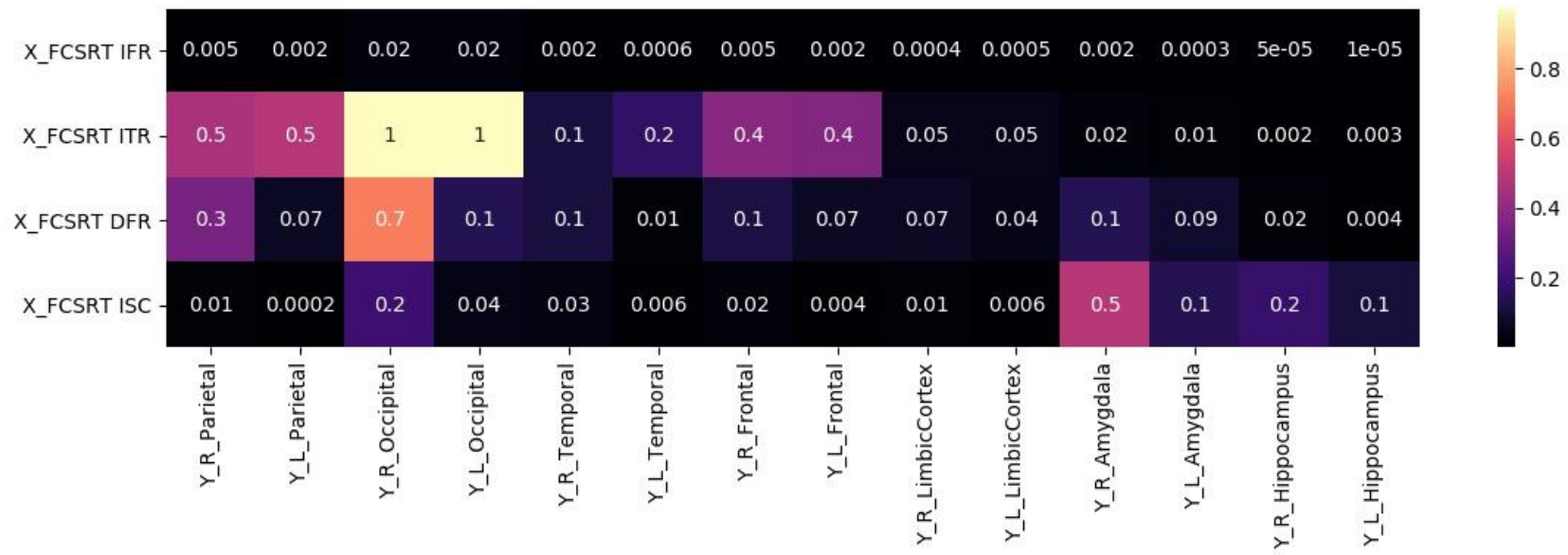
Correzione con Benjamini-Hochberg

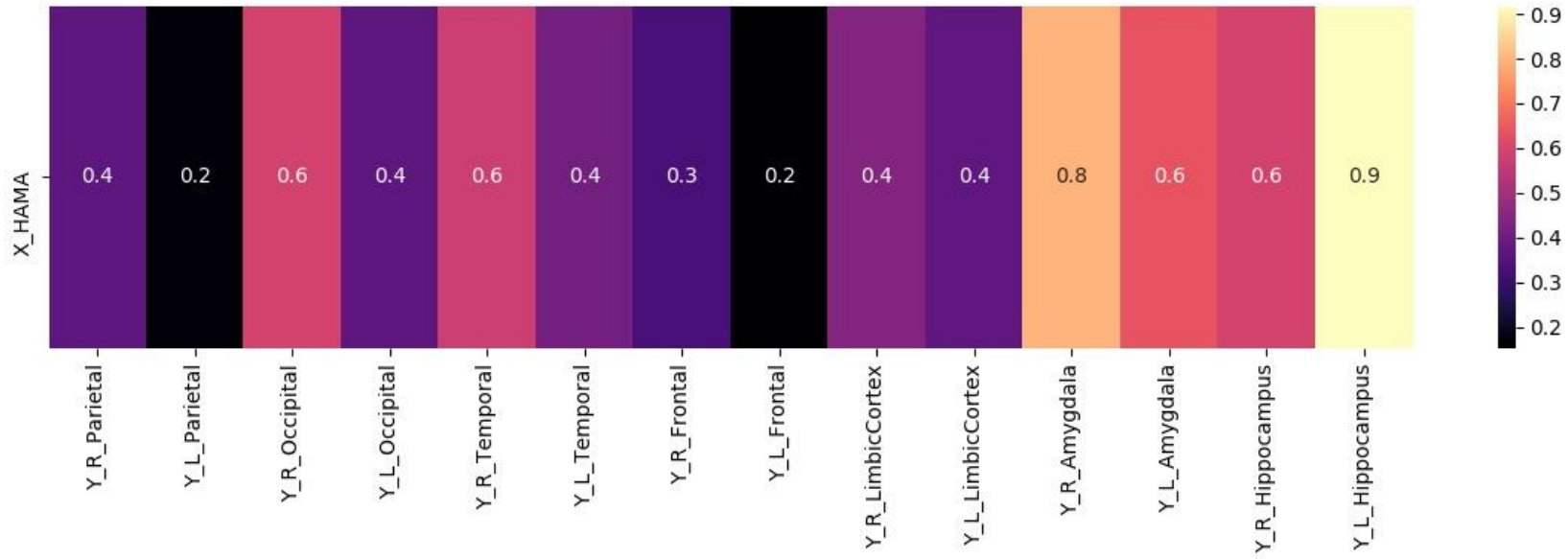


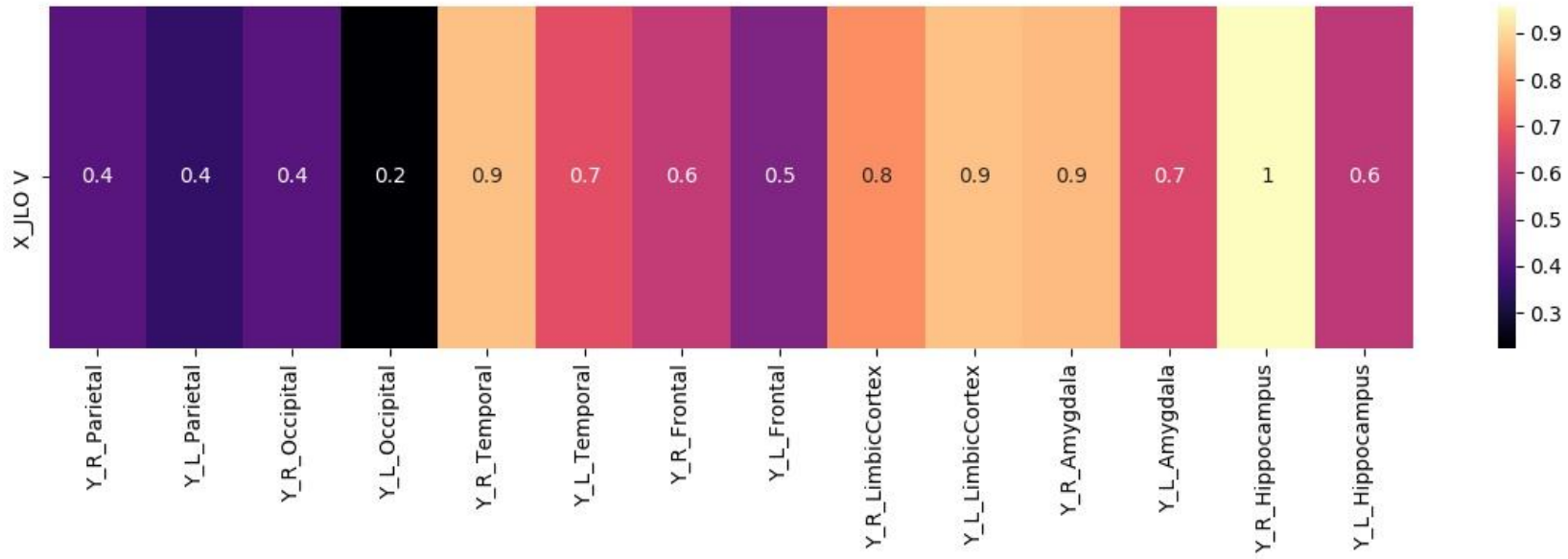


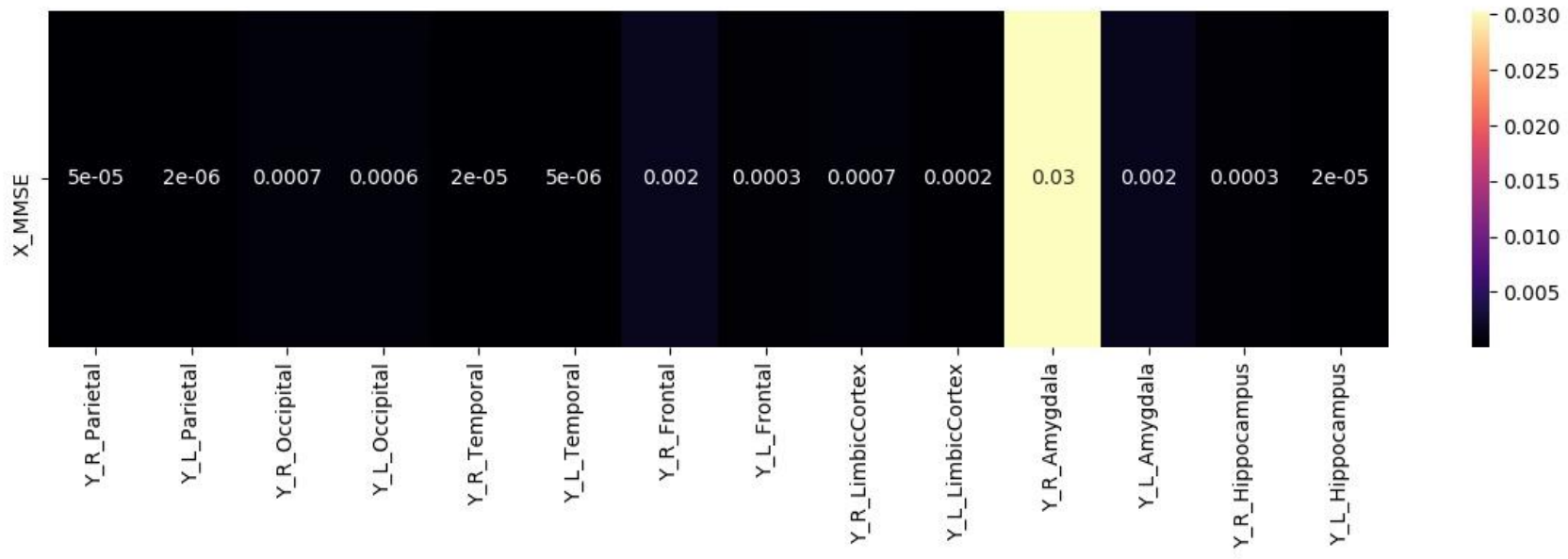


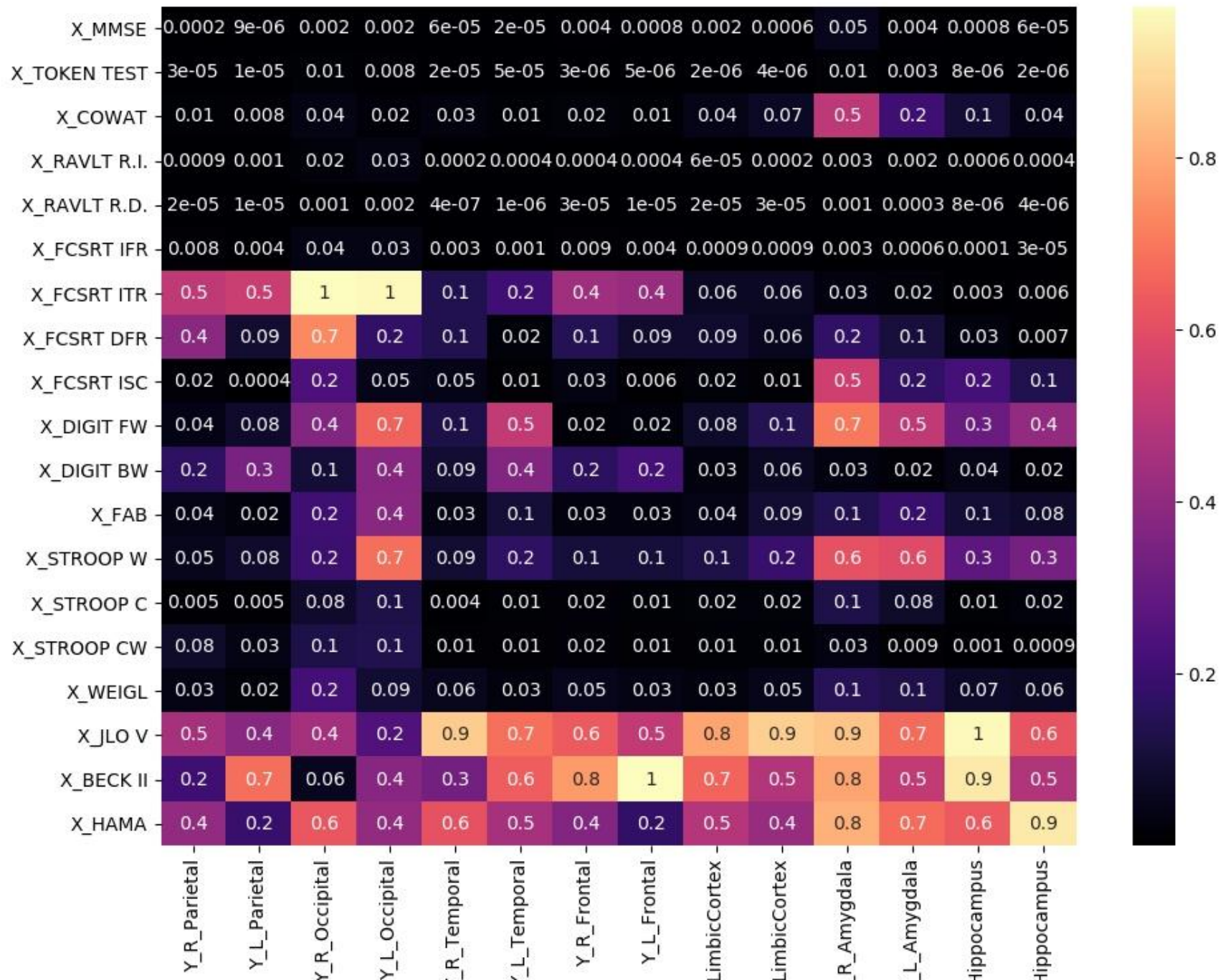


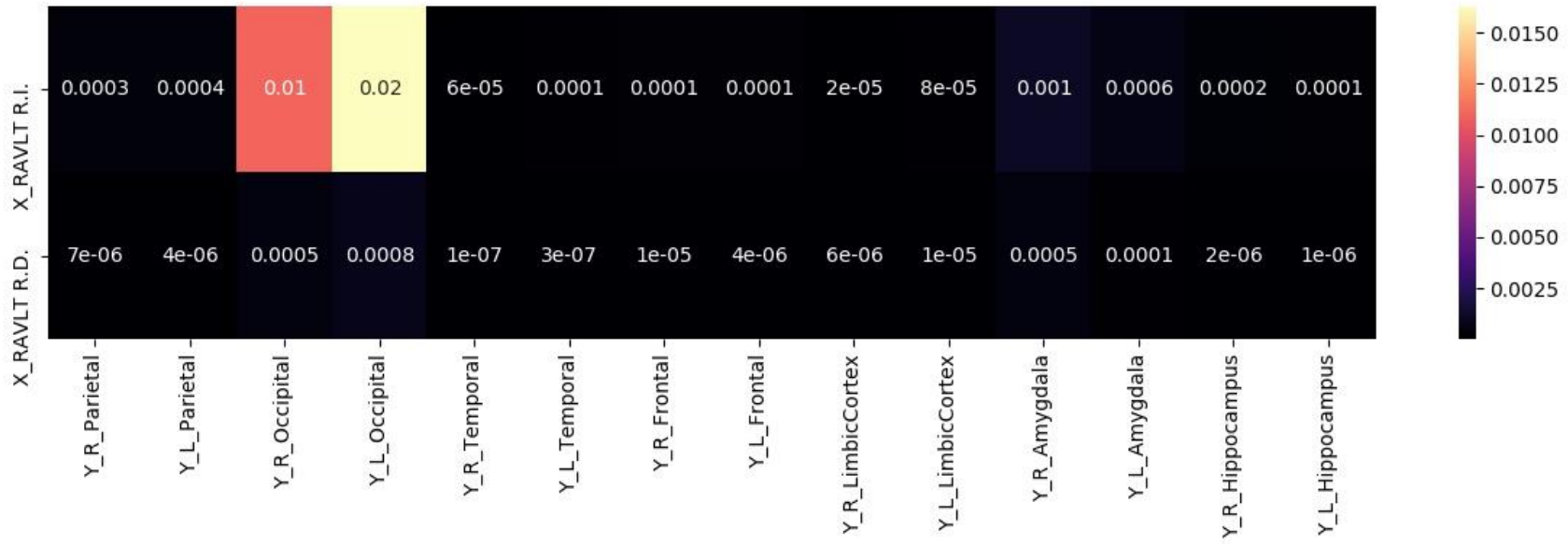


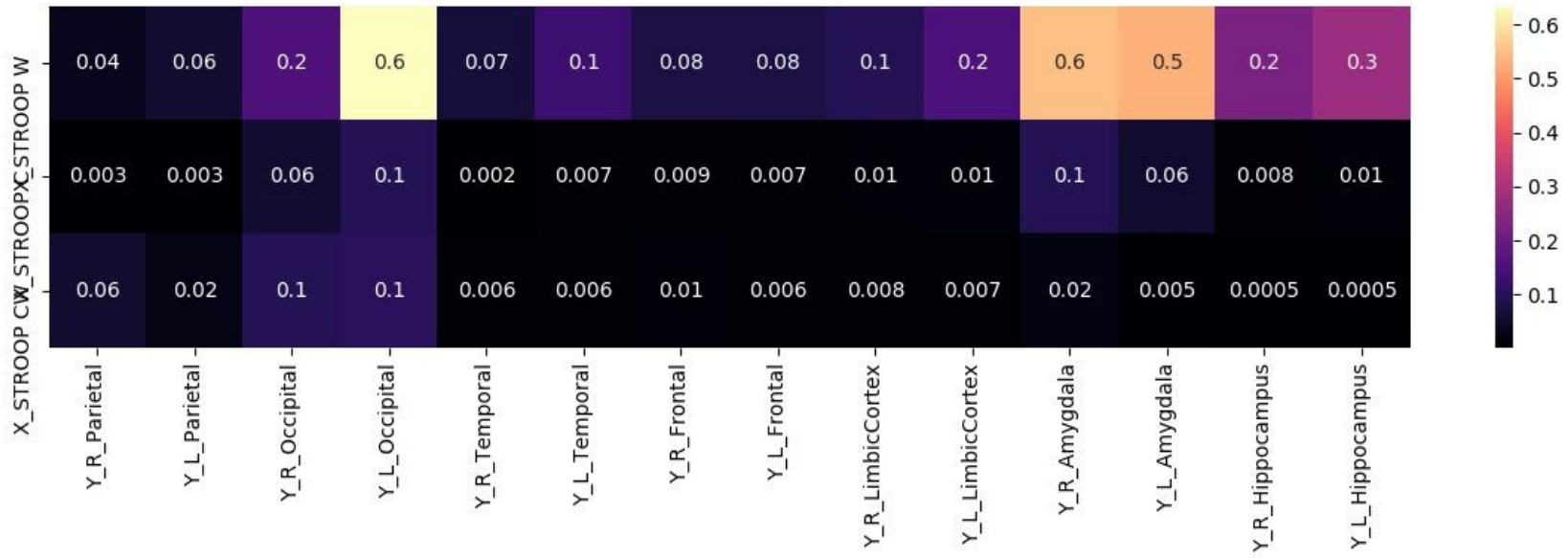


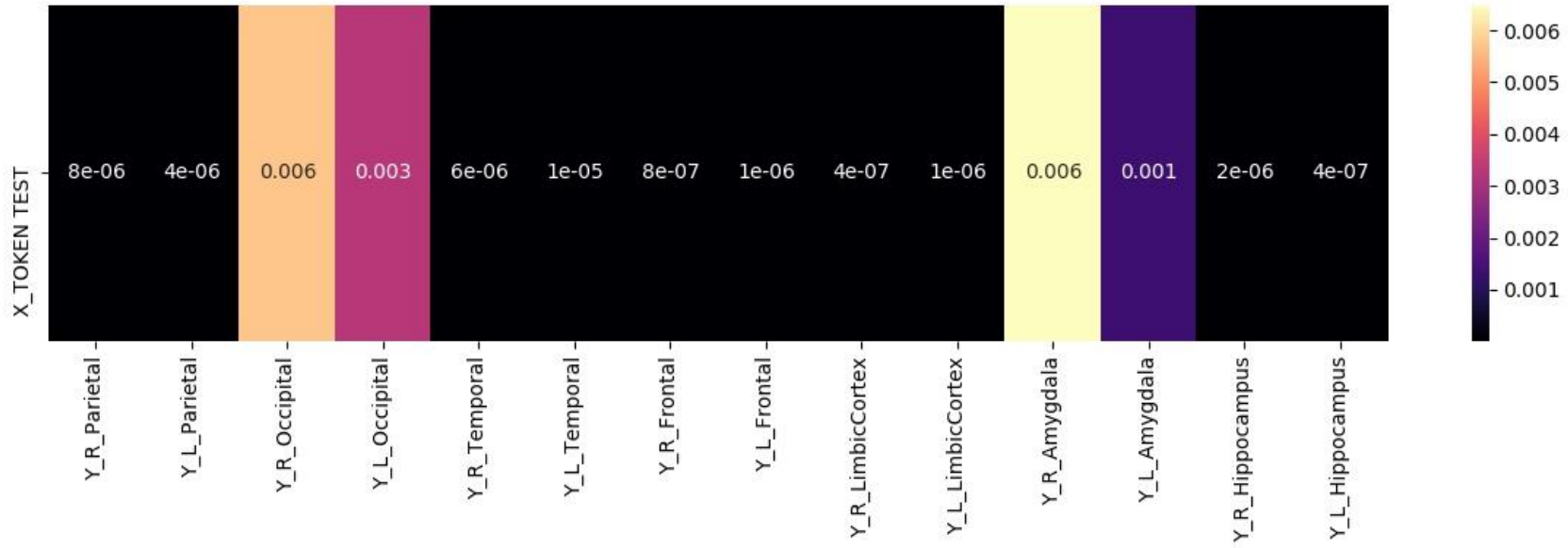


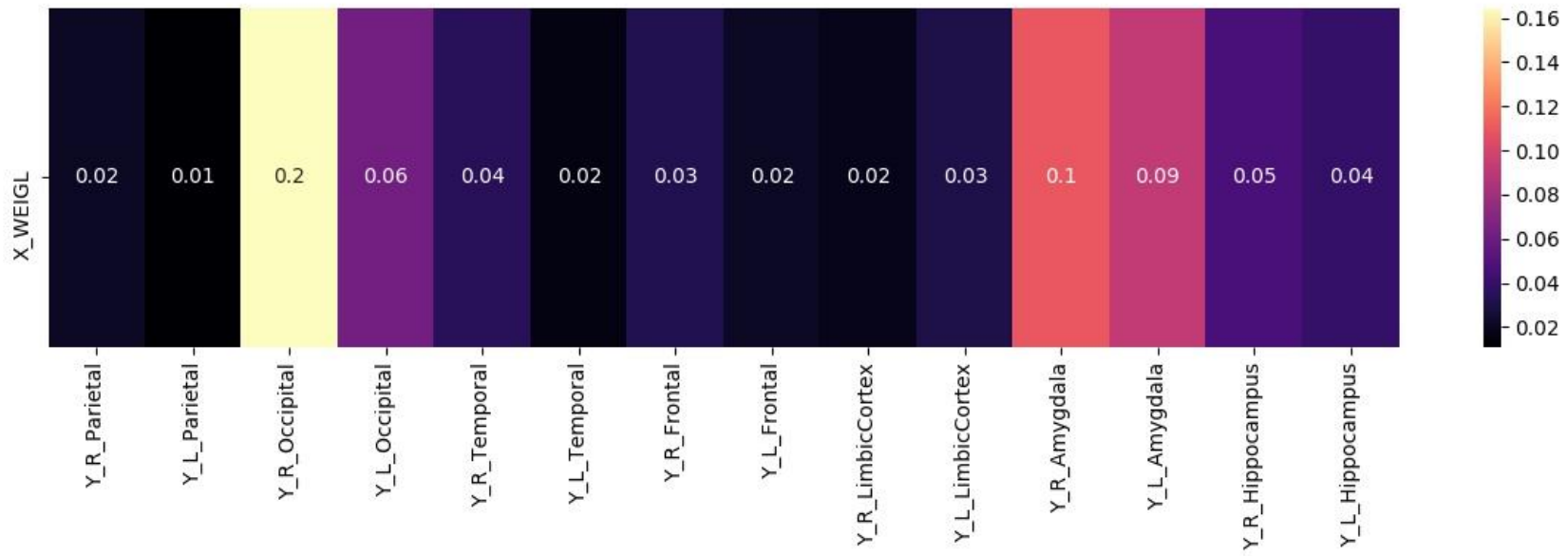




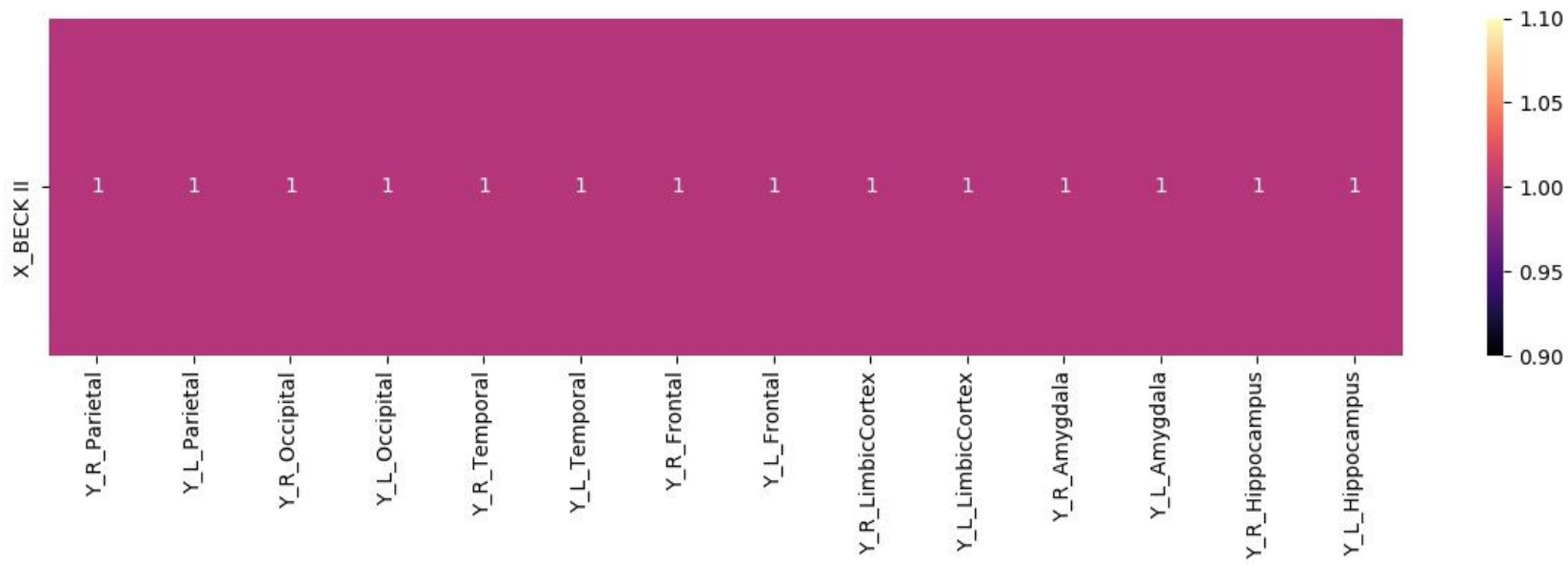


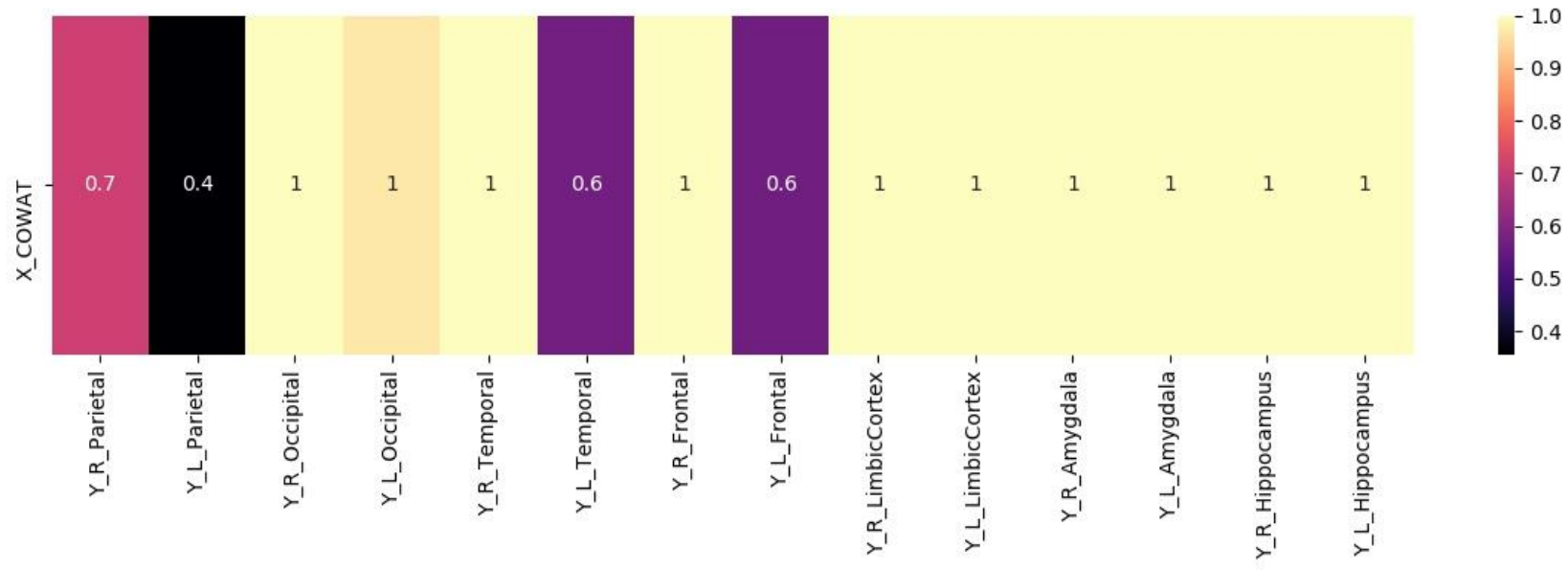


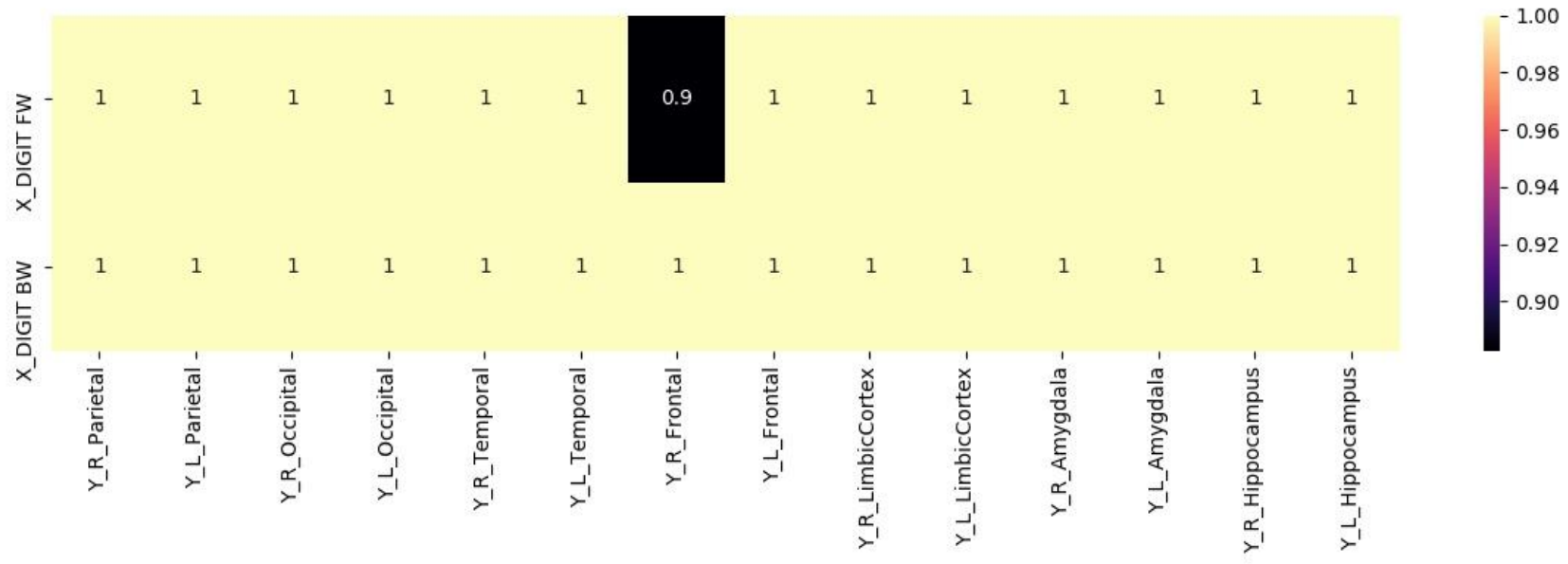


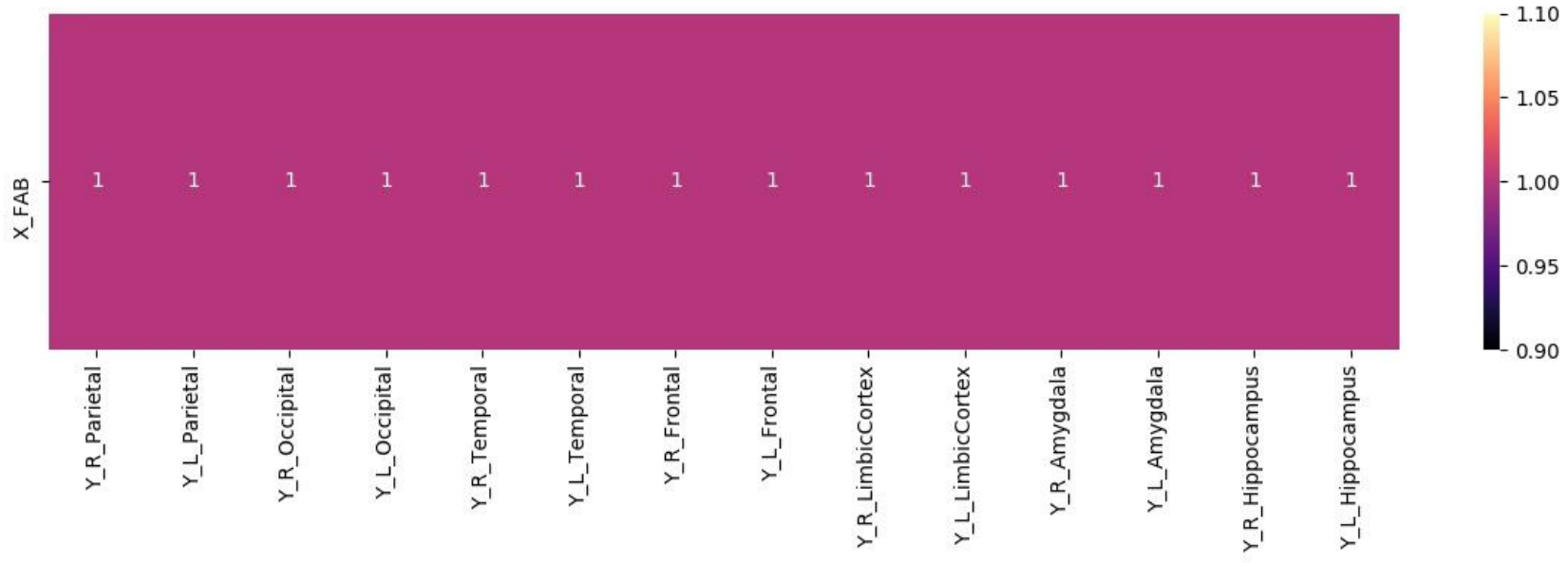


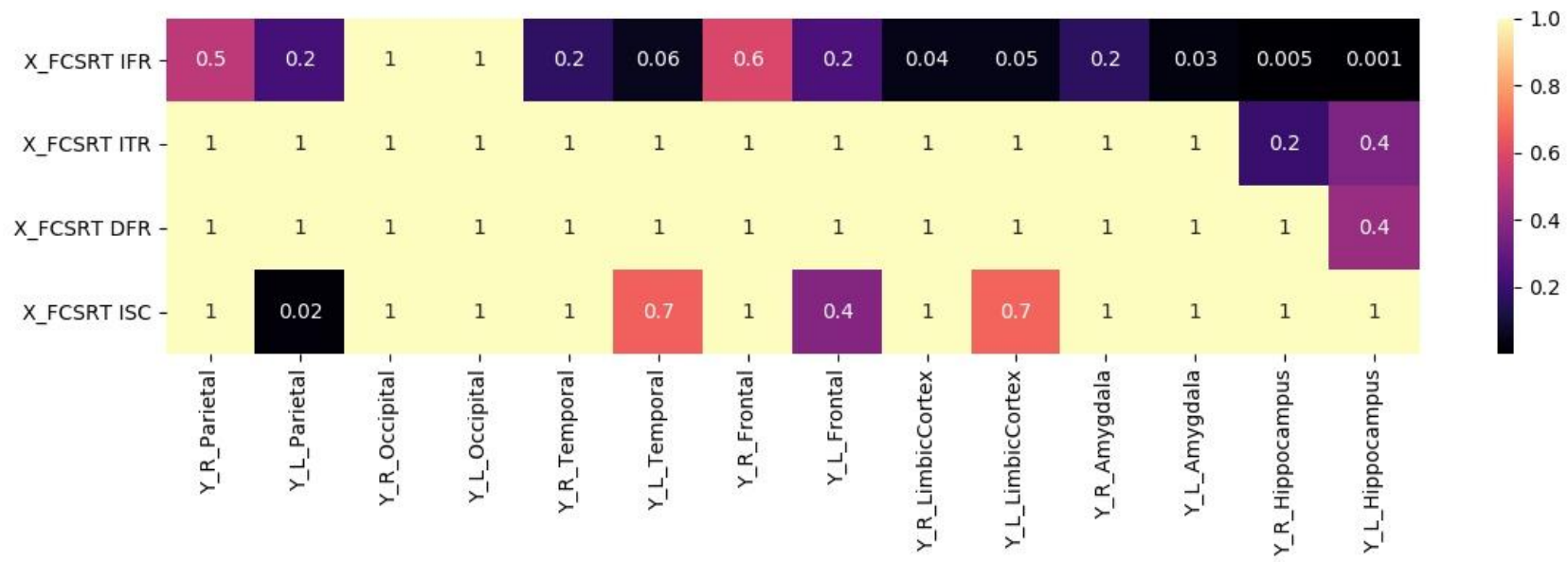
Correzione con Bonferroni

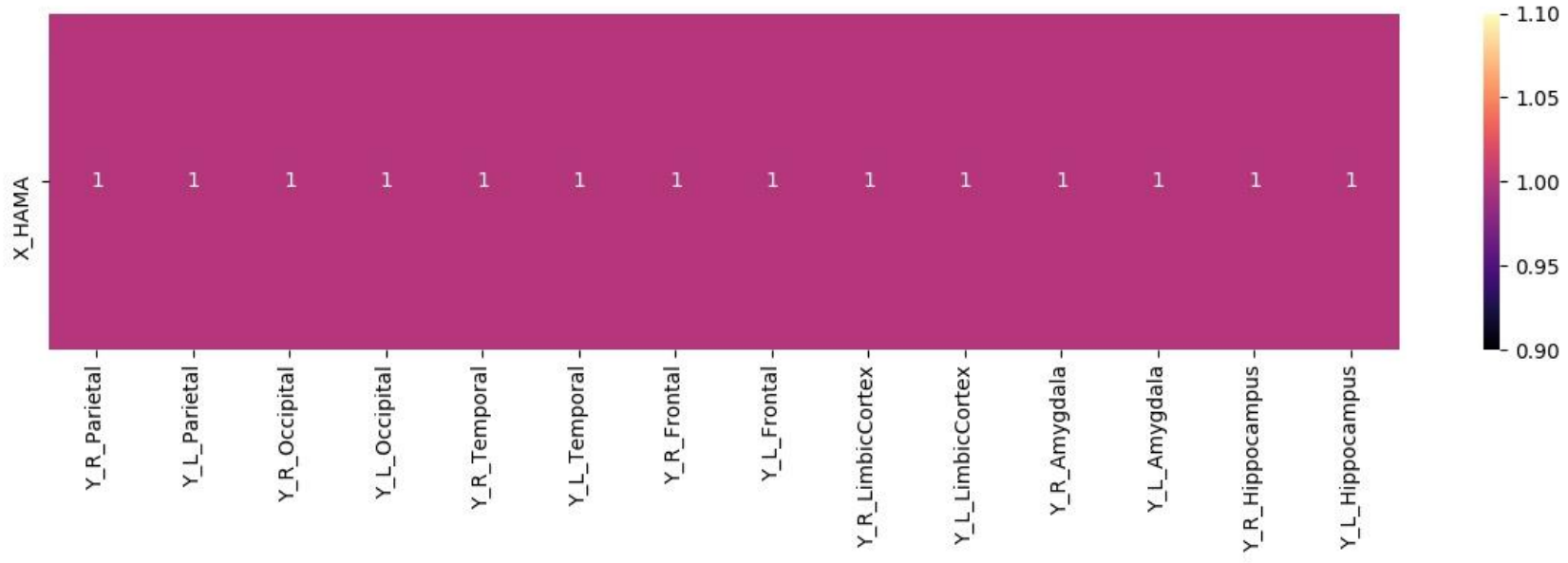


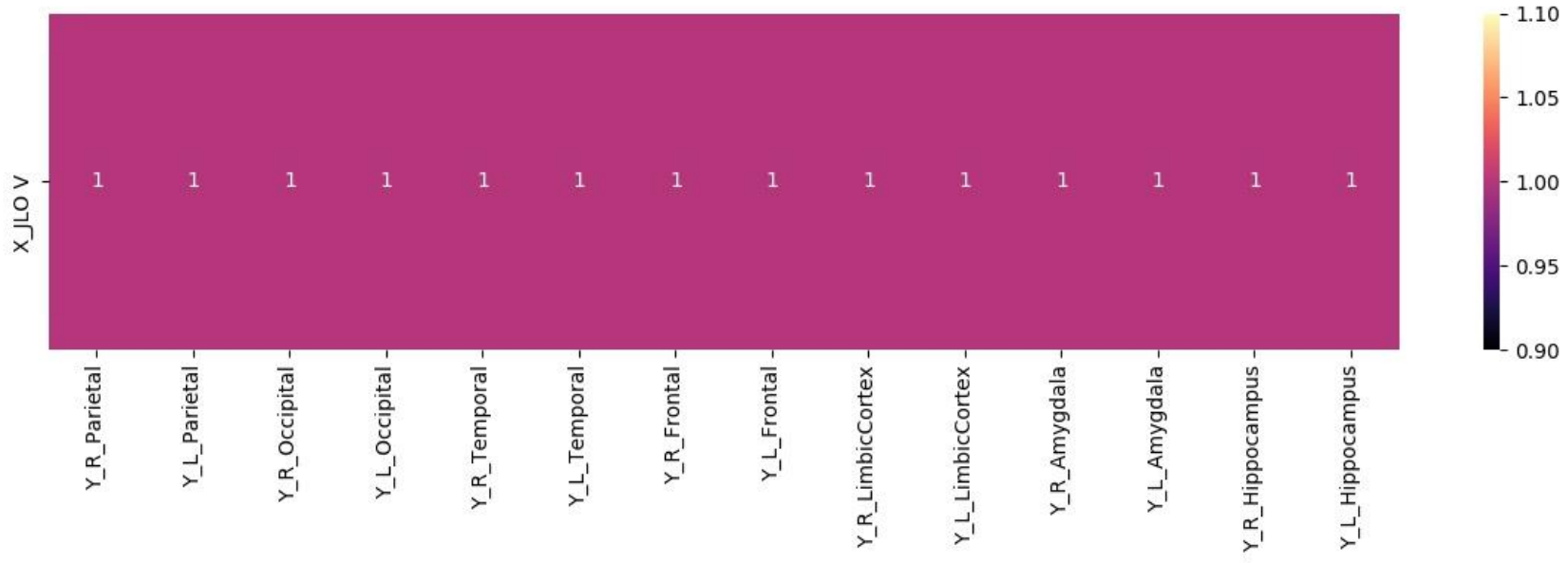


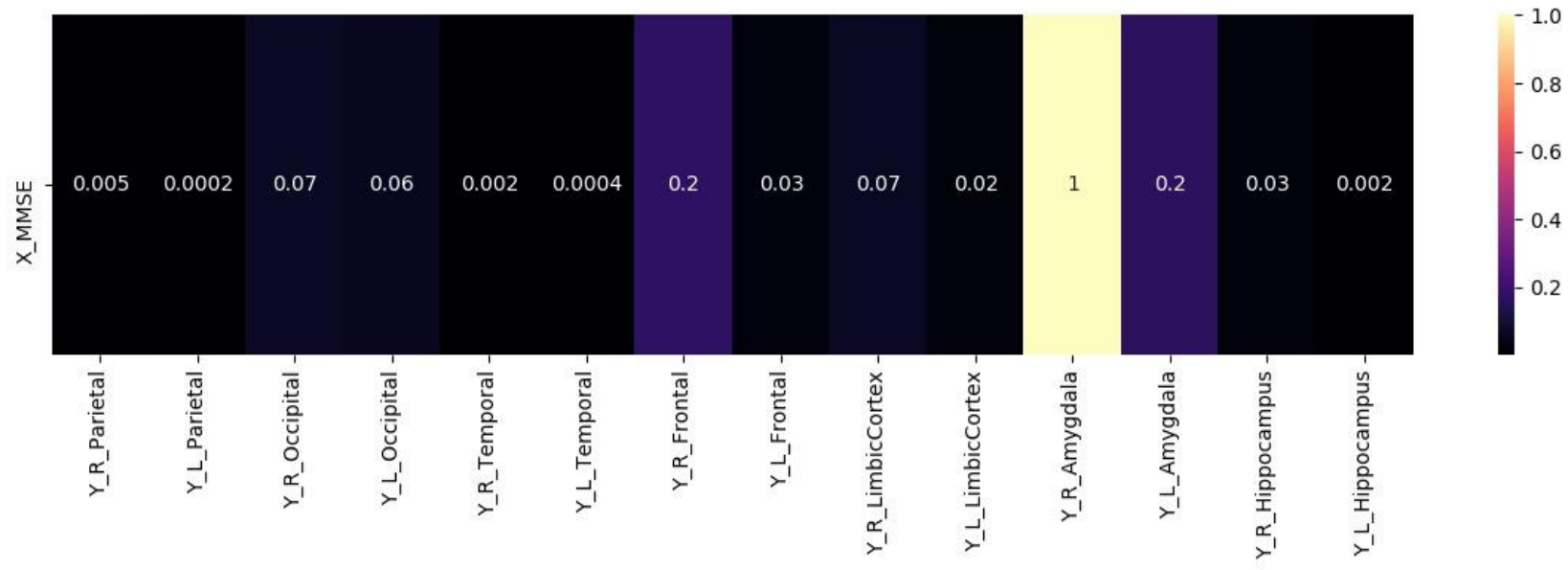


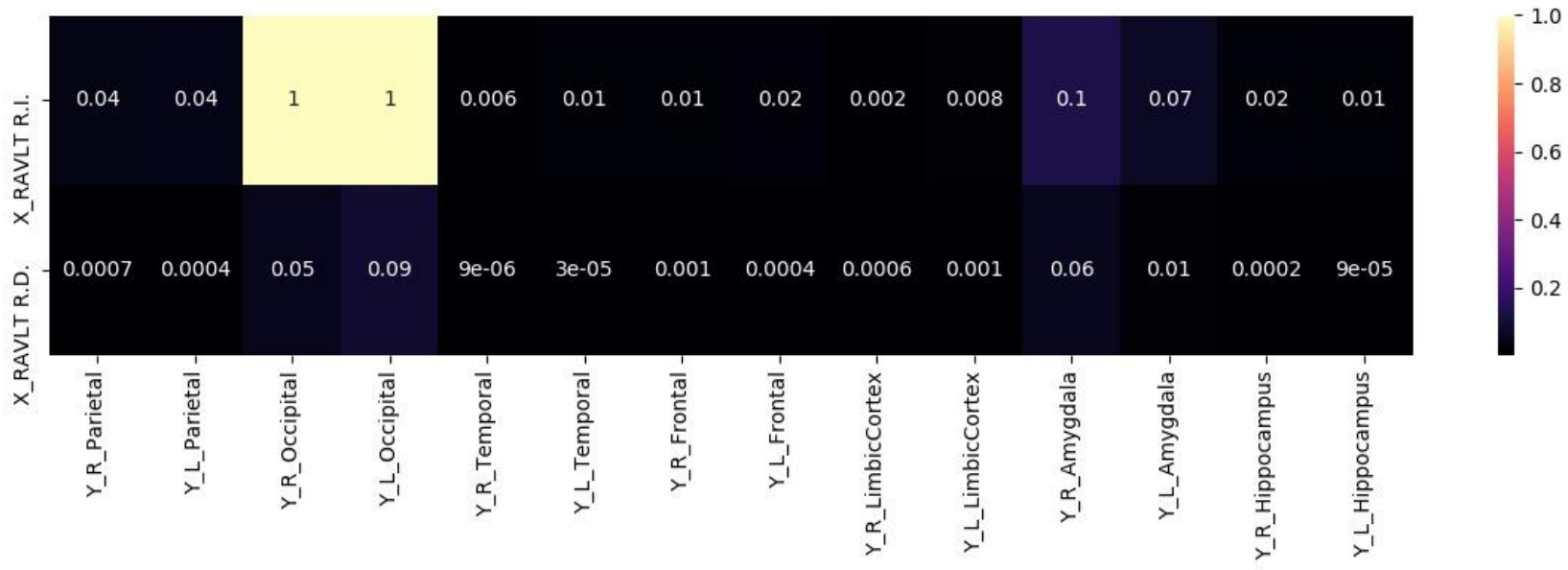


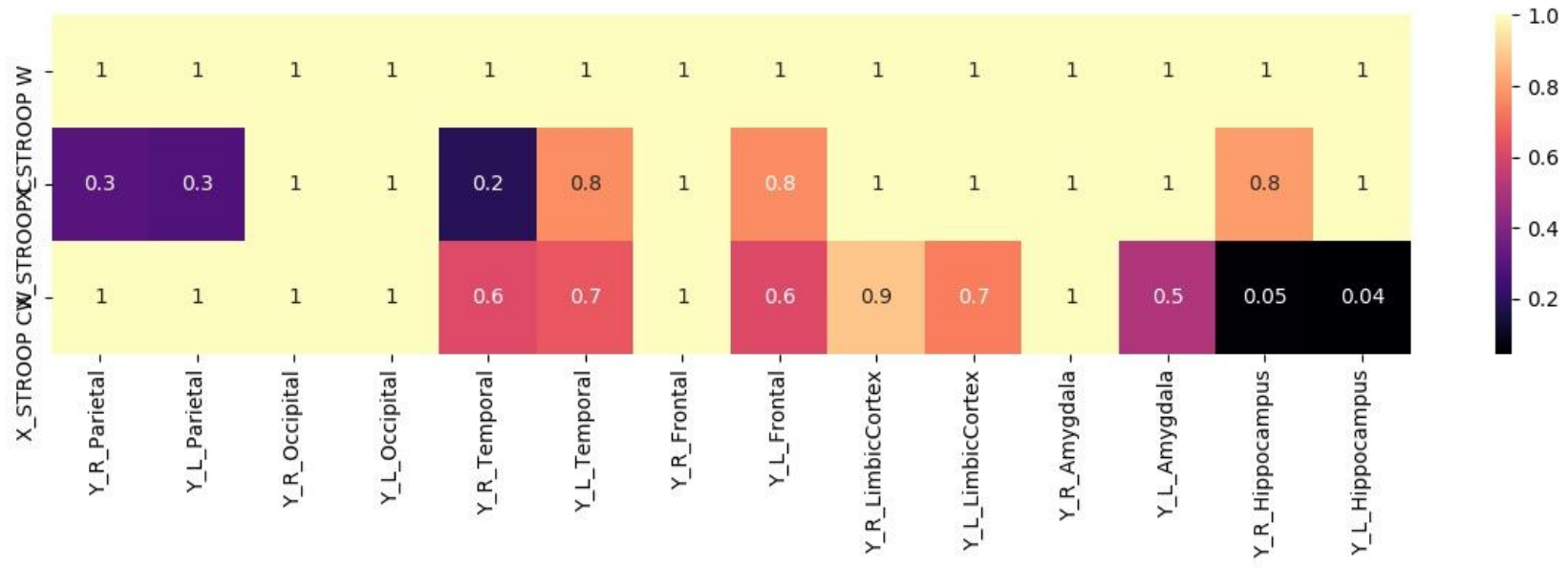


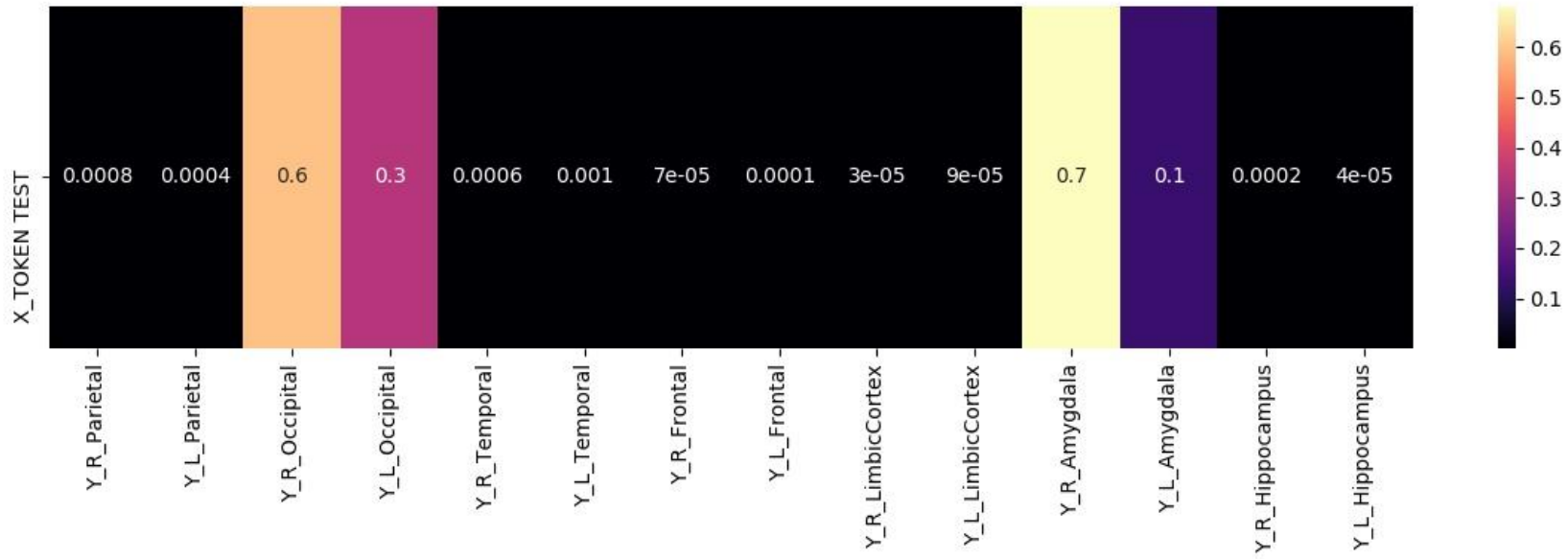


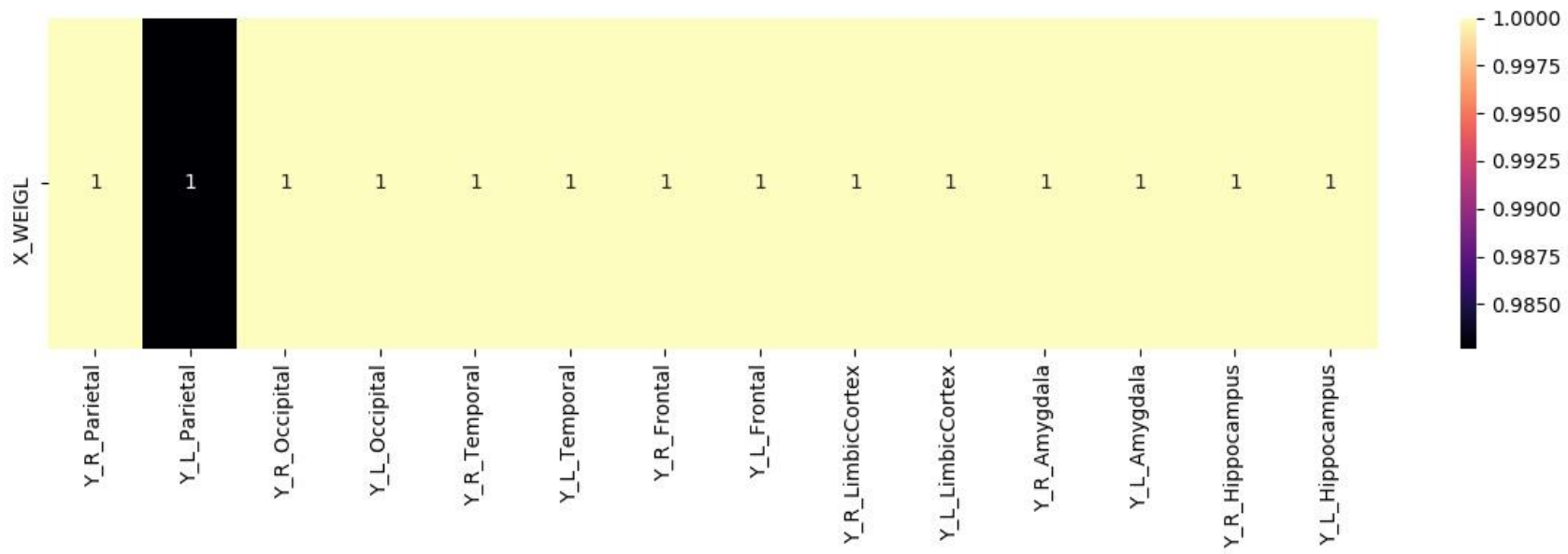










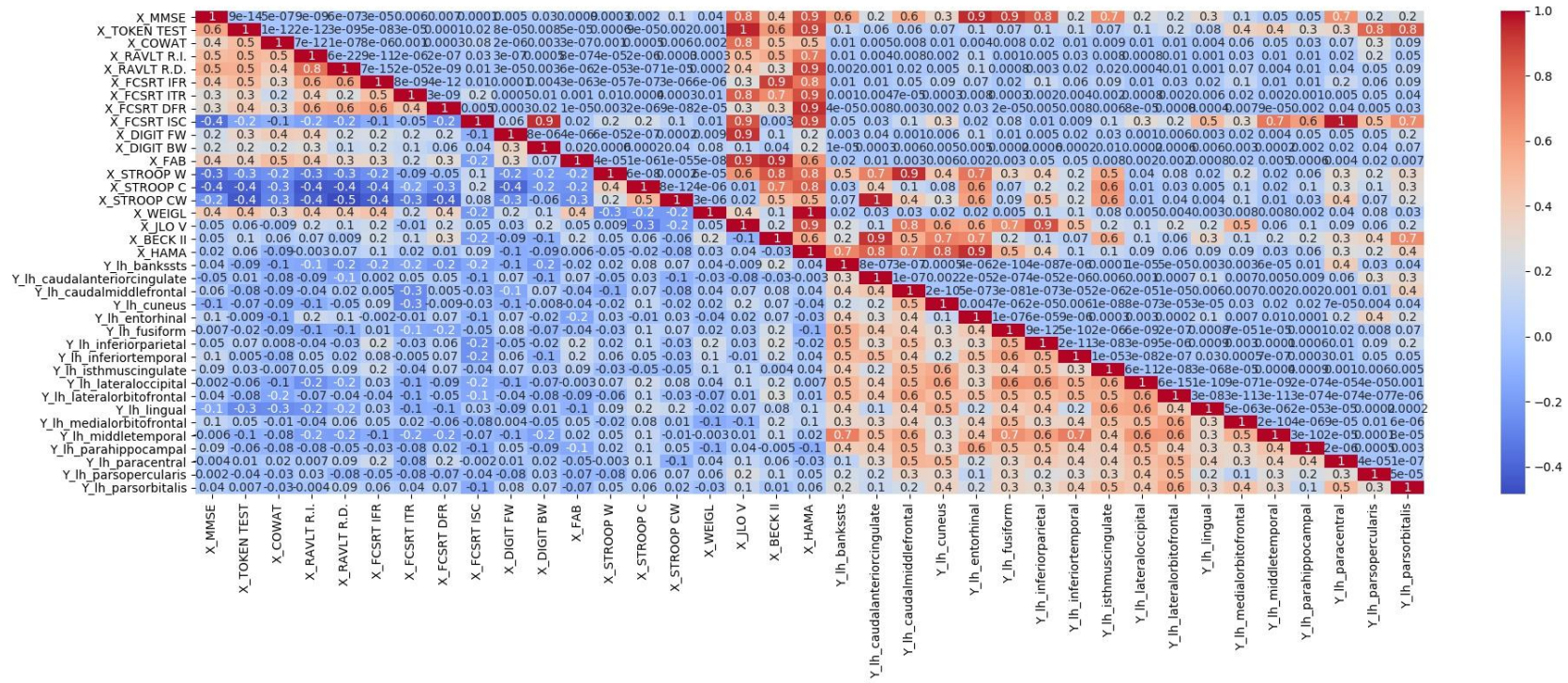


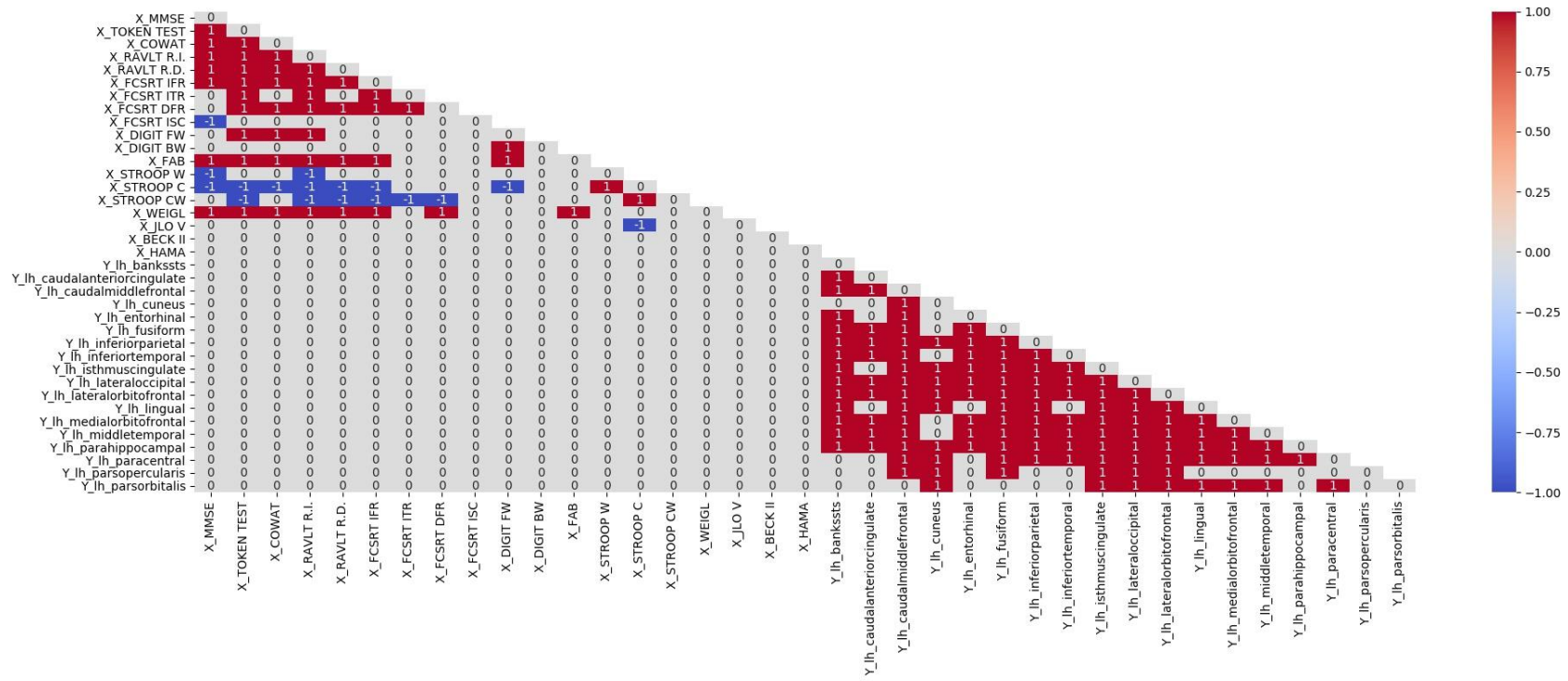
Allegato 2

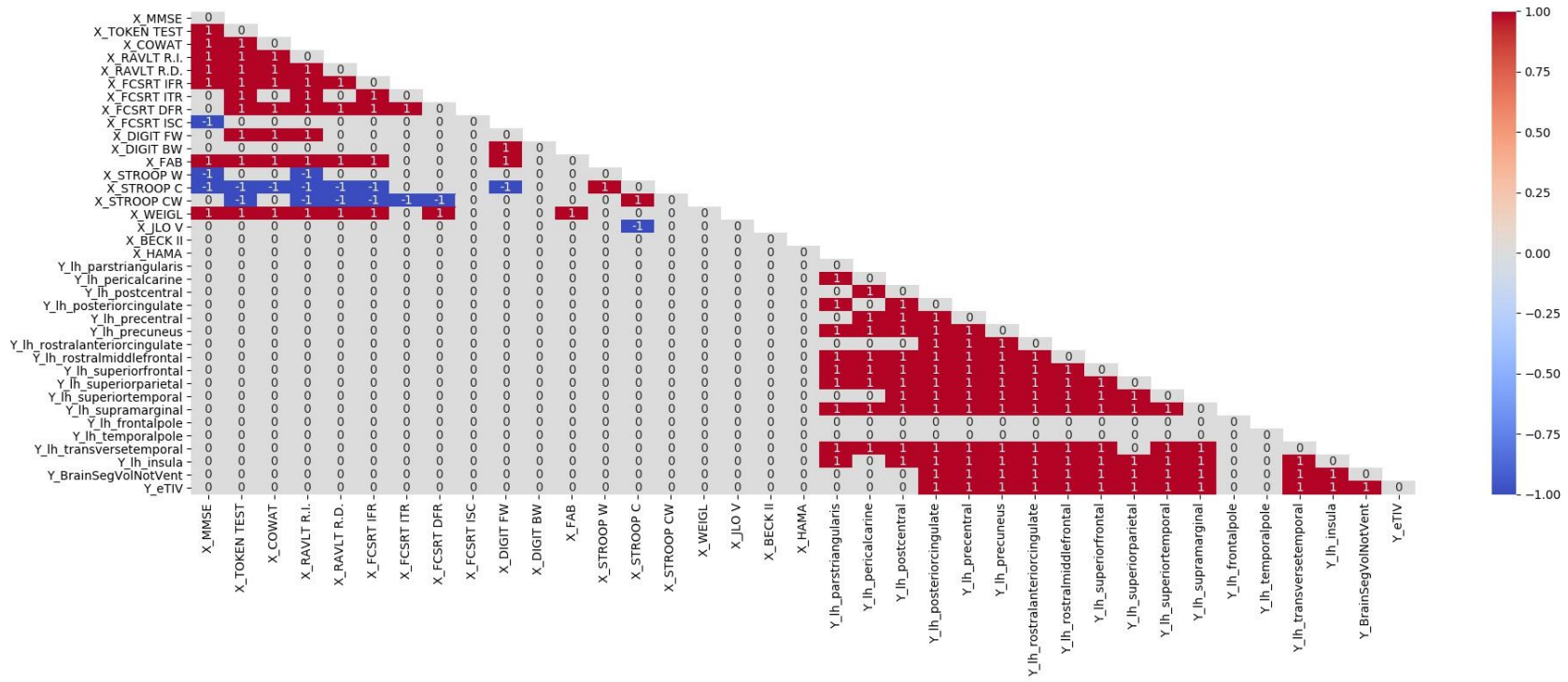
Morbo di Alzheimer

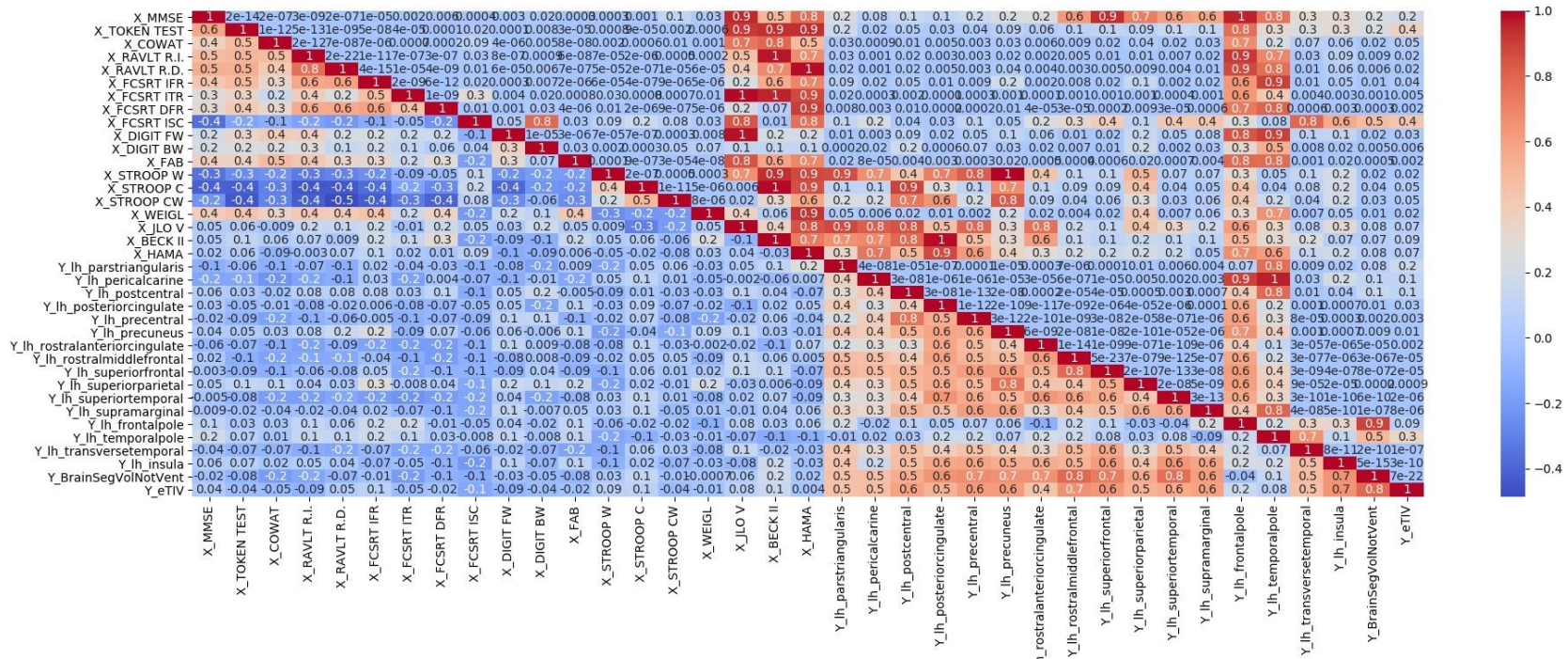
Test Cognitivi vs Volume

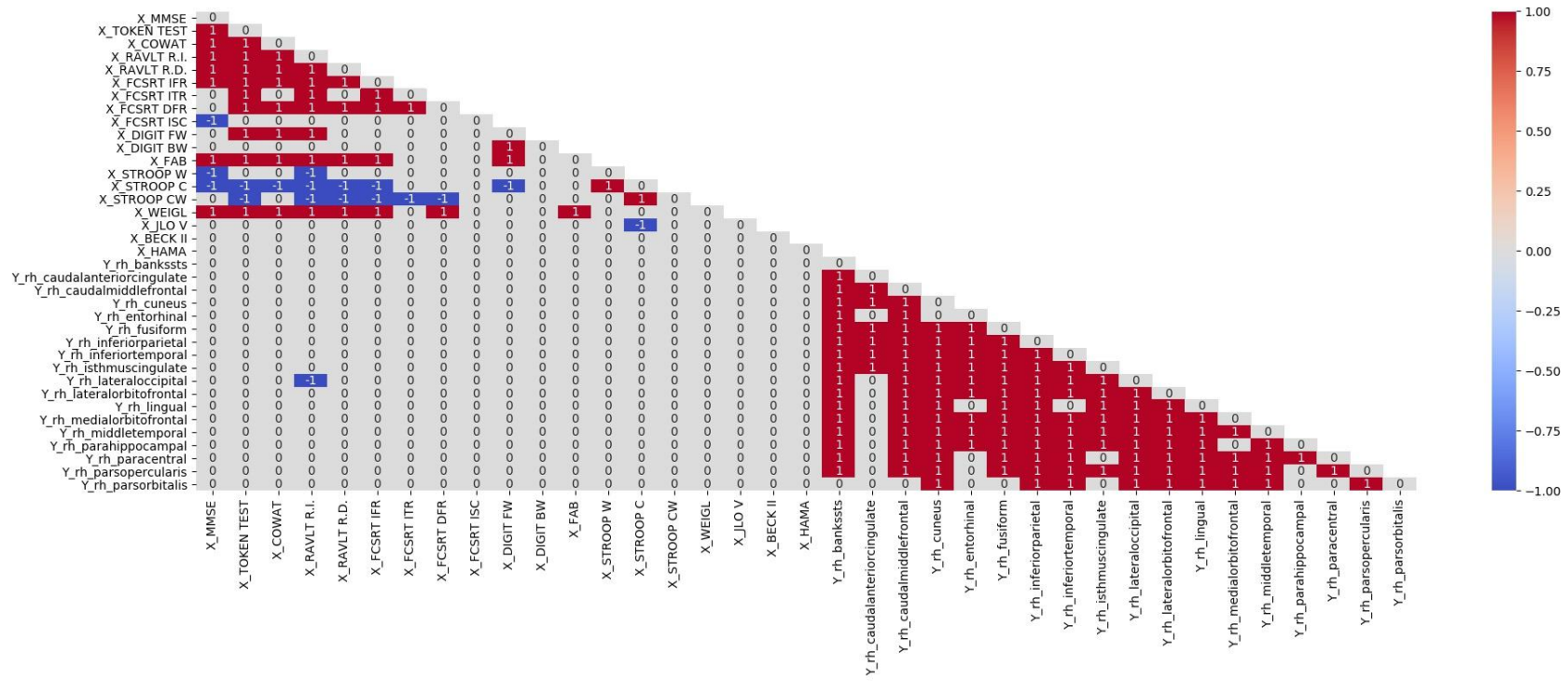
Correlazione di Pearson

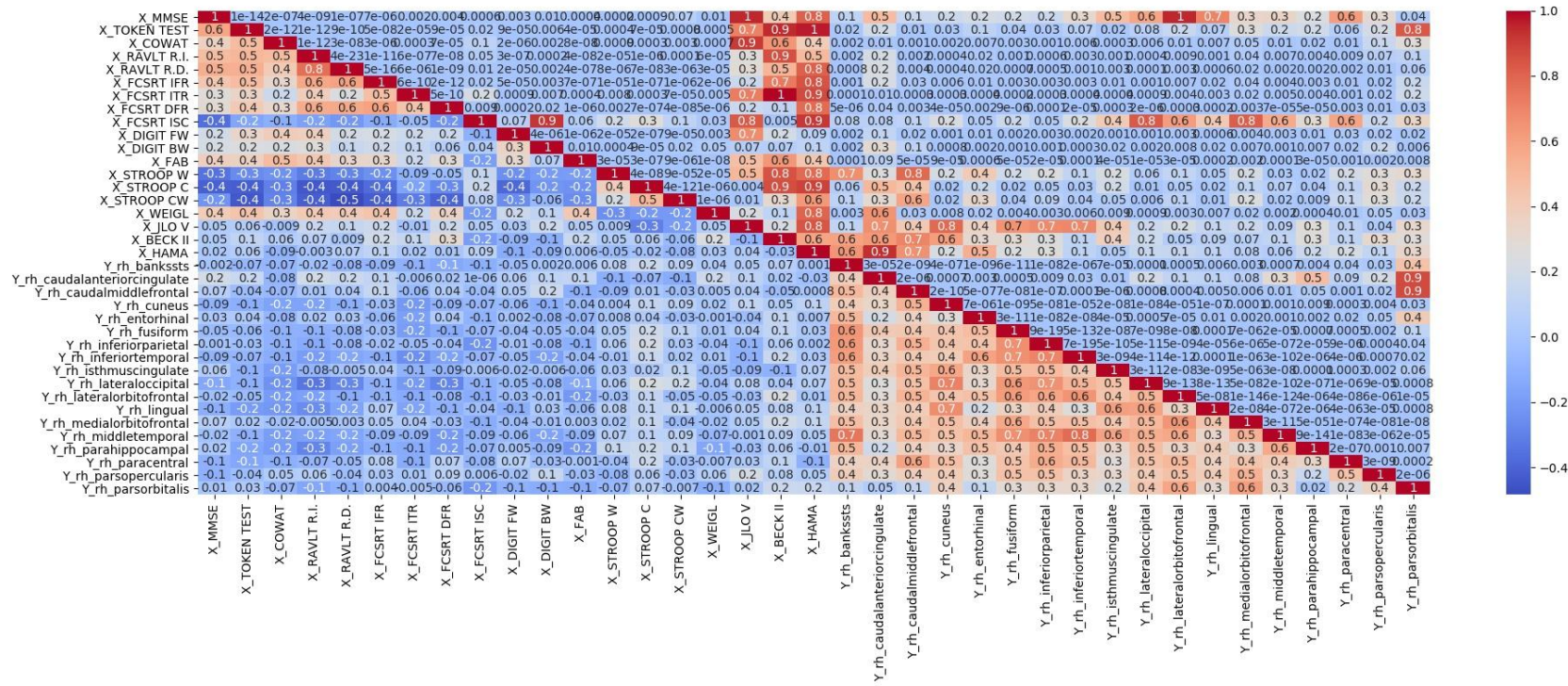


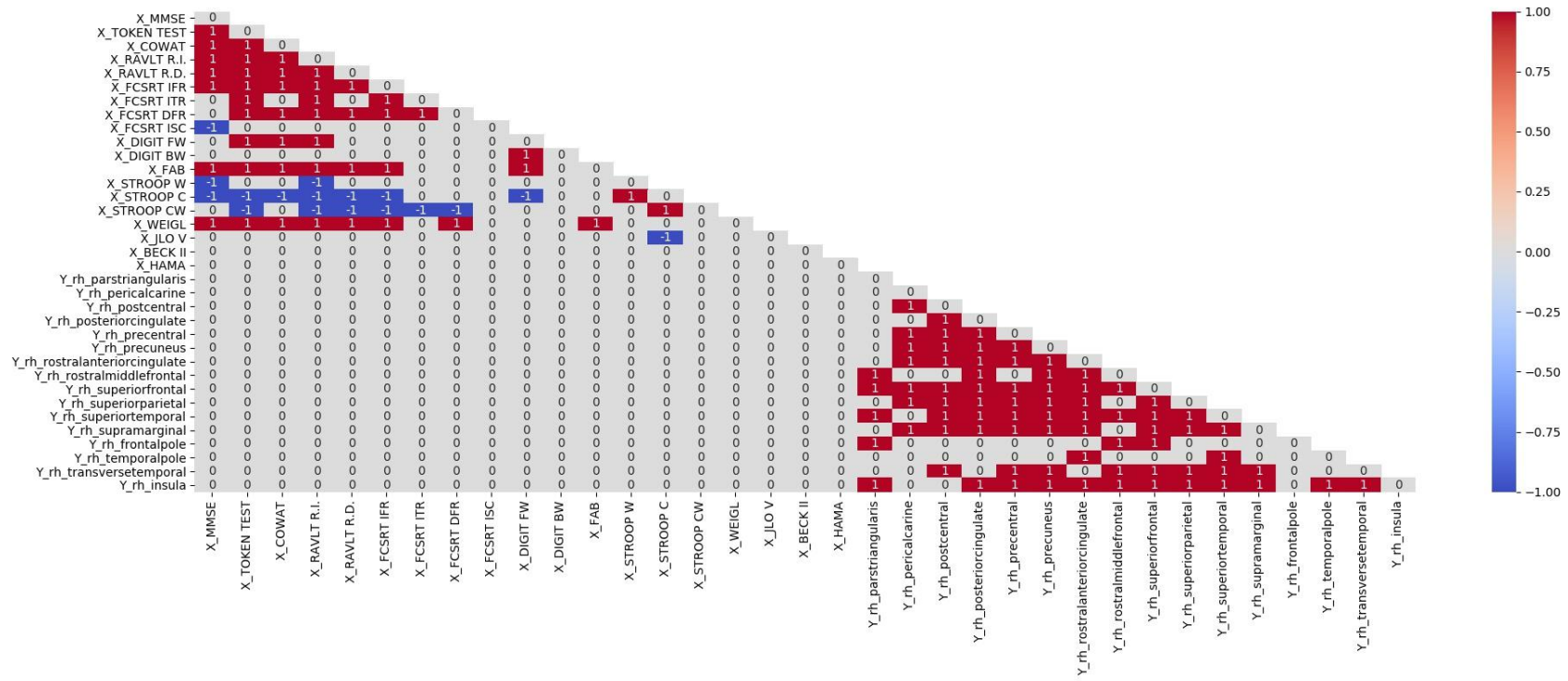


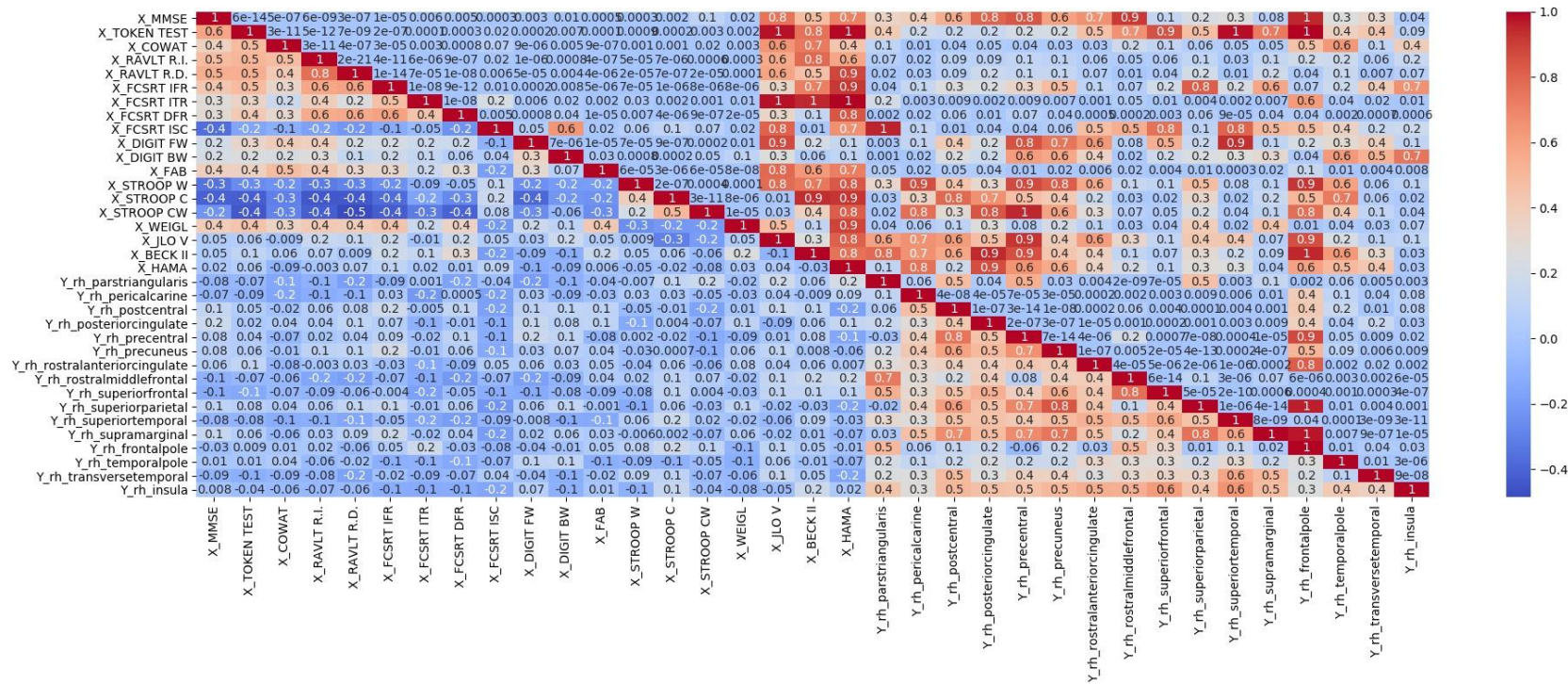




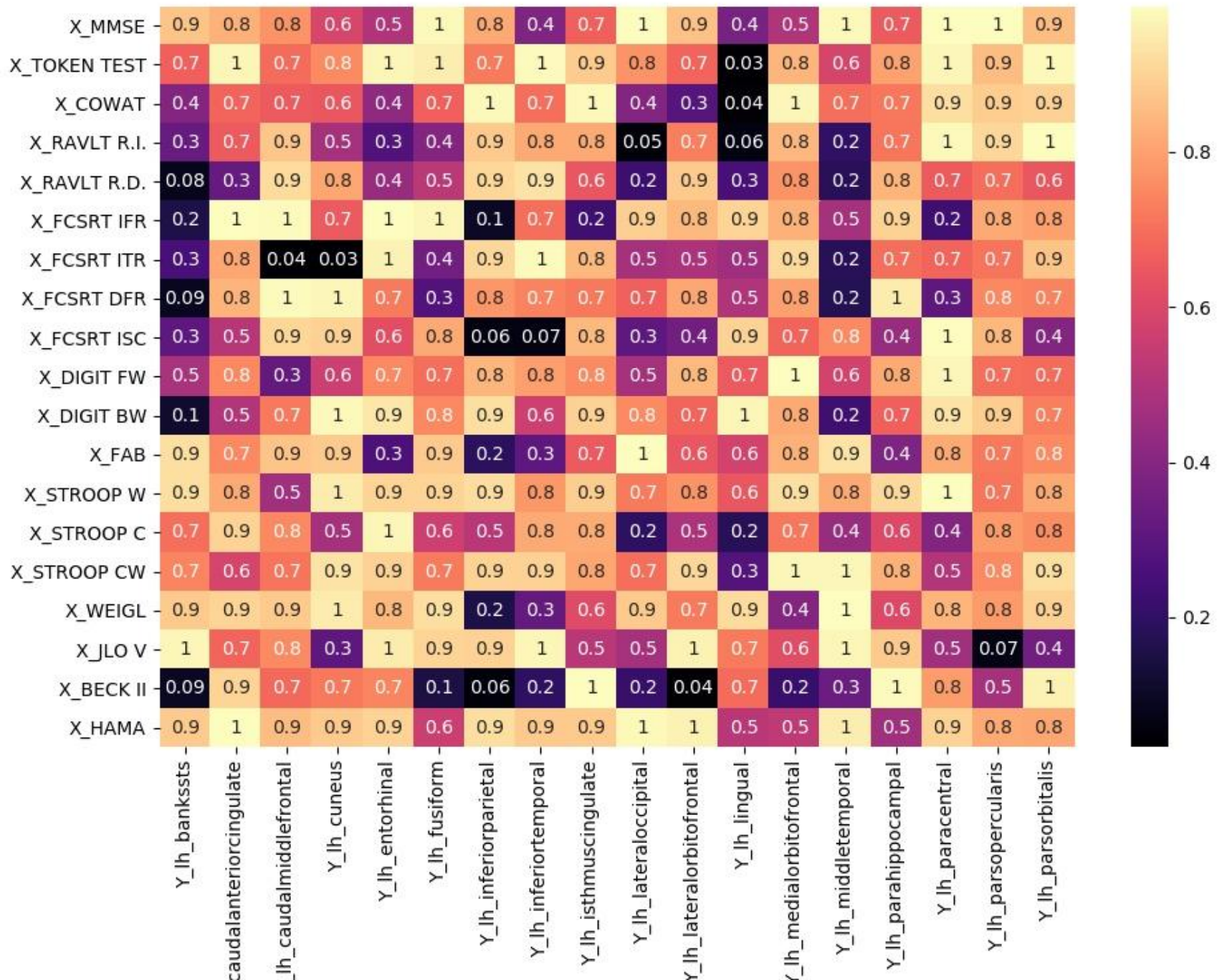


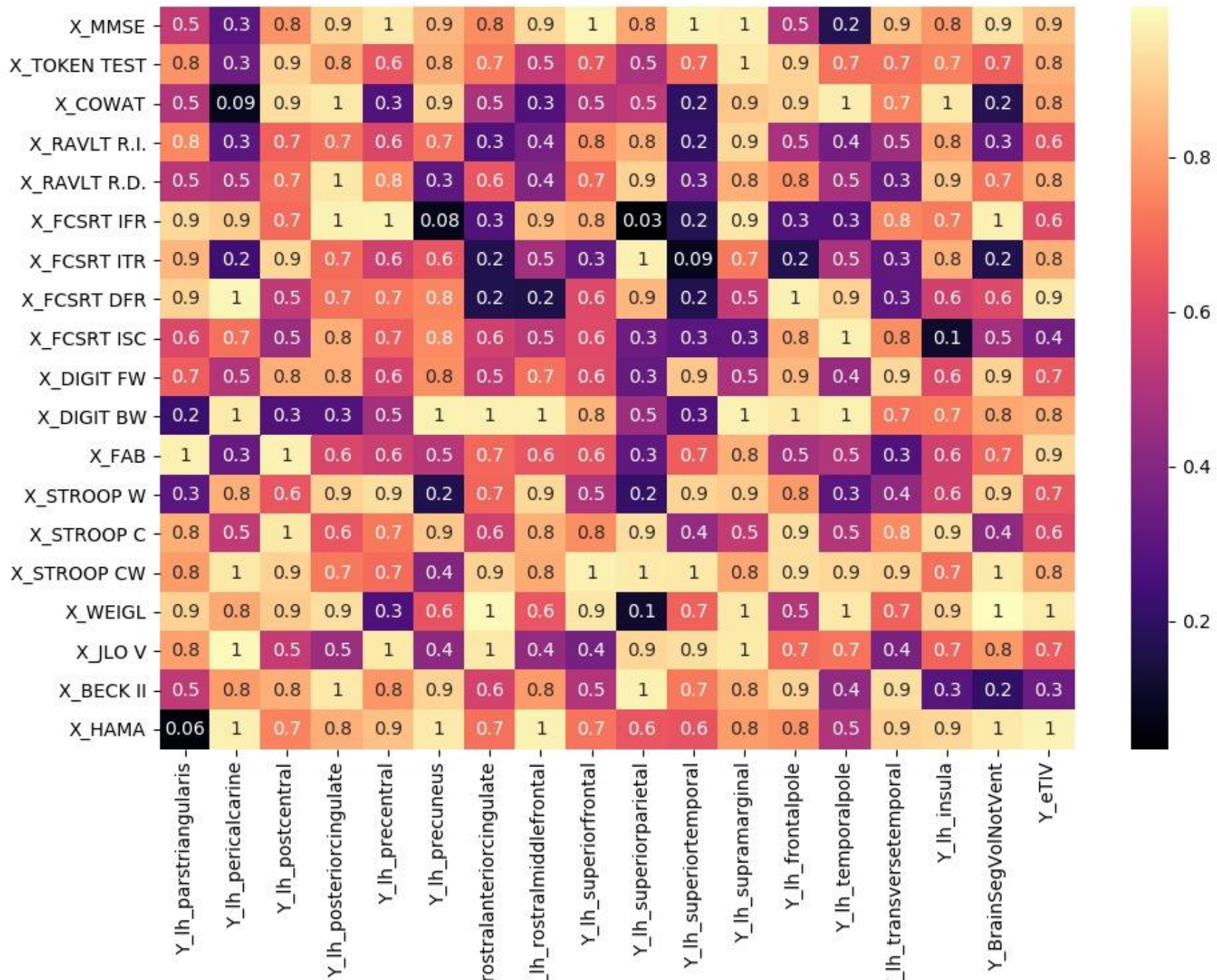


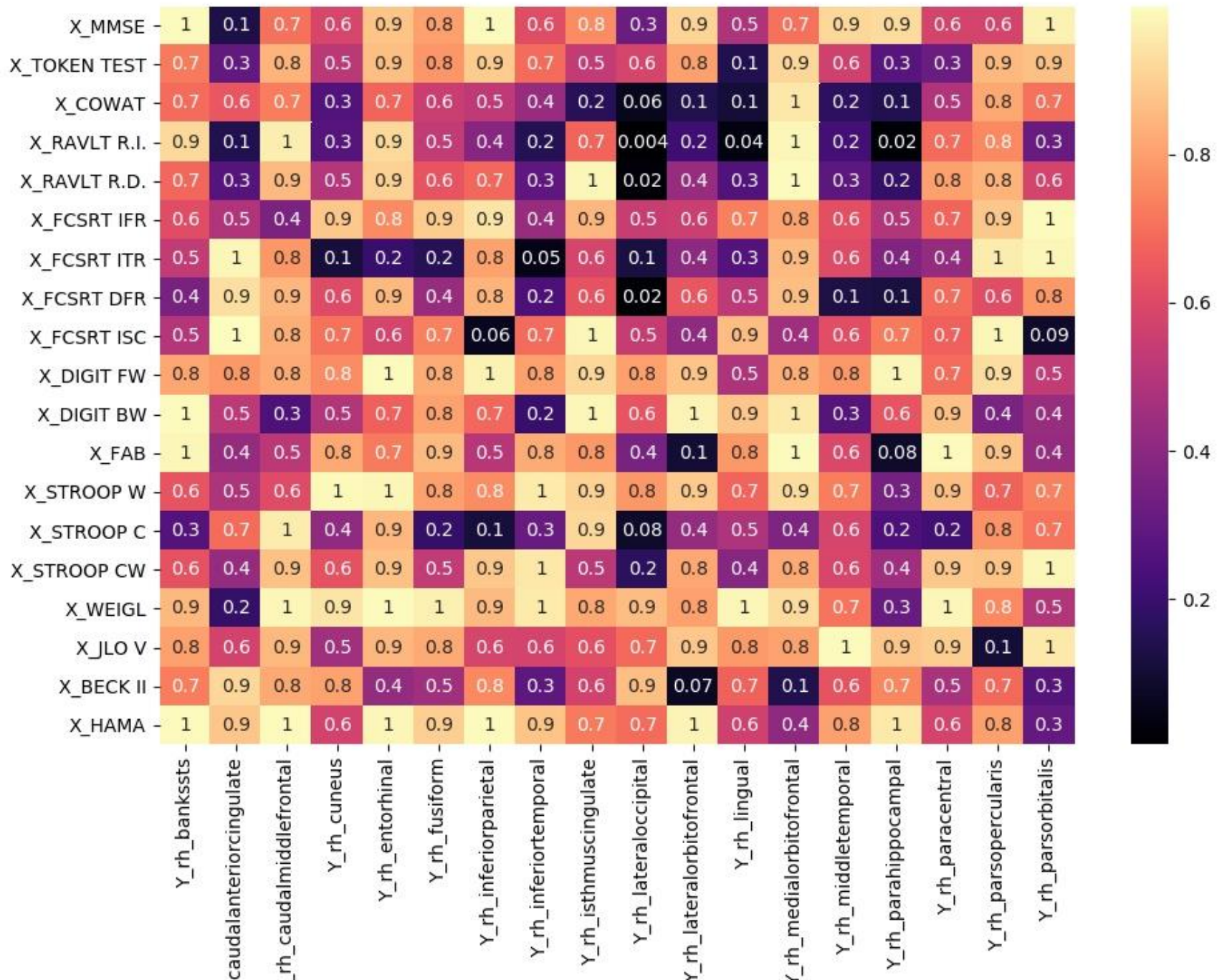


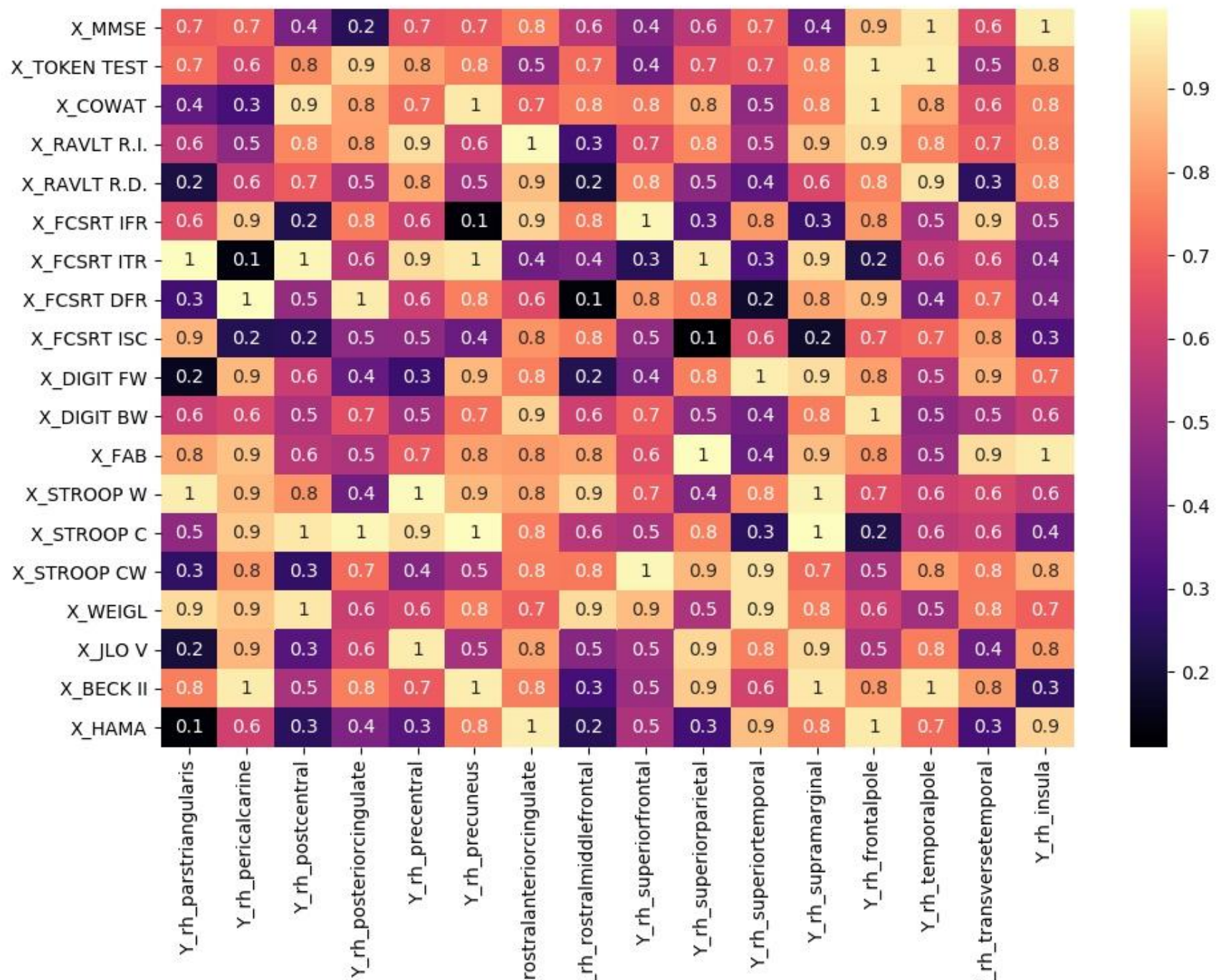


Correzione con Benjamini-Hochberg









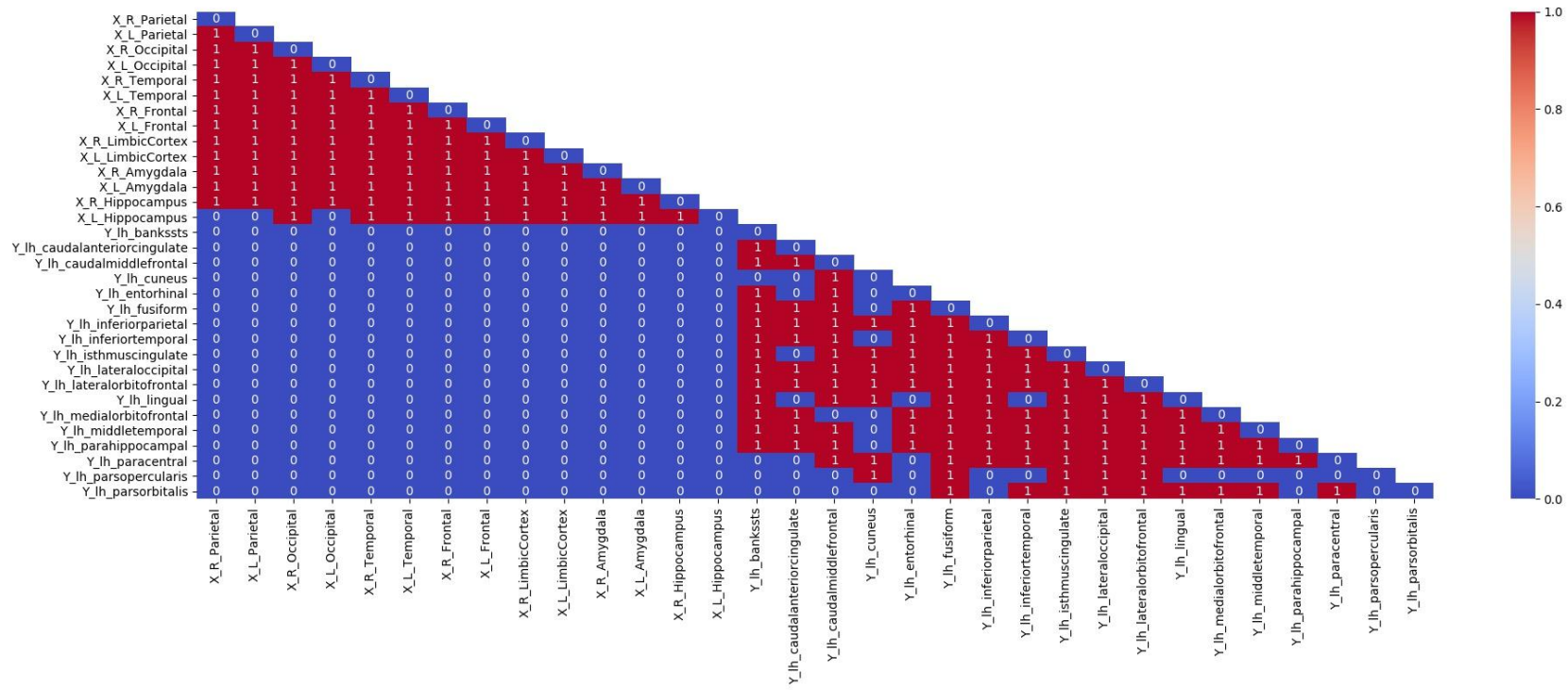
Correzione con Bonferroni

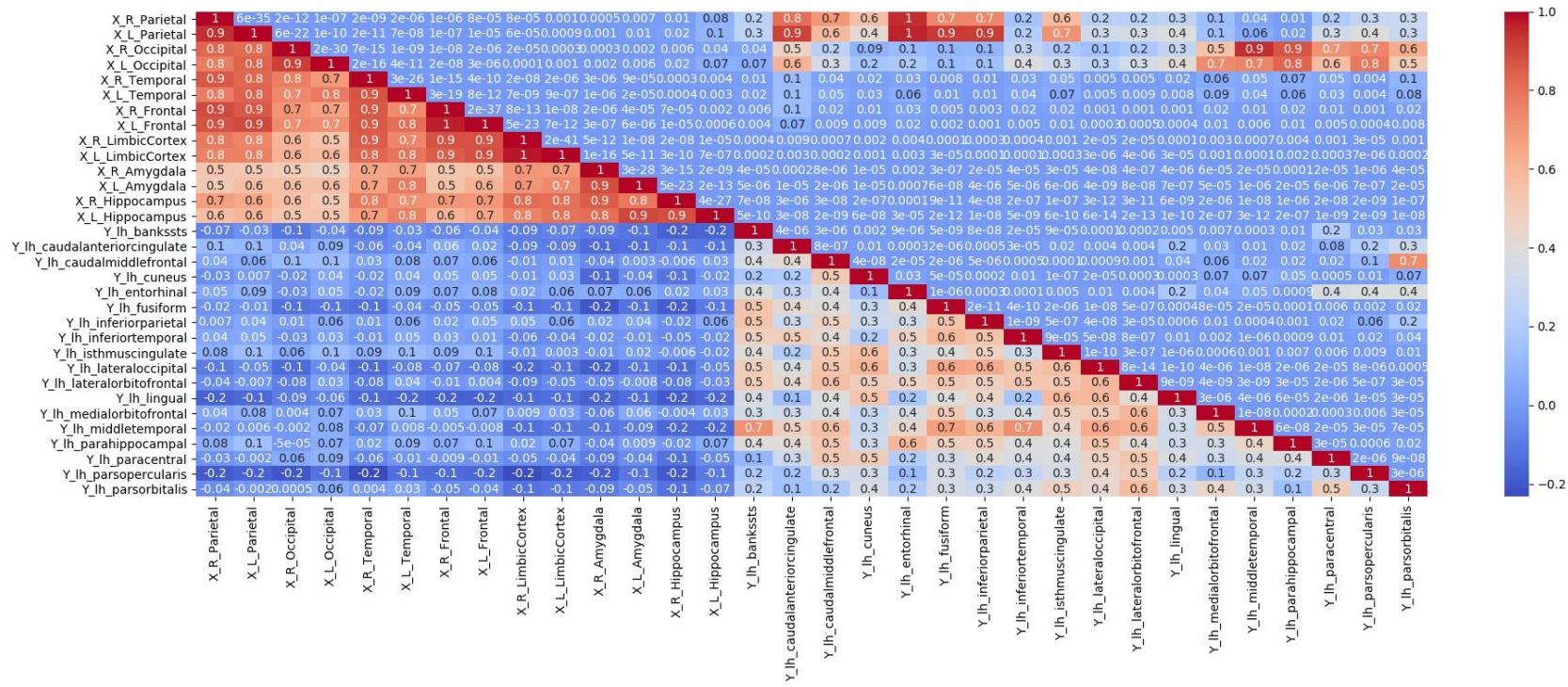
Allegato 3

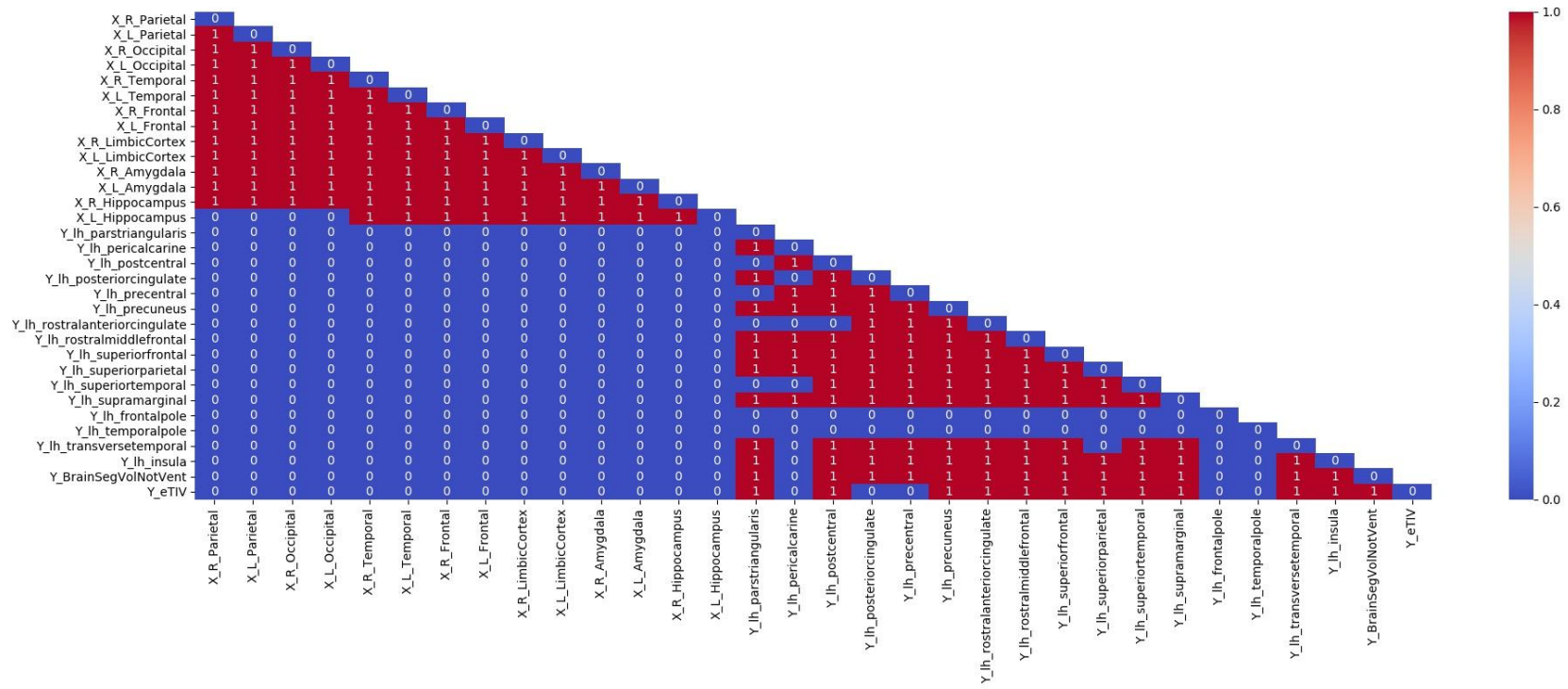
Morbo di Alzheimer

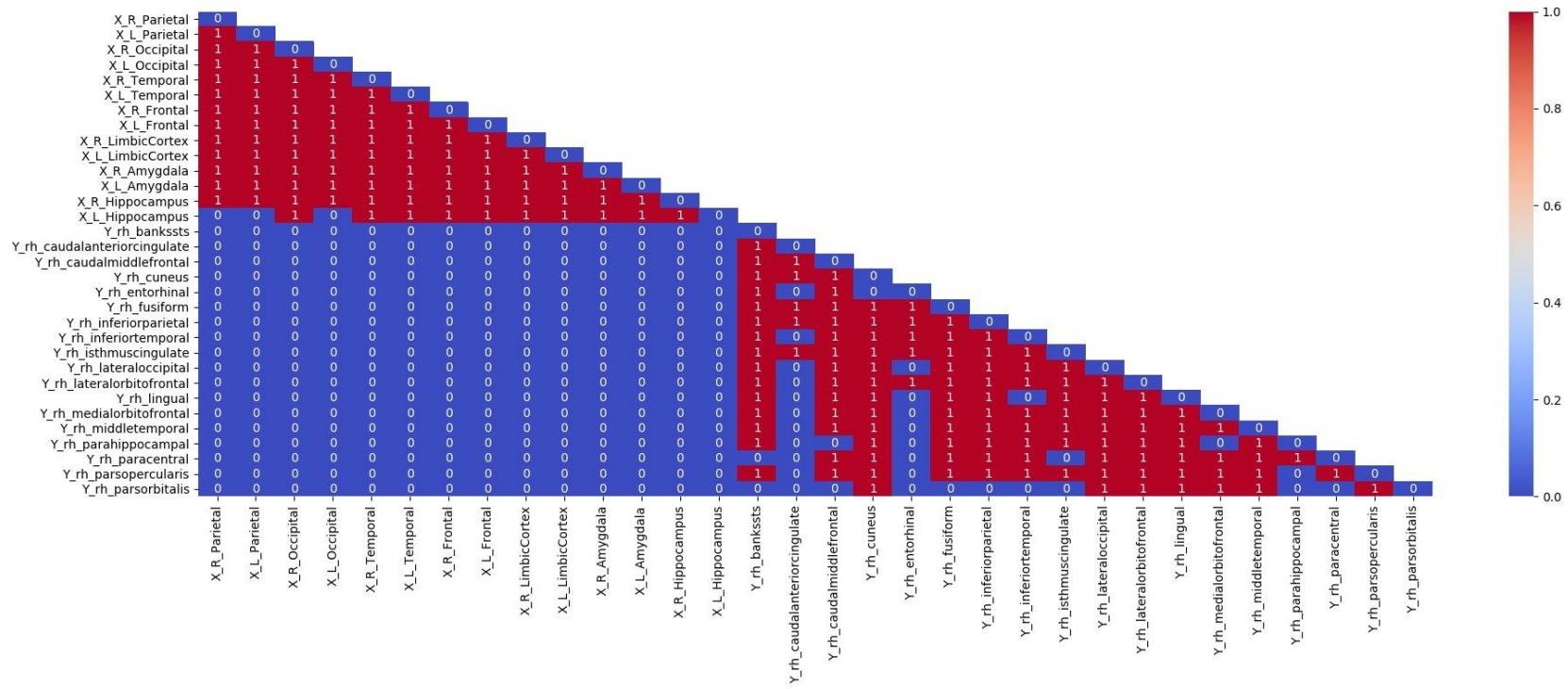
Dati di Uptake vs Volume

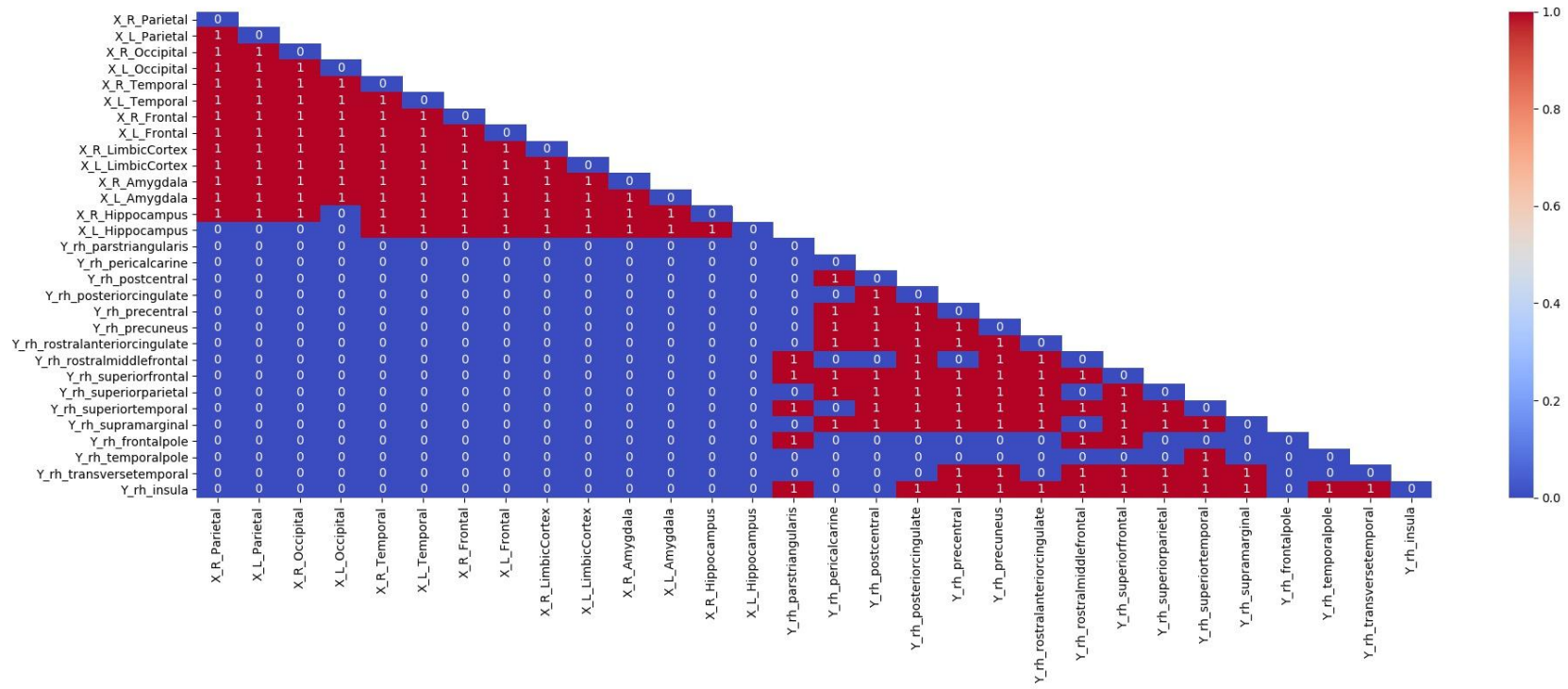
Correlazione di Pearson

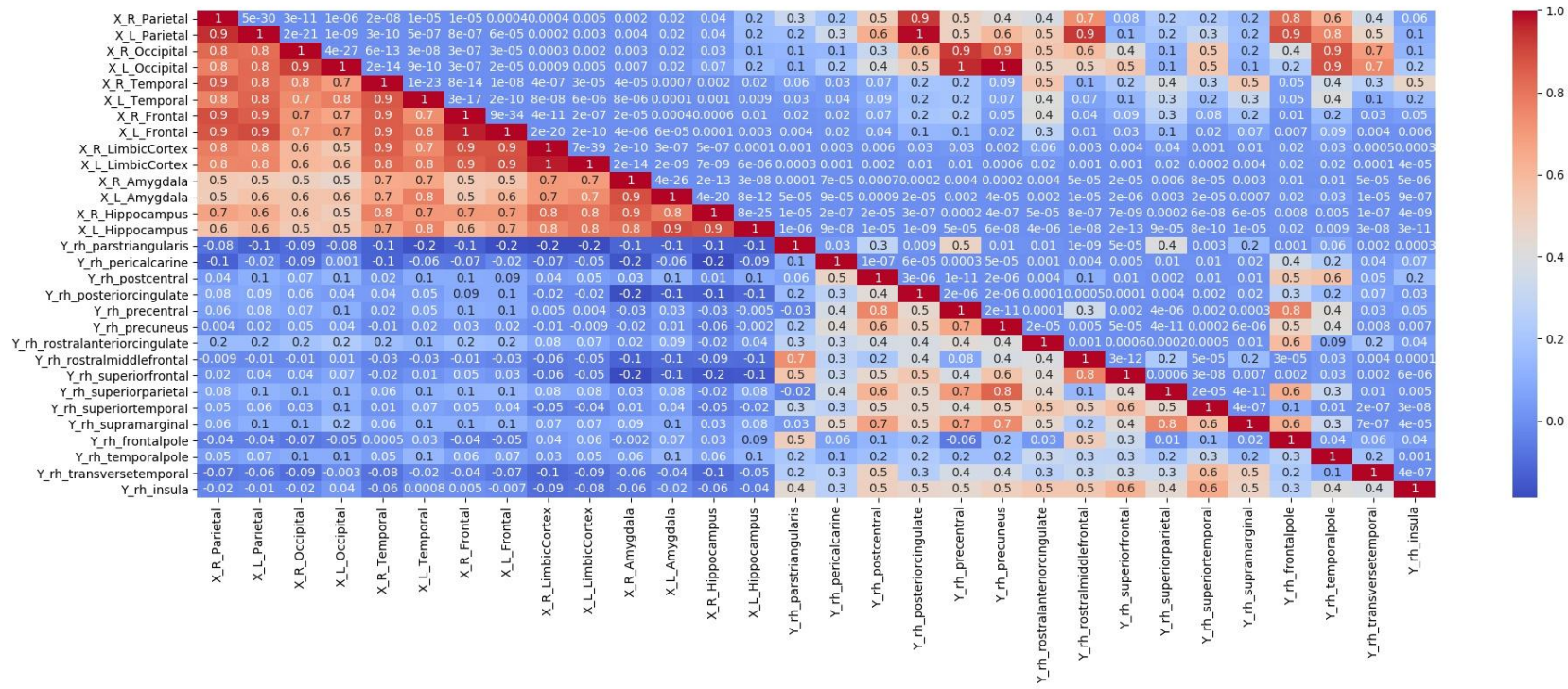




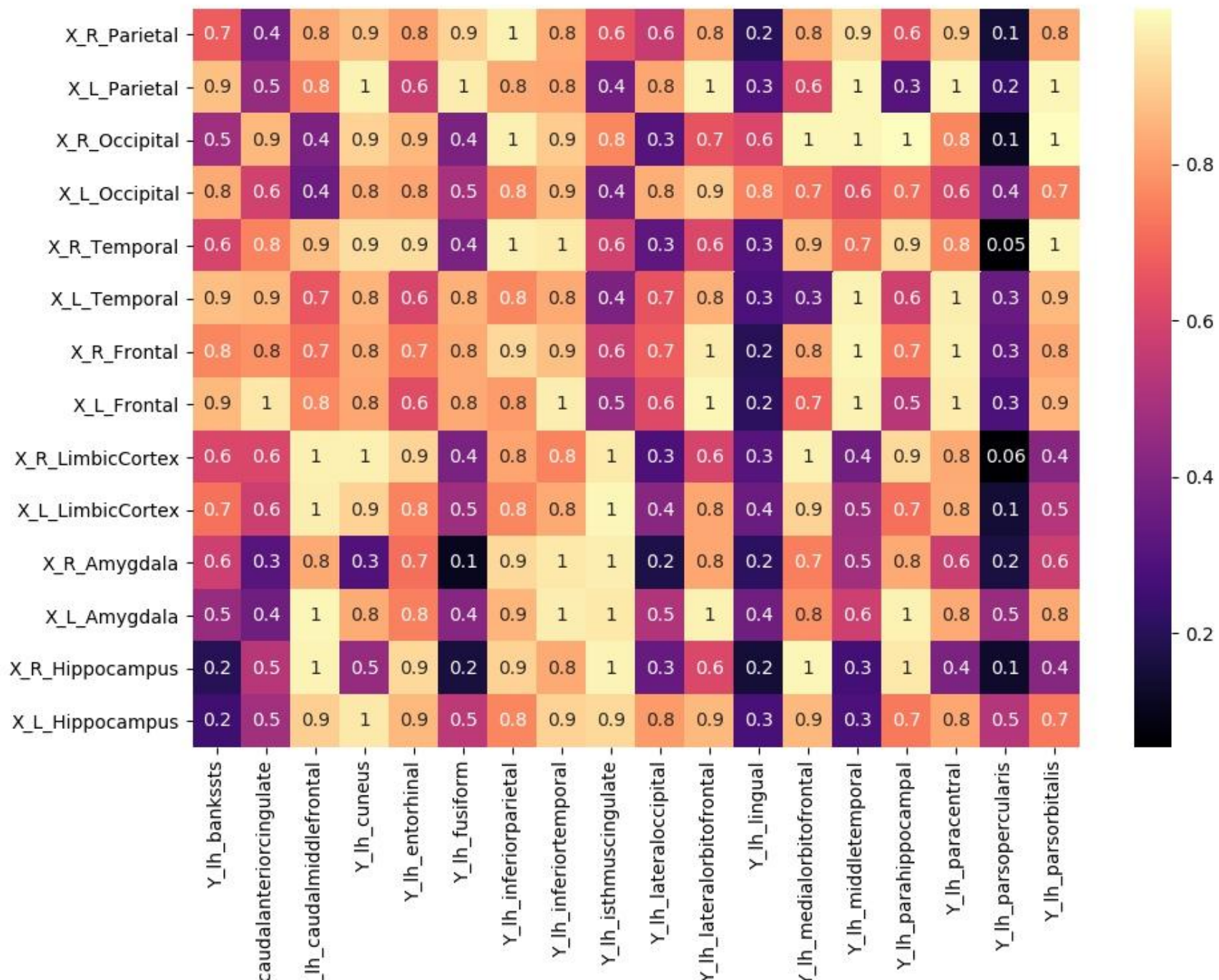


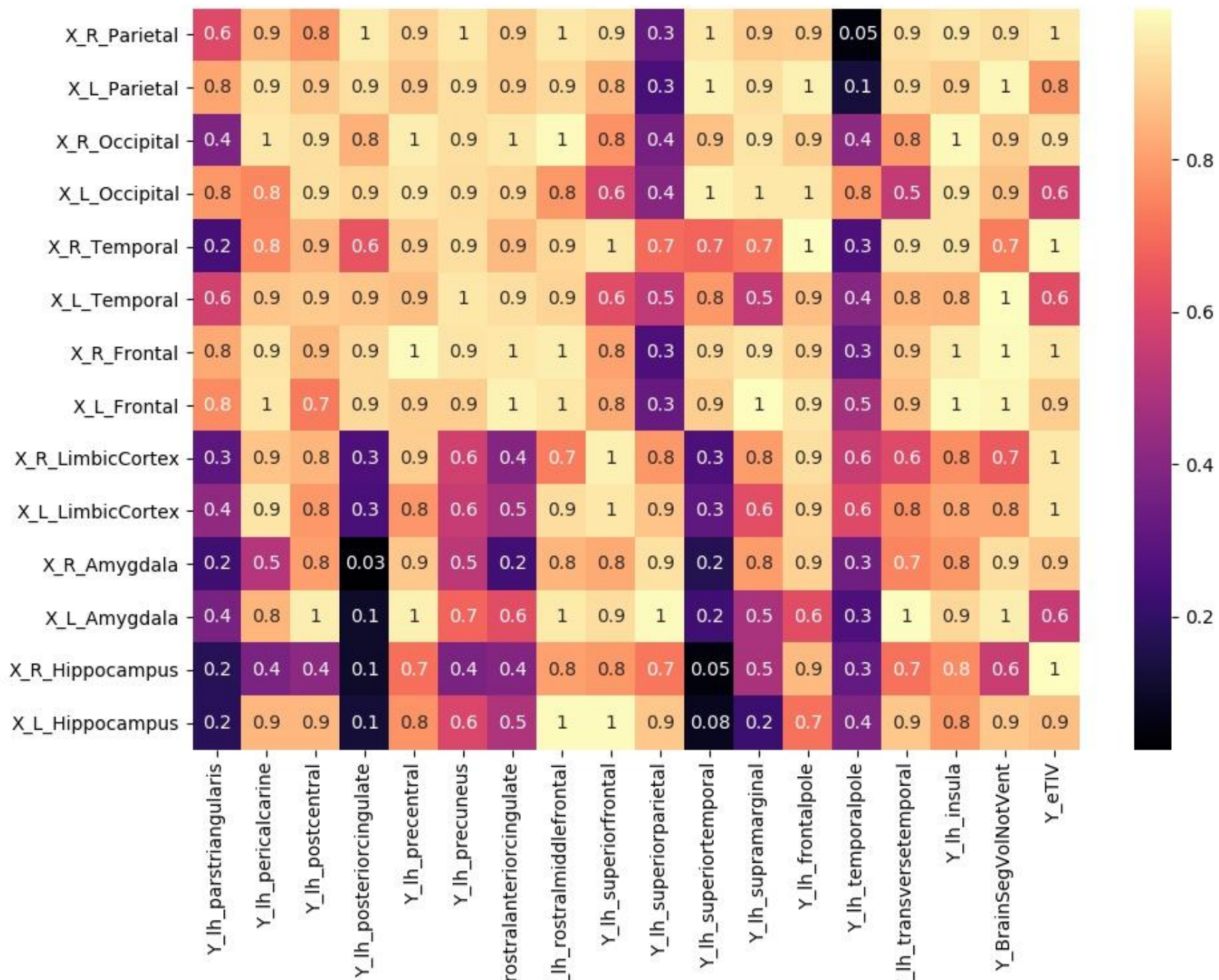


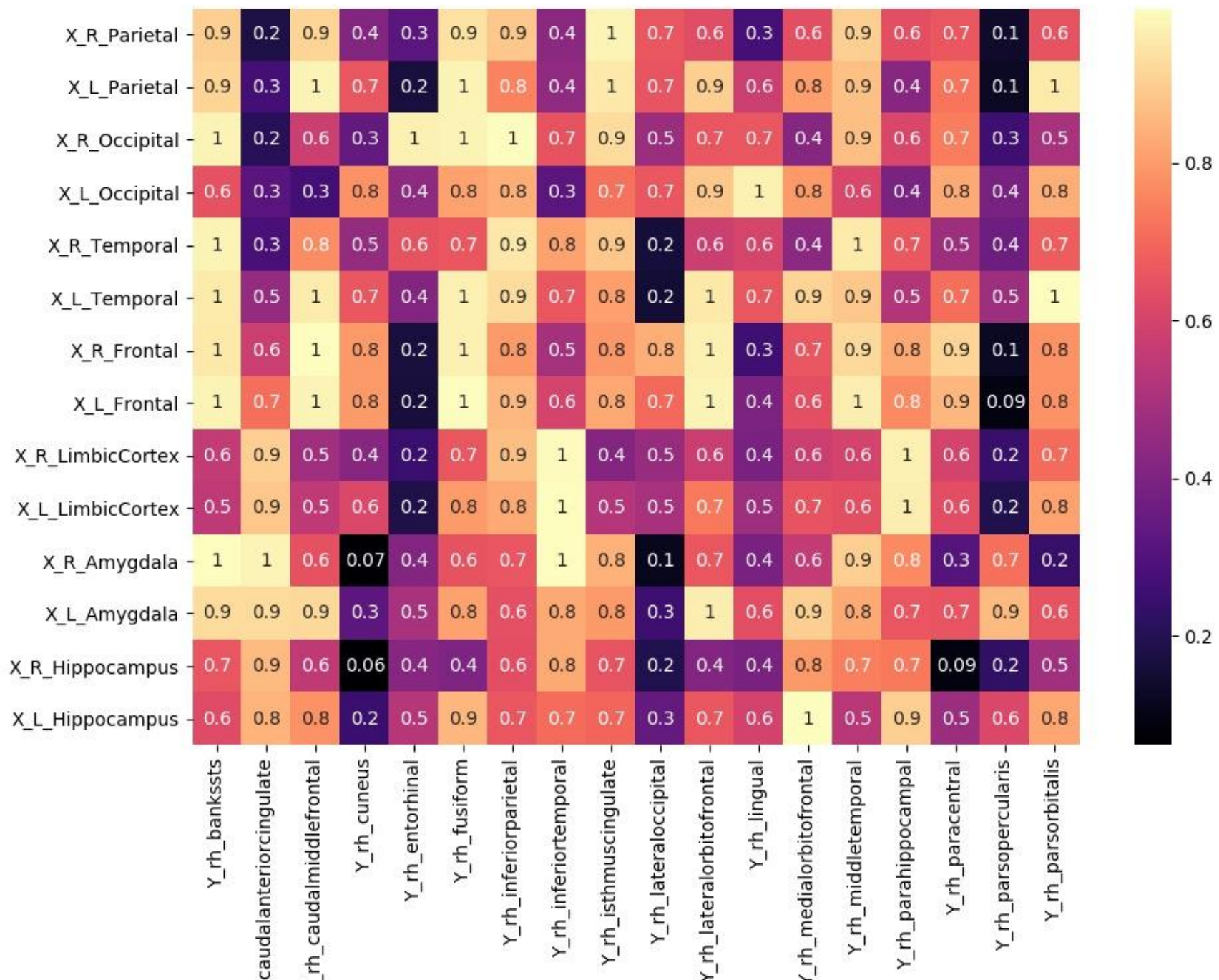


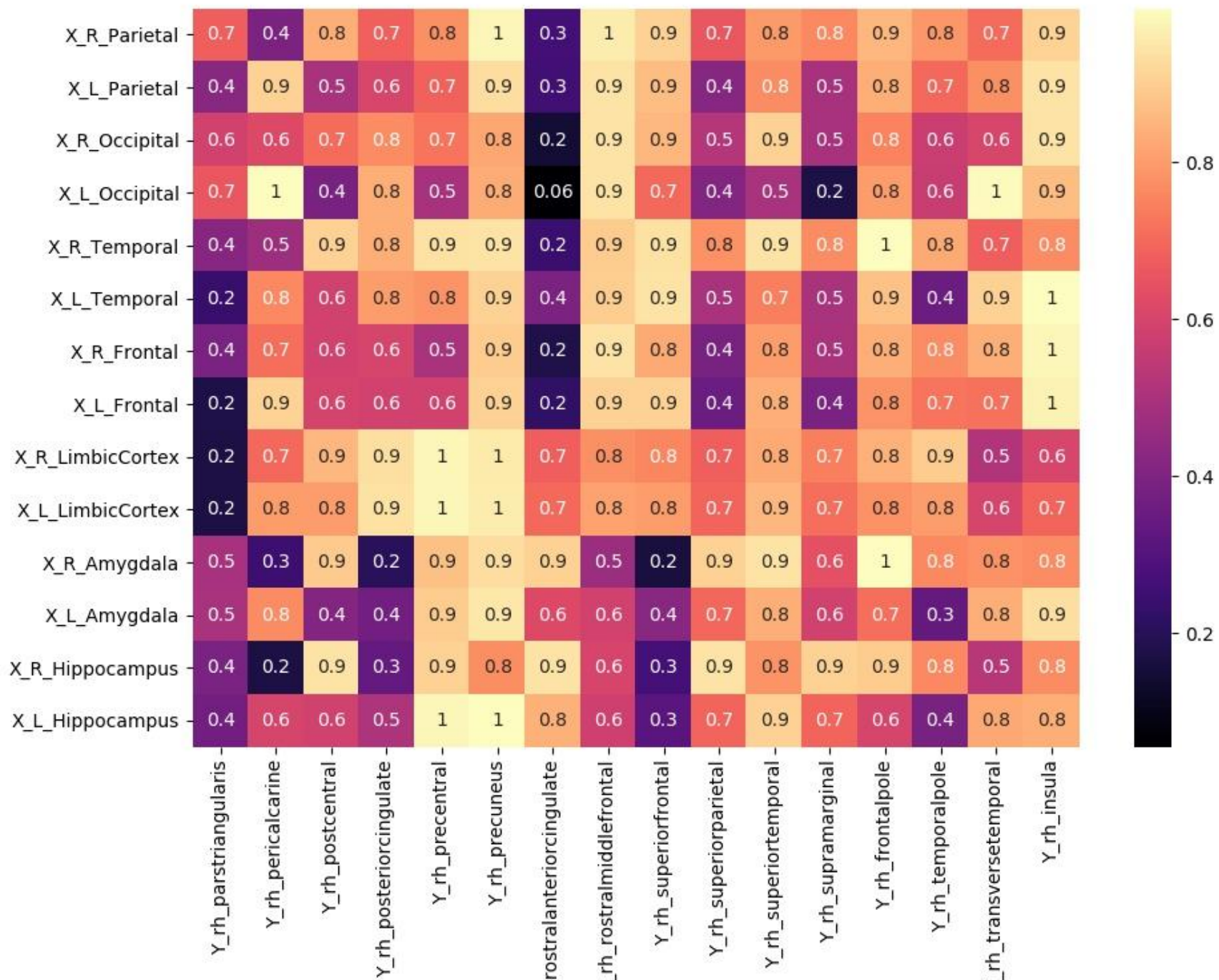


Correzione con Benjamini-Hochberg









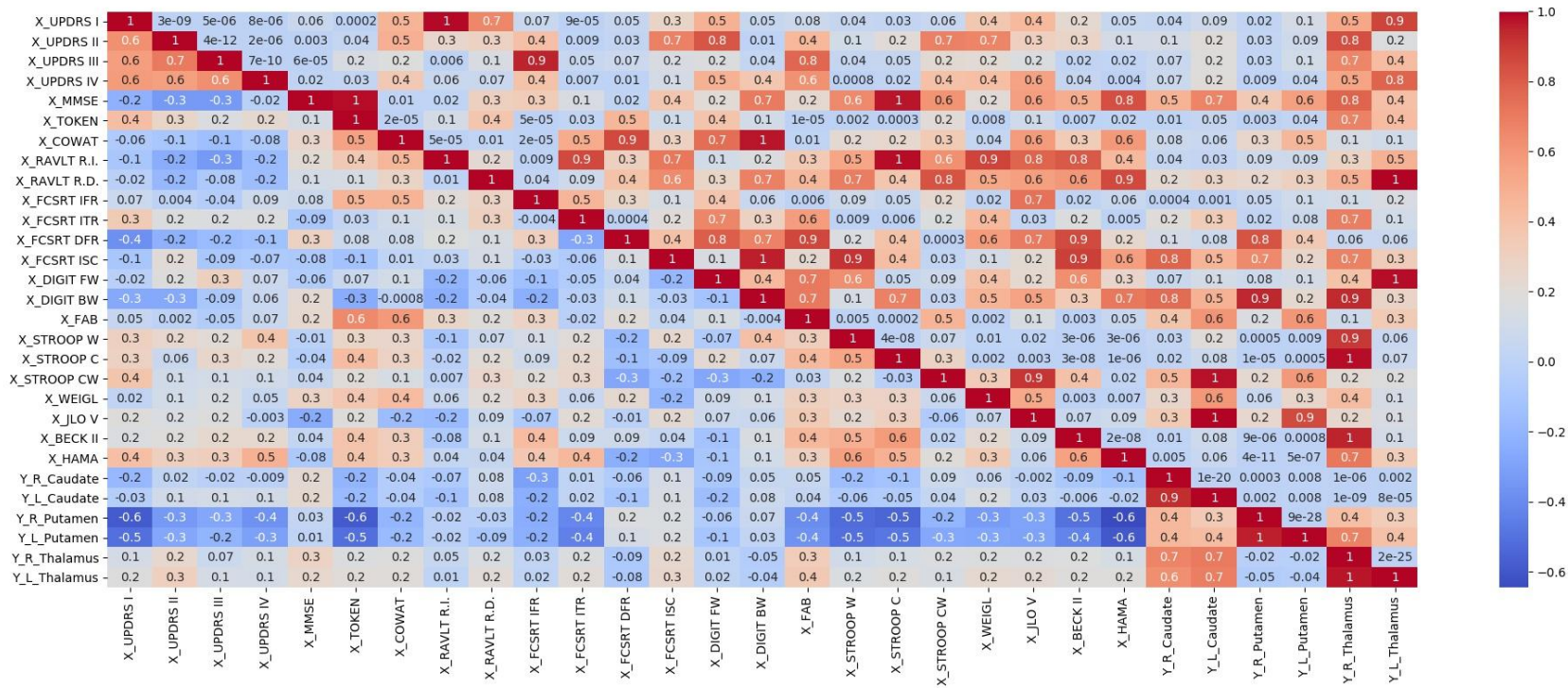
Correzione con Bonferroni

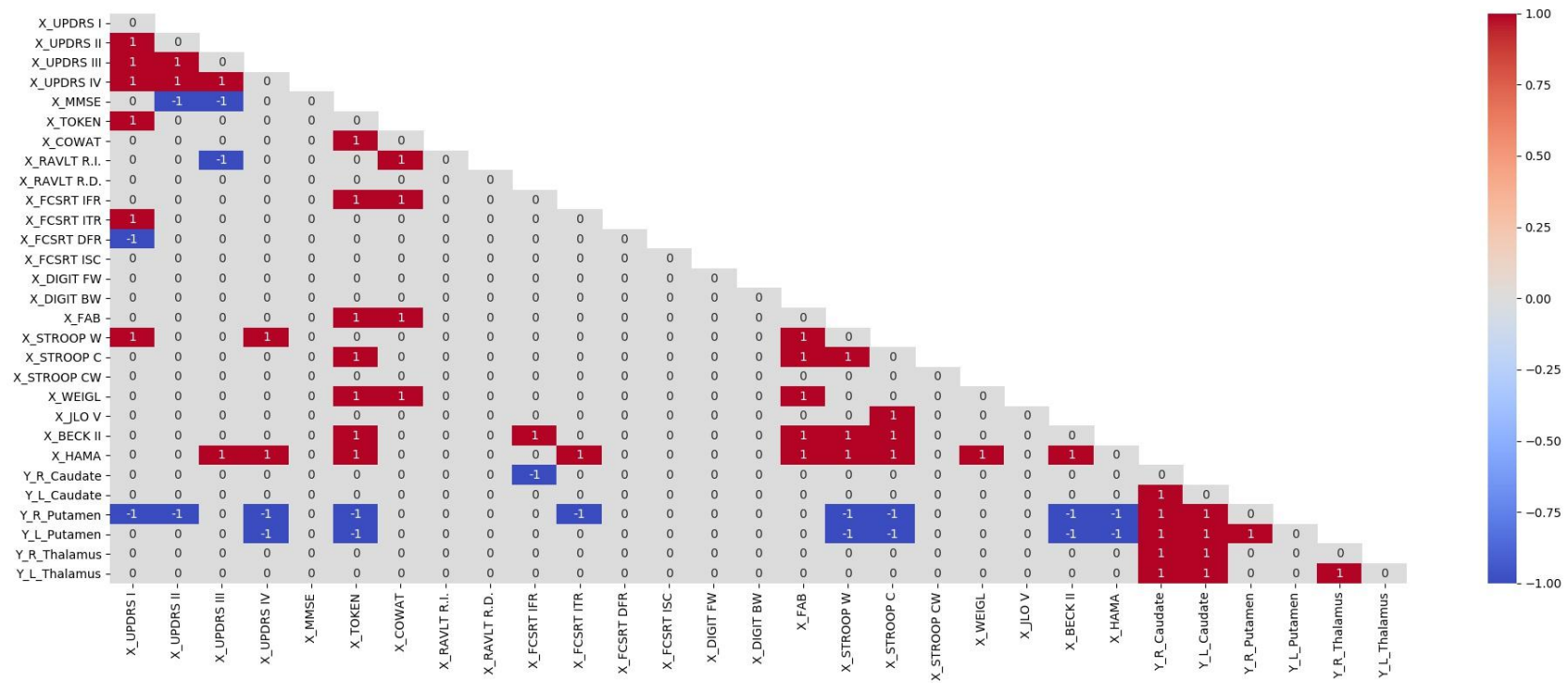
Allegato 4

Morbo di Parkinson

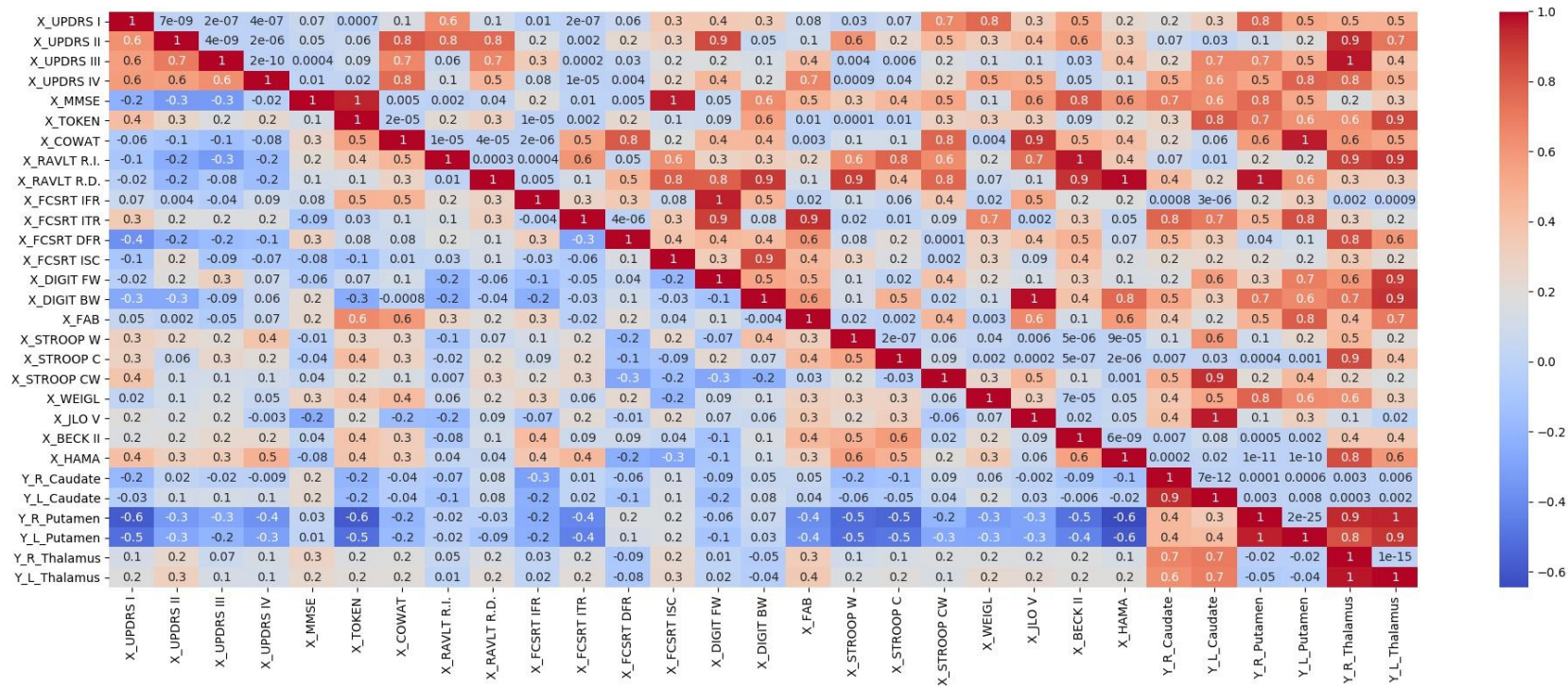
Test Cognitivi vs Dati di Uptake

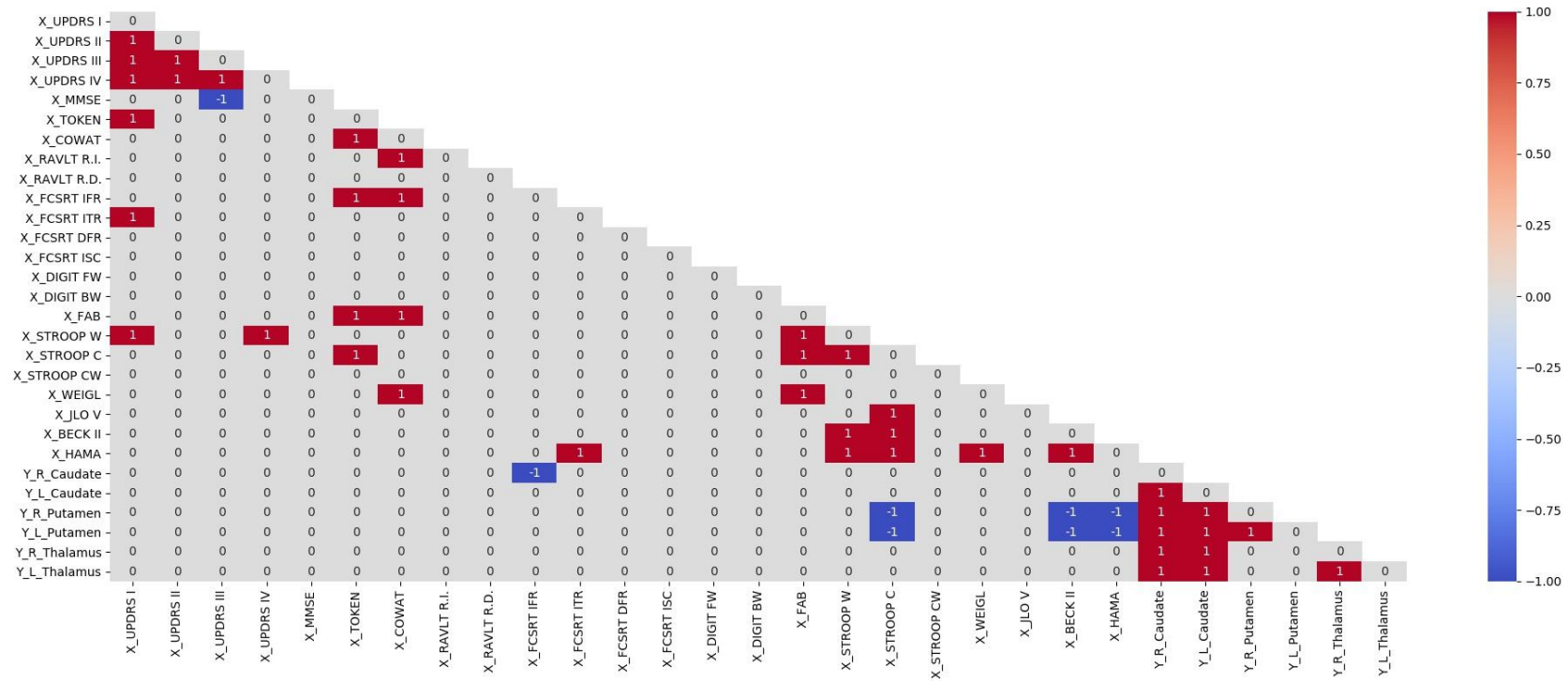
Correlazione di Pearson



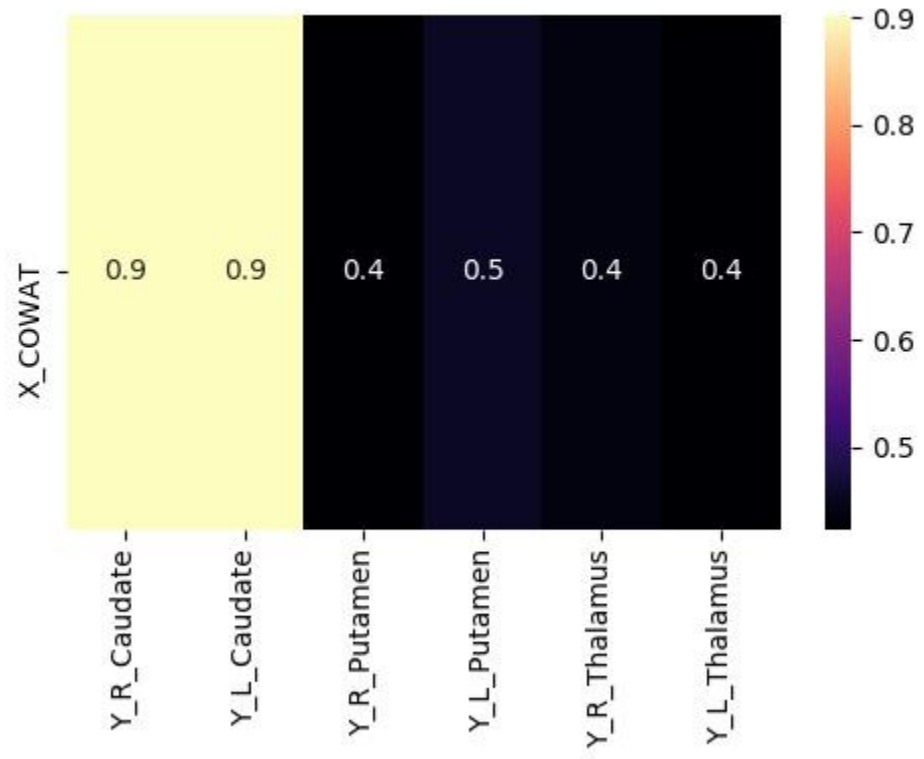


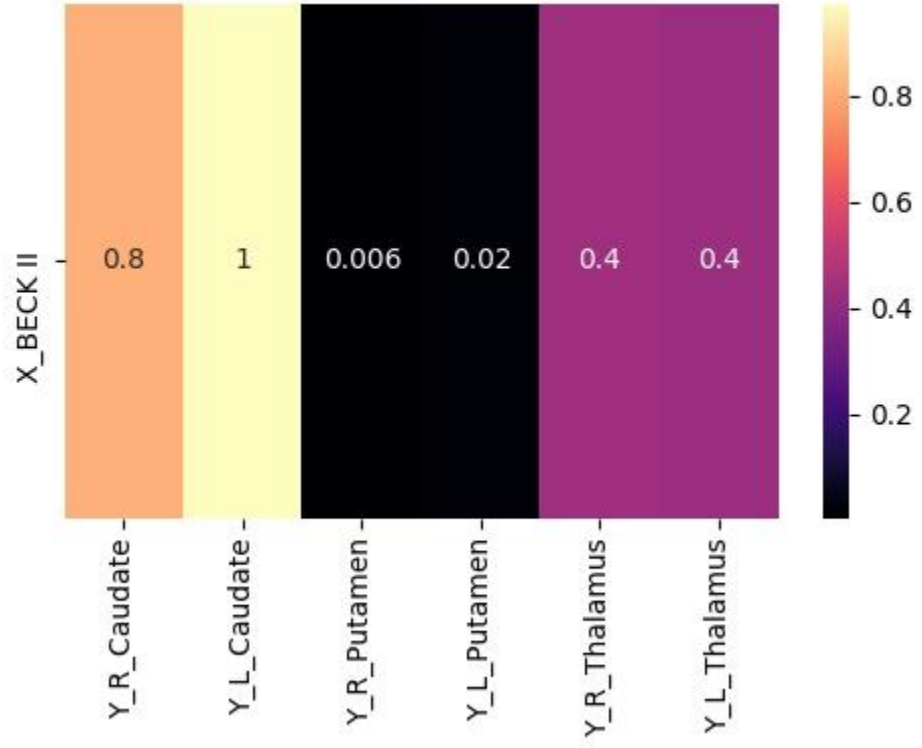
Correlazione di Spearman

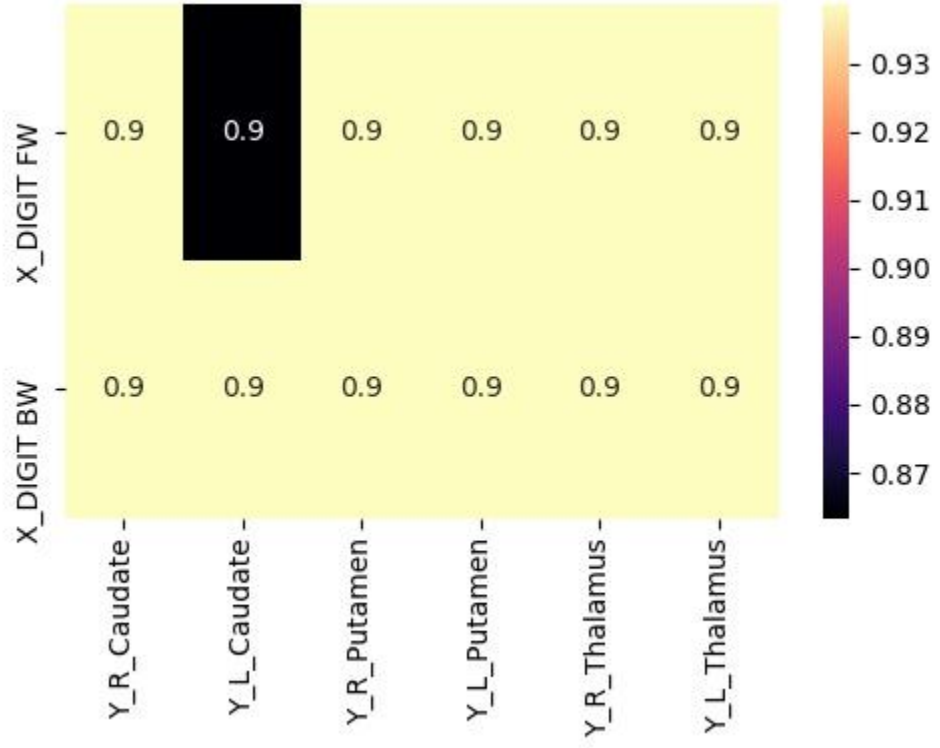


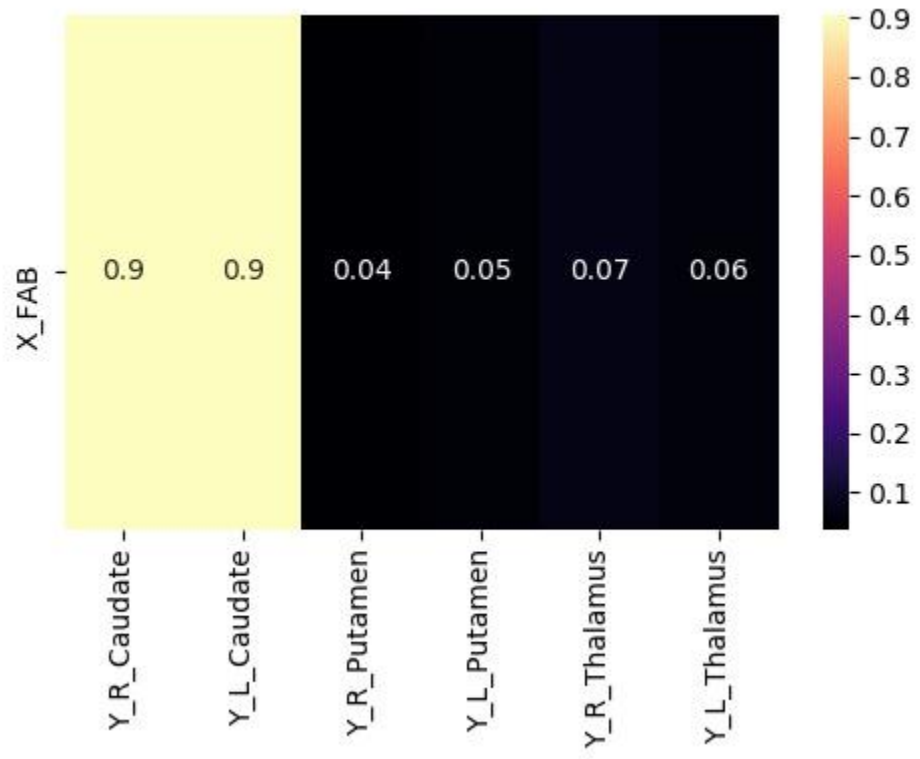


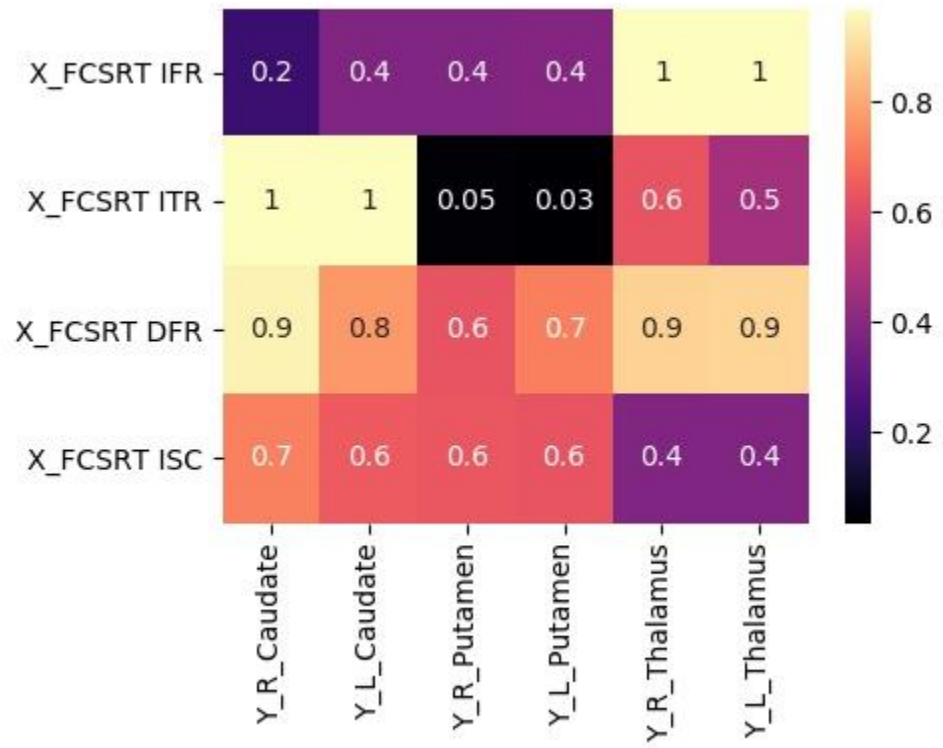
Correzione con Benjamini-Hochberg

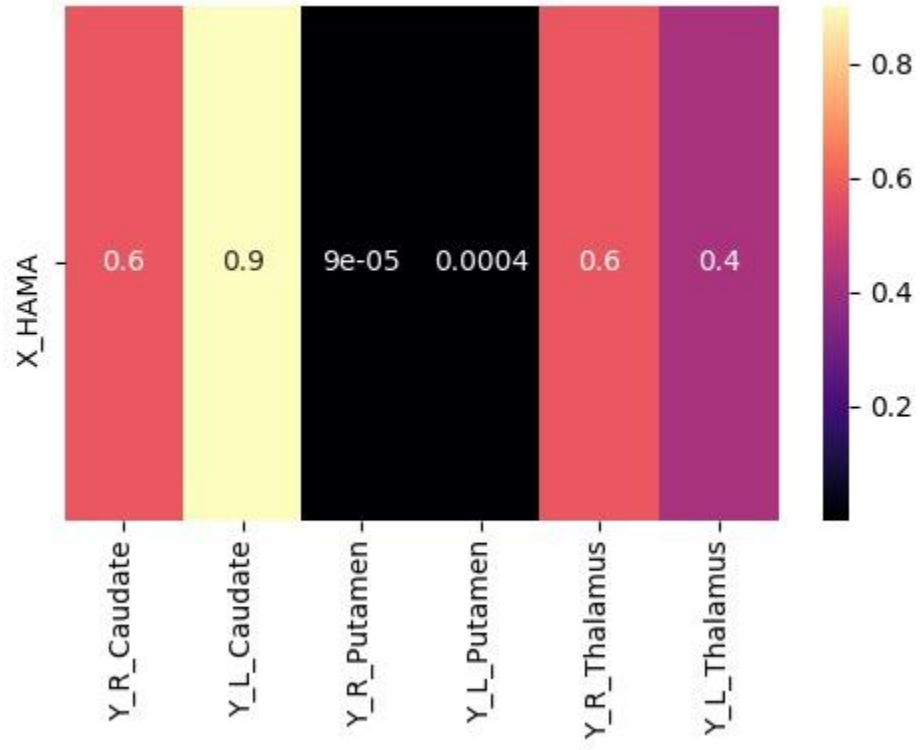


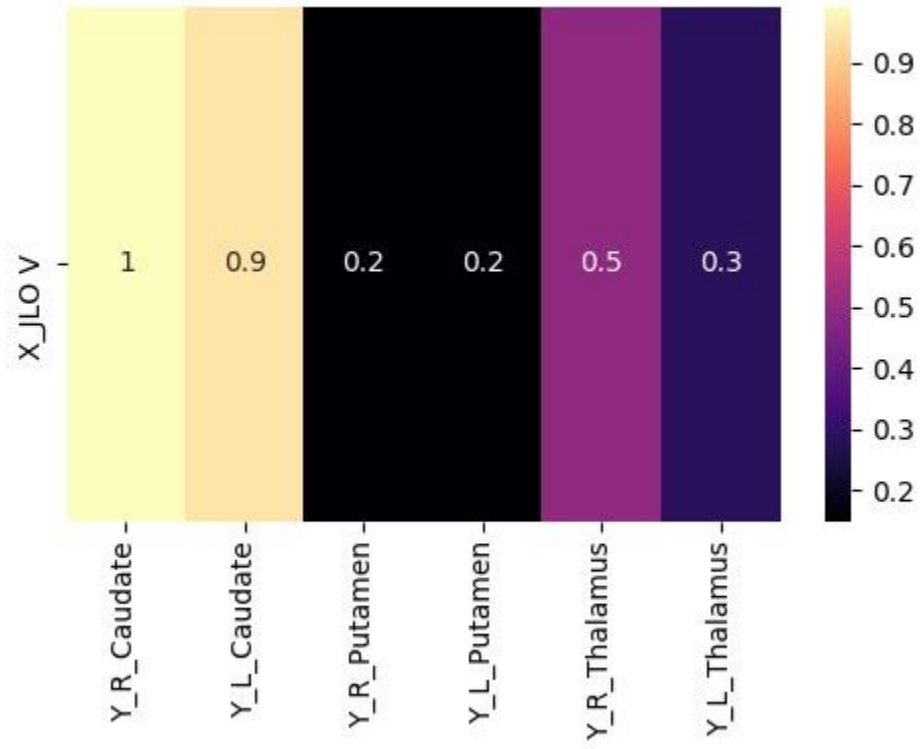


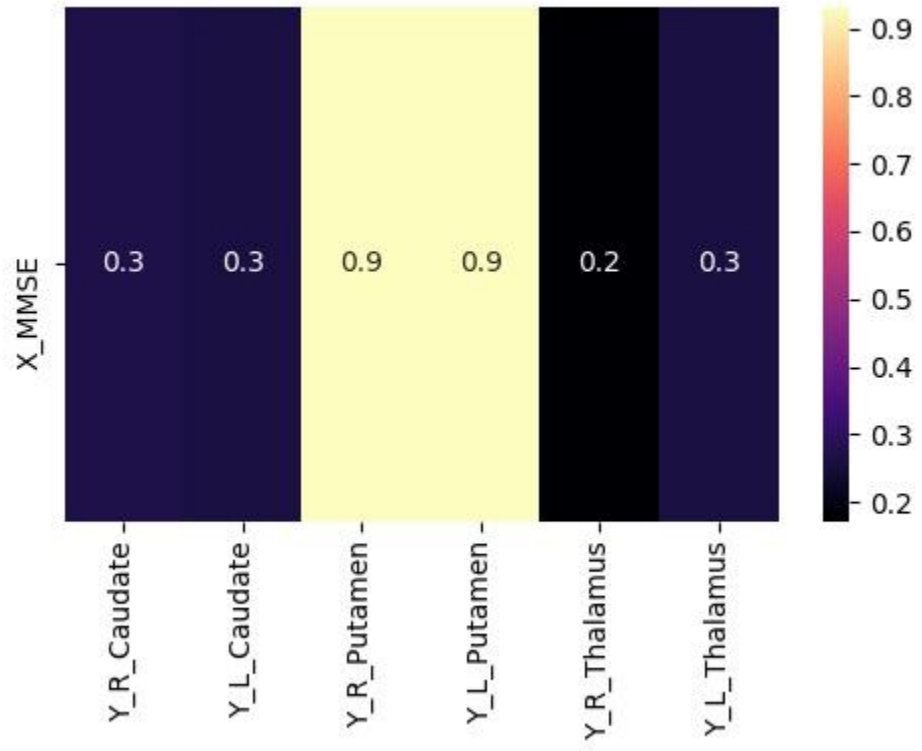


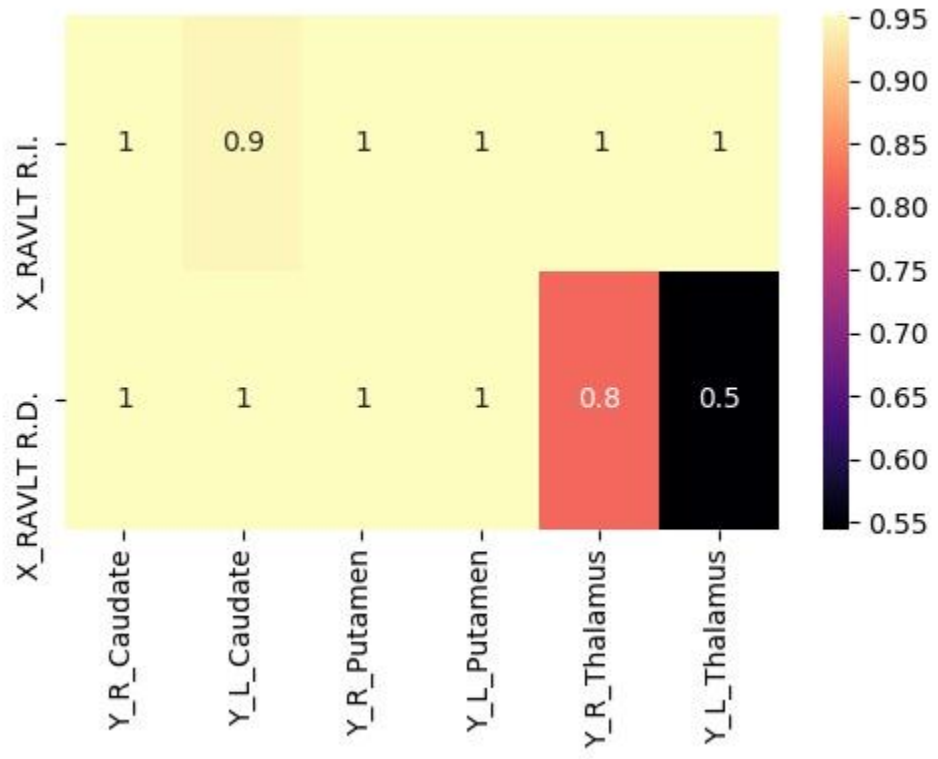


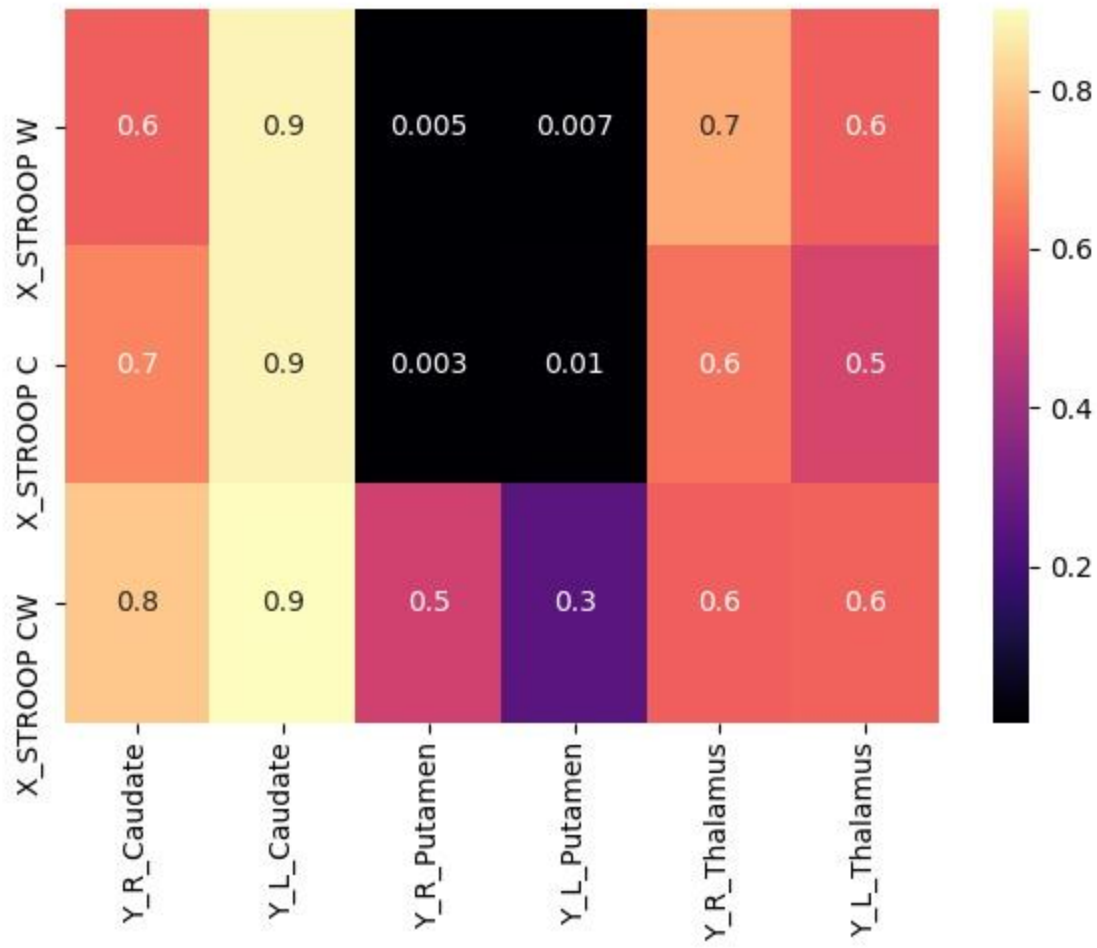


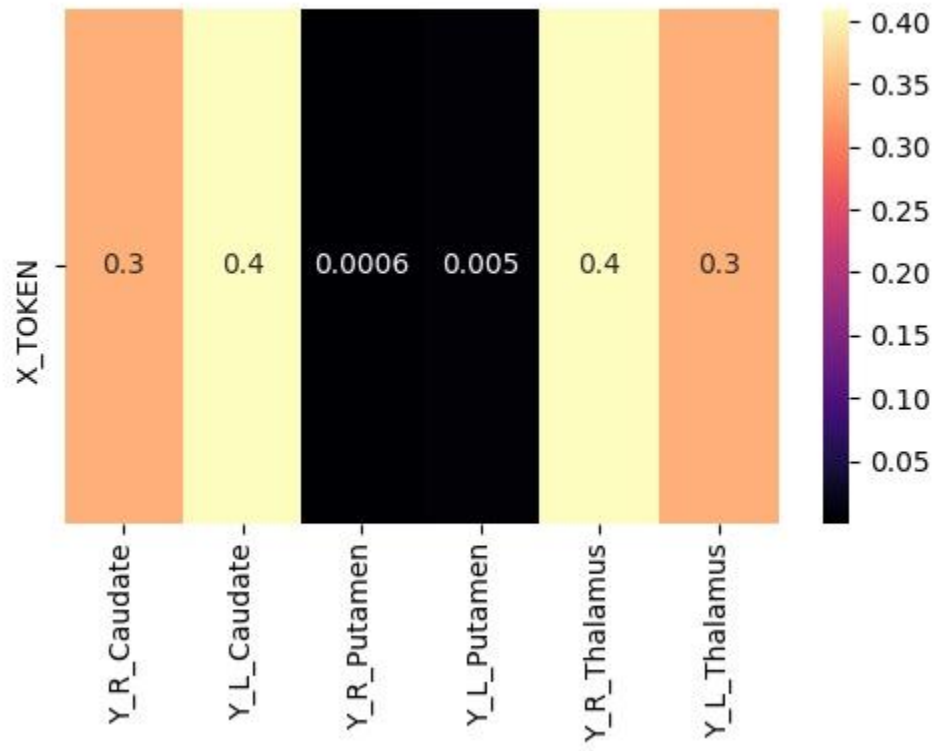


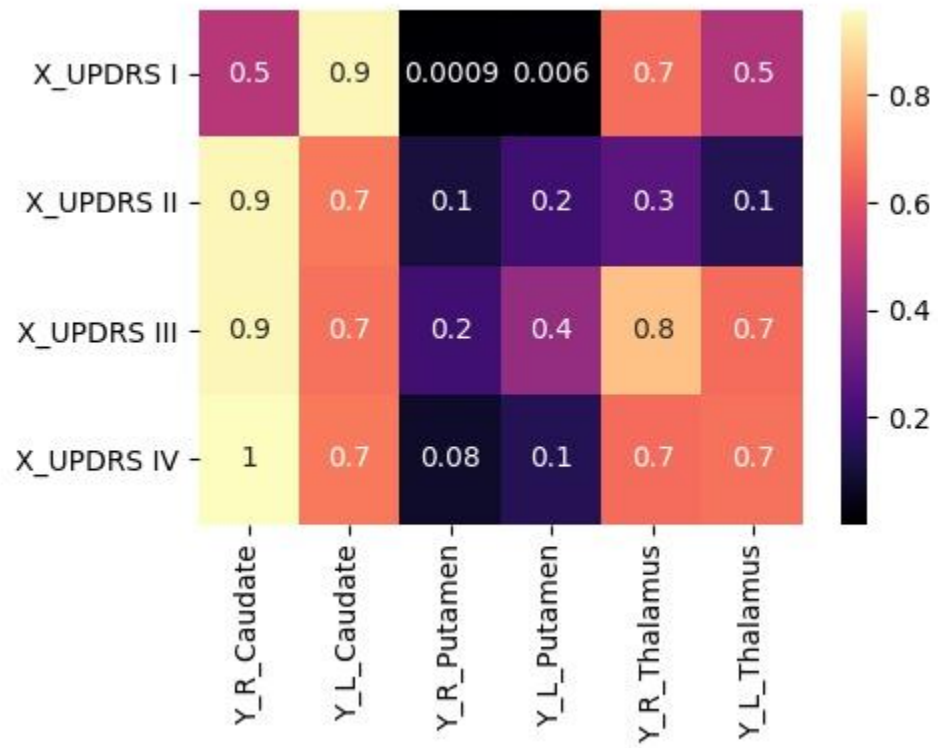


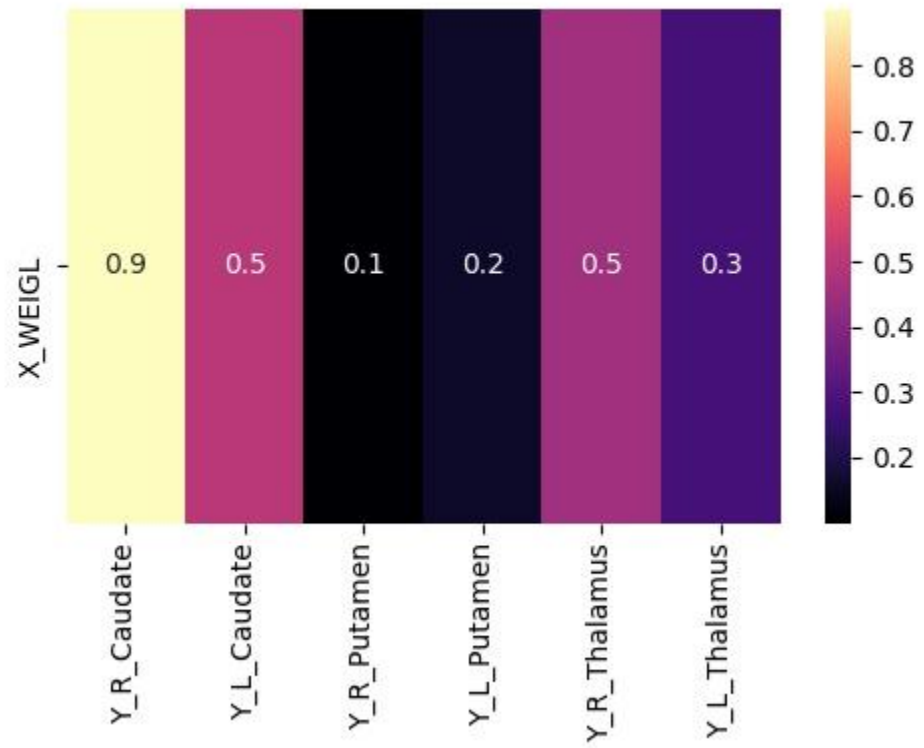




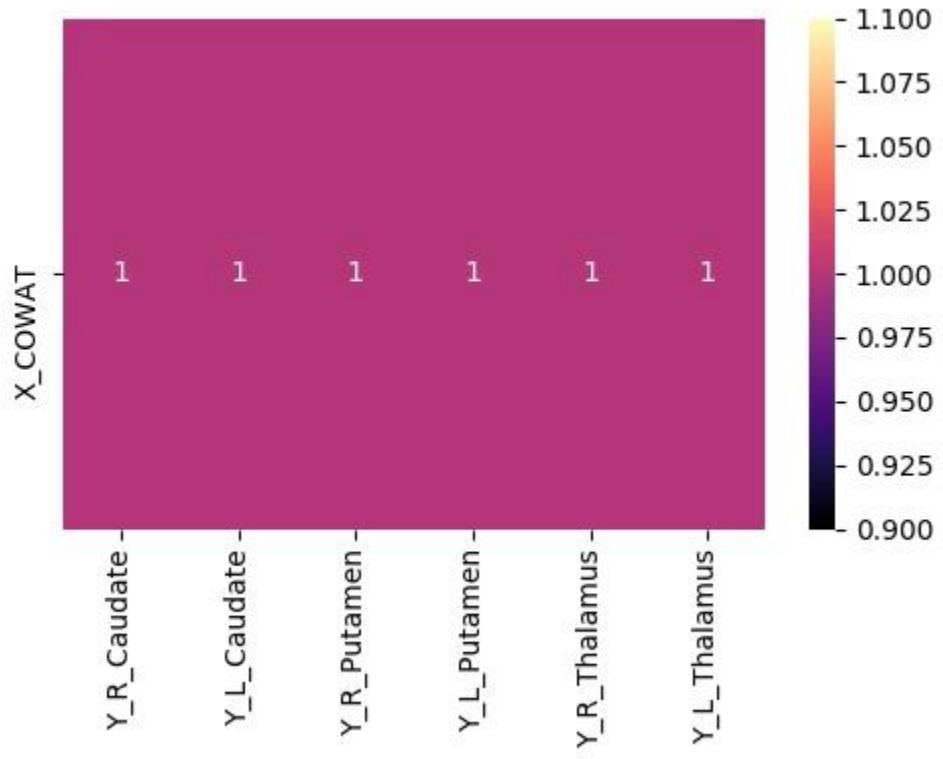


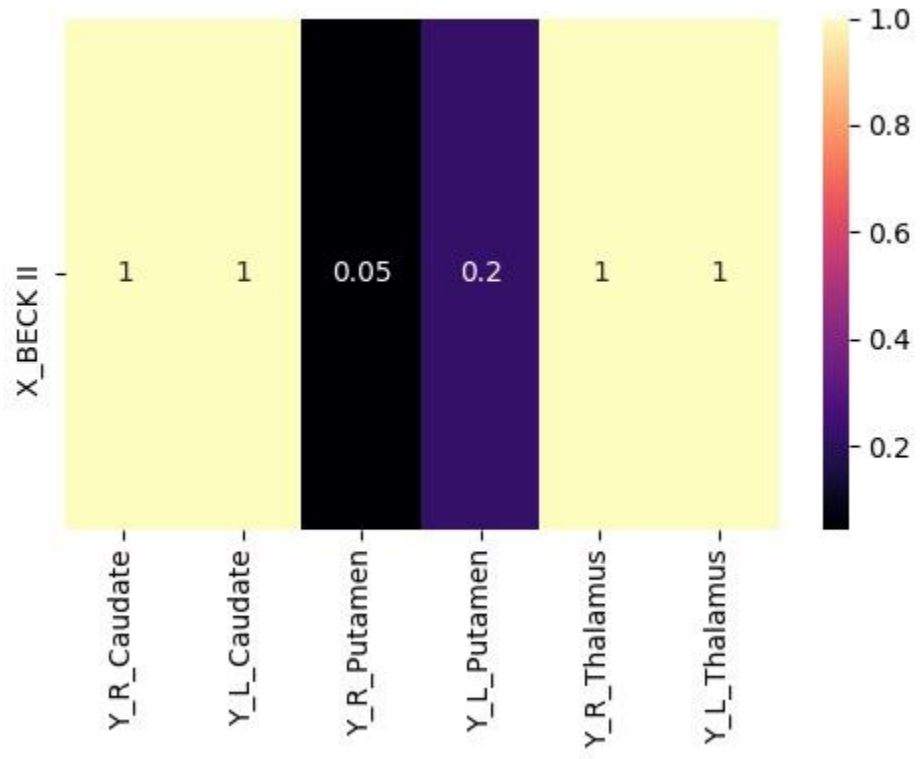


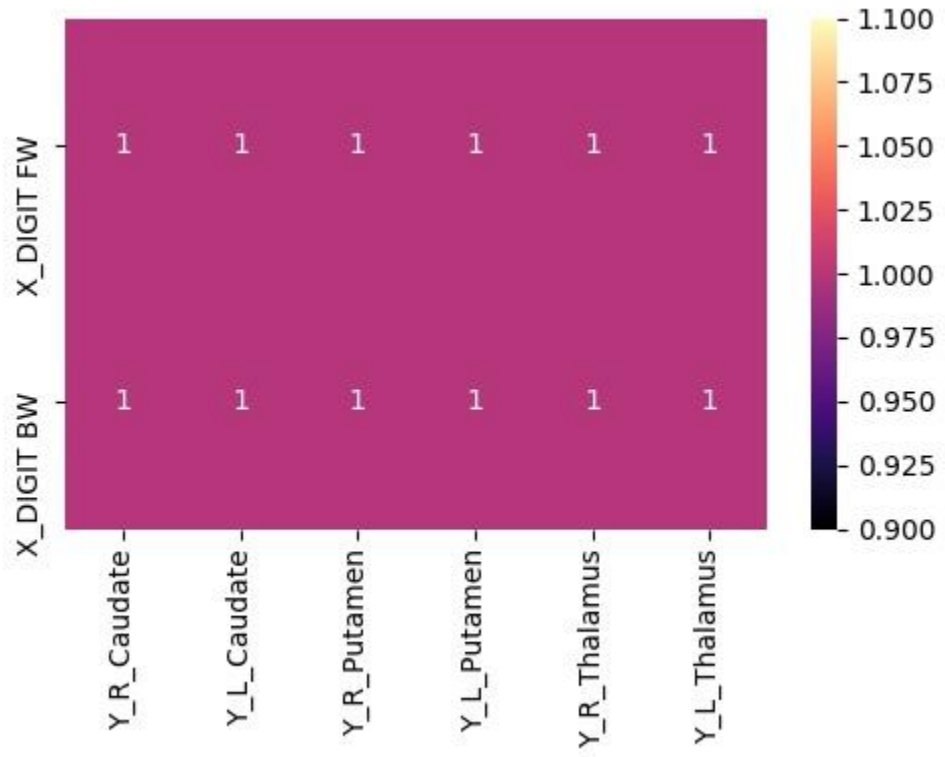


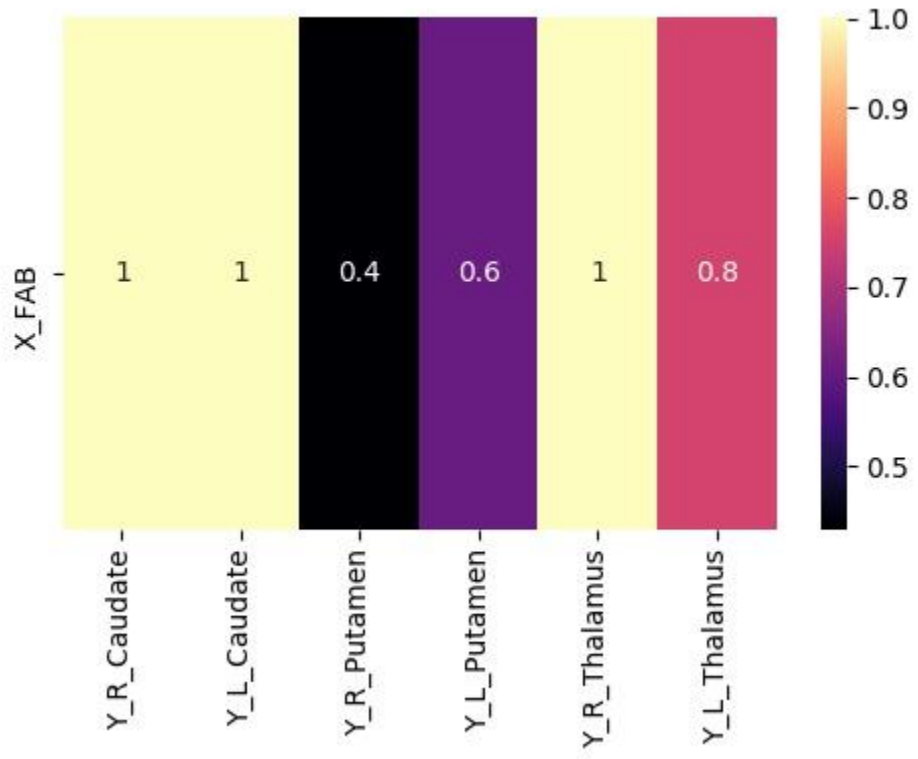


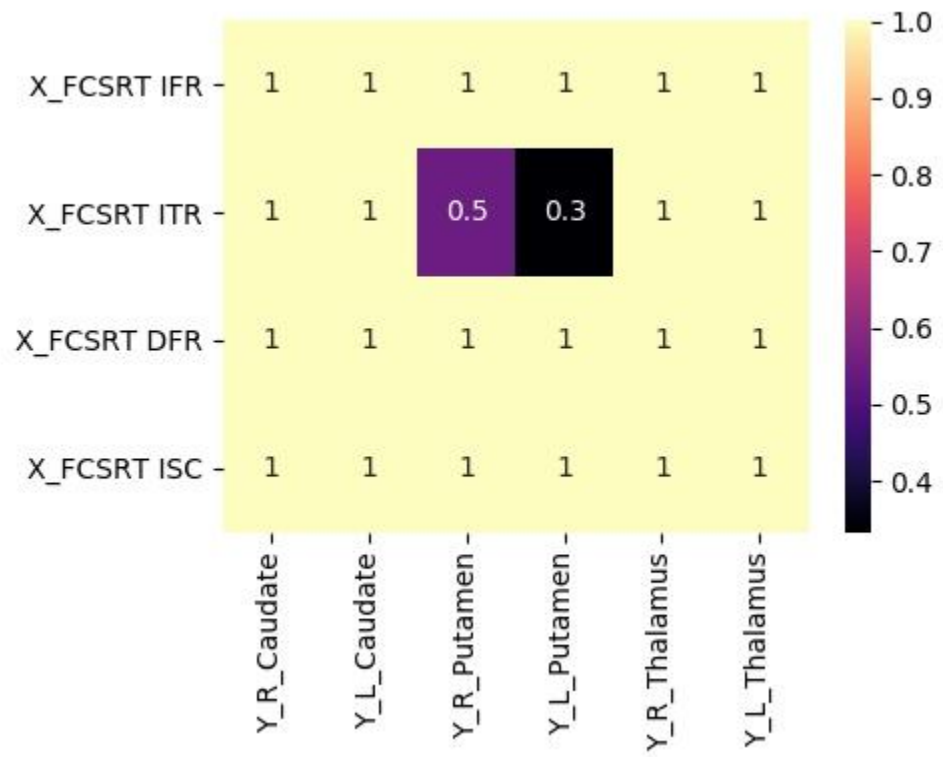
Correzione con Bonferroni

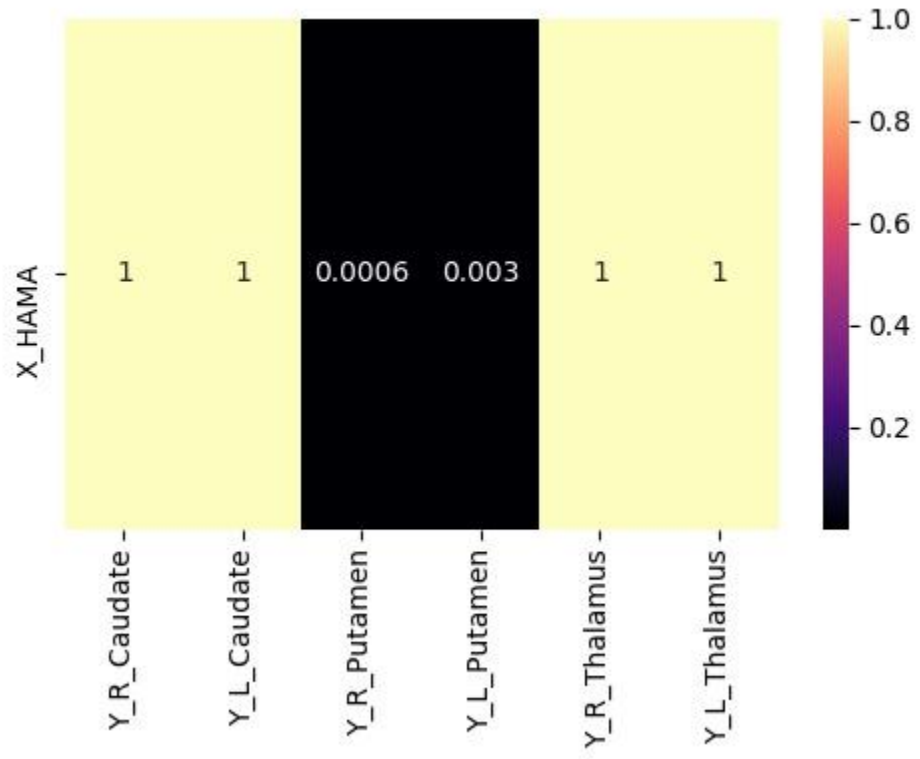


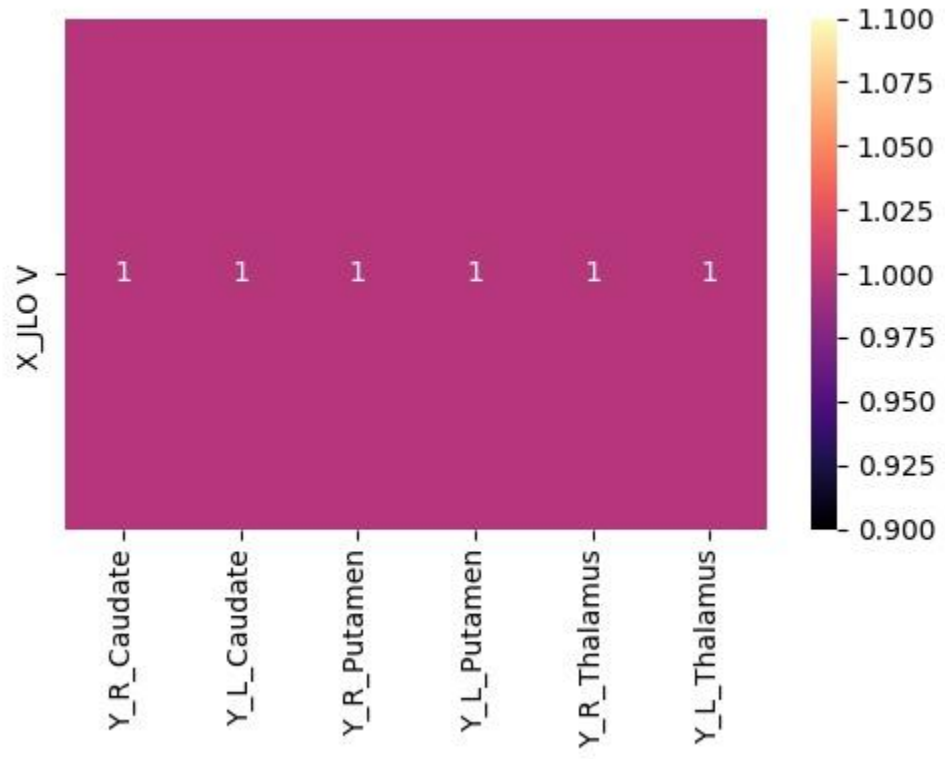


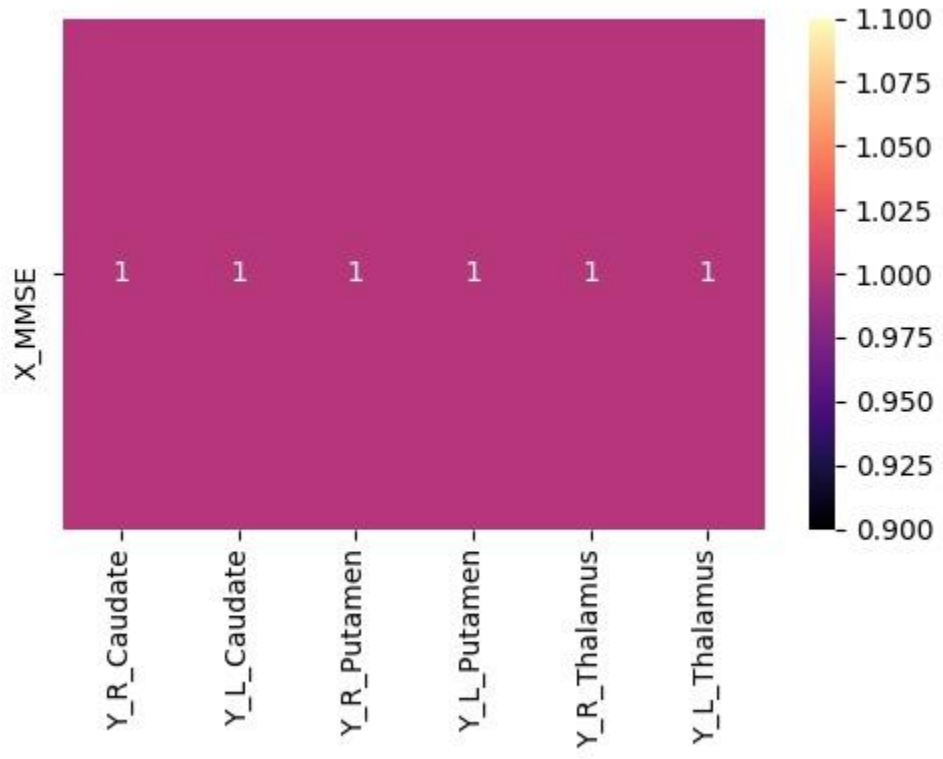


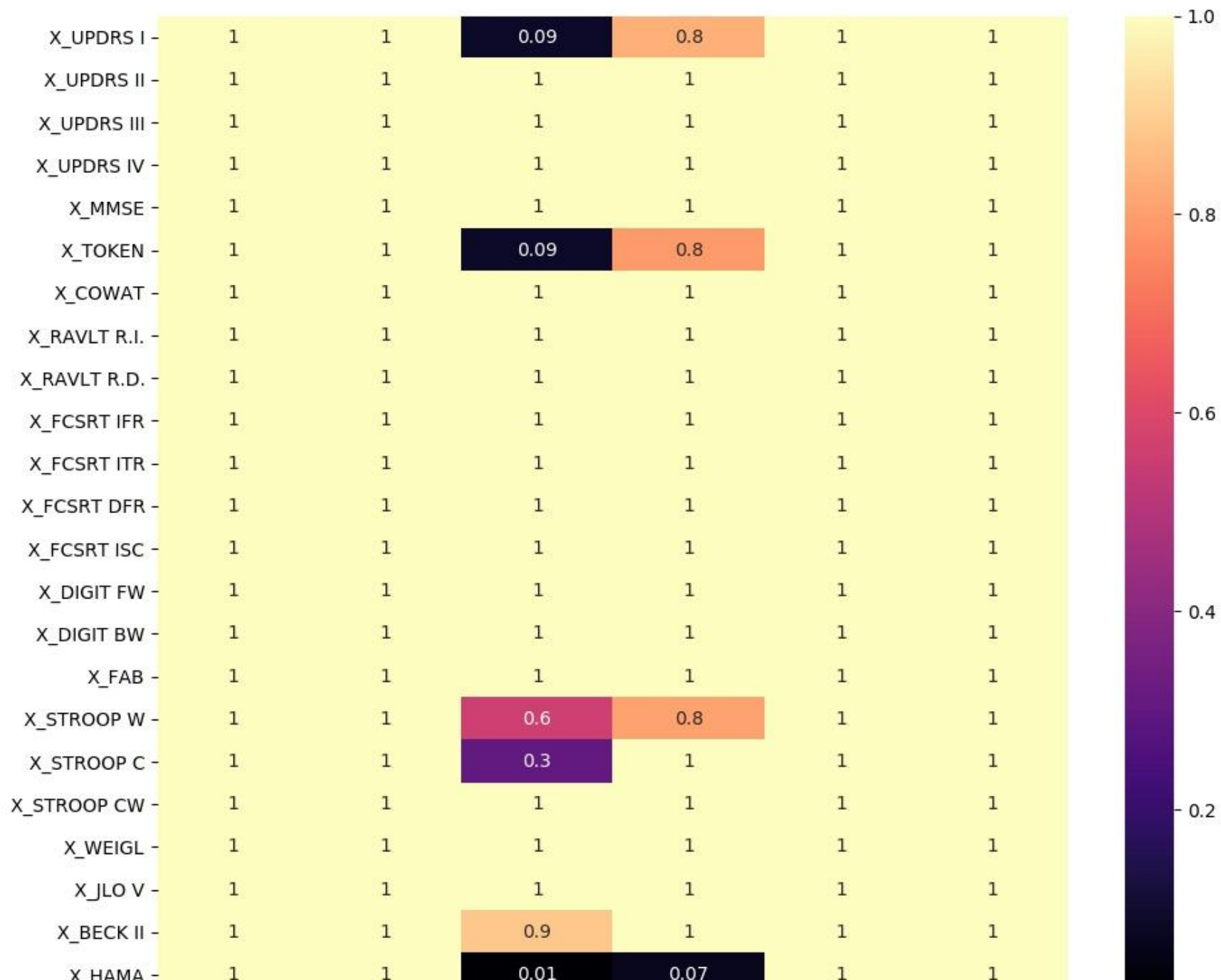


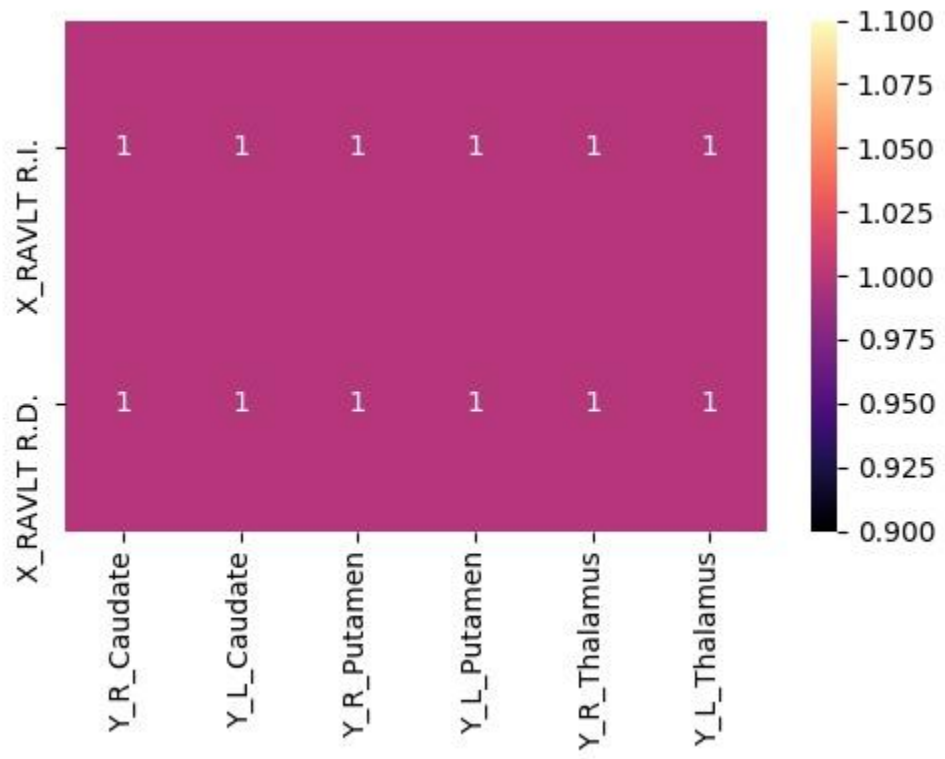


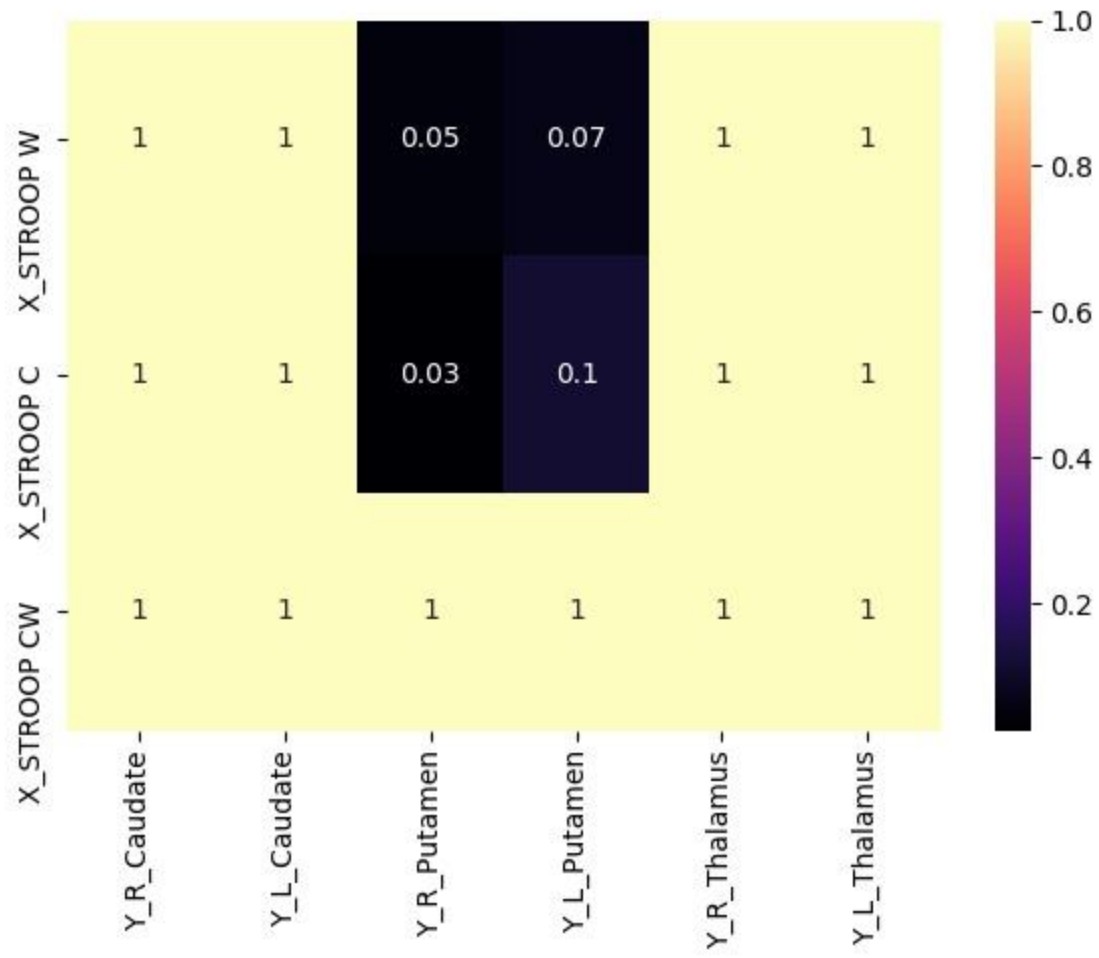


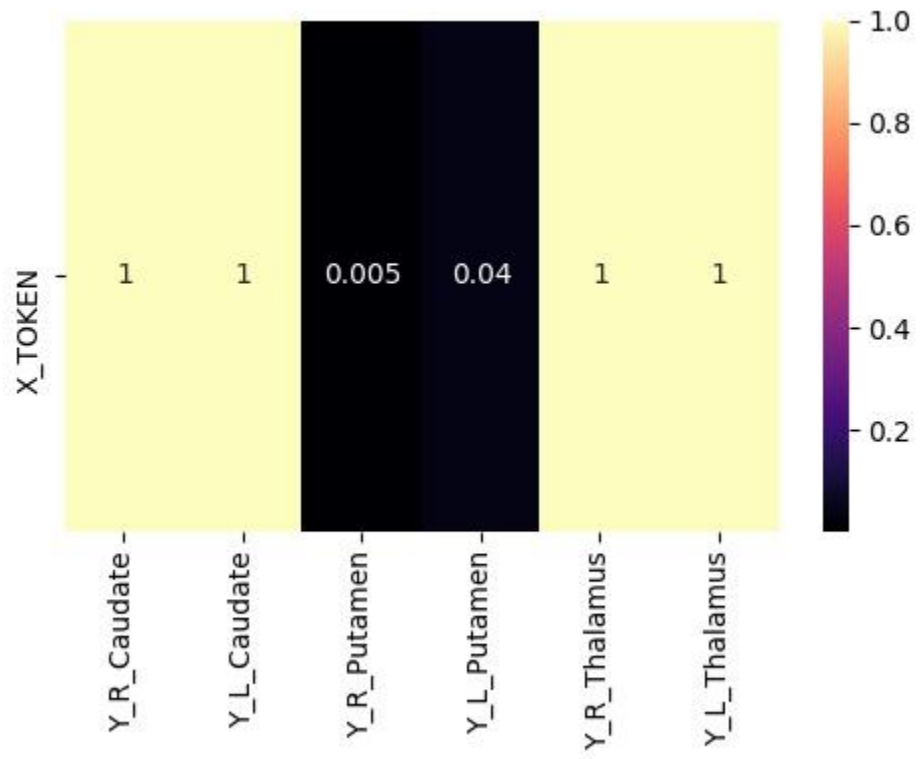


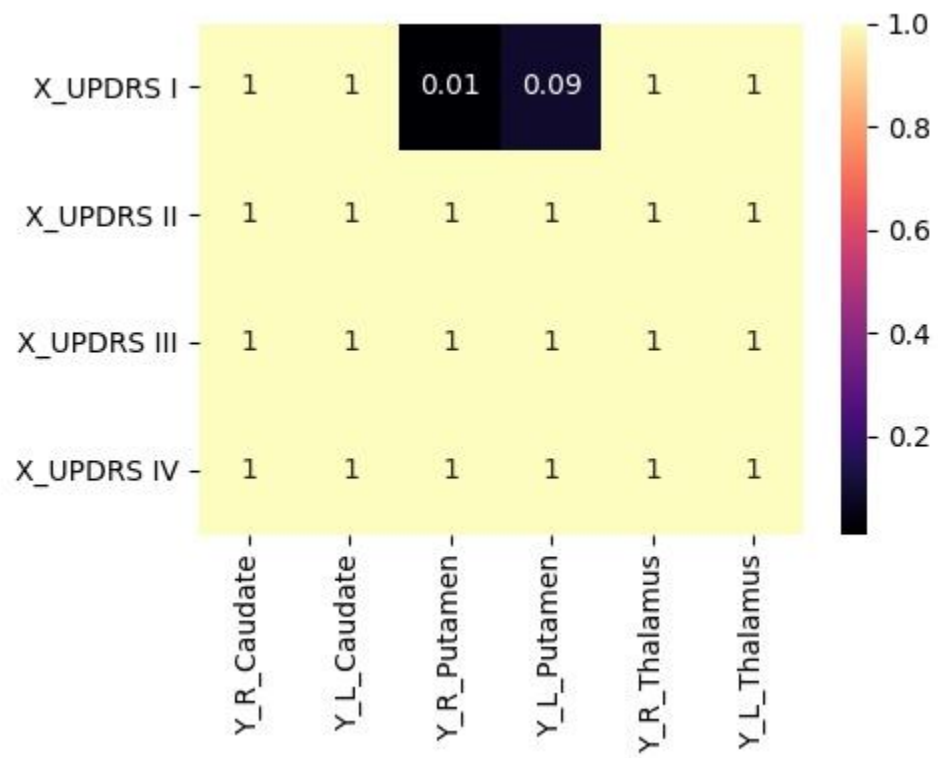


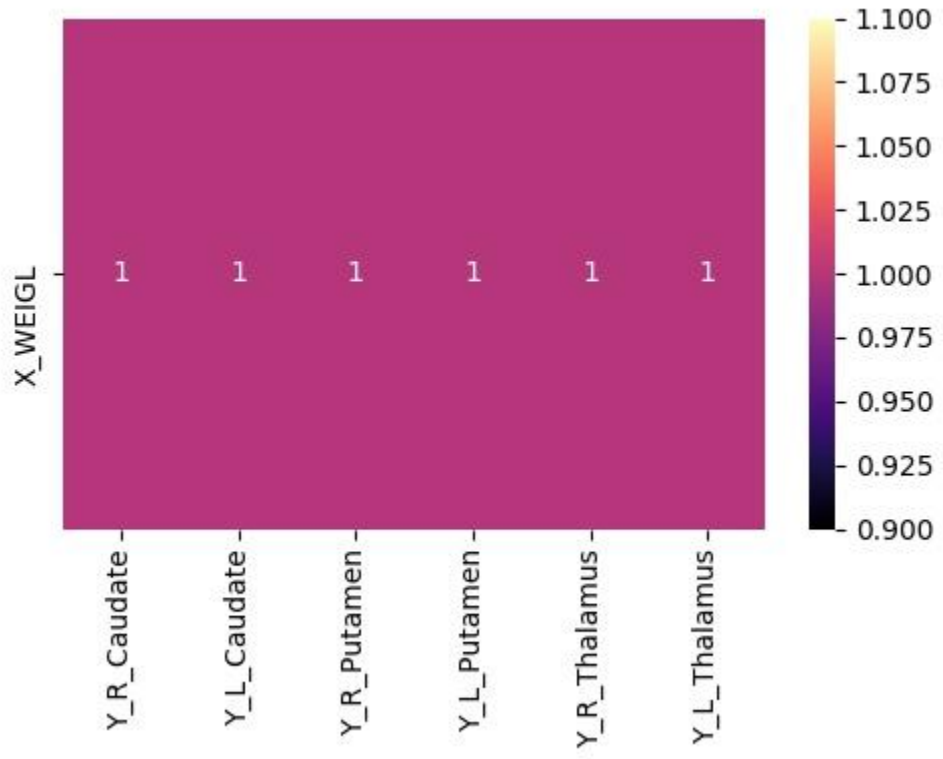










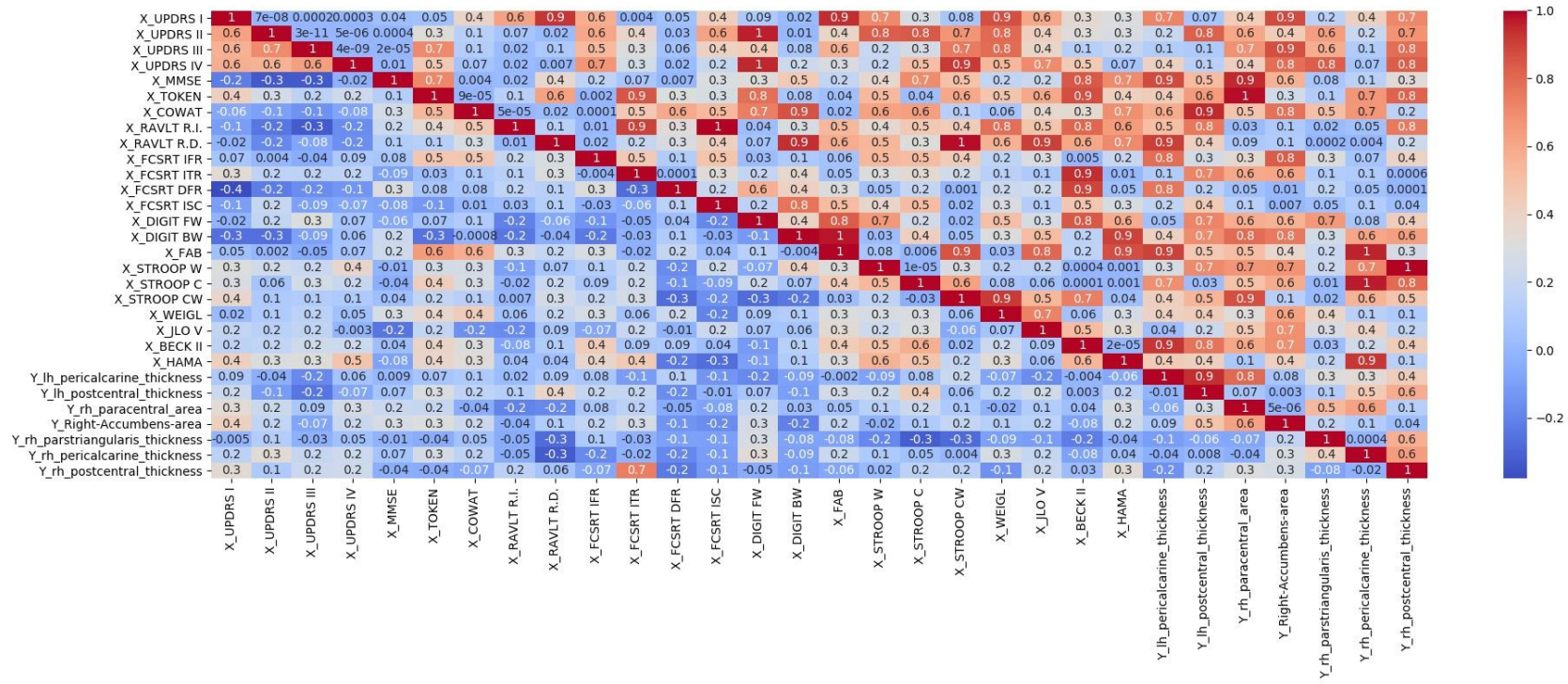


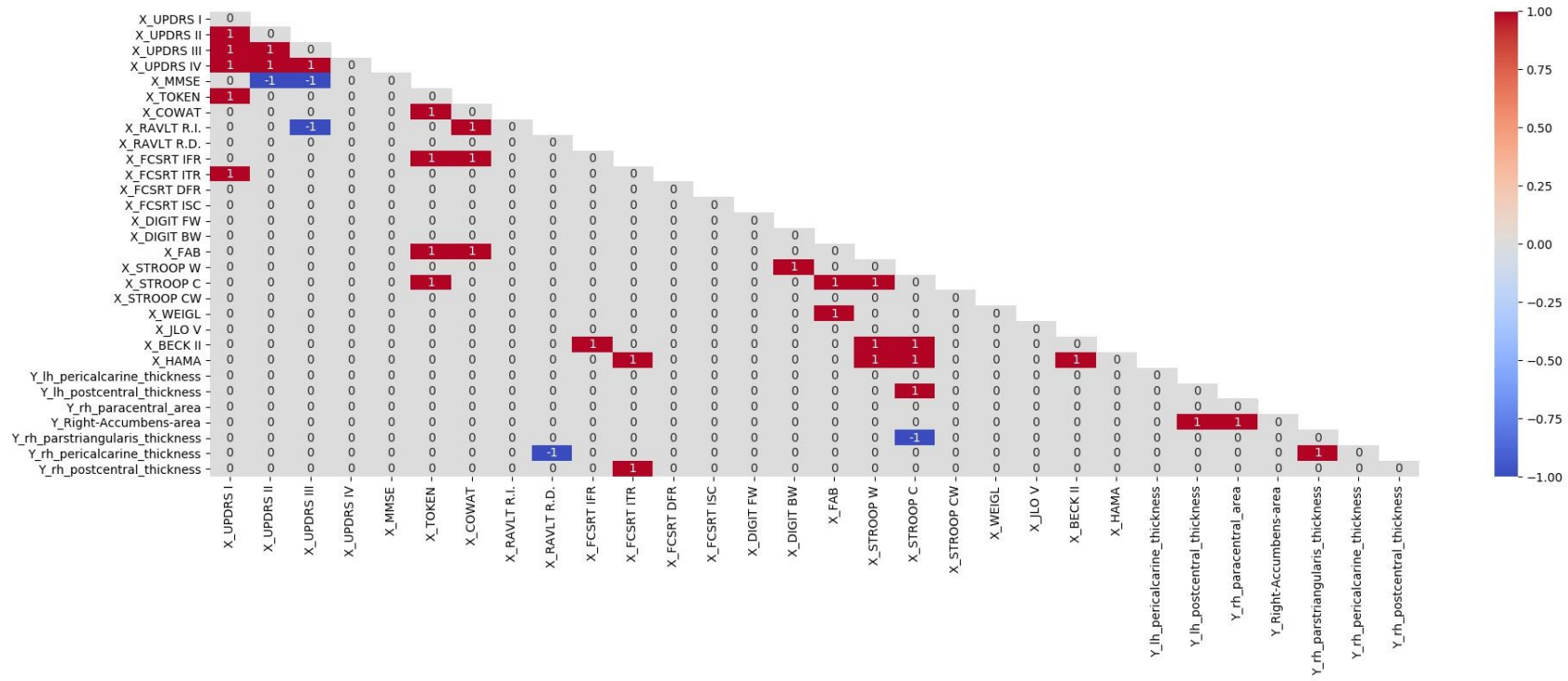
Allegato 5

Morbo di Parkinson

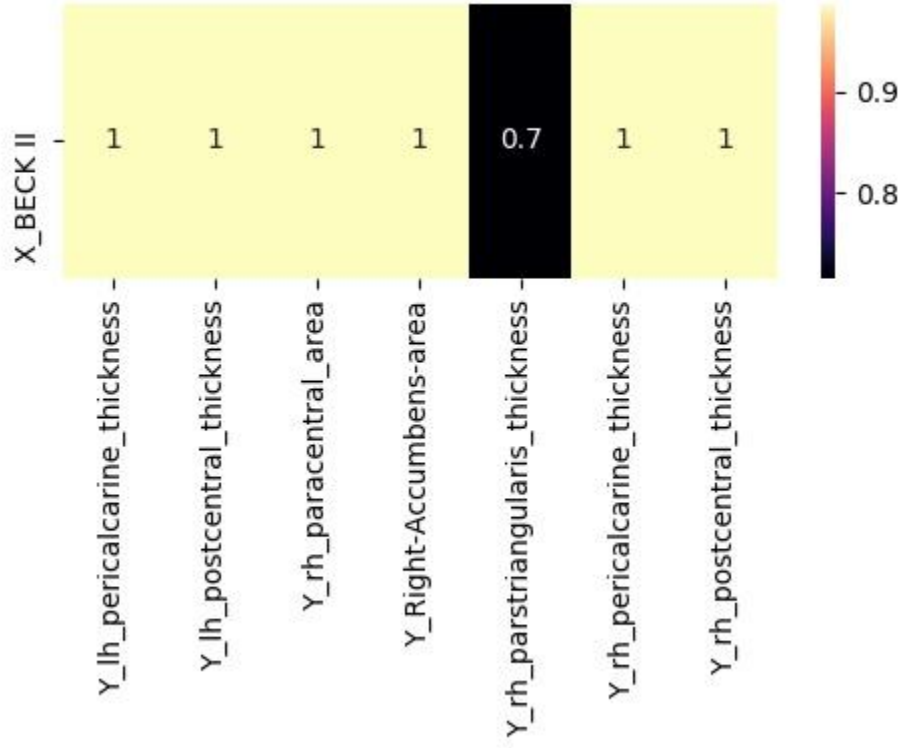
Test Cognitivi vs Freesurfer Features

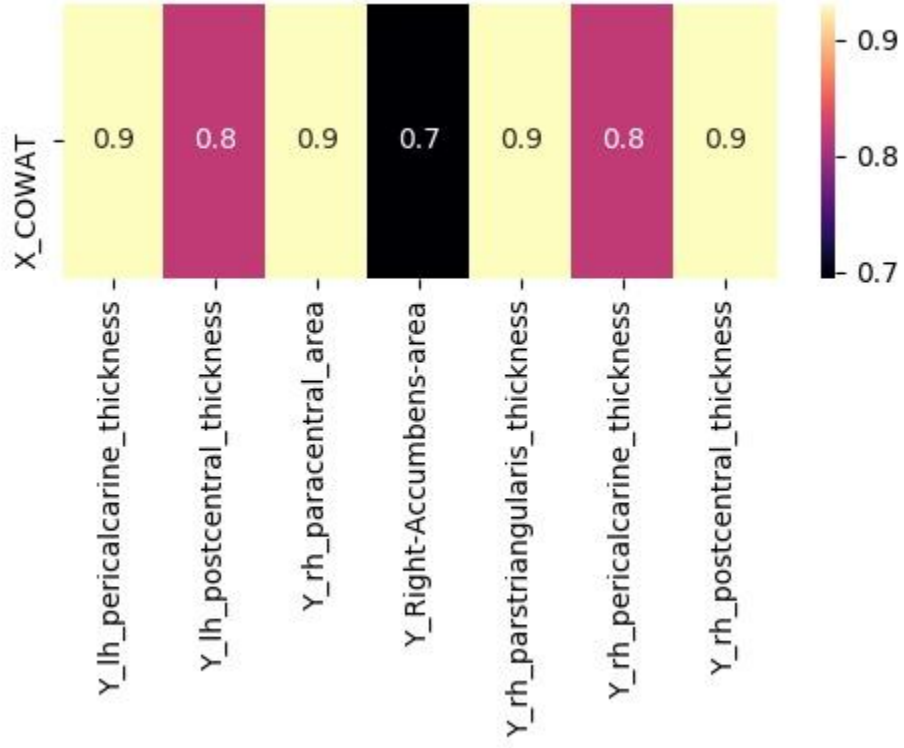
Correlazione di Pearson

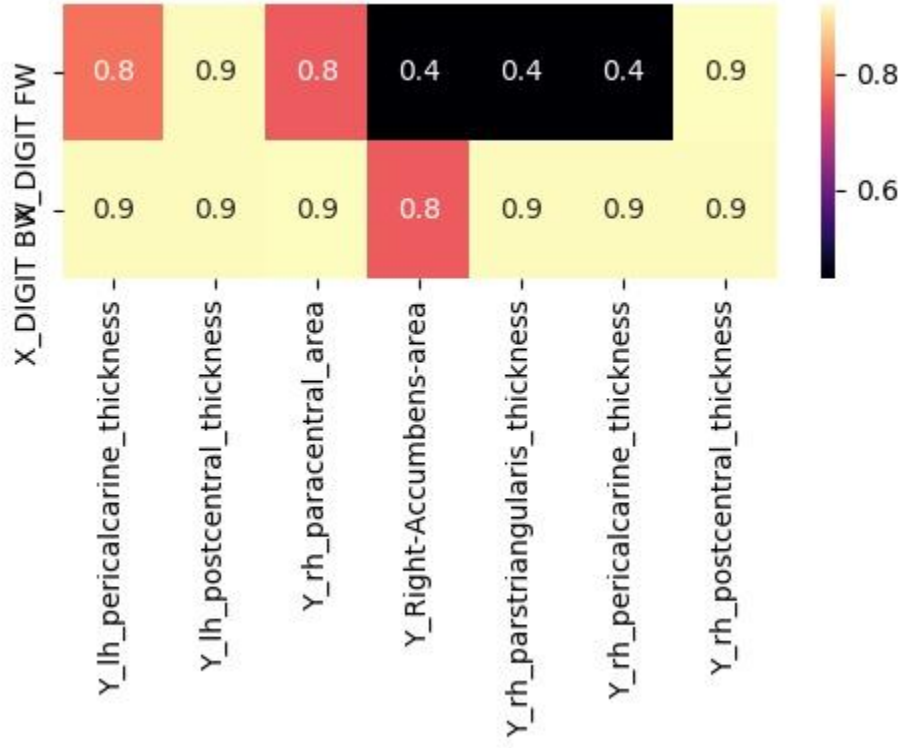


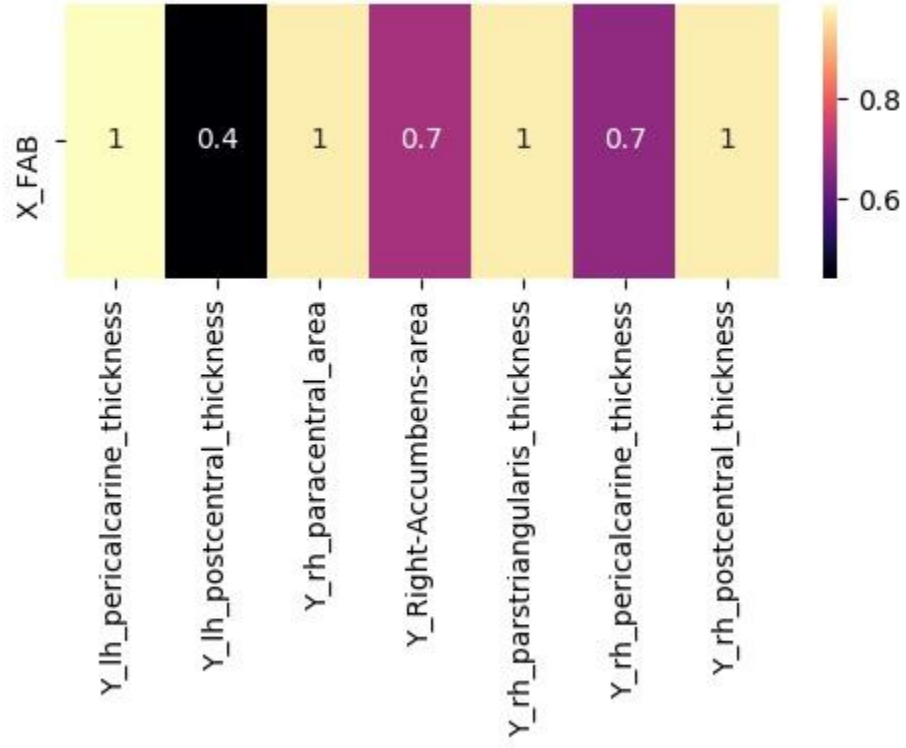


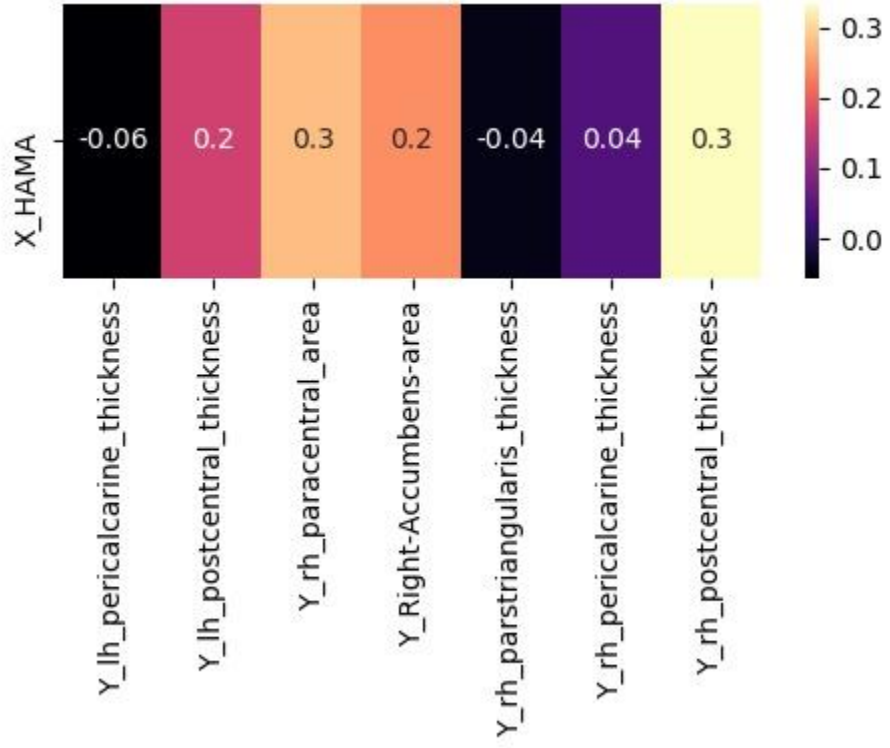
Correzione con Benjamini-Hochberg

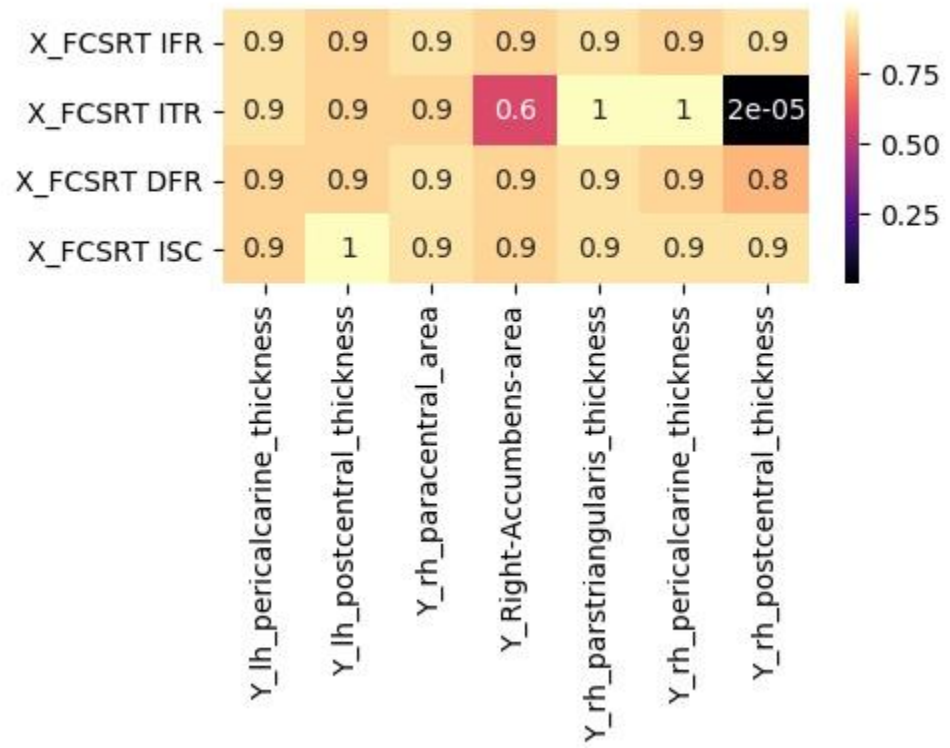


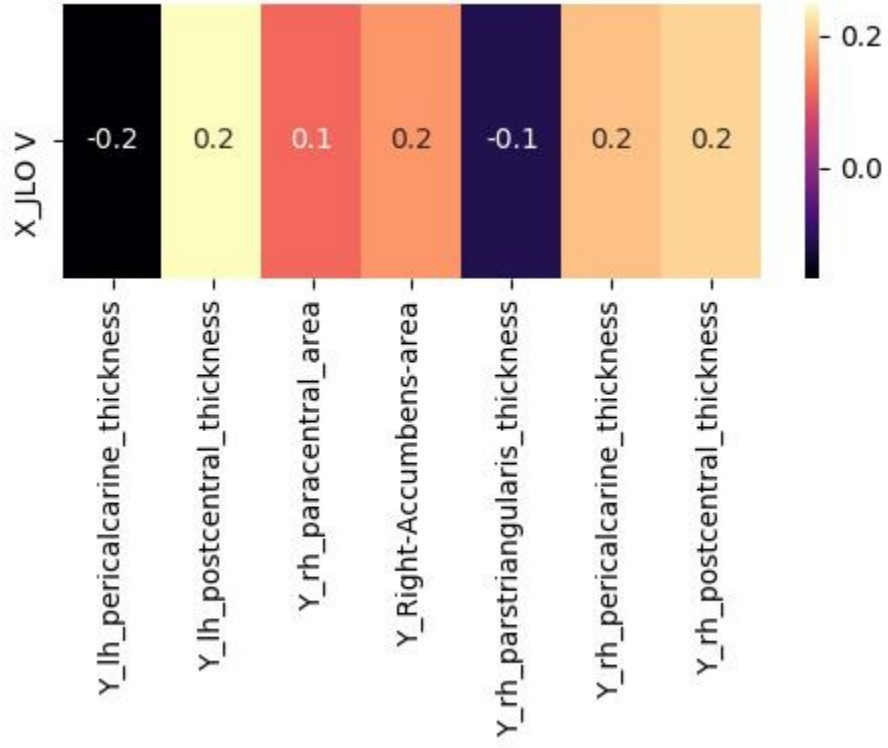


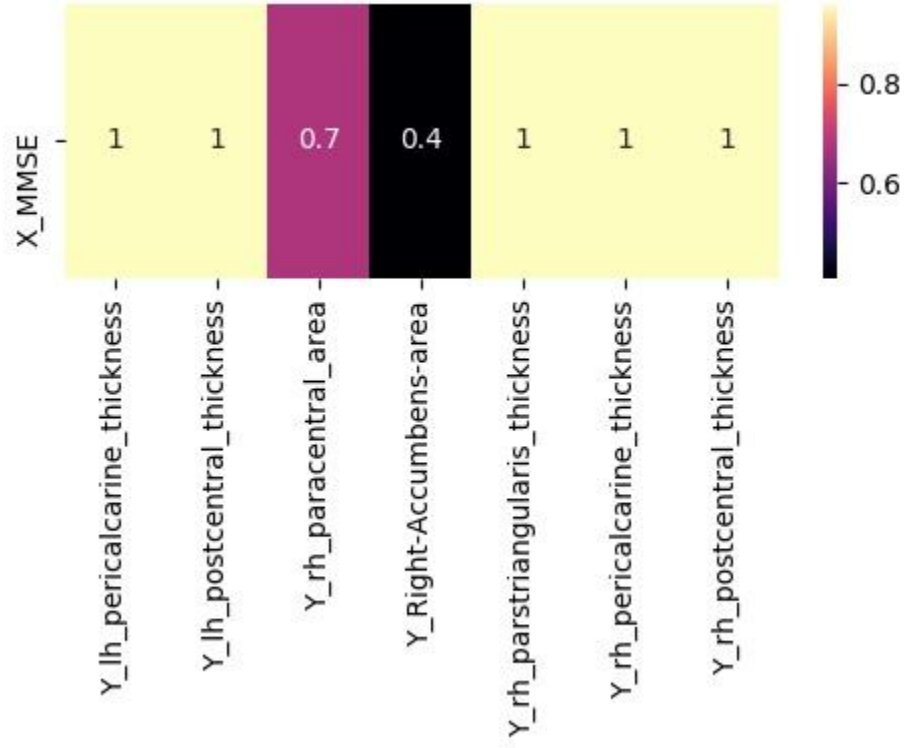


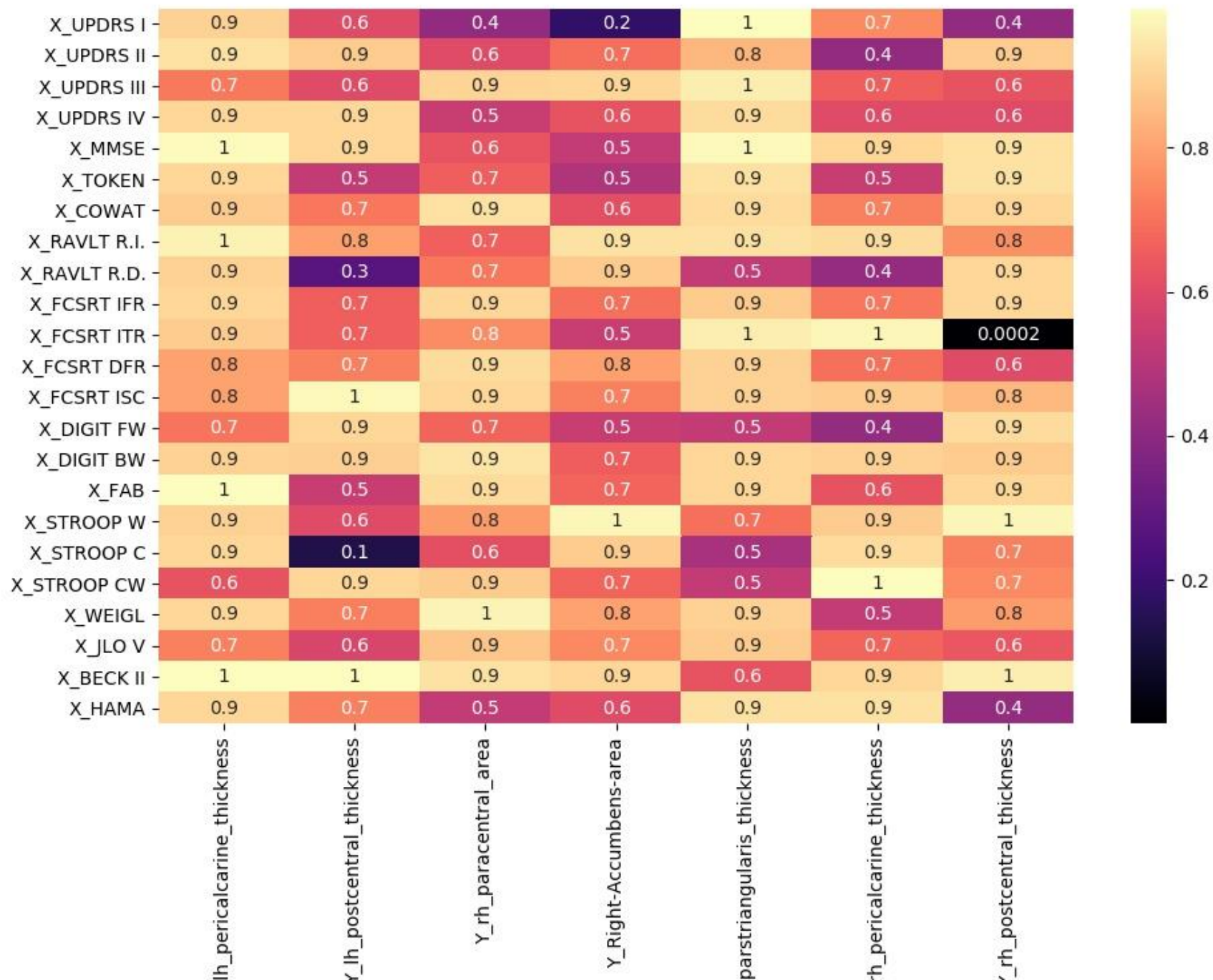


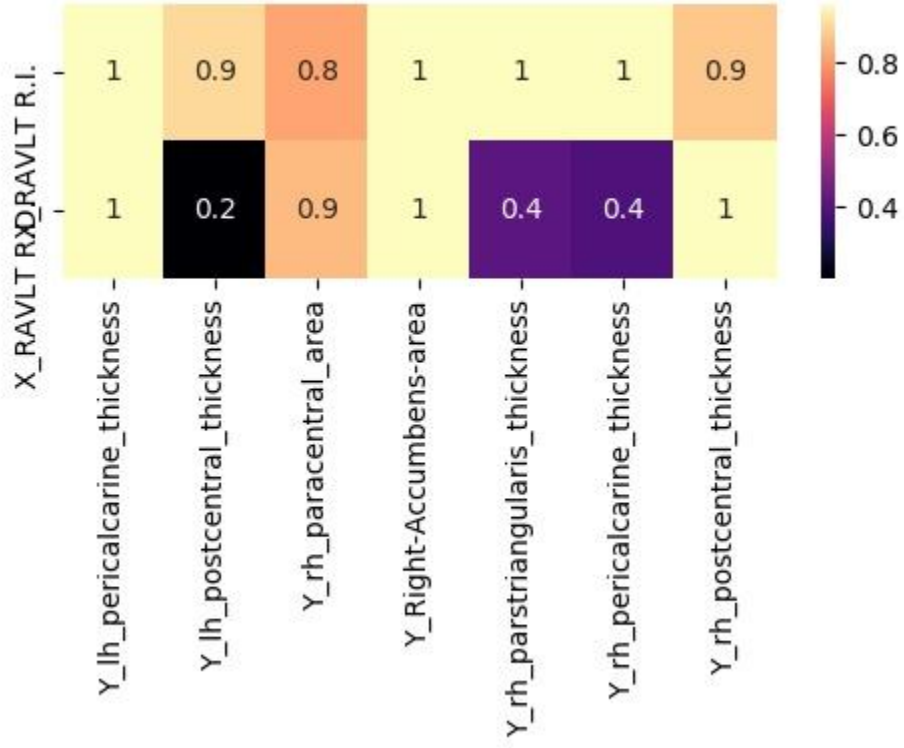


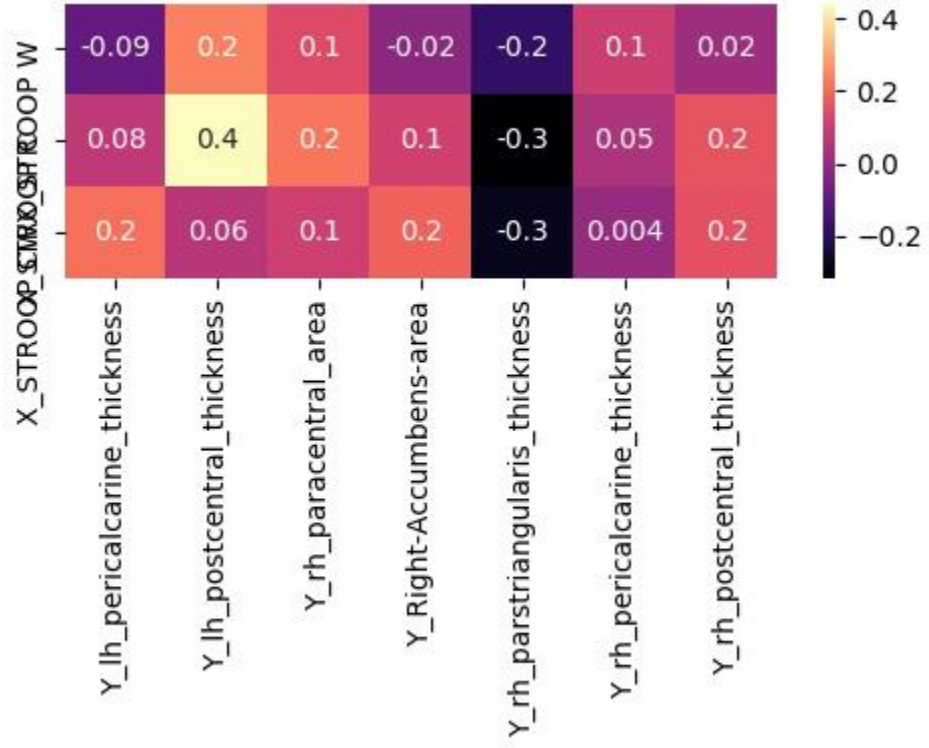


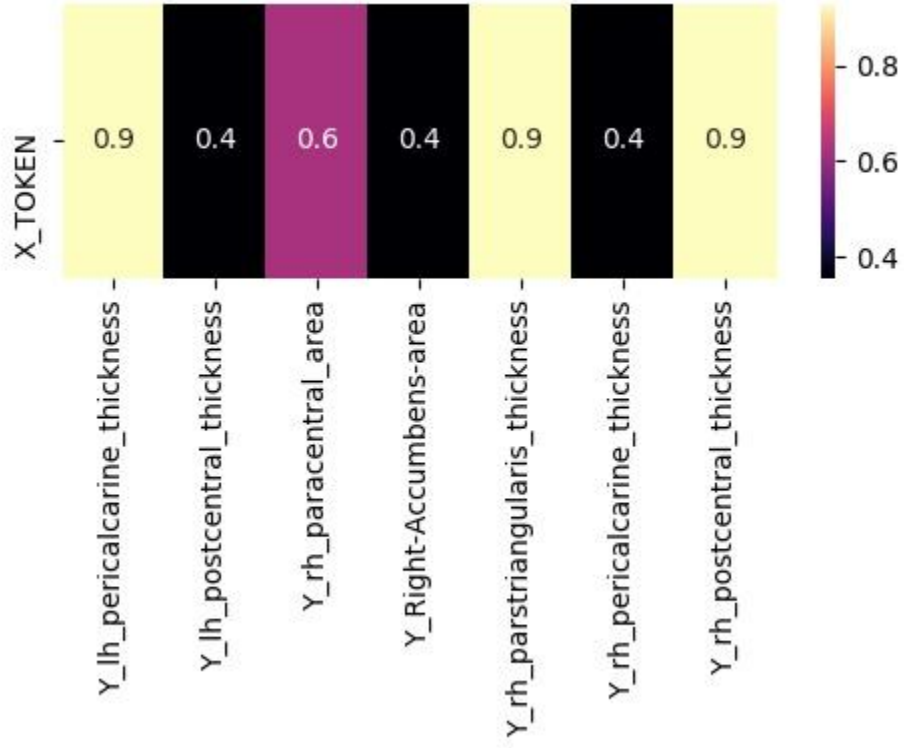


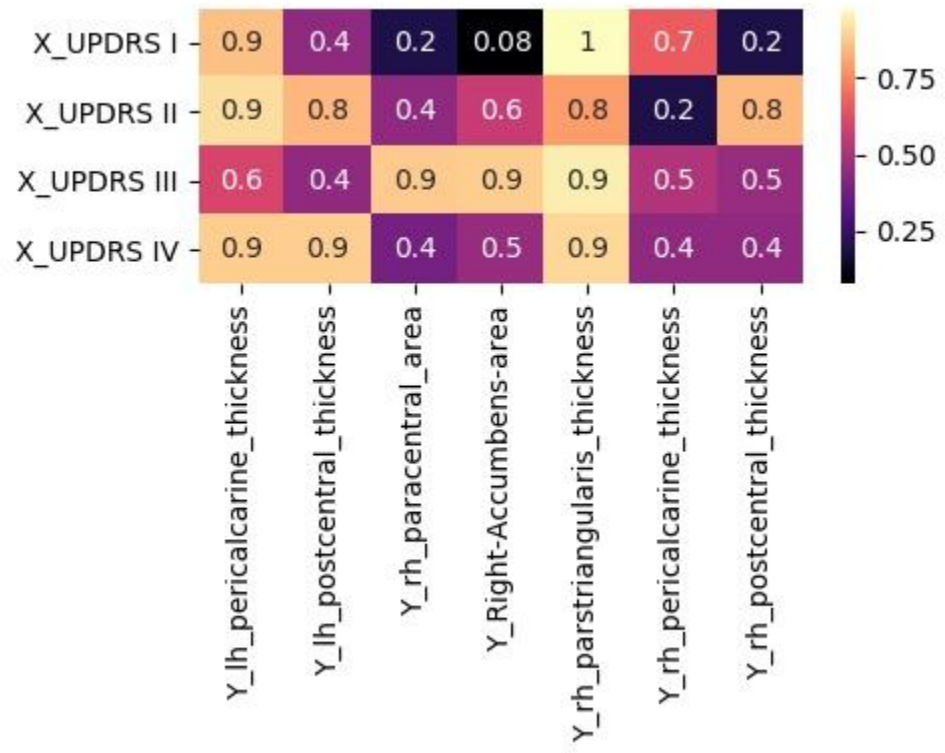


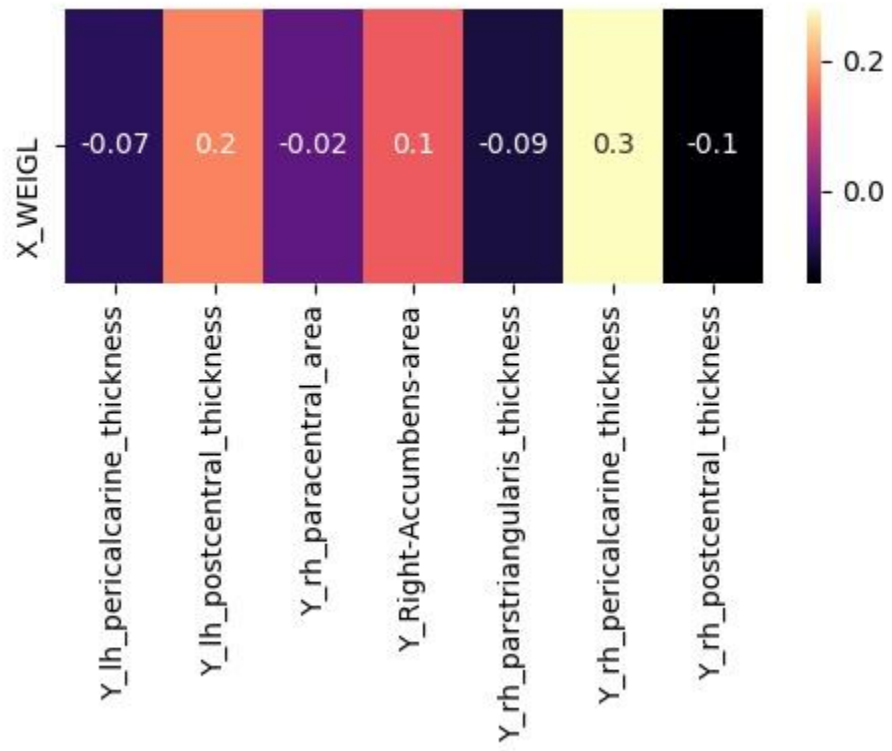




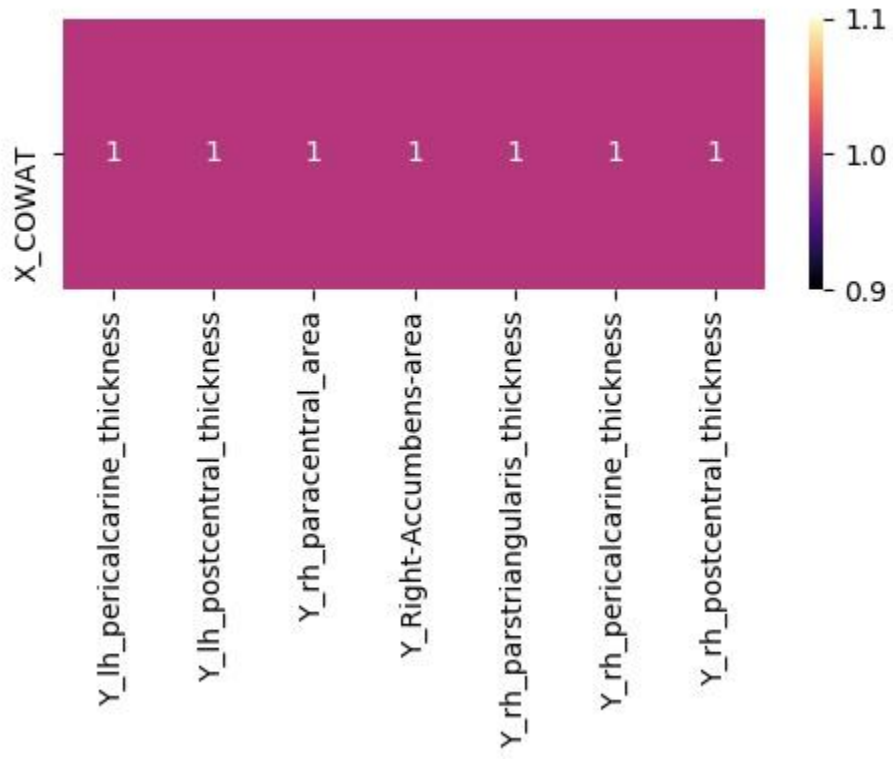


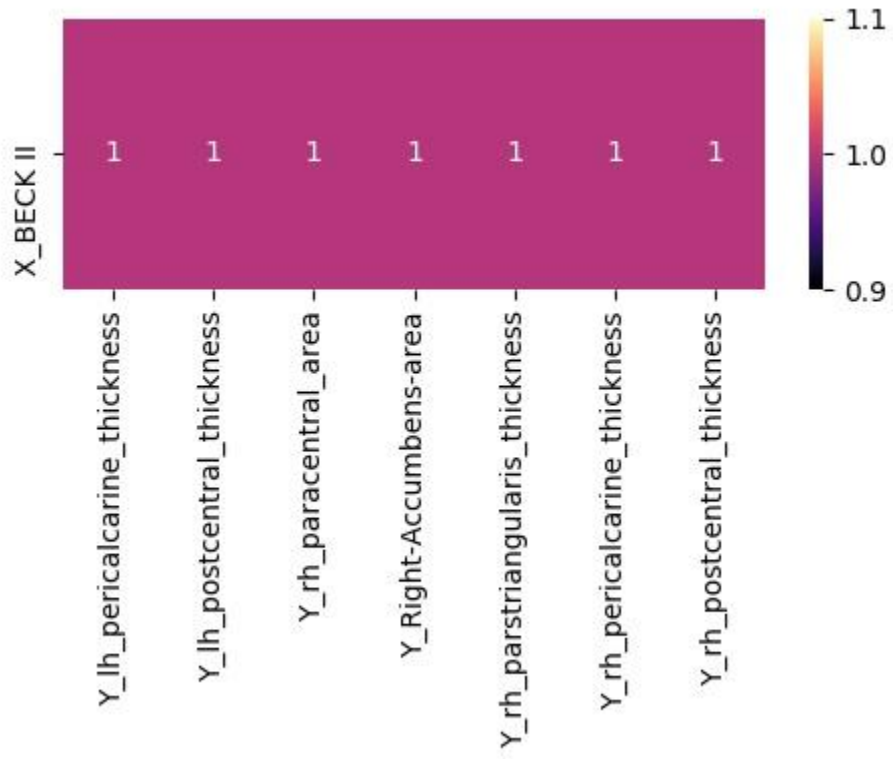


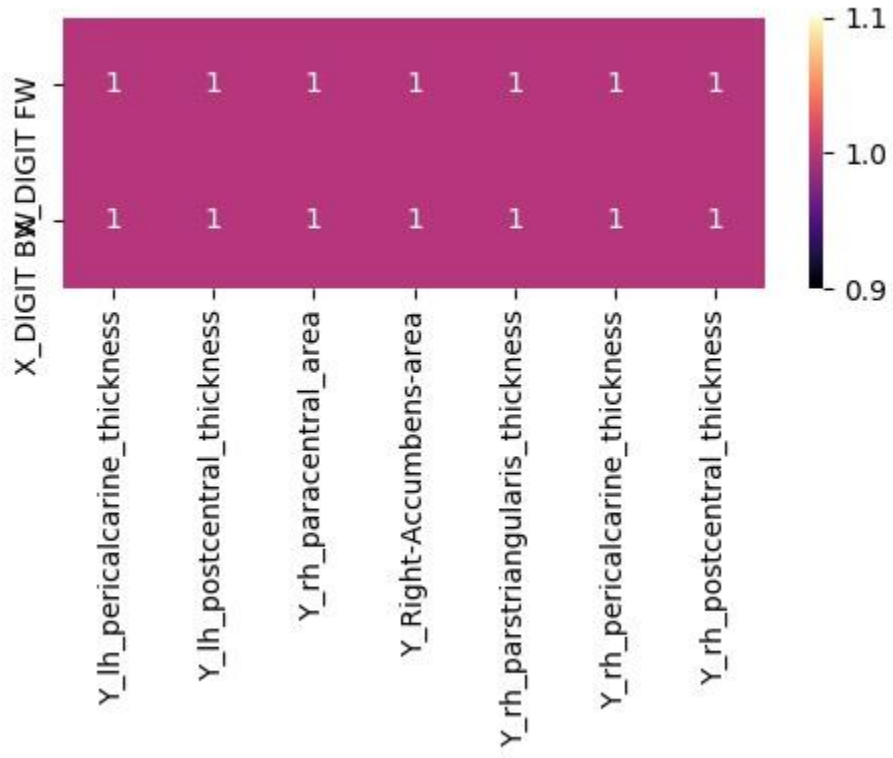


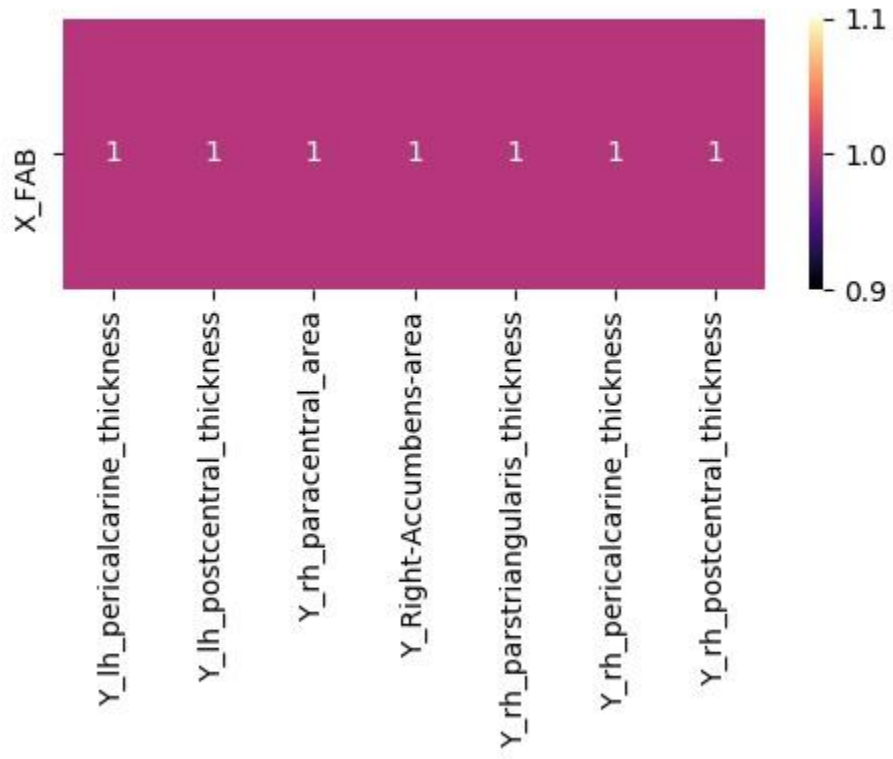


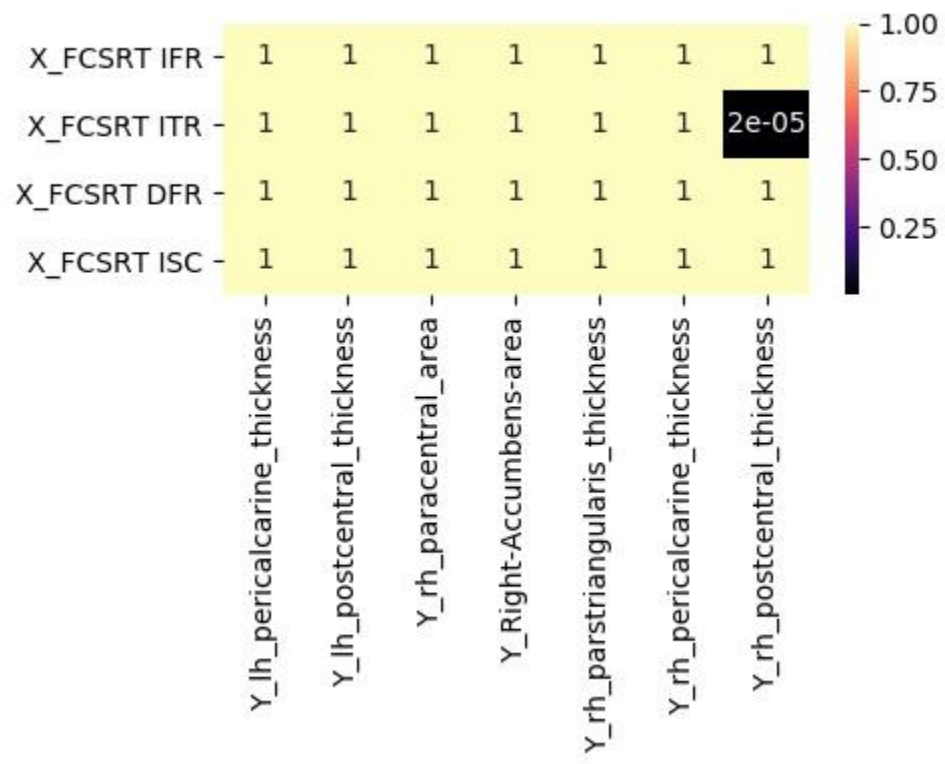
Correzione con Bonferroni

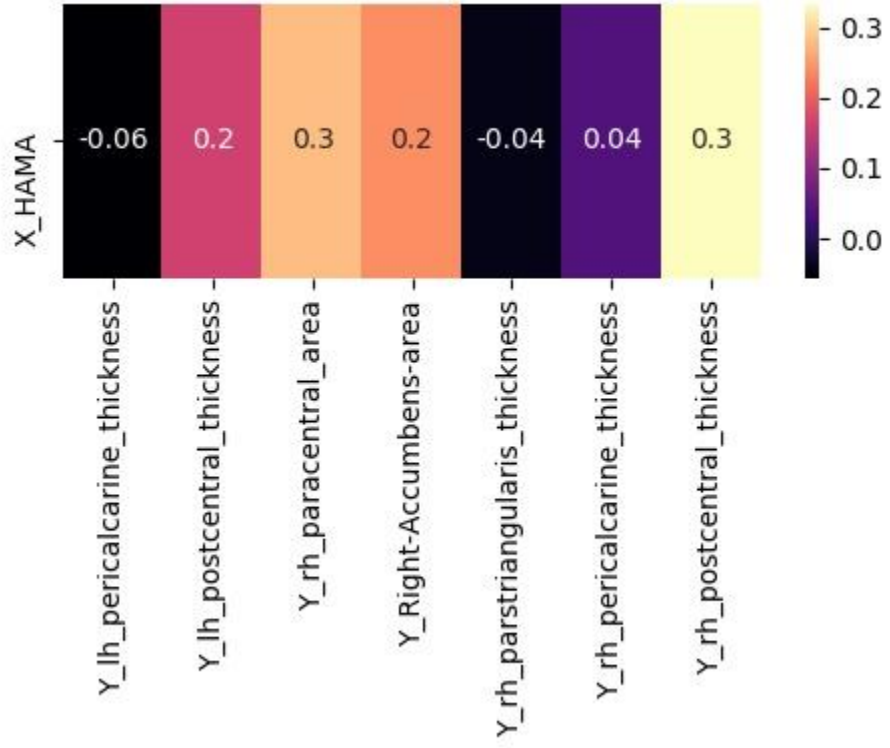


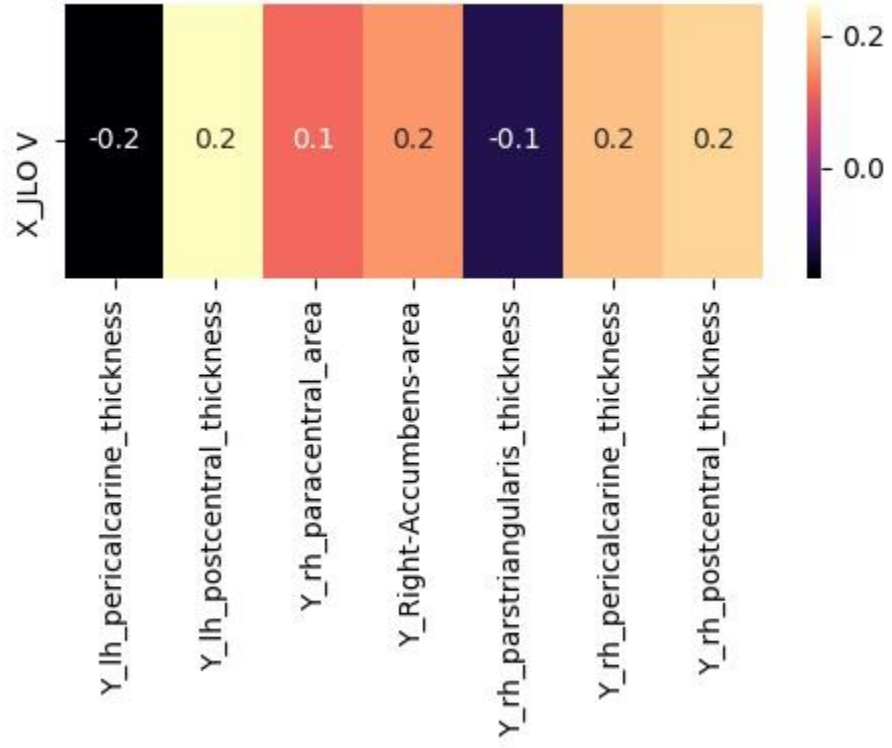


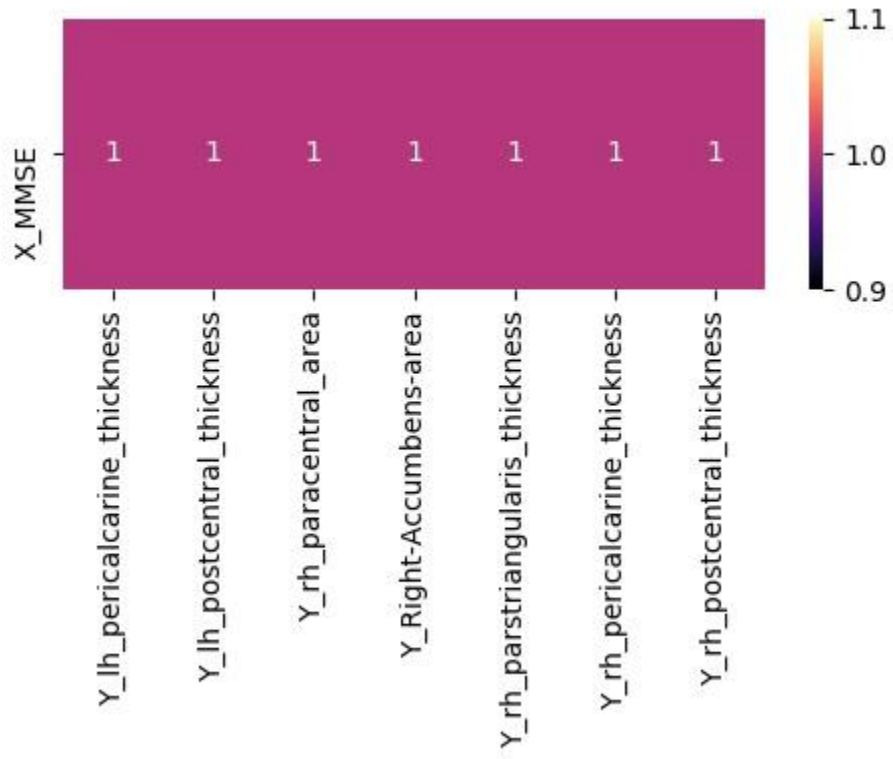


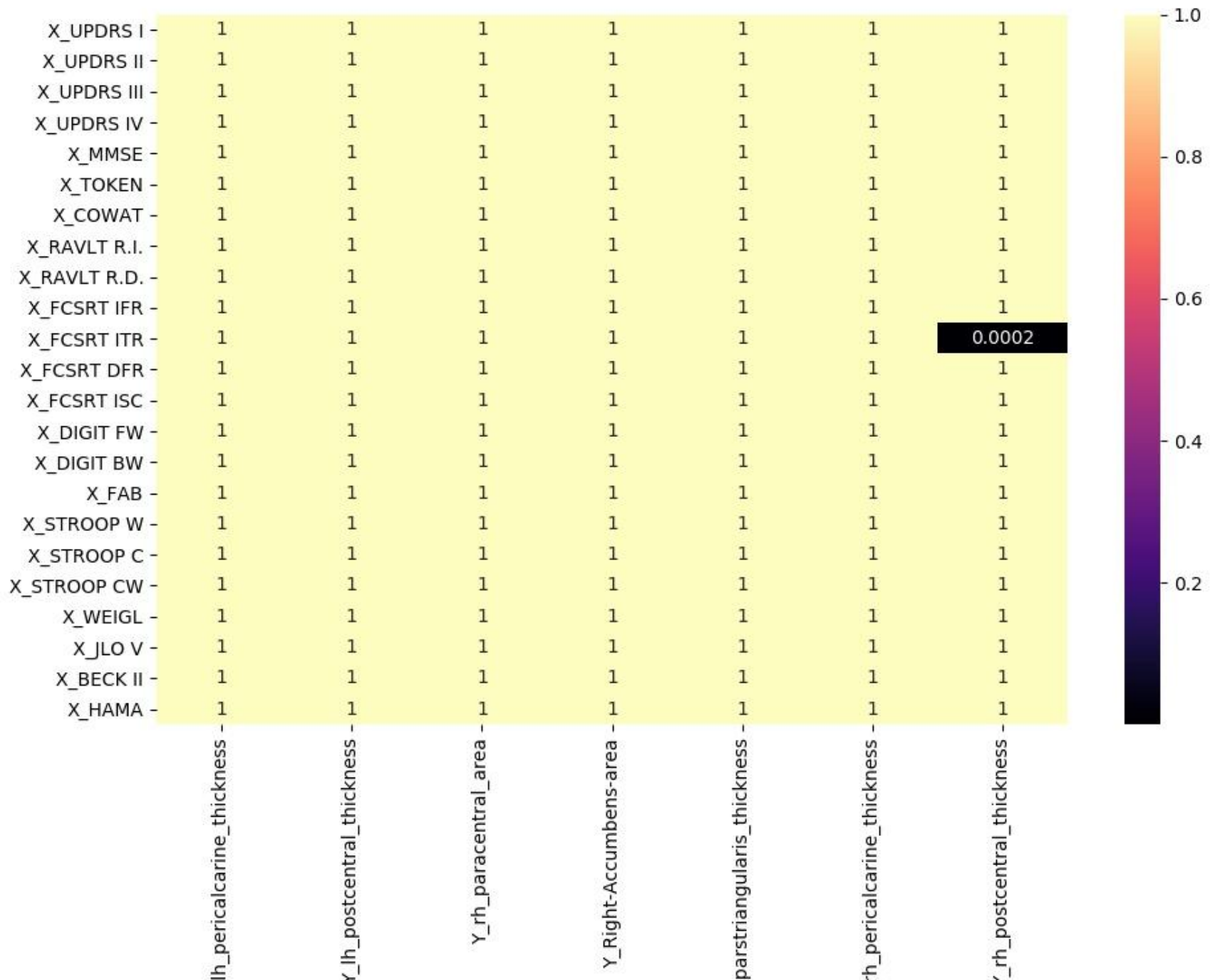


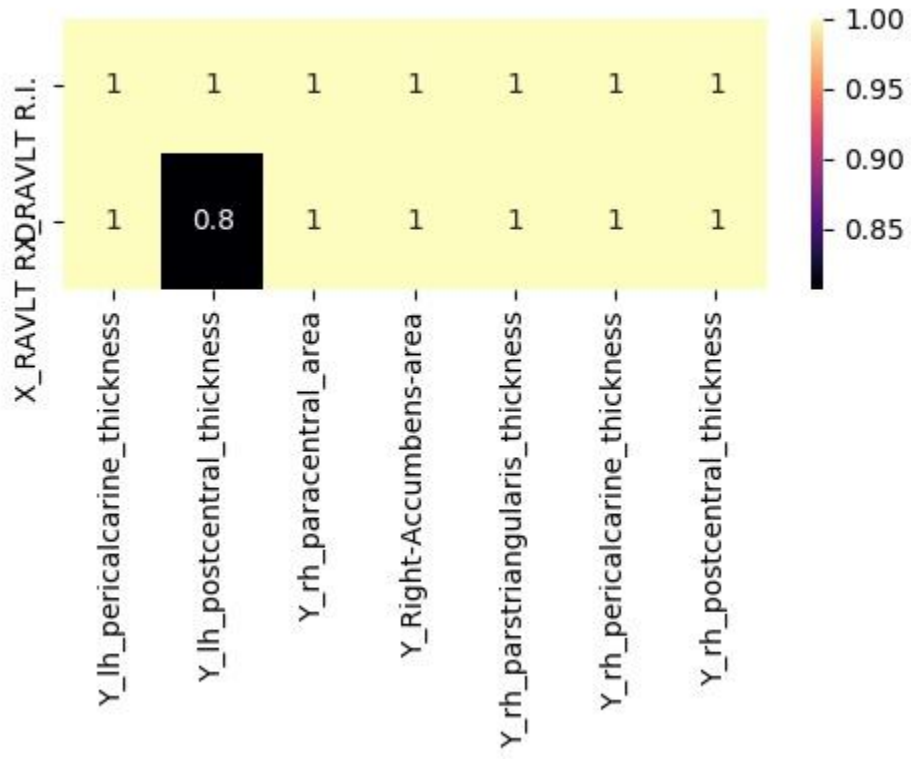


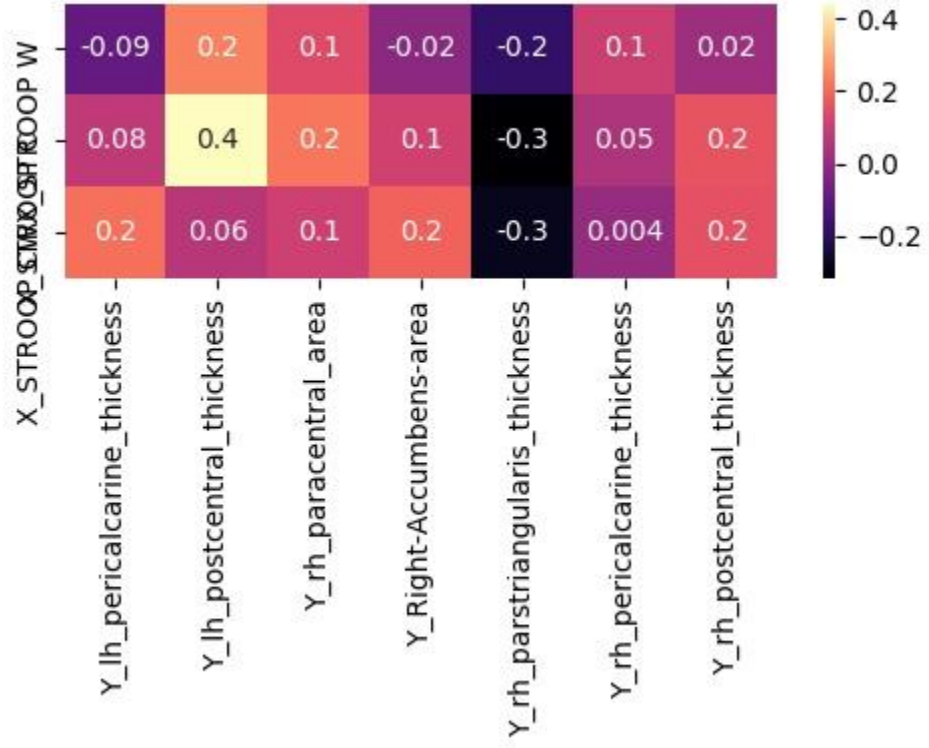


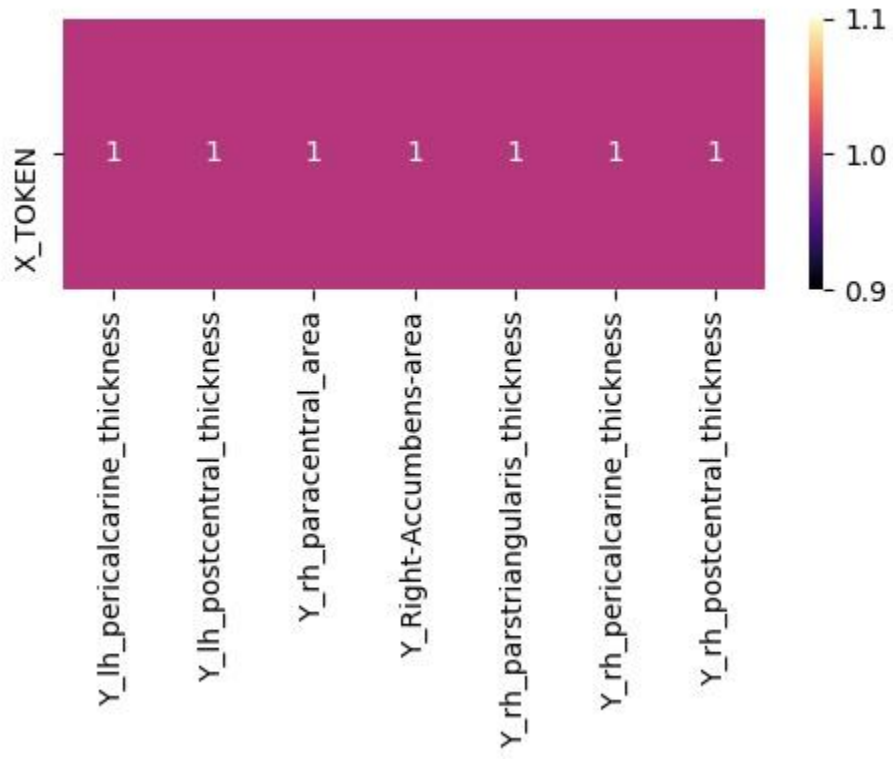


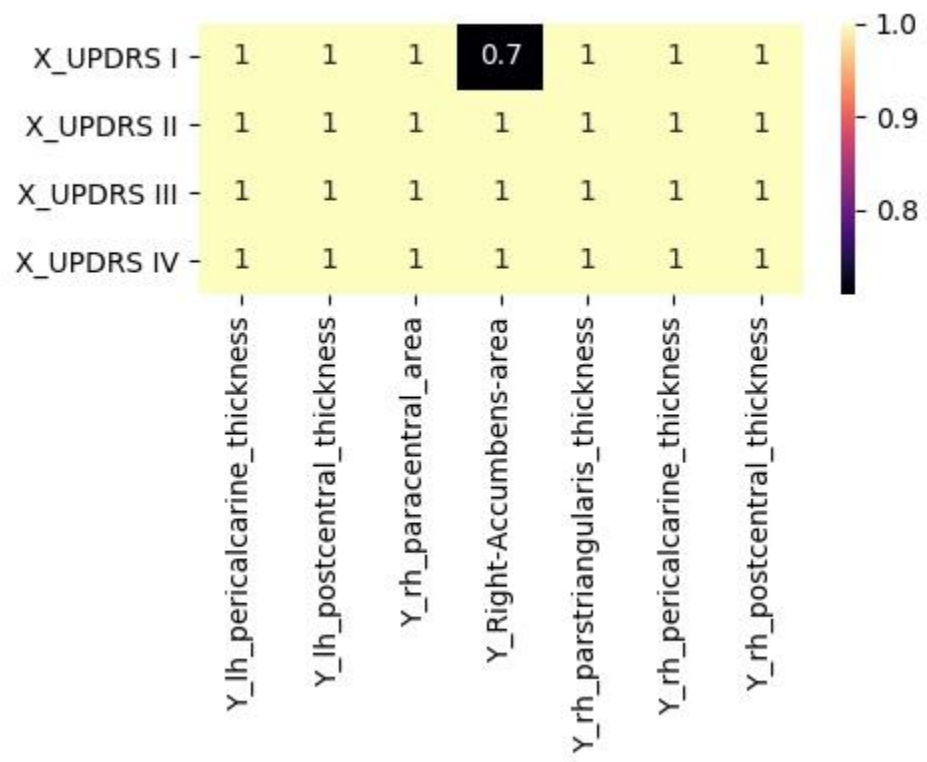


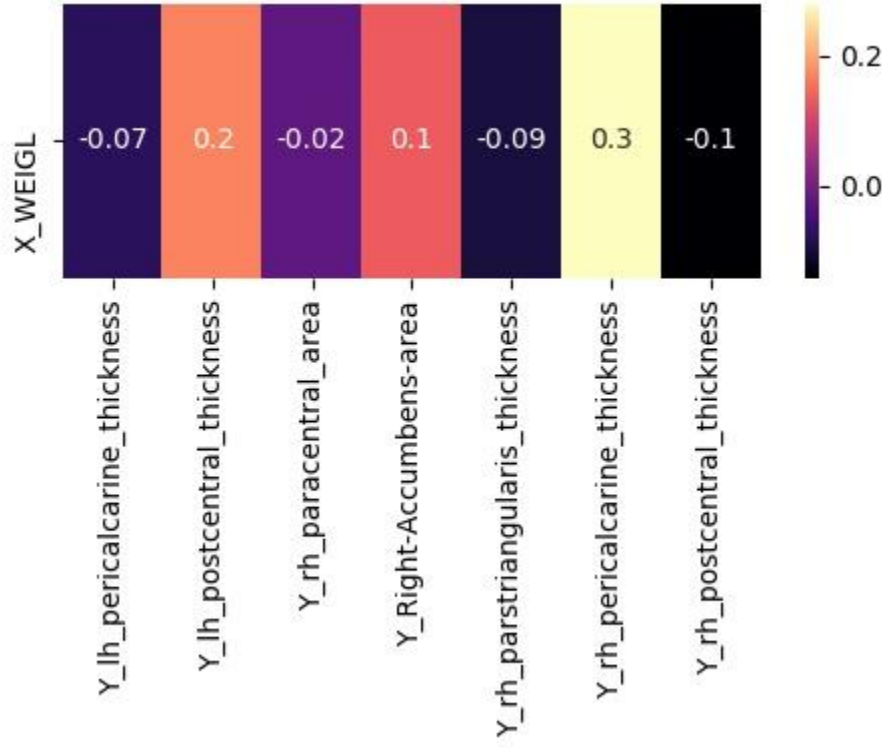










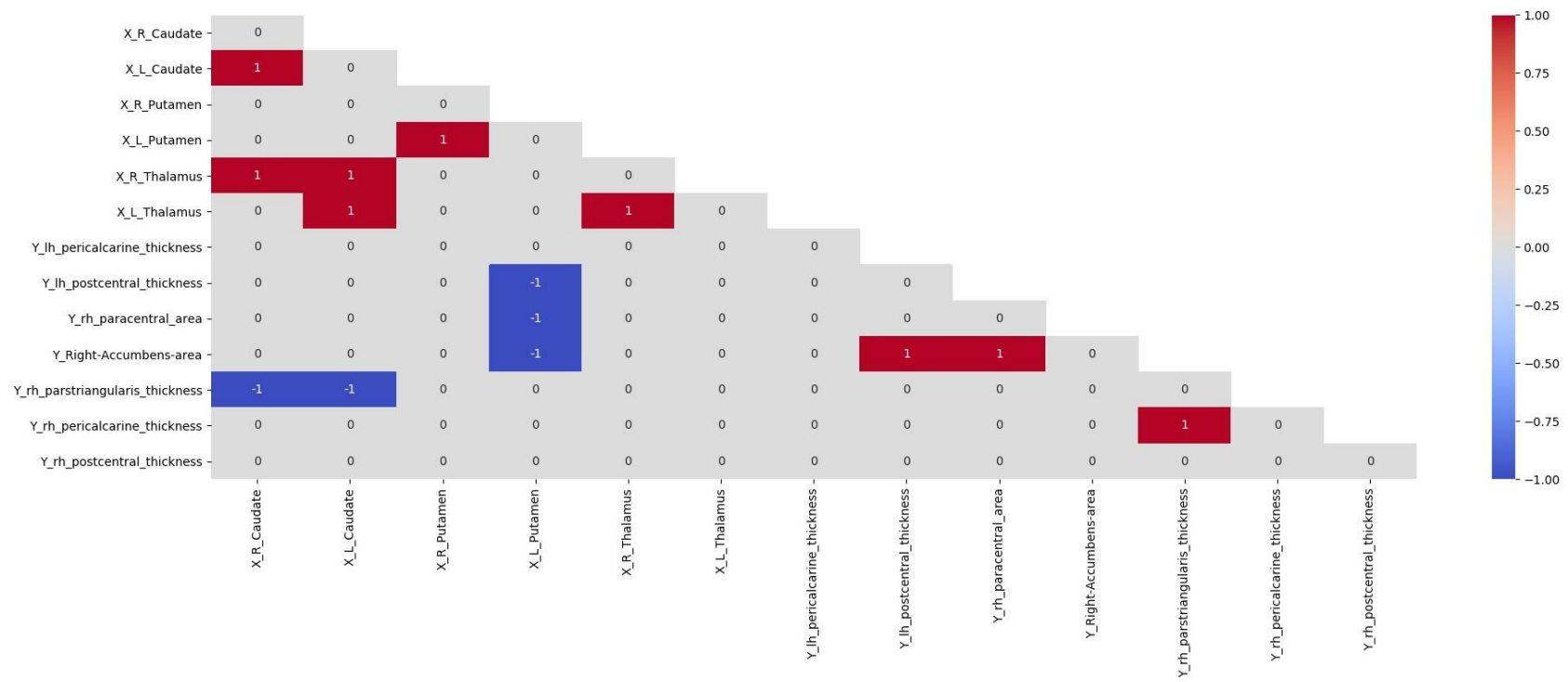


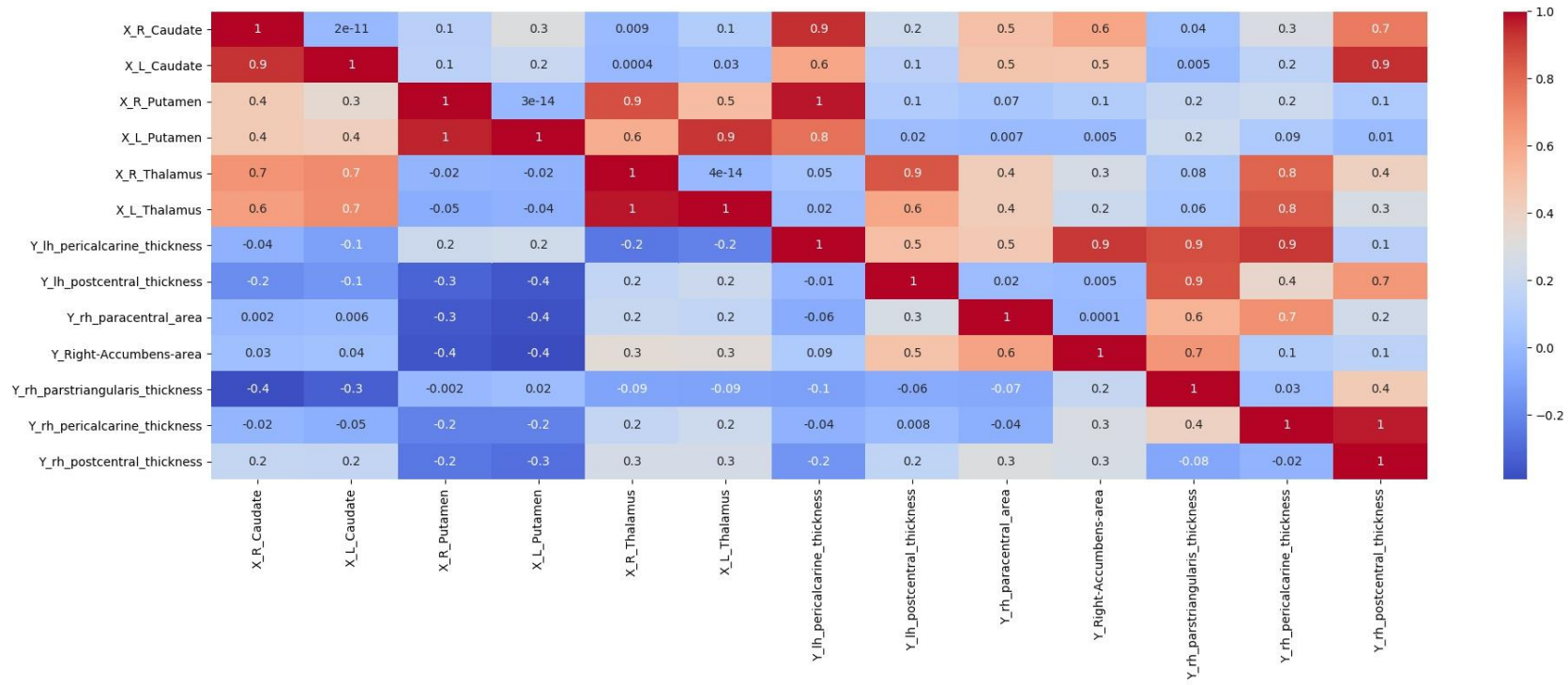
Allegato 6

Morbo di Parkinson

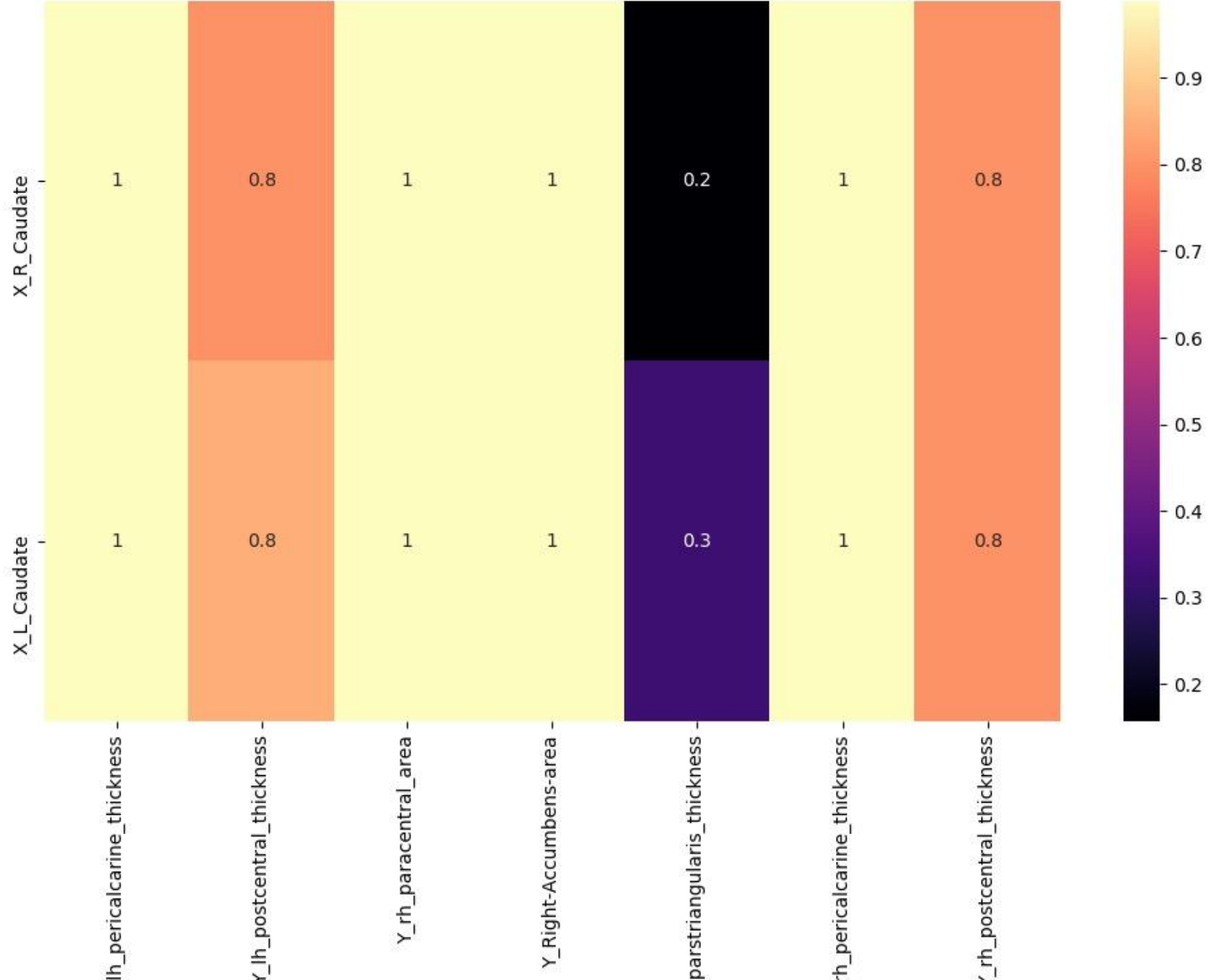
Dati di Uptake vs Freesurfer Features

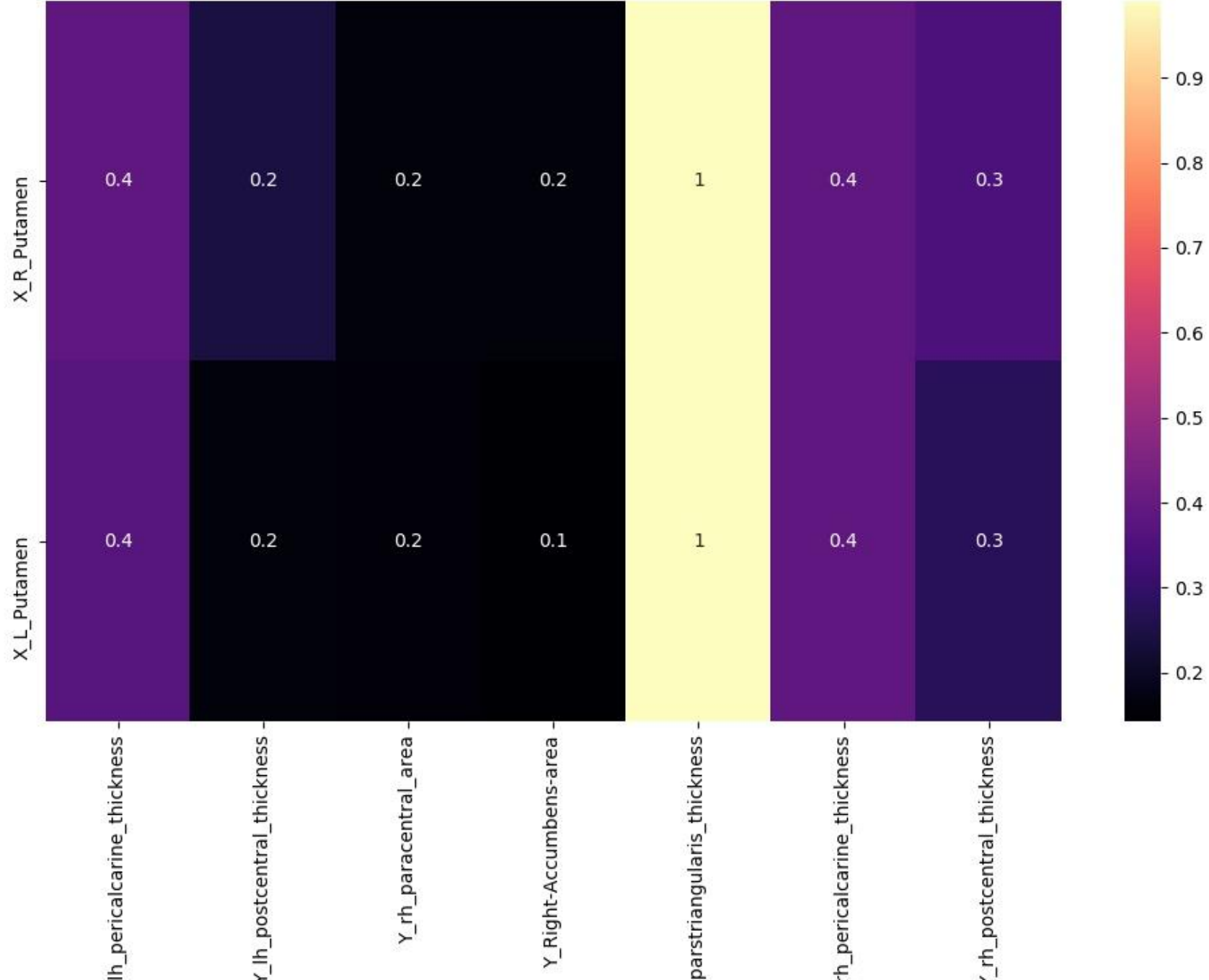
Correlazione di Pearson

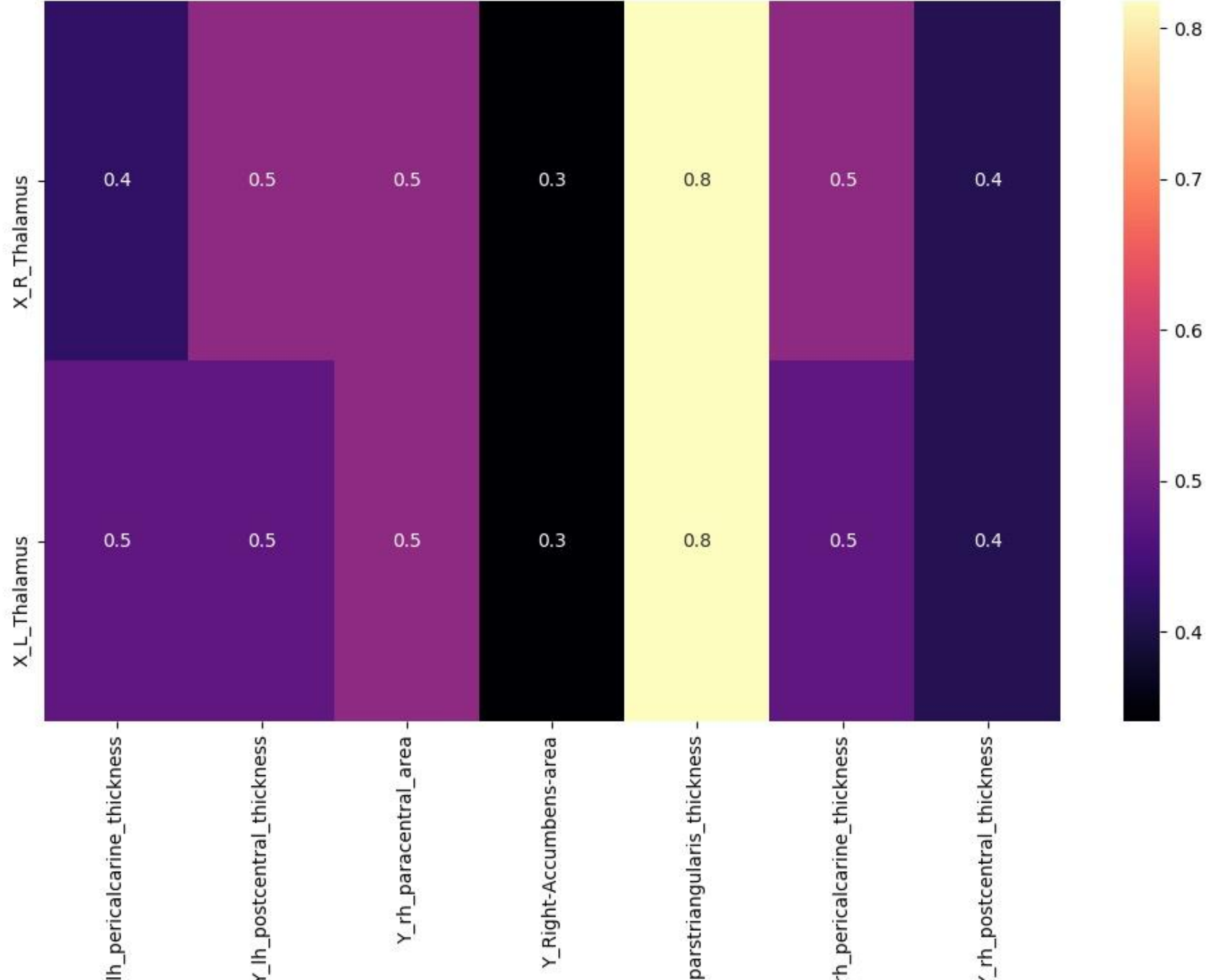




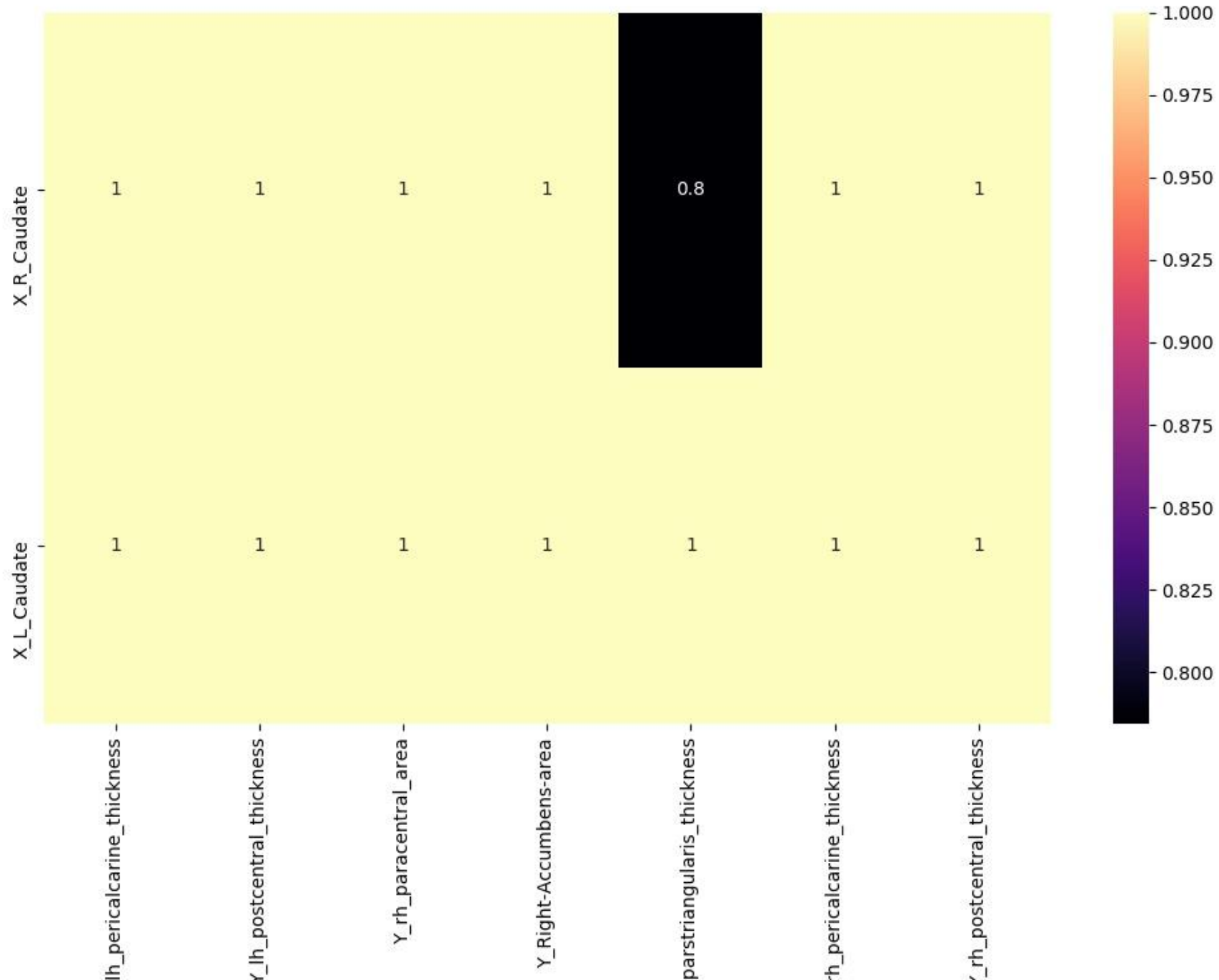
Correzione con Benjamini-Hochberg

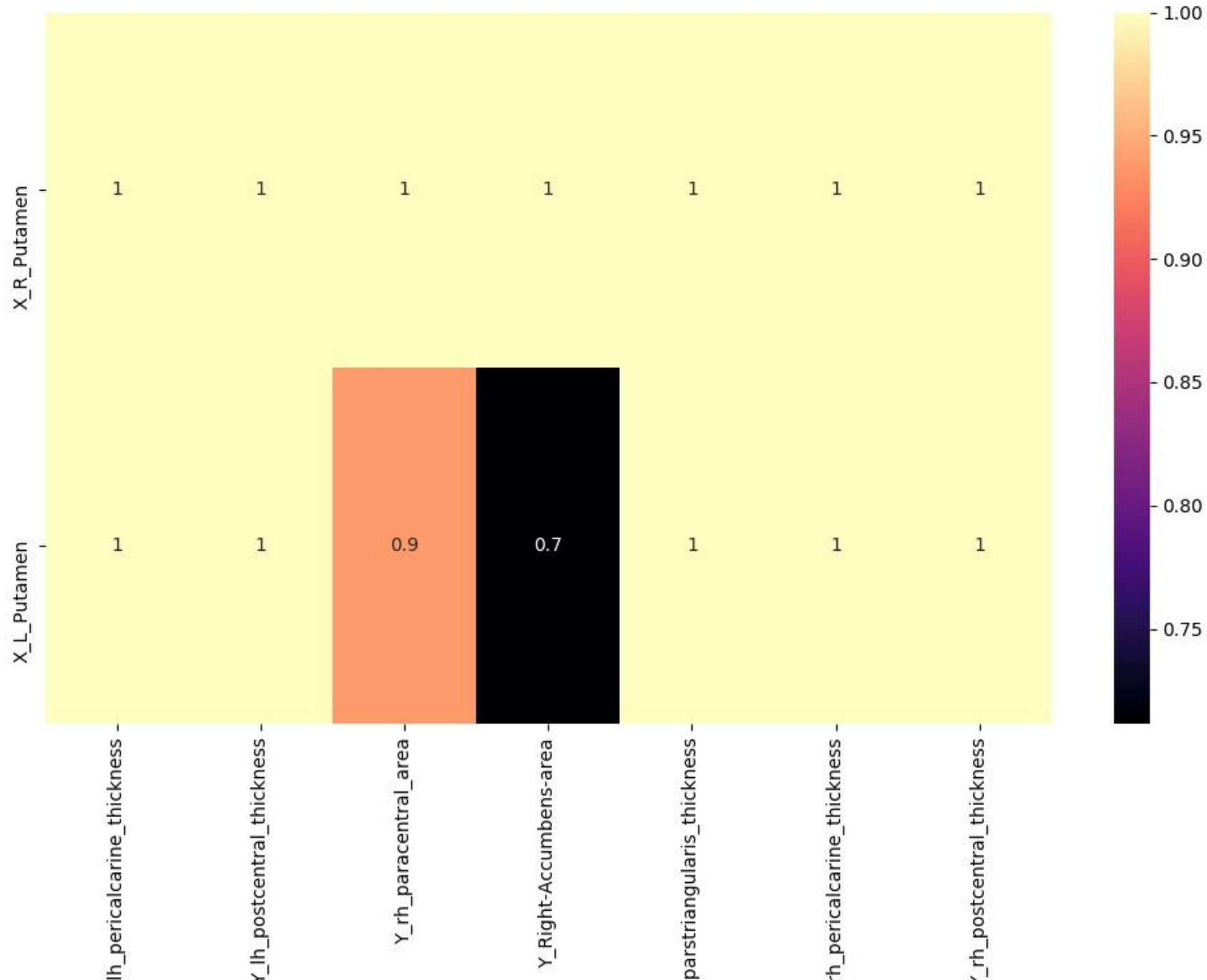






Correzione con Bonferroni







Allegato 7

Marker ER positivo/negativo. Totale pazienti=80 (P=62.5%,N=37.5%). Primo passo di Feature selection													
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var	
radiomiche da singola immagine	ADC	svm	no_alg	all	0,5625	0,02265625	0,5510714286	0,03361549036	0,5566666667	0,02756666667	0,5319191919	0,02629151565	
		mip	fisher	45	0,6125	0,02640625	0,6455952381	0,03054139739	0,6033333333	0,01832222222	0,5811255411	0,02259080102	
		rf	fisher	45	0,6	0,0275	0,5901190476	0,04238521825	0,5733333333	0,02595555556	0,5637085137	0,03012497788	
	PC	dt	no_alg	all	0,6375	0,01078125	0,6125	0,01540625	0,5966666667	0,01143333333	0,5958441558	0,01249283934	
		dt	relieff	20	0,65	0,03375	0,6325	0,03539513889	0,6333333333	0,04066666667	0,624011544	0,03677453654	
		svm	relieff	5	0,6125	0,02328125	0,6125	0,02533680556	0,6166666667	0,02761111111	0,6029365079	0,02443336483	
	SUB	dt	no_alg	all	0,525	0,030625	0,5116666667	0,03237777778	0,5	0,02644444444	0,4998340548	0,02875478351	
		mip	relieff	20	0,625	0,053125	0,6025	0,06319513889	0,6	0,06488888889	0,5913924964	0,06158192734	
		knn	gini_index	5	0,6	0,02125	0,6191666667	0,02584791667	0,6133333333	0,0256	0,5901587302	0,02184658604	
	T2	mip	no_alg	all	0,7	0,035	0,6901190476	0,05956974206	0,6666666667	0,04355555556	0,6562409812	0,04674273892	
mip		fisher	35	0,725	0,015	0,7248809524	0,0216211593	0,7	0,01955555556	0,6946897547	0,01849924435		
cliniche/radiomiche e da singola immagine	ADC	rf	no_alg	all	0,5875	0,02828125	0,5611309524	0,04152678926	0,55	0,02316666667	0,5352941503	0,02725956297	
		dt	gini_index	30	0,6	0,03375	0,6083333333	0,03811111111	0,6	0,03777777778	0,582950938	0,03549871634	
		mip	chi	50	0,6125	0,02953125	0,6323809524	0,04904075964	0,59	0,02512222222	0,566984127	0,02922350718	
	PC	dt	no_alg	all	0,575	0,0225	0,5176190476	0,05231428571	0,52	0,03115555556	0,4992424242	0,03396019988	
		dt	chi	50	0,65	0,030625	0,6523809524	0,04645181406	0,6333333333	0,036	0,622012987	0,03600061057	
		dt	relieff	15	0,6625	0,01578125	0,6467857143	0,03553005952	0,63	0,02032222222	0,6187229437	0,02321396947	
	SUB	rf	no_alg	all	0,625	0,00625	0,634047619	0,01191133787	0,6066666667	0,00595555556	0,5914718615	0,005432808606	
		dt	relieff	5	0,6625	0,06890625	0,6583333333	0,067875	0,6566666667	0,07245555556	0,6498556999	0,07111224906	
		mip	chi	25	0,625	0,046875	0,605	0,05648888889	0,6066666667	0,05951111111	0,5956637807	0,05473652673	
	T2	mip	no_alg	all	0,6875	0,02265625	0,6892857143	0,04350396825	0,6633333333	0,02676666667	0,649018759	0,03007645534	
mip		fisher	25	0,7	0,01	0,6975	0,01575069444	0,6866666667	0,01337777778	0,6767532468	0,01250013306		
cliniche/radiomiche	ALL	mip	no_alg	all	0,5625	0,04140625	0,5317857143	0,06635744048	0,5233333333	0,04712222222	0,5117748918	0,04952634238	
		mip	chi	50	0,675	0,035	0,66	0,04324722222	0,6666666667	0,04022222222	0,6473809524	0,04354462586	
		svm	chi	50	0,6125	0,05765625	0,6225	0,05575069444	0,63	0,0601	0,6086507937	0,05736929327	
	radiomiche	ALL	mip	no_alg	all	0,575	0,0475	0,5377380952	0,07054961735	0,54	0,0544	0,5305627706	0,05852337766
			mip	chi	25	0,625	0,034375	0,6298809524	0,04976798469	0,6	0,03777777778	0,5876767677	0,03805249444
	mip		relieff	15	0,625	0,021875	0,6330952381	0,03058752834	0,5933333333	0,02128888889	0,5865367965	0,02027796435	

Marker ER positivo/negativo (P=62.5%,N=37.5%) - Secondo passo di Feature selection												
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
radiomiche da singola immagine	T2	mip	lr	11	0,7375	0,01078125	0,7391666667	0,01598680556	0,7233333333	0,01267777778	0,7164141414	0,01300277278
		mip	lr	43	0,7375	0,02953125	0,71875	0,04469878472	0,73	0,0301	0,7074875125	0,03940666349

Costruzione del modello con approccio leave one out												
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	precision	recall	f1score				
radiomiche da singola immagine	T2	svm	no_alg	11	0,8125	0,8723404255	0,82	0,8453608247				
		mip	no_alg	11	0,8	0,8541666667	0,82	0,8367346939				

TEST

DATI	numero pazienti	per classe (N/P)	accuracy	precision	recall	f1score
UNIME	19	P=16	0,7895	0,8333	0,9375	0,8823
		N=3		0	0	0

Marker HER2 positivo/negativo. Totale pazienti=80 (P=29%,N=71%)												
Primo passo di Feature selection												
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
radiomiche da singola immagine	ADC	dt	no_alg	all	0,6375	0,01703125	0,6080952381	0,02638117914	0,5716666667	0,013725	0,5625685426	0,0154932303
		rf	reliefF	30	0,7625	0,01390625	0,7164285714	0,03331570295	0,7066666667	0,0266222222	0,6954312354	0,0278661841
		mlp	gini_index	5	0,7125	0,03140625	0,6748809524	0,04903981009	0,6733333333	0,0554	0,6527561328	0,0479028803
	PC	nb	no_alg	all	0,6625	0,01578125	0,5226190476	0,04252380952	0,57	0,0309333333	0,5248684649	0,0296096989
		svm	reliefF	5	0,6625	0,03140625	0,6348809524	0,03596600057	0,66	0,0401777777	0,6176390276	0,0397884415
		mlp	fisher	15	0,6375	0,04203125	0,5819047619	0,0592723356	0,6116666667	0,0626694444	0,5793939394	0,0577131313
	SUB	nb	no_alg	all	0,7125	0,03453125	0,5660714286	0,08685126134	0,6133333333	0,0655444444	0,5785680986	0,0733480055
		knn	fisher	5	0,7	0,035	0,6886904762	0,03985352891	0,6916666667	0,0399027777	0,6686069486	0,0380448140
		dt	fisher	15	0,6875	0,01953125	0,6569047619	0,03253905896	0,655	0,0247805555	0,6231002331	0,0273418689
	T2	mlp	no_alg	all	0,675	0,031875	0,6595238095	0,05028231293	0,5933333333	0,0217333333	0,5921328671	0,0281099517
		dt	gini_index	40	0,7	0,01	0,6655952381	0,02672178288	0,6433333333	0,0107333333	0,6303363303	0,0140052424
		dt	reliefF	15	0,7125	0,01890625	0,665952381	0,04676196145	0,6366666667	0,0324333333	0,6287062937	0,0342591978
cliniche/radiomiche e da singola immagine	ADC	dt	no_alg	all	0,7375	0,02015625	0,6782142857	0,03592631803	0,6816666667	0,029525	0,6706116106	0,0329617331
		nb	reliefF	10	0,725	0,015	0,6958333333	0,02379513889	0,7366666667	0,0325444444	0,6861760462	0,0257313218
		svm	reliefF	25	0,75	0,015625	0,691547619	0,03776600057	0,7066666667	0,0335666666	0,6702164502	0,0325240456
	PC	nb	no_alg	all	0,675	0,019375	0,530952381	0,04693849206	0,5783333333	0,0359472222	0,536989677	0,0359857067
		mlp	gini_index	20	0,6625	0,00640625	0,5604761905	0,02352063492	0,5866666667	0,0204888888	0,5614452214	0,0202507071
		dt	chi	5	0,6625	0,02203125	0,5766666667	0,05415623583	0,5583333333	0,0274583333	0,5541025641	0,0351274819
	SUB	nb	no_alg	all	0,7125	0,03453125	0,5660714286	0,08685126134	0,6133333333	0,0655444444	0,5785680986	0,0733480055
		nb	reliefF	5	0,6375	0,02328125	0,6525	0,02050069444	0,6866666667	0,0302111111	0,619992785	0,0230917403
		mlp	chi	30	0,6875	0,02578125	0,6592857143	0,04887465986	0,6433333333	0,0397333333	0,6176456876	0,0381463049
	T2	nb	no_alg	all	0,675	0,038125	0,5861904762	0,06253577098	0,6083333333	0,0472361111	0,5676257076	0,0510603830
		nb	fisher	50	0,675	0,025625	0,7171428571	0,01931269841	0,7	0,0155	0,6423154623	0,0229476486
		dt	reliefF	50	0,7125	0,01890625	0,6652380952	0,04080493197	0,655	0,0306138888	0,6657021867	0,0299026662
cliniche/radiomiche e	ALL	nb	no_alg	all	0,7	0,019375	0,588452381	0,0502046627	0,63	0,0428777777	0,5892840493	0,0417361929
		mlp	gini_index	10	0,725	0,015	0,7098809524	0,0216009212	0,7066666667	0,0205111111	0,6806349206	0,0187515926
		dt	reliefF	50	0,7125	0,01578125	0,6627380952	0,03248697562	0,6466666667	0,0199333333	0,6397957598	0,0238493188
radiomiche	ALL	dt	no_alg	all	0,6625	0,01578125	0,6032142857	0,01708901644	0,6183333333	0,0193583333	0,5980863581	0,0197331426
		rf	reliefF	30	0,7	0,01	0,6678571429	0,03477692744	0,6616666667	0,0271694444	0,6177022977	0,0231493269
		knn	chi	5	0,6375	0,03578125	0,6283333333	0,03182222222	0,6266666667	0,0402888888	0,6028282828	0,0355994128

Marker HER2 positivo/negativo (P=29%,N=71%) - Secondo passo di Feature selection												
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
cliniche/radiomiche da singola immagine	ADC	knn	mi	27	0,7375	0,03265625	0,6891666667	0,04719513889	0,7133333333	0,0592666666	0,6885137085	0,0505318541
		svm	mi	42	0,725	0,030625	0,6983333333	0,038525	0,715	0,0418583333	0,6837806638	0,0404793688
cliniche/radiomiche e *	ALL	dt	cfs	5	0,8	0,019375	0,7532142857	0,04037473073	0,7733333333	0,0357333333	0,7493994894	0,0372179752

* in realtà sono selezionate solo radiomiche

Costruzione del modello con approccio leave one out												
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	precision	recall	f1score				
cliniche/radiomiche e *	ALL	rf	no_alg	5	0,8	0,6666666667	0,6086956522	0,6363636364				
		svm	no_alg	5	0,7375	0,5277777778	0,8260869565	0,6440677966				

TEST						
DATI	numero pazienti	per classe (N/P)	accuracy	precision	recall	f1score
UNIME	19	N=16	0,8421	0,8421	1	0,91
		P=3		0	0	0

Marker KI67 positivo/negativo. Totale pazienti=78 (P=14%,N=86%)												
Primo passo di Feature selection												
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
radiomiche da singola immagine	ADC	svm	no_alg	all	0,7803571429	0,0290210455	0,637797619	0,0599188810	0,644047619	0,0649503968	0,6118315018	0,0597826832
		nb	chi	10	0,8589285714	0,0107429846	0,6806547619	0,0560906498	0,7297619048	0,0561238662	0,6906560107	0,0537727090
		rf	relieff	45	0,8214285714	0,0253188775	0,6766071429	0,0582107461	0,7083333333	0,0711238662	0,6767365967	0,0619753294
	PC	svm	no_alg	all	0,8964285714	0,0058545918	0,6669642857	0,0618152104	0,7273809524	0,0595365646	0,6872860473	0,0578977600
		svm	chi	10	0,8857142857	0,0139795918	0,7616071429	0,0567680874	0,8047619048	0,0627607709	0,7717149517	0,0576762662
		rf	fisher	5	0,8589285714	0,0232429846	0,7670833333	0,0481126736	0,8142857143	0,0560714285	0,7709090909	0,0516275482
	SUB	dt	no_alg	all	0,8178571429	0,0144005102	0,6461309524	0,0590250318	0,6654761905	0,0639299886	0,636959707	0,0547512565
		rf	fisher	30	0,8839285714	0,0015465561	0,6982142857	0,0366198979	0,7880952381	0,0390816326	0,7157109557	0,0306010421
		dt	relieff	50	0,8982142857	0,0119929846	0,6907738095	0,0722318771	0,7285714286	0,0681632653	0,7041025641	0,0688497041
	T2	svm	no_alg	all	0,7589285714	0,0256536989	0,5826190476	0,0553844104	0,6119047619	0,0652267573	0,5727072927	0,0582746003
		nb	chi	25	0,8089285714	0,0222608418	0,6769047619	0,0550646825	0,7023809524	0,0709183673	0,6652347652	0,0559198166
		dt	relieff	40	0,85	0,02125	0,6504166667	0,0651369791	0,7071428571	0,0678061224	0,6623443223	0,0637383515
cliniche/radiomiche e da singola immagine	ADC	svm	no_alg	all	0,7928571429	0,0265433673	0,6420833333	0,0638418402	0,65	0,0760544217	0,6351282051	0,0679050624
		dt	chi	20	0,8089285714	0,0169674744	0,68125	0,0420182291	0,7607142857	0,0535274943	0,6821328671	0,0449815911
		nb	relieff	15	0,8214285714	0,0221938775	0,6741071429	0,0575666985	0,7083333333	0,0639810090	0,6715617716	0,0599584766
	PC	svm	no_alg	all	0,8714285714	0,0094260204	0,6491071429	0,0537546945	0,7119047619	0,0602267573	0,6683449883	0,0530251933
		mip	gini_index	5	0,8482142857	0,0211894132	0,7420833333	0,0420918402	0,8083333333	0,0541510771	0,7510955711	0,0460727533
		nb	fisher	10	0,8571428571	0,0210459183	0,7432738095	0,0550202699	0,7726190476	0,0663109410	0,7389160839	0,0558746704
	SUB	rf	no_alg	all	0,8196428571	0,0198628826	0,6226190476	0,0711748866	0,6476190476	0,0760260771	0,6299450549	0,0713347237
		rf	fisher	30	0,8464285714	0,0054974489	0,6806547619	0,0364477926	0,7666666667	0,0402551020	0,6862604063	0,0310040930
		dt	fisher	5	0,8089285714	0,0097608418	0,669047619	0,0335544217	0,7607142857	0,0560501700	0,68002664	0,0383377583
	T2	dt	no_alg	all	0,7946428571	0,0039700255	0,5735119048	0,0213329967	0,6488095238	0,0478812358	0,5875241425	0,0251965244
		rf	chi	5	0,7964285714	0,0229719387	0,7211904762	0,0434321995	0,7952380952	0,0427380952	0,7053096903	0,0477524458
		rf	fisher	30	0,7964285714	0,0198469387	0,6629166667	0,0425223958	0,7357142857	0,0555215419	0,6651631702	0,0458356139
cliniche/radiomiche	ALL	dt	no_alg	all	0,8214285714	0,0065688775	0,5443452381	0,0237943771	0,6476190476	0,0462641723	0,5719014319	0,0256080839
		dt	chi	50	0,8464285714	0,0095790816	0,7041666667	0,0327083333	0,8071428571	0,0317970521	0,7173076923	0,0309727975
		dt	chi	35	0,8571428571	0,0198341836	0,7036904762	0,0663571995	0,7285714286	0,0665759637	0,7058874459	0,0657578456
radiomiche	ALL	all	no_alg	all	0,7678571429	0,0227678571	0,5227380952	0,0347392857	0,5476190476	0,0469387755	0,5261804862	0,0377020552
		dt	fisher	20	0,8732142857	0,0219036989	0,8007142857	0,0481984693	0,8238095238	0,0605952381	0,7934398934	0,0515532430
		dt	fisher	40	0,8482142857	0,0149394132	0,7011904762	0,0491819019	0,7654761905	0,0460898526	0,7046753247	0,0480883716

Marker KI67 positivo/negativo (P=14%,N=86%) - Secondo passo di Feature selection												
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
radiomiche	ALL	svm	lr	21	0,8446428571	0,0157557398	0,6949404762	0,0580379287	0,7380952381	0,0683956916	0,7012604063	0,0595681710
		mip	cfs	7	0,8178571429	0,0247321428	0,6957738095	0,0530803890	0,7476190476	0,0615589569	0,6968897769	0,0565075096
radiomiche da singola immagine	PC	mip	cfs	1	0,8714285714	0,0228826530	0,78375	0,0472654513	0,8392857143	0,0546839569	0,7894289044	0,0513299211
		mip	cfs	2	0,8571428571	0,0292091836	0,7795833333	0,0483973958	0,830952381	0,0554081632	0,7777622378	0,0558082622

Costruzione del modello con approccio leave one out								
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	precision	recall	f1score
radiomiche da	PC	knn/svm	no_alg	1	0,8846153846	0,9677419355	0,8955223881	0,9302325581
		mip	no_alg	1	0,8205128205	0,9649122807	0,8208955224	0,8870967742

singola immagine	PC	svm *	no_alg	2	0,8846153846	0,9677419355	0,8955223881	0,9302325581
		mlp	no_alg	2	0,8461538462	0,9661016949	0,8507462687	0,9047619048

			TEST			
DATI	numero pazienti	per classe (N/P)	accuracy	precision	recall	f1score
UNIME	18	N=14	0,778	0,75	0,92	0,85
		P=4		0,5	0,75	0,6

Marker PR positivo/negativo. Totale pazienti=80 (P=50%,N=50%)												
Primo passo di Feature selection												
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
radiomiche da singola immagine	ADC	svm	no_alg	all	0,6	0,030625	0,6252380952	0,0392009070	0,6	0,030625	0,5884271284	0,0313224371
		mlp	fisher	25	0,65	0,015	0,6633333333	0,0182666666	0,65	0,015	0,643015873	0,0157678004
		svm	fisher	30	0,6375	0,02640625	0,6719047619	0,0329453514	0,6375	0,02640625	0,6185858586	0,0285255223
	PC	nb	no_alg	all	0,6125	0,03890625	0,6266666667	0,0576709750	0,6125	0,03890625	0,5868109668	0,0454461725
		knn	fisher	5	0,7375	0,02953125	0,7652380952	0,0303675737	0,7375	0,02953125	0,726046176	0,0329099998
		knn	fisher	25	0,7	0,02875	0,6969047619	0,0469195578	0,7	0,02875	0,6695382395	0,0414071458
	SUB	mlp	no_alg	all	0,5375	0,05328125	0,5383333333	0,0598916666	0,5375	0,05328125	0,5254761905	0,0556531179
		knn	fisher	5	0,625	0,040625	0,6164285714	0,0559687641	0,625	0,040625	0,6076695527	0,0481511029
		svm	fisher	15	0,6125	0,04203125	0,6383333333	0,0507805555	0,6125	0,04203125	0,5986507937	0,0427284013
	T2	mlp	no_alg	all	0,5625	0,01640625	0,5716666667	0,0211138888	0,5625	0,01640625	0,5503968254	0,0172692428
		mlp	fisher	50	0,7	0,0225	0,6833333333	0,0411666666	0,7	0,0225	0,6788095238	0,0328255291
		mlp	gini_index	35	0,6875	0,03203125	0,71	0,0369	0,6875	0,03203125	0,6731746032	0,0354781809
cliniche/radiomiche e da singola immagine	ADC	svm	no_alg	all	0,6125	0,03578125	0,6402380952	0,0449687641	0,6125	0,03578125	0,6007287157	0,0366596041
		mlp	fisher	25	0,65	0,015	0,6633333333	0,0182666666	0,65	0,015	0,643015873	0,0157678004
		svm	fisher	30	0,6375	0,02640625	0,6719047619	0,0329453514	0,6375	0,02640625	0,6185858586	0,0285255223
	PC	nb	no_alg	all	0,6125	0,03890625	0,6266666667	0,0576709750	0,6125	0,03890625	0,5868109668	0,0454461725
		knn	fisher	5	0,7375	0,02953125	0,7652380952	0,0303675737	0,7375	0,02953125	0,726046176	0,0329099998
		dt	chi	5	0,6625	0,00953125	0,69	0,0150666666	0,6625	0,00953125	0,6523809524	0,0099761904
	SUB	mlp	no_alg	all	0,5	0,046875	0,5	0,0516111111	0,5	0,046875	0,4885714286	0,0481043335
		knn	fisher	5	0,625	0,040625	0,6164285714	0,0559687641	0,625	0,040625	0,6076695527	0,0481511029
		mlp	reliefF	5	0,6125	0,02640625	0,6316666667	0,0422472222	0,6125	0,02640625	0,5991269841	0,0292464663
	T2	mlp	no_alg	all	0,5625	0,01640625	0,5869047619	0,0250830498	0,5625	0,01640625	0,5415223665	0,0171838122
		mlp	fisher	50	0,7	0,0225	0,6833333333	0,0411666666	0,7	0,0225	0,6788095238	0,0328255291
		mlp	gini_index	25	0,6625	0,03140625	0,67	0,0334333333	0,6625	0,03140625	0,6536507937	0,0336935248
cliniche/radiomiche e	ALL	mlp	no_alg	all	0,6375	0,02328125	0,6416666667	0,0381181972	0,6375	0,02328125	0,6181601732	0,0303445640
		mlp	reliefF	15	0,6875	0,06328125	0,6916666667	0,0755138888	0,6875	0,06328125	0,6727777778	0,0707130322
		mlp	fisher	50	0,6375	0,04203125	0,635	0,0635737528	0,6375	0,04203125	0,6135569986	0,0525080744
radiomiche	ALL	nb	no_alg	all	0,5875	0,02203125	0,5533333333	0,0624820861	0,5875	0,02203125	0,5301443001	0,0355374824
		mlp	reliefF	15	0,6875	0,06328125	0,6916666667	0,0755138888	0,6875	0,06328125	0,6727777778	0,0707130322
		mlp	fisher	50	0,6375	0,04203125	0,635	0,0635737528	0,6375	0,04203125	0,6135569986	0,0525080744

Marker PR positivo/negativo (P=50%,N=50%) - Secondo passo di Feature selection												
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
radiomiche da singola immagine	PC	knn	lr	3	0,75	0,021875	0,7871428571	0,0177088435	0,75	0,021875	0,7357431457	0,0259008714
		nb	cfs	6	0,725	0,04	0,7433333333	0,0444555555	0,725	0,04	0,7142857143	0,0436422272
radiomiche	ALL	knn	no_alg	18	0,725	0,199375	0,725	0,199375	0,725	0,199375	0,725	0,199375

Costruzione del modello con approccio leave one out												
Tipo features	Tipo immagine	algorithm	feature alg	n_features	accuracy	precision	recall	f1score				
radiomiche da singola immagine	PC	mlp	no_alg	3	0,725	0,725	0,725	0,725				
		knn	no_alg	3	0,7375	0,7714285714	0,675	0,72				
		nb	no_alg	6	0,725	0,7045454545	0,775	0,7380952381				
		rf	no_alg	6	0,675	0,6666666667	0,7	0,6829268293				

TEST							
DATI	numero pazienti	modello	per classe (N/P)	accuracy	precision	recall	f1score
UNIME	19	Mlp/3 features	N=6	0,6842	0,684	0	0
			P=13				
		nb/6 features	N=6	0,4736	0,334	0,667	0,445
			P=13				

Allegato 8

Selected Features

CANCRO AL SENO			
ER	HER2	KI67	PR
T2_original_shape_Flatness	T2_original_glrIm_RunVariance	PC_wavelet-LHH_gldm_DependenceEntropy	PC_original_shape_Flatness
T2_original_glcm_Idmn	T2_original_glrIm_LongRunEmphasis	PC_wavelet-HHL_gldm_SmallDependenceLowGrayLevelEmphasis	PC_wavelet-LLH_glcm_Correlation
T2_wavelet-LHH_glszm_ZoneEntropy	T2_original_glszm_ZonePercentage		PC_wavelet-HHH_ngtdm_Busyness
T2_wavelet-LHH_gldm_LargeDependenceLowGrayLevelEmphasis	T2_original_gldm_DependenceNonUniformityNormalized		
T2_wavelet-LLH_glszm_SmallAreaLowGrayLevelEmphasis	ADC_wavelet-LLH_firstorder_Minimum		
T2_wavelet-LLH_gldm_SmallDependenceLowGrayLevelEmphasis			
T2_wavelet-LLH_firstorder_Skewness			
T2_log-sigma-4-0-mm-3D_glcm_lmc2			
T2_log-sigma-4-0-mm-3D_firstorder_Skewness			
T2_log-sigma-5-0-mm-3D_glszm_SmallAreaEmphasis			
T2_log-sigma-5-0-mm-3D_firstorder_Skewness			

Allegato 9

Predizione di presenza di linfonodi. Totale pazienti 52 (positivi 81%, negativi 19%)											
Primo passo di feature selection con metodi ranker.											
tipo features	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
radiomiche	mlp	no_alg	168	0,7866666667	0,0176	0,55	0,061875	0,6025	0,05768125	0,568968254	0,0574623582
	dt	reliefF	38	0,8066666667	0,0377333333	0,6658333333	0,0860895833	0,69	0,086775	0,6711507937	0,0853384369
	dt	no_alg	167	0,76	0,0464	0,6525	0,0879868055	0,6625	0,09078125	0,6438492063	0,0895898289
cliniche/radiomiche	mlp	reliefF	30	0,8033333333	0,0161	0,62	0,060475	0,69	0,06115	0,6384920635	0,0563057445
Secondo passo di feature selection su features aggregate tramite min e rimozione features correlate											
tipo features	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
radiomiche	dt	mi	16	0,8233333333	0,0195666666	0,7441666667	0,0457229166	0,815	0,041775	0,7466269841	0,0450349978
	mlp	mi	19	0,82	0,0196	0,7091666667	0,0563395833	0,775	0,049375	0,7132936508	0,0529582782
radiomiche senza features correlate	dt	cfs	8	0,8833333333	0,0091666666	0,78	0,0486	0,8525	0,03393125	0,7952380952	0,0417460317
	mlp	mi	10	0,76	0,0544	0,75	0,0586805555	0,775	0,068125	0,7167857143	0,0745715561
Terzo passo di feature selection											
tipo features	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
radiomiche senza features correlate	mlp	mi	7	0,88	0,0256	0,755	0,0901	0,775	0,0775	0,7638888889	0,0841049382
	mlp	mi	6	0,82	0,0276	0,6966666667	0,075475	0,7375	0,06703125	0,6984126984	0,0719285399
Risultati leave one out											
tipo features	algorithm	feature alg	n_features	accuracy	precision	recall	f1score				
radiomiche senza	mlp	no_alg	7	0,8	0,897	0,854	0,875				

Allegato 10

Selected Features

TUMORE AL COLON

COINVOLGIMENTO DI LINFONODI

Fmorph_a_dens_mvee
Fmorph_v_dens_conv_hull
Fstat_cov
Fih_P10
Fivh_V90
Fcm_joint_max
Fcm_diff_var

Allegato 11

Paziente sano/malato e AD monolaterale/bilaterale con approccio 10-fold CV. Totale pazienti 94. (S=54%; ADM=28%, ADB= 18%)

Prima fase di feature selection

Descrizione	Numero	Tipo features	Algoritmo	Algoritmo	Numero	accuratezza	acc_var	precisione	prec_var	recall	rec_var	f1score	f1_var
controllo/AD mono/AD bi	3	img/cliniche	rf	no_alg	all	0,7588888889	0,01449506173	0,7343650794	0,02774326657	0,6844444444	0,02565925926	0,6782010582	0,02617210045
	3	img/cliniche	rf	gini_index	5	0,7988888889	0,00277654321	0,7484126984	0,01736016629	0,7388888889	0,01240123457	0,7152641803	0,009851890224
	3	img/cliniche	mlp	fisher	15	0,7344444444	0,00445555556	0,6415079365	0,01708718191	0,6733333333	0,00815308642	0,6293338143	0,01039498932
controllo/AD	2	img/cliniche	svm	no_alg	all	0,87	0,01449506173	0,8866666667	0,01237777778	0,8766666667	0,01310833333	0,8682864358	0,01486025126
	2	img/cliniche	dt	gini_index	20	0,8833333333	0,009734567901	0,8875	0,009670138889	0,8866666667	0,009975	0,8822727273	0,009957769615
	2	img/cliniche	dt	reliefF	15	0,8822222222	0,00817777778	0,8973809524	0,006801814059	0,8841666667	0,007964583333	0,8797763348	0,0084808225
AD mono/AD bi	2	img/cliniche	rf	no_alg	all	0,625	0,055125	0,5875	0,07585069444	0,6083333333	0,05701388889	0,5730952381	0,06489801587
	2	img/cliniche	knn	gini_index	45	0,715	0,045025	0,7166666667	0,06833333333	0,7083333333	0,05729166667	0,685	0,059025
	2	img/cliniche	knn	fisher	15	0,665	0,051525	0,65	0,07055555556	0,6583333333	0,063125	0,6383333333	0,063225
controllo/AD mono/AD bi	3	img	knn	no_alg	all	0,65	0,01845061728	0,5672222222	0,03610154321	0,62	0,02668641975	0,550012025	0,02872277706
	3	img	knn	chi	40	0,6588888889	0,01135925926	0,5680555556	0,0233599537	0,5866666667	0,02256296296	0,5470527621	0,01884824635
	3	img	svm	reliefF	35	0,6555555556	0,0135308642	0,5493650794	0,03505124717	0,5588888889	0,01972962963	0,5361568062	0,02308977498
controllo/AD	2	img	svm	no_alg	all	0,86	0,007387654321	0,8926190476	0,002963095238	0,8675	0,00600625	0,8563780664	0,008181492778
	2	img	svm	chi	10	0,8211111111	0,02262839506	0,8338690476	0,02538807044	0,8166666667	0,02695833333	0,8064466089	0,0296984036
	2	img	knn	gini_index	30	0,8211111111	0,01522098765	0,8416666667	0,01481597222	0,8116666667	0,019225	0,8040531691	0,02204648463
AD mono/AD bi	2	img	nb	no_alg	all	0,655	0,055725	0,5383333333	0,1056555556	0,5916666667	0,06590277778	0,5451190476	0,08148822279
	2	img	dt	fisher	45	0,66	0,0674	0,6208333333	0,1091840278	0,65	0,0775	0,6095238095	0,08830839002
	2	img	svm	gini_index	5	0,655	0,066225	0,5916666667	0,09194444444	0,625	0,06701388889	0,579047619	0,07721632653

Paziente sano/malato con approccio 10-fold CV. Totale pazienti: 94 (S=54%, AD=46%)

Seconda fase di feature selection

Numero classi	Tipo features	Algoritmo di classificazi	Algoritmo di feature selection	Numero di features	accuratezza	acc_var	precisione	prec_var	recall	rec_var	f1score	f1_var
2	img/cliniche	rf dt	cfs mi	9	0,893333333	0,002029629	0,912380952	0,001141496	0,894166666	0,00218125	0,890566378	0,002207370
				1	0,893333333	0,011906172	0,8975	0,01205625	0,891666666	0,012833333	0,890544733	0,012784315
2	img	rf svm/mlp	cfs cfs	6	0,851111111	0,013116049	0,862857142	0,013299263	0,8525	0,01293125	0,849187756	0,013140190
				6	0,85	0,011956790	0,857023809	0,012505513	0,8475	0,01268125	0,846319791	0,012682650

Costruzione del modello con approccio leave one out

Numero	Tipo	Algoritmo	Algoritmo	Numero di	accuratezza	precisione	recall	f1score
2	img/cliniche	dt	no_alg	1	0,893617021	0,902439024	0,860465116	0,880952381
	img	mlp	no_alg	6	0,851063829	0,853658536	0,813953488	0,833333333

Classificazione AD monolaterale/Bilaterale. Totale pazienti: 43 (ADM=60%, ADB=40%)

Seconda fase di feature selection

Numero	Tipo features	Algoritmo	Algoritmo	Numero di	accuratezza	acc_var	precisione	prec_var	recall	rec_var	f1score	f1_var
2	img/cliniche (da cini_index)	knn	lr	4	0,67	0,0731	0,65	0,1052777777	0,6833333333	0,0747222222	0,6366666667	0,0893777777
		svm	mi	9	0,675	0,049125	0,6591666667	0,0823118056	0,6583333333	0,0561805556	0,6220238095	0,0668932114
	img/cliniche (da fisher)	svm	lr	9	0,72	0,0206	0,7166666667	0,0429861111	0,7333333333	0,0233333333	0,6873809524	0,0288571995
2	img	mlp	lr	10	0,715	0,032525	0,675	0,0672916667	0,7333333333	0,0358333333	0,675	0,0509583333
		mlp	cfs	7	0,695	0,037225	0,7333333333	0,0382638889	0,7	0,0322222222	0,684047619	0,0359841836
		svm	cfs	7	0,69	0,0614	0,7083333333	0,0684027777	0,7083333333	0,0586805556	0,6816666667	0,0661361111

Costruzione del modello con approccio leave one out

Numero	Tipo features	Algoritmo	Algoritmo	Numero di	accuratezza	precisione	recall	f1score
2	img/cliniche	rf	no_alg	23	0,790697674	0,8148148148	0,8461538462	0,830188679
	img	svm	no_alg	12	0,6744186047	0,75	0,6923076923	0,72

Marker amiloide positivo/negativo. Totale pazienti=27 (P=77.78%,N=22.22%)											
Primo passo di Feature selection											
Features	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
no clinical	mlp	fisher	15	0,7777777778	0,008230452675	0,7331349206	0,02239425863	0,6785714286	0,003401360544	0,6877289377	0,007773685411
	mlp	chi	40	0,8518518519	0,002743484225	0,7546296296	0,06688314472	0,6666666667	0,01388888889	0,6791666667	0,02920138889
Marker amiloide positivo/negativo (P=77.78%,N=22.22%) - Secondo passo di Feature selection											
Tipo immagine	algorithm	feature alg	n_features	accuracy	acc_var	precision	prec_var	recall	rec_var	f1score	f1_var
no clinical	dt	lr	6	0,8148148148	0,0109739369	0,75	0,01388888889	0,8214285714	0,02295918367	0,7692307692	0,01704142012
	mlp	mi	20	0,8148148148	0,002743484225	0,7648809524	0,0148986678	0,7023809524	0,001133786848	0,719047619	0,003276643991
Costruzione del modello con approccio leave one out											
Tipo immagine	algorithm	feature alg	n_features	accuracy	precision	recall	f1score				
no clinical	dt *	no_alg	6	0,8571428571	0,9444444444	0,8947368421	0,9189189189				
	mlp	no_alg	6	0,7142857143	0,8823529412	0,7894736842	0,8333333333				
TEST											
DATI	numero pazienti	accuracy	precision	recall	f1score						
UNIME	4	1	1	1	1						

Allegato 12

Selected Features

MORBO DI ALZHEIMER		
SANO_vs_MALATO	MONO_vs_BI	PRESENZA_AMILOIDE
Left-Inf-Lat-Vent	lh_paracentral_area	lh_cuneus_thickness
Left-Amygdala	lh_cuneus_thickness	lh_transversetemporal_thickness
Right-Inf-Lat-Vent	lh_fusiform_thickness	lh_cuneus_volume
rh_entorhinal_thickness	lh_inferiorparietal_thickness	rh_bankssts_thickness
left_Lateral-nucleus	lh_inferiortemporal_thickness	rh_temporalpole_thickness
left_Basal-nucleus	lh_isthmuscingulate_thickness	rh_temporalpole_volume
	lh_lateraloccipital_thickness	
	lh_precentral_thickness	
	lh_superiorparietal_thickness	
	lh_insula_thickness	
	rh parahippocampal_area	
	rh_bankssts_thickness	

Allegato 13

Allegato 14

Morbo di Parkinson. Classificazione sano/malato. Totale pazienti=51 (S=27.45%, PD=72.55%)											
Primo passo di Feature selection											
Tipo features	Algoritmo	Algoritmo di	Numero di	accuratezza	acc_var	precisione	prec_var	recall	rec_var	f1score	f1_var
clinici/uptake/immagini	knn	chi	5	0,98	0,0036	0,94	0,0324	0,95	0,0225	0,9444444444	0,0277777777
	rf	fisher	15	0,96	0,0064	0,9275	0,03230625	0,925	0,025625	0,9206349206	0,0302343159
clinici/immagini	mlp	fisher	20	0,7133333333	0,0542666666	0,6633333333	0,0748916666	0,7083333333	0,0697916666	0,6507142857	0,0770734693
	knn	reliefF	50	0,67	0,0521	0,6666666667	0,0503472222	0,6916666667	0,0570138888	0,638452381	0,0581326672
uptake/immagini	knn/svm/rf	chi	5	0,98	0,0036	0,94	0,0324	0,95	0,0225	0,9444444444	0,0277777777
	knn	reliefF	5	0,98	0,0036	0,94	0,0324	0,95	0,0225	0,9444444444	0,0277777777
solo immagini	knn	reliefF	50	0,73	0,0361	0,7416666667	0,0311805555	0,7791666667	0,0350868055	0,7138095238	0,0362730158
	knn	reliefF	35	0,71	0,0369	0,7166666667	0,0322222222	0,7541666667	0,0380034722	0,6921428571	0,0367630952

Classificazione sano/malato Parkinson's (S=27.45%, PD=72.55%) - Secondo passo di Feature selection											
Tipo features	Algoritmo	Algoritmo di	Numero di	accuratezza	acc_var	precisione	prec_var	recall	rec_var	f1score	f1_var
clinici/uptake/immagini	knn/svm	lr	2	0,98	0,0036	0,94	0,0324	0,95	0,0225	0,9444444444	0,0277777777
	knn/svm	mi	2	0,98	0,0036	0,94	0,0324	0,95	0,0225	0,9444444444	0,0277777777
uptake/immagini	knn/svm	lr	2	0,98	0,0036	0,94	0,0324	0,95	0,0225	0,9444444444	0,0277777777
	knn/svm	mi	2	0,98	0,0036	0,94	0,0324	0,95	0,0225	0,9444444444	0,0277777777
solo immagini	knn	lr	9	0,73	0,0281	0,75	0,0236111111	0,7833333333	0,0249305555	0,7142857143	0,0295986394
	svm	lr	5	0,77	0,0401	0,6966666667	0,0802527777	0,7166666667	0,0679861111	0,6994047619	0,0746400935

Costruzione del modello con approccio leave one out							
Tipo features	Algoritmo	Algoritmo di	Numero di	accuratezza	precisione	recall	f1score
uptake/immagini	knn/svm	no_alg	2	0,9803921569	0,9736842105	1	0,9866666667
	mlp	no_alg	2	0,9215686275	0,9714285714	0,9189189189	0,9444444444
solo immagini	svm	no_alg	27	0,9215686275	0,9459459459	0,9459459459	0,9459459459
	mlp	no_alg	27	0,8431372549	0,8918918919	0,8918918919	0,8918918919

Selected Features

MORBO DI PARKINSON	
SANO_vs_MALATO	
CON DATI UPTAKE	SENZA DATI UPTAKE
R_Putamen	Left-Thalamus-Proper
L_Putamen	3rd-Ventricle
	Right-Thalamus-Proper
	Right-Accumbens-area
	lh_precentral_area
	lh_frontalpole_meancurv
	lh_cuneus_thickness
	lh_inferiorparietal_thickness
	lh_middletemporal_thickness
	lh_postcentral_thickness
	lh_precuneus_thickness
	rh_paracental_area
	rh_precentral_area
	rh_superiortemporal_area
	rh_transversetemporal_area
	rh_lingual_meancurv
	rh_parstriangularis_thickness
	rh_pericalcarine_thickness
	rh_postcentral_thickness
	rh_transversetemporal_volume
	right_Anterior-amygdaloid-area-AAA
	left_presubiculum-body
	left_CA4-head
	right_hippocampal-fissure
	Left-VA
	Left-PuL
	Right-MDm