

# Intelligenza artificiale e Neuroimaging @UNISA: Applicazioni allo studio di malattie neurodegenerative

Paola Galdi<sup>1</sup>, Angela Serra<sup>2</sup>, Fabrizio Esposito<sup>3</sup>, Roberto Tagliaferri<sup>4,5</sup>

<sup>1</sup>University of Edinburgh, <sup>2</sup>Tampere University, <sup>3</sup>Dipartimento di Medicina, Chirurgia e Odontoiatria “Scuola Medica Salernitana”, Università degli studi di Salerno, <sup>4</sup>Neurone Lab, Dipartimento di Scienze Aziendali - Management and Innovation Systems, Università degli Studi di Salerno, <sup>5</sup>Nodo AI3S (Artificial Intelligence & Intelligent Innovative Systems) - Università degli Studi di Salerno, Laboratorio Nazionale CINI di Artificial Intelligence and Intelligent Systems  
paola.galdi@gmail.com, angela.serra@gmail.com, faesposito@unisa.it, robtag@unisa.it

## Abstract

In questo contributo sono sintetizzate le attività svolte dal gruppo interdisciplinare di Intelligenza Artificiale e Computazionale e Neuroimaging presso l'Università di Salerno negli ultimi anni. Sono stati sviluppati metodi, pipeline e applicazioni di modelli computazionali per il preprocessing di dati di imaging del cervello e per la classificazione non-supervisionata e supervisionata di dati a singola e multipla vista, allo scopo di evidenziare caratteristiche o biomarker che differenzino la struttura o l'attività cerebrale in soggetti affetti da malattie neurodegenerative rispetto a controlli sani. Tutti i modelli proposti sono stati validati su dati reali e pubblicati su riviste internazionali.

## 1 Introduzione

Negli ultimi anni si è assistito allo sviluppo di molte e diversificate pipeline per l'analisi di dati di neuroimaging. Lo scopo è quello di riuscire a identificare differenti modalità di funzionamento in soggetti caratterizzati da diverse malattie o da particolari performance di task cognitivi. L'uso di tecniche di Intelligenza Artificiale e Computazionale è ormai un punto fermo come evidenziato in molte survey, come ad esempio, in [Serra *et al.*, 2018; Serra *et al.*, 2019b]. I contributi più rilevanti, frutto dell'attività di ricerca svoltasi presso l'Università degli studi di Salerno, includono:

- Tecniche automatiche per l'estrazione di caratteristiche basate su consensus clustering e aggregazione stocastica di liste ordinate, progettate per supportare la diagnostica di malattie neurodegenerative [Galdi *et al.*, 2018; Galdi *et al.*, 2019].
- Tecniche di integrazione di dati multi-vista applicate a immagini ottenute da risonanza magnetica funzionale (fMRI) e diffusion tensor imaging (DTI) per la classificazione di pazienti affetti da sclerosi laterale amiotrofica (SLA) e morbo di Parkinson (MP) [Fratello *et al.*, 2017].
- Un algoritmo per la selezione di archi da grafi cotruttivi a partire da dati fMRI e DTI, ideato allo scopo di identificare le connessioni più rilevanti in modelli di reti

neuronal per lo studio della sclerosi laterale amiotrofica [Serra *et al.*, 2019a].

## 2 Materiali e Metodi

In questa sezione sono illustrati gli algoritmi ed i dati usati nelle sperimentazioni e presenti nelle pubblicazioni.

### 2.1 Estrazione di caratteristiche da dati di risonanza magnetica funzionale

L'alta dimensionalità che caratterizza i dati fMRI richiede l'applicazione di tecniche di estrazione e selezione delle caratteristiche che siano appropriate per il problema in esame. In [Galdi *et al.*, 2018] è introdotta una metodologia basata sul consensus clustering, un particolare approccio al problema della classificazione non supervisionata che prevede la combinazione di varie partizioni di uno stesso insieme di dati per produrre una soluzione finale che sia accurata e stabile. Questa tecnica è stata applicata a mappe di attivazione estratte tramite analisi delle componenti indipendenti (ICA) da un campione di soggetti sani e pazienti affetti da SLA e MP, allo scopo di suddividere la mappa del cervello di ogni soggetto in un dato numero di regioni di interesse. Estruendo una caratteristica (ad esempio il valore medio o mediano) da ognuna delle regioni individuate è possibile comprimere la dimensione dell'input dal numero di voxel (pixel volumetrici che compongono una mappa spaziale del cervello) al numero di regioni selezionato. Uno degli obiettivi dell'applicazione di tecniche di intelligenza artificiale e computazionale a dati fMRI è quello di individuare regioni che mostrino un'attività anomala rispetto a un cervello sano, e questo obiettivo è in genere perseguito con analisi statistiche comparative che evidenzino differenze significative tra regioni di interesse. In [Galdi *et al.*, 2019] è proposto invece un approccio basato sull'aggregazione stocastica di liste ordinate per l'identificazione di regioni del cervello che esibiscano un comportamento coerente in gruppi di soggetti affetti dalla stessa patologia. Le due tecniche sono state validate su dati reali e i risultati ottenuti sono consistenti con la letteratura precedente, dimostrando che l'approccio proposto ha il potenziale per supportare il processo di diagnosi di malattie neurodegenerative.

## 2.2 Classificazione multivista basata su metodi Ensemble per dati di connettività cerebrale

La diagnosi di malattie neurodegenerative è un procedimento complesso che in genere include una serie di test clinici standard mirati a escludere patologie con un comportamento simile. La diagnosi di SLA e MP è ulteriormente complicata dall'eterogeneità delle manifestazioni cliniche di tali malattie. Per questo motivo, lo sviluppo di test diagnostici basati su apprendimento automatico è un'area di ricerca che suscita grande interesse. In [Fratello *et al.*, 2017] è stato introdotto un framework per la classificazione di dati multi-vista per la diagnosi di pazienti affetti da SLA o MP. L'apprendimento multi-vista sfrutta il contenuto informativo di diversi insiemi di caratteristiche misurate sugli stessi campioni al fine di ottenere una rappresentazione dei dati che sia più completa e che di conseguenza permetta di rivelare meccanismi complessi che potrebbero altrimenti passare inosservati [Serra *et al.*, 2019b]. Il modello proposto prevede l'estrazione di caratteristiche tramite clustering e l'addestramento di un insieme di classificatori basati su alberi di decisione (Random Forest) a partire da due tipi di dati di connettività cerebrale: mappe spaziali della default-mode-network estratte tramite ICA da dati fMRI, e mappe delle fibre nervose estratte da dati DTI (fractional anisotropy maps). Il clustering è applicato su entrambe le viste (fMRI e DTI) per ottenere una rappresentazione compressa dei dati: le mappe sono partizionate in cluster di voxel, e ogni cluster è rappresentato dalla sua mediana. I rappresentanti dei cluster diventano quindi l'input di due classificatori multi-vista: nel primo modello due classificatori distinti sono addestrati sulle due viste, e l'esito della classificazione è stabilito da una votazione a maggioranza tra tutti gli alberi di decisione; nel secondo modello le due viste vengono integrate prima dell'addestramento di un unico classificatore. Entrambi gli approcci sperimentati hanno dimostrato un potere predittivo maggiore rispetto a classificatori basati su singola vista.

## 2.3 Strong-weak pruning: un algoritmo per identificare connessioni significative in modelli di reti neuronali

Negli ultimi anni la teoria dei grafi e l'analisi delle reti complesse hanno avuto un ruolo di notevole importanza nel settore delle neuroscienze. Siccome le reti di connessioni inter-neuronali sono tipicamente completamente connesse, uno dei problemi principali è quello di individuare ed eliminare connessioni a basso contenuto informativo. Generalmente, questo problema viene risolto utilizzando dei test statistici (come il t-test) che classificano le connessioni in forti e deboli considerando solo il peso della connessione. L'algoritmo Strong-Weak pruning, invece, identifica le connessioni rilevanti non solo in base al loro peso, ma anche in base alla loro funzione strutturale a livello globale nella rete.

Questo metodo è stato utilizzato per analizzare le reti di un campione di soggetti sani e un campione di soggetti affetti da SLA e individuare una rete di connessioni rilevanti per ognuno dei due gruppi di soggetti. Successivamente è stato effettuato uno studio sulla propagazione dell'informazione nelle due categorie di reti, utilizzando modelli basati su ran-

dom walk. Tali modelli permettono di estrarre caratteristiche globali e locali delle reti, al fine di quantificare le differenze tra gruppi di reti di soggetti sani e di soggetti malati di SLA. Lo studio dei cammini minimi e delle misure di centralità su tali network ha rivelato sostanziali differenze tra reti di soggetti sani e malati e anche tra reti di pazienti malati di SLA a diversi stadi della malattia. Tali risultati sono compatibili con risultati precedenti che supportano la teoria che la SLA è una malattia multisistemica che colpisce diverse regioni tra le quali l'area sensoria e motoria, e suggerisce l'esistenza di un meccanismo compensatorio che influenza la topologia delle connessioni funzionali delle regioni motorie.

## 2.4 Discussione e Conclusioni

I metodi sviluppati presso l'Università degli Studi di Salerno hanno prodotto innovazione metodologica e hanno contribuito a individuare nuovi risultati con valenza neurobiologica. In particolare, permettono di individuare aree del cervello che differenziano soggetti sani e malati in malattie neurodegenerative che necessitano di un supporto alla diagnosi soprattutto ai primi stadi della malattia. Uno degli obiettivi futuri è l'integrazione di modelli basati su deep learning nelle pipeline e il loro confronto con i modelli basati su feature estratte con modelli computazionali o bioingegneristiche.

## Riferimenti bibliografici

- [Fratello *et al.*, 2017] Michele Fratello, Giuseppina Caiazzo, Francesca Trojsi, Antonio Russo, Gioacchino Tedeschi, Roberto Tagliaferri, e Fabrizio Esposito. Multi-view ensemble classification of brain connectivity images for neurodegeneration type discrimination. *Neuroinformatics*, 15(2):199–213, 2017.
- [Galdi *et al.*, 2018] Paola Galdi, Michele Fratello, Francesca Trojsi, Antonio Russo, Gioacchino Tedeschi, Roberto Tagliaferri, e Fabrizio Esposito. Consensus-based feature extraction in rs-fMRI data analysis. *Soft Computing*, 22(11):3785–3795, 2018.
- [Galdi *et al.*, 2019] Paola Galdi, Michele Fratello, Francesca Trojsi, Antonio Russo, Gioacchino Tedeschi, Roberto Tagliaferri, e Fabrizio Esposito. Stochastic rank aggregation for the identification of functional neuromarkers. *Neuroinformatics*, pages 1–18, 2019.
- [Serra *et al.*, 2018] Angela Serra, Paola Galdi, e Roberto Tagliaferri. Machine learning for bioinformatics and neuroimaging. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(5):e1248, 2018.
- [Serra *et al.*, 2019a] Angela Serra, Paola Galdi, Emanuele Pesce, Michele Fratello, Francesca Trojsi, Gioacchino Tedeschi, Roberto Tagliaferri, e Fabrizio Esposito. Strong-weak pruning for brain network identification in connectome-wide neuroimaging: Application to amyotrophic lateral sclerosis disease stage characterization. *International Journal of Neural Systems*, 2019.
- [Serra *et al.*, 2019b] Angela Serra, Paola Galdi, e Roberto Tagliaferri. Multiview learning in biomedical applications. In *Artificial Intelligence in the Age of Neural Networks and Brain Computing*, pages 265–280. Elsevier, 2019.