

# Azione IV.4 - Dottorati su tematiche dell'innovazione: Explainable artificial intelligence per lo studio dell'impatto dei fattori ambientali sulla salute

Donato Romano

## Sommario

Negli ultimi decenni, lo sviluppo tecnologico ha permesso l'immagazzinamento e l'elaborazione di enormi quantità di dati. Questa vasta disponibilità di risorse è stata il terreno fertile per lo sviluppo e l'applicazione di tecniche di machine learning. I calcolatori utilizzano i dati a disposizione per scoprire ed imparare eventuali relazioni, in modo tale da poter fare previsioni o classificazioni su dati non ancora osservati. Le tecniche di machine learning vengono utilizzate ormai in qualsiasi settore, dalle scienze sociali alla biologia, dalla fisica alla medicina. In generale, questo tipo di elaborazione è simile ad una "scatola nera": si conoscono i dati in input, i risultati, ma non come ci si è giunti. In medicina o in giurisprudenza questo tipo di risultati non possono essere accettati poiché sarebbero ignote le cause che hanno maggiormente influenzato la decisione del calcolatore. In quest'ottica, risulta molto efficace l'utilizzo dell'eXplainable Artificial Intelligence (XAI), ossia un insieme di processi di machine learning che permettono di comprendere le cause che hanno portato il calcolatore a restituire un determinato risultato [1]. In questo progetto, si vuole proporre uno studio che combini reti complesse e metodi XAI. Partendo da dati di soggetti sani e affetti da malattie neurodegenerative (che possono essere dati di imaging o dati genetici), vengono costruite delle reti complesse per estrarre features come il degree, la betweenness media e il clustering coefficient. Queste grandezze, insieme a dati ambientali, saranno utilizzate dagli algoritmi di machine learning supervisionati per l'assegnazione dei soggetti ad una delle classi di diagnosi e valutare quanto ognuna dei features abbia influito sulla classificazione del singolo soggetto. Queste informazioni possono essere utilizzate per studiare dei fattori di rischio personalizzati per singolo soggetto.

Questo approccio integrativo è del tutto innovativo in quanto consentirebbe di evidenziare pattern di connettività associati a stati di salute o di rischio in determinate coorti di soggetti.

## 1. Presupposti scientifici e stato dell'arte

L'intelligenza artificiale è caratterizzata da algoritmi di apprendimento automatico di difficile interpretazione. La diffusione di questa tecnologia in settori come la medicina e la giurisprudenza rende necessario conoscere le cause che hanno condotto il calcolatore a restituire un valore in output. La crescente necessità di comprendere meglio perché un modello di intelligenza artificiale genera certe decisioni, ha portato allo sviluppo di un insieme di metodi di machine learning chiamato "eXplainable Artificial Intelligence" [2]. Questa nuova frontiera dell'AI si basa su alcuni principi generali: deve essere in grado di spiegare le sue capacità e

comprensioni; spiegare cosa ha fatto e cosa accadrà dopo; rendere trasparenti le informazioni salienti che sta manipolando [2]. Essa, inoltre, permette di considerare le decisioni algoritmiche come eque ed etiche [1].

Tra i metodi più utilizzati nell'ambito della XAI troviamo gli algoritmi LIME e SHAP. Lo stato dell'arte si arricchisce sempre di più con applicazioni in ambiti diversi tra loro. In particolare, nell'ambito medico, si cerca di spiegare correlazioni tra l'interazione dell'essere umano con l'ambiente, la genetica e l'insorgenza di malattie [4][5].

L'utilizzo di una intelligenza artificiale spiegabile è reso necessario anche dal Regolamento generale sulla protezione dei dati (GDPR), in particolare il "diritto alla spiegazione" entrato in vigore in tutta l'UE il 25 maggio 2018 [3].

## 2. Scopo, metodologie e aspettative della ricerca

Da quanto esposto nella sezione precedente, risulta evidente l'utilità e la necessità dell'utilizzo dell'eXplainable Artificial Intelligence in ambito medico. In particolare, questi metodi sono molto promettenti per studiare come i fattori ambientali influiscano insieme al genoma umano sullo stato di salute delle persone e come questo possa essere correlato all'insorgenza di malattie.

Nello studio delle malattie neurodegenerative, ad esempio, le reti complesse risultano un ottimo strumento per caratterizzare la struttura delle connessioni cerebrali [8]. Da Immagini di risonanze magnetiche cerebrali si possono calcolare matrici di adiacenza che rappresentano un sistema modellizzabile con algoritmi di reti complesse [8]. Poiché malattie come l'Alzheimer e il Parkinson modificano le connessioni tra neuroni, tali cambiamenti si ripercuotono sulla topologia delle reti complesse. Il punto di partenza è l'approfondimento delle proprietà delle reti complesse, delle misure di centralità dei suoi elementi (nodi) e dei relativi legami (edges).

Negli ultimi decenni, la teoria delle reti complesse è stata applicata allo studio di sistemi con un grande numero di campi, che vanno dalla biologia alla fisica, dall'economia alla medicina e alle scienze sociali. La prima sfida delle reti complesse è stata quella di definire nuove grandezze per caratterizzare la topologia delle reti reali e la loro organizzazione. Successivi studi hanno portato alla definizione di diversi tipi di rete come quella ponderata, necessaria per descrivere collegamenti con diverse capacità trasmissive, o reti spaziali dove è necessario considerare i vincoli spaziali tra nodi e collegamenti.

Per scopi matematici le reti sono indicate come grafici, mentre il nome rete è preferito quando si presta attenzione agli esempi e alla caratterizzazione del mondo reale. Un grafo indiretto  $G$ , con  $N$  nodi e  $L$  connessioni è rappresentato dalla coppia  $(N, L)$ , dove con  $N$  si indica l'insieme dei nodi  $n_1, \dots, n_N$  del grafo  $G$ , mentre  $L$  rappresenta l'insieme dei suoi collegamenti.

Per descrivere un grafo, è necessario introdurre il concetto di adiacenza. Due nodi si dicono adiacenti se esiste un link che li connette, o possono essere connessi attraverso una sequenza alternata di nodi e link chiamata cammino.

Un grafo  $G_{N,L}$  può essere completamente descritto da una matrice di adiacenza  $A$ , una matrice  $N \times N$  il cui elemento  $a_{ij}$  è uguale a 1 se i nodi  $i$  e  $j$  sono connessi, 0 altrimenti. Per caratterizzare le proprietà topologiche delle reti reali, diverse metriche

e proprietà sono state proposte. Queste definizioni sono utili per esplorare e valutare il comportamento delle diverse reti, sia da un punto di vista locale che globale.

- Il grado  $k_i$  di un nodo è definito come il numero di link ad esso incidenti

$$k_i = \sum_{j=1}^N a_{ij}$$

- La betweenness  $b_i$  di un nodo  $i$  è dato dalla quantità

$$b_i = \sum_{k,j \in N, k \neq j} \frac{n_{jk}(i)}{n_{jk}}$$

Dove  $n_{jk}(i)$  è il numero di cammini più brevi che connettono i nodi  $j$  e  $k$  e coinvolgono il nodo  $i$ , mentre  $n_{jk}$  è il numero totale di cammini più brevi che connettono i nodi  $j$  e  $k$ .

- Il coefficiente di clustering,  $C_i$ , definisce la probabilità che due nodi, adiacenti al nodo  $i$ , siano collegati tra loro

$$C_i = \frac{\sum_{j,m} a_{ij}a_{jm}a_{mi}}{k_i(k_i - 1)}$$

Oltre allo studio di imaging, le reti complesse sono utilizzate per cercare cluster di geni, ossia geni che interagiscono tra loro.

Per le analisi delle reti, Python con i suoi numerosi pacchetti dedicati allo studio delle reti complesse (NewtorkX, Graphviz), risulta essere uno strumento tra i più adatti.

Una volta estratte le features che andranno ad arricchire i dati ambientali, si passerà alla fase di addestramento dell'algoritmo di machine learning. L'apprendimento automatico, o machine learning (ML), è una branca dell'intelligenza artificiale che si occupa di progettare algoritmi in grado di apprendere delle informazioni da un insieme di dati di partenza, detto training set, per poi effettuare dei riconoscimenti, classificazioni, prendere decisioni su dati non ancora osservati. Le reti neurali (Neural Networks), sono dei modelli computazionali di ML che ripropongono il comportamento semplificato dei neuroni del cervello. Una rete neurale è composta fondamentalmente da tre o più strati di nodi: il primo riceve i segnali in input, il secondo strato, detto nascosto, riceve le informazioni in ingresso, assegna un peso a ciascuna di esse ed effettua una somma; l'ultimo strato, riceve gli output dello strato nascosto, ne moltiplica ciascuno per un peso e ne fa la somma. Il risultato ottenuto sarà confrontato con il risultato atteso, al fine di calcolare l'errore commesso dalla rete, e correggere i pesi utilizzati dall'algoritmo di apprendimento e migliorare il risultato. Questa procedura viene ripetuta per ogni campione del training set. La rete così addestrata verrà utilizzata per classificare un insieme di dati non utilizzati in fase di addestramento, detto test set, per valutare l'accuratezza del modello.

In previsione di un'analisi multidimensionale, saranno utilizzati metodi di federate learning, modelli di apprendimento automatico che utilizzano dati distribuiti su più nodi senza la necessità dello scambio degli stessi. Lo studio utilizzerà dati raccolti da ARPA e da Copernicus Atmosphere Monitoring Service.

È nota la correlazione di alcune patologie con fattori ambientali. Ad esempio, le allergie a vari tipi di polline nell'aria sono esacerbate dall'inquinamento atmosferico e dalle condizioni meteorologiche estreme, causando diverse malattie croniche. In questo progetto verrà utilizzato il servizio di monitoraggio dell'atmosfera di Copernicus per estrarre informazioni sui rapporti di miscelazione della massa dell'aerosol e sulle specie chimiche (ad esempio monossido di carbonio, biossido di azoto, PM10). Copernicus è un programma dell'UE volto a sviluppare servizi di informazione europei basati sull'osservazione della Terra da satellite e dati in situ (non spaziali). È supportato dalla Commissione Europea (CE), dall'Agenzia Spaziale Europea (ESA) per la componente spaziale e dall'Agenzia Europea dell'Ambiente (EEA) per la componente in situ. L'obiettivo di Copernicus è monitorare e prevedere lo stato dell'ambiente a terra, in mare e in atmosfera, al fine di supportare strategie di mitigazione e adattamento ai cambiamenti climatici, la gestione efficiente delle situazioni di emergenza e il miglioramento della sicurezza di ogni cittadino. Le informazioni fornite da Copernicus hanno lo scopo di migliorare la sicurezza delle persone, ad es. fornendo informazioni sui disastri naturali come gli incendi boschivi o le inondazioni, aiutando così a prevenire la perdita di vite umane e proprietà e danni all'ambiente. Con questo progetto si vuole studiare l'impatto dell'ambiente sulla salute utilizzando i dati sull'ambiente forniti dal programma Copernicus insieme a quelli messi a disposizione dall'ARPA.

Il Programma Copernicus è servito da satelliti dedicati (le famiglie Copernicus Sentinel) e da una serie di Missioni Contribuenti aggiuntive (satelliti gestiti da varie agenzie commerciali e nazionali). Dal lancio di Sentinel-1A nel 2014, l'Unione Europea ha avviato un processo per mettere in orbita una costellazione di quasi 20 satelliti in più prima del 2030. Questi dati satellitari sono integrati e convalidati da dati in situ.

I servizi di informazione forniti da Copernicus sono disponibili su base completa, aperta e gratuita.

Estrarremo i dati locali relativi al singolo soggetto che verranno aggiunti agli altri descrittori per fornire un modello più accurato dello stato di salute di un soggetto.

Questo approccio integrativo è del tutto innovativo in quanto consentirebbe di evidenziare pattern di connettività associati a stati di salute o di rischio in determinate coorti di soggetti. La fase finale dello studio verterà sull'applicazione dell'algoritmo SHAP [6], un metodo di intelligenza artificiale interpretabile, ai risultati ottenuti con a rete neurali sui campioni del test set.

Questo metodo mira a stimare il contributo delle singole features alla predizione perturbando una data istanza e osservando l'effetto di queste perturbazioni sull'output del modello. Molti modelli di apprendimento automatico, come alberi decisionali, support vector machine e metodi basati su reti neurali vengono applicati per analizzare dati ad alta dimensionalità per molte applicazioni. Di solito una maggiore complessità consente una maggiore precisione, ma a scapito dell'interpretabilità. Saranno studiati approcci di interpretazione ML basati sulla spiegazione della scatola nera. Sistemi XAI saranno applicati agli algoritmi ML al fine di fornire un quadro chiaro delle caratteristiche rilevanti che influenzano le prestazioni dei modelli, le loro relazioni con i risultati e tra di loro e sia i loro effetti locali/globali sul problema in esame.

L'interpretabilità dei valori di rischio e delle predizioni per ciascun soggetto, si traducono nella possibilità da parte degli specialisti di comprendere le principali cause di una previsione e individuare i fattori che influiscono maggiormente sugli score di rischio clinico a livello individuale. Questo approccio si inserisce in un innovativo filone di ricerca per la definizione di un nuovo concetto di Medicina Personalizzata con l'ausilio di tecniche di intelligenza artificiale ed è particolarmente conforme alle linee guida presentate dall'Unione Europea in materia di Etica ed Intelligenza Artificiale. L'Unione Europea ha infatti nominato una Commissione di Esperti in AI che ha presentato delle linee guide in materia di "Trustworthy Artificial Intelligence" nell'Aprile 2019. Esse pongono in primo piano la necessità di introdurre metodi per rendere trasparenti le decisioni di modelli predittivi in diversi contesti applicativi e per poter introdurre l'interazione umana nelle decisioni prese da sistemi artificiali.

L'obiettivo del progetto è sviluppare nuovi metodi che integrano dati fisiologici e dati ambientali al fine di definire un valore multidimensionale robusto e sensibile, che identifichi un fattore di rischio personalizzato per ciascun soggetto.

Alla fine di questo progetto, ci si aspetta una maggiore comprensione dei fattori di rischio che influiscono sull'insorgenza di malattie neurodegenerative e alcuni tipi di cancro al livello dei singoli soggetti.

## Bibliografia

[1] Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI) - Amina Adadi, Mohammed Berrada.

[2] XAI—Explainable artificial intelligence - Gunning David; Stefik Mark; Choi Jaesik; Miller Timothy; Stumpf Simone; Yang Guang-Zhong.

[3] European Union General Data Protection Regulation (GDPR), *Jun. 2016*.

[4] Effects of environmental pollutants on gut microbiota - Yuanxiang Jin, Sisheng Wu, Zhaoyang Zeng, Zhengwei Fu.

[5] Gene-by-environment interactions in urban populations modulate risk phenotypes - Marie-Julie Favé, Fabien C. Lamaze, David Soave, Alan Hodgkinson, Héloïse Gauvin, Vanessa Bruat, Jean-Christophe Grenier, Elias Gbeha, Kimberly Skead, Audrey Smargiassi, Markey Johnson, Youssef Idaghdour & Philip Awadalla.

[6] A unified approach to interpreting model predictions, in *Advances in Neural Information Processing Systems*, Vol. 30, eds Guyon I., Luxburg U. V., Bengio S., Wallach H., Fergus R., Vishwanathan S., Garnett R. (Long Beach, CA: Curran Associates, Inc.), 4765–4774. - Lundberg S. M., Lee S.-I. (2017).

[7] Why should I trust you?": explaining the predictions of any classifier, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16* (New York, NY: Association for Computing Machinery), 1135–1144. 10.1145/2939672.2939778 - Ribeiro M. T., Singh S., Guestrin C. (2016)

[8] Nicola Amoroso, Marianna La Rocca, Alfonso Monaco, Roberto Bellotti, Sabina Tangaro - Complex networks reveal early MRI markers of Parkinson's disease.