



Simone Mastri

Dottore in Bioengineering



+39 331 106 9540



s.mastri98@gmail.com



[linkedin.com/in/simone-mastri](https://www.linkedin.com/in/simone-mastri)

Lingue

Italiano ●●●●●

Inglese ●●●●○

Esperienze all'estero

Scambio culturale

Woodrow Wilson High School
(Washington D.C) | 2015

Viaggio di studio

Bell Cambridge Language School
(Cambridge) | 2015

Soft skills

Presentazione ●●●●○

Comunicazione ●●●●○

Team working ●●●●○

Leadership ●●●●○

Gestione conflitti ●●●●○

Problem solving ●●●●○

Puntualità ●●●●○

Responsabilità ●●●●○

Hard skills

Python ●●●●○

MATLAB ●●●●○

SQL ●●●●○

R ●●●●○

Prompting ●●●●○

Microsoft 365 ●●●●○

VS Code ●●●●○

Hobby e interessi

Beach Volleyball

Rugby

Cucina

Profilo

Sono una persona seria, puntuale e precisa. Preferisco studiare/lavorare in gruppo perché credo che il confronto con gli altri porti a risultati migliori e renda l'esperienza più stimolante, specialmente quando si collabora con persone con background diversi. Durante gli studi universitari, ho affinato le mie capacità comunicative orali e scritte, anche in inglese. Avendo collaborato con professionisti del settore sanitario, ho imparato a spiegare concetti tecnici in modo chiaro anche a chi non ha una formazione specifica.

Mi considero una persona curiosa e vogliosa di imparare. Al momento, sono affascinato dal mondo della Data Analysis e dell'Artificial Intelligence.

Istruzione e formazione

Laurea Magistrale in Bioengineering - Neuroengineering and Neurotechnologies

Università di Genova

Settembre 2021 - Marzo 2024

110/110 e lode

Tesi: "*Machine Learning based approaches to identify non-obvious relationships between clinical and biological features in Chronic Lymphocytic Leukemia*".

Laurea Triennale in Ingegneria Biomedica

Università Politecnica delle Marche

Settembre 2017 - Dicembre 2020

110/110 e lode

Tesi: "*Caratterizzazione tempo-frequenza dell'attività muscolare nella malattia di Parkinson*".

Diploma di Liceo Scientifico Tradizionale

Liceo Scientifico Livio Cambi

2012 - 2017

93/100

Cambridge English FCE (B2)

2016

Grade B

Tesi magistrale - *Machine Learning based approaches to identify non-obvious relationships between clinical and biological features in Chronic Lymphocytic Leukemia.*

Ho svolto il mio lavoro di tesi in collaborazione con un *medico* del dipartimento di medicina sperimentale dell'università di Genova. Questo mi ha permesso di:

- *collaborare con professionisti* con background diverso dal mio e comprendere come interagire con loro per soddisfare le loro esigenze.
- *imparare a spiegare concetti specifici e tecnici*, come metodologie di analisi dei dati e machine learning, in modo semplice, chiaro e conciso, anche a chi non ha familiarità con questi argomenti.
- *esplorare un argomento completamente nuovo*, sia dal punto di vista biologico che metodologico, quale la leucemia linfatica cronica (CLL), acquisendo maggiore *autonomia nelle decisioni* relative alle metodologie da adottare.

Ho analizzato un *dataset* contenente *features cliniche e biologiche* relative a 32 pazienti affetti da CLL in diverse condizioni sperimentali. L'obiettivo era identificare relazioni non-ovvie tra features che individuassero gruppi di pazienti omogenei e che evidenziassero potenziali marker biologici regolati da specifiche condizioni di stimolazione. Ho implementato una pipeline di Data Analysis e Machine Learning che ha incluso i seguenti passaggi principali:

1. *data management, categorical feature encoding e data imputation.*
2. *data exploration*: data distribution visualization e principali descrittori statistici, normality assessment, non-parametric statistical difference testing, non-parametric correlation testing e outlier detection.
3. *data preprocessing*: tecniche di normalization e scaling.
4. *dimensionality reduction techniques*: PCA e t-SNE, considerando 2 o 3 dimensioni. Per t-SNE ho operato una grid-search sugli hyperparameters.
5. *clustering techniques*: k-means e k-medoids. Ho trovato k usando una euristica data driven basata sulla massimizzazione del silhouette score.
6. *analisi dei cluster trovati*: cluster quality assessment, intra-condition cluster characterization.

La metodologia che ho implementato si è dimostrata promettente, poiché, abbiamo constatato che, può, effettivamente, rilevare relazioni non ovvie tra features, sulla base delle quali è possibile stratificare i pazienti in gruppi omogenei e identificare marker biologici associati a specifiche stimolazioni. Questo lavoro ha gettato le basi per un articolo scientifico che è attualmente in fase di preparazione.

Ottimizzazione di un encoding ASP

Ho compiuto un'attenta analisi dell'output di un *encoding ASP*, utilizzato per pianificare le sessioni di fisioterapia in una struttura sanitaria, con l'obiettivo di migliorarne la *performance*. Ho, quindi, introdotto nuove rules e constraints basate su criteri di assegnazione impliciti adottati dal coordinatore della struttura. Questi sono stati individuati confrontando parametri significativi dell'assegnazione operatore-paziente nello scheduling finale, cioè ottenuto a seguito delle modifiche manuali del coordinatore, e dello scheduling iniziale, cioè prodotto tramite AI.

2023-2024

2023

2023

Simulazione dell'attività elettrica di neuroni

Ho implementato il *modello di Izhikevich* per riprodurre l'attività elettrica di regular spiking neurons e intrinsically bursting neurons. Quindi, ho valutato al variare della corrente di stimolazione: 1) il loro comportamento in condizioni under- e supra-threshold; 2) la voltage threshold necessaria alla produzione di spikes; 3) l'attività mostrata (gain). Questo è stato fatto considerando step currents e noisy step currents.

Inoltre, nel corso di Computational Neuroscience ho implementato il *Leaky Integrate-and-Fire neuron model* (e una sua variazione), il *modello di Hodgkin e Huxley* e un *modello per l'interazione sinaptica*.

2022

Implementazione di un algoritmo di Super-Burst detection

Durante l'attività di *Research Track* ho implementato l'*algoritmo di superbust detection* proposto da Wagenaar et al.

Principali steps adottati sul segnale neurale acquisito in vitro da culture di neuroni corticali: i) Spike detection; ii) Burst detection; iii) Network burst detection; iv) Subburst detection; v) Superburst detection.

Infine ho *valutato l'evoluzione temporale* dei superbusts tramite indici di similarità e return plots.

2022

Progetti-Esercitazioni del corso Artificial Intelligence in Medicine

Nel corso di *Artificial Intelligence in Medicine* ho preso parte in esercitazioni/progetti riguardanti:

- random sampling (stratified shuffle split)
 - data visualization
 - dimensionality reduction techniques (PCA e t-SNE)
 - ik-means clustering e clustering results visualization
 - model assessment (external and internal cluster evaluation metrics)
 - silhouette analysis
 - resampling stability assessment
 - multi-class classification
 - train, test and validation sets creation
 - model training (SVM) and evaluation (f1 score and confusion matrix)
-

2022

Studio dell'impatto di una lesione sulla connettività tra regioni cerebrali

Ho studiato l'impatto di una lesione indotta tra la rostral forelimb area (RFA) e la primary somatosensory area (S1). I segnali neurali acquisiti tramite elettrodi da queste regioni sono stati preprocessati usando Brainstorm. Quindi, la *connettività intra-area e inter-aree* è stata valutata in condizioni pre- e post- lesione utilizzando il *Phase Locking Value* (PLV) calcolato su bande di interesse. Infine, ho valutato la differenza statistica dei valori PLV nei casi considerati e li ho visualizzati tramite RainCloud plots. I risultati ottenuti mostrano come questa metrica possa essere utilizzata per valutare variazioni nella connettività (fenomeni di plasticità neuronale).

Progetti-Esercitazioni del corso di Neural Signal Analysis

Per il corso di *Neural Signal Analysis* ho implementato algoritmi e metriche per lo studio di segnali neurali, come:

- Algoritmi di spike detection e di burst detection
- Post Stimulus Time Histogram e Cross Correlation tra spike trains
- Phase Locking Value
- Event Related Potential in Global Field Potentials

Implementazione e confronto di algoritmi di classificazione supervisionata

Ho confrontato la *performance di vari modelli di classificazione supervisionata* (Naive Bayes - NB e Logistic Regression - LR) applicati ad un dataset di immagini (FEI dataset) utilizzando metriche come error rate, sensitivity, specificity, area under ROC. L'obiettivo della classificazione era quello di distinguere tra il volto di uomini e donne e tra espressioni facciali serie e sorridenti. A questo scopo ho preprocessato le immagini, applicato PCA per esplorare la struttura del dataset e per verificare la presenza di features importanti nella classificazione. Inoltre, ho applicato anche la Linear Discriminant Analysis (LDA) per evidenziare le features più importanti (eigenfaces) nel distinguere le espressioni e il genere. Il classificatore che ha mostrato le migliori performance è il LDA NB.

Progetti-Esercitazioni del corso Analysis of Biomedical Data and Signals

Nel corso di *Analysis of Biomedical Data and Signals* ho compiuto esercitazioni pratiche in:

- data visualization
- linear regression, correlation coefficient e coefficient of determination (variance accounted for)
- density estimation (multivariate normal distribution)
- Principal Component Analysis (Pareto diagram, Lorenz diagram), standardization, biplot e PCA loadings.
- spike sorting tramite cluster analysis: spike detection, data segmentation, PCA visualization, cluster analysis tramite K-means e tramite Gaussian Mixtures model.
- Bayes (full e naive), Linear Discriminant Analysis e Logistic classifiers. Area Under ROC.
- Hidden Markov Models
- black-box models (system identification toolbox): ARX, ARMAX e PEM.

Processing ed elaborazione di immagini biomediche

Grazie alle esercitazioni pratiche svolte per il corso di *Biomedical Imaging*, ho imparato le *tecniche base per il processing e l'elaborazione di immagini* come le operazioni matematiche tra immagini, l'equalizzazione, la regolazione e la sogliaatura dell'istogramma, il filtraggio (sia in Fourier che non), l'erosione e l'apertura, la definizione di regioni di interesse, etc. *ma anche operazioni più complesse* come la registrazione di immagini e volumi di MRI e l'algoritmo di back-propagation per la ricostruzione.

Progetti del corso Biomedical robotics

Durante il corso di *Biomedical robotics* ho avuto modo di:

- modellare e simulare la dinamica di un *motore DC* usando Simulink.
- simulare la *cinematica diretta e inversa* di un manipolatore planare.
- creare un semplice *ambiente virtuale 3D* in Simulink per la neuro-riabilitazione. Nello specifico l'ambiente ha una serie di target interattivi con cui l'utente interagisce attraverso il robot Phantom Omini. L'interazione con i bersagli può essere ostacolata o assistita attraverso dei campi di forza e/o viscosi.

2021

Progetti corso Udemy “100 Days of Code: The Complete Python Pro Bootcamp

In questo corso su Udemy ho compiuto dei piccoli *progetti in python* per consolidare quanto appreso durante le lezioni.

2021

Tesi Triennale - Caratterizzazione tempo-frequenza dell'attività muscolare nella malattia di Parkinson.

In questo lavoro di tesi ho analizzato *segnali EMG*, acquisiti da muscoli di interesse delle gambe, e *dati di stereofotogrammetria* acquisiti su pazienti affetti dalla malattia di Parkinson. In particolare, ho usato i dati stereofotogrammetrici per determinare gli istanti di tempo relativi al ciclo del passo. Da quest'ultimi ho determinato le porzioni di segnale EMG di interesse che ho poi analizzato al fine di estrarre le singole attivazioni muscolari. Ho filtrato queste porzioni, quindi ho compiuto su esse una *analisi multi-risoluzione nel dominio tempo-frequenza* utilizzando la trasformata *Wavelet*. Dallo *spettrogramma* dei segnali EMG ho implementato un algoritmo per l'*estrazione delle bande in frequenza e degli istanti temporali di inizio e fine della singola attivazione muscolare* all'interno del ciclo del passo considerato. Quindi ho confrontato questi risultati con quelli ottenuti su un gruppo di controllo per determinare possibili differenze.

2020

Aspetti tecnici

Large Language Models (“modelli di AI”) che utilizzo / con cui ho lavorato:

- Google Gemini Advanced
- ChatGPT
- GitHub Copilot

Linguaggi di programmazione:

- Python
- MATLAB
- R
- SQL
- C e C ++
- HTML e CSS
- ASP e PDDL
- Simulink

Pacchetti / librerie con cui ho lavorato:

- Python:
 - Pandas, NumPy, SciPy, Statsmodels, Scikit-learn e Scikit-learn-extra.
 - Matplotlib, Seaborn, Plotly e PtitPrince.
 - Tkinter, QtDesigner, smtplib, datetime, beautifulsoup e os.
- MATLAB:
 - Brainstorm e Field Trip (conoscenza basilare)
 - SpyCodePro (toolbox UNIGE per il processing di segnali neurali)
 - principali toolbox per signal processing, image processing, statistica e machine learning.

Softwares che utilizzo / con cui ho lavorato:

- Microsoft 365 (ex Office 365)
- Google Workspace
- Microsoft Teams
- Visual Studio Code
- Pycharm
- Anaconda
- R studio
- MongoDB
- Clingo
- Simulink
- GitHub Desktop
- Miro
- Canva