

## Lingue

Inglese	● ● ● ● ●
Italiano	● ● ● ● ●
Portoghese	● ● ● ● ●
Olandese	● ● ● ● ●

## Abilità

Python
R
Data Analysis
Data Visualization
Feature Extraction
Survival and Time-to-event Analysis
Time-Series and signal processing
Wavelets and Empirical Mode Decomposition
Monte Carlo Methods
Machine Learning
Artificial Neural Networks and Deep Learning
Explainable Artificial Intelligence
Sampling Design Analysis
Microsoft Office

## Hobbies

Recitazione

## Studi

09/2013 -  
10/2016  
10/2016 -  
09/2018  
10/2018 -  
12/2022

**Laurea Triennale – Fisica**

96/110

**Laurea Magistrale – Fisica**

Physics of Complex Systems and Big Data, 110/110 cum laude

**PhD**

Matematica Applicata

Università di Roma Tor Vergata

University Roma Tor Vergata

Universiteit Utrecht

## Ricercatore PostDoc

Da aprile 2023 a Marzo 2024 sono stato PostDoc di un anno in *Sampling Design e Data Analysis* presso il Dipartimento di Matematica dell'*Università di Genova*. In collaborazione con la *Regione Liguria* (uffici del Settore Programmazione e Statistica Finanziaria) ho sviluppato un modello per migliorare l'individuazione, il campionamento e lo studio delle popolazioni rare in Liguria (ad esempio, famiglie numerose, nuclei familiari con più di due donne di età compresa tra 15 e 55 anni insieme ad altre). Pertanto grandi set di dati come quelli EU-SILC vengono ispezionati attraverso algoritmi di selezione delle caratteristiche per il Machine Learning (ML) per determinare le classi della popolazione ligure che necessitano di essere ricampionate. Successivamente, la fase di campionamento viene realizzata attraverso un adattamento ad hoc della tecnica Multi Frame Sampling (MFS).

## Altre collaborazioni

Nel marzo 2023 ho collaborato per un breve periodo con il *Dipartimento di Scienze Economiche e Aziendali* dell'*Università di Pavia*. In questo progetto ho contribuito allo sviluppo di un modello ML per quantificare il processo di trasformazione per un insieme di ben oltre 100 aziende italiane. Nello specifico abbiamo provato a costruire un punteggio basato su come le aziende hanno cambiato la propria posizione nel mercato, come il personale di un'azienda ha cambiato la propria posizione all'interno dell'azienda stessa e come sono cambiate le vendite delle aziende nel corso degli anni 2019-2021. Questi risultati sono stati presentati lo scorso 31 marzo in occasione del Kick-off day dell'"Institute for Transformative Innovation Research" (ITIR).

## Programma di dottorato

Durante il mio dottorato ho avuto l'opportunità di migliorare le mie capacità nell'analisi, gestione ed estrazione di enormi set di dati. In particolare mi sono concentrato sull'implementazione di metodi all'avanguardia di reti neurali artificiali (ANN) per trattare dati complessi. Utilizzando gli algoritmi Explainable Artificial Intelligence (XAI) sono riuscito a rendere spiegabile e completamente interpretabile modelli di previsione basati sulle reti neurali.

In sintesi, ho contribuito allo sviluppo dei seguenti progetti:

1. Algoritmi ML e Deep Learning (DL) sono stati utilizzati per classificare gli spettri Raman del DNA genomico di cellule tumorali come nel caso del adenocarcinoma del colon e il melanoma. Questo lavoro è stato sviluppato nell'ambito del progetto DIANA (<https://www.sbai.uniroma1.it/progetti/diana/index.php>). Per andare oltre l'uso standard del DL come modello predittivo, l'algoritmo XAI (ad esempio, Vanilla Gradient) è stato utilizzato anche per studiare le caratteristiche fisico-chimiche dei campioni cellulari che il modello 1-D CNN ha catturato durante la fase di apprendimento, ad esempio, un evidente  $CH_3$  stiramento simmetrico e antisimmetrico nelle cellule maligne. Il codice Python è disponibile su GitHub: <https://github.com/glancia93/Cancer-diagnosis-via-Raman-Spect>
2. Le potenzialità delle Reti Neurali Convolutionali (CNN) sono state studiate nell'ambito di classificazioni di serie temporali associate ad eventi di fisica climatica come *El Niño* e *La Niña*. Ho analizzato principalmente le serie temporali generate come le simulazioni Zebiak-Cane (ZC). Questo lavoro mirava a spiegare l'attività di riconoscimento della CNN quando il modello viene alimentato con le simulazioni ZC; lo studio dell'attività di riconoscimento dei pattern è stato effettuato tramite algoritmi XAI. Le analisi hanno mostrato la forte capacità della CNN 1-D di catturare i processi di regolazione e feedback delle onde durante la fase di apprendimento. Codice Python disponibile su GitHub: <https://github.com/glancia93/Physics-captured-by-data-based-methods-in-El-Nino-pre> PyCODE.

3. Ho contribuito allo sviluppo di una nuova metodologia statistica in due fasi per prevedere le infezioni acquisite in terapia intensiva con un tempo di risposta di 24 o 48 ore. Questo lavoro è stato condotto in collaborazione con il Centro medico universitario di Utrecht (UMCU). Per stimare la probabilità desiderata di infezione è stato utilizzato un approccio basato sull'analisi della sopravvivenza (modello Landmark, modello Competing Risk, modello Cox). Mentre una CNN 1-D è stata progettata e addestrata per effettuare condensare come punteggio di rischio le informazioni provenienti dalle registrazioni di 1 minuto dei monitor presenti nella terapia intensiva. I punteggi restituiti dalla CNN 1-D sono stati incorporati nei dati utilizzati durante l'adattamento del modello di sopravvivenza; lo scopo era dimostrare come il DL possa fornire nuove variabili esplicative e migliorare il modo comune di estrarre informazioni in ambito di terapia intensiva. Il vasto set di dati (con una dimensione di circa 20 GB) che ho utilizzato includeva le storie cliniche di tutti i pazienti (5538 pazienti, caratteristiche cliniche e basali, registrazioni giornaliere di circolazione, parametri respiratori e di temperatura, valori di laboratorio, punteggi di coscienza e 1-registrazioni dei minuti dai monitor dell'unità di terapia intensiva). Oltre a implementare e convalidare una strategia di modellazione predittiva, ho anche pulito, organizzato e gestito il set di dati; selezionare ed estrarre nuove caratteristiche promettenti da quell'enorme quantità di dati è stato un altro passaggio fondamentale di questo lavoro. Le informazioni estratte dalla CNN 1-D sono risultate avere un potere predittivo marginale quando si considerano episodi infettivi ricorrenti (AUC = 0,72, aumento relativo dell'AUC = 2%); prestazioni migliori sono state ottenute quando sono stati inclusi solo i primi episodi di infezione (AUC = 0,75, aumento relativo dell'AUC = 8%). Come negli altri progetti, l'interpretazione della CNN 1-D è stata realizzata mediante algoritmi XAI; dall'analisi è emerso che la CNN 1-D tende a catturare caratteristiche cliniche legate all'insorgenza di una Sindrome da Risposta Infiammatoria Sistemica (SIRS). I risultati di questo lavoro sono stati presentati anche a Riga alla *31a Conferenza biometrica internazionale*. Codici Python e R sono disponibili su GitHub: <https://github.com/glancia93/ICUAI-dynamic-prediction>.

## Altre Esperienze

### Durante la Laurea Magistrale

Durante il mio Master, ho sviluppato le mie capacità nella progettazione, programmazione e risoluzione di molti problemi applicativi, ad esempio:

- Integrazione di sistemi caotici e complessi basati su equazioni differenziali ordinarie o equazioni differenziali stocastiche.
- Implementazione e simulazione di sistemi spin-glass tramite metodi Monte Carlo. Uno studio del comportamento critico (ad esempio, modello Ising 2-D con diverse condizioni al contorno e primi vicini o interazione del campo medio).
- Analisi di decomposizione in modalità wavelet ed empirica di un record di consumo elettrico di 30 anni. Esempio di analisi tempo-frequenza di un processo non stazionario. Uno studio basato sull'ispezione delle frequenze fondamentali tramite trasformata di Fourier o Hilbert.
- Simulazione della dinamica di modelli epidemiologici, quali SIS e SIR, su diversi grafici di connessione: uno studio sui comportamenti stazionari o critici a seconda del valore del tasso di infezione.
- Implementazione della BEMD (Bidimensional empirical mode decomposition) per la correzione dei fronti d'onda 2-D acquisiti durante un'osservazione della subgigante gialla Gliese777. Dati acquisiti dal Large Binocular Telescope. Lo scopo di questo lavoro mirava a presentare la BEMD come una strategia innovativa e adattiva per identificare, analizzare e rimuovere il rumore da dati non processati.

### Abilità Comunicative

- Durante la mia esperienza post-dottorato ho spesso fatto parte alle commissioni d'esame (sia scritte che orali) di corsi di laurea come *Sampling Design Analysis*, *Inferential Statistic*.
- Sono stato assistente didattico durante il mio dottorato di ricerca (Corso di laurea in Probabilità e Statistica).

- Durante il dottorato ero abituato a lavorare in team; nei regolari incontri settimanali, mi veniva spesso chiesto di raccontare i principali progressi della mia ricerca in modo semplice, diretto e convincente.
- Sono un membro attivo del Korego Theatre Group di Amsterdam. Sotto la guida di Carmelinda Gentile ho notevolmente sviluppato le mie capacità di comunicazione verbale e non verbale. Ho anche recitato in drammi classici e moderni come *Ti ho sposato per allegria*, *Il berretto a sonagli*, *Medea* e molti altri.

## Publicazioni

- "Physics captured by data-based methods in El Niño prediction"  
Chaos 32, 103115 (2022)  
<https://doi.org/10.1063/5.0101668>
- "Two-step interpretable modeling of Intensive Care Acquired Infections"  
Artificial Intelligence in Medicine (2024)  
<https://www.sciencedirect.com/science/article/pii/S0933365724001040>
- "Learning models for classifying Raman spectra of genomic DNA from tumor subtypes"  
Scientific Reports (2023)  
<https://doi.org/10.1038/s41598-023-37303-w>
- "Interpretable predictions with Convolutional Neural Networks for complex data"  
Utrecht University Repository (2023)  
<https://doi.org/10.33540/1770>