

# Applicazione di metodi di machine learning per la identificazione delle correlazioni tra le variabili biometriche di un individuo e la sintomatologia associata a SARS-CoV-2

*Metodi di machine learning per analisi dati*

WORKSHOP Dipartimento di Ingegneria – INFN Perugia

Sara Cutini, Paolo Valigi, Gabriele Costante e Alberto Garinei

11 Gennaio 2021



UNIVERSITÀ DEGLI STUDI  
DI PERUGIA



## Il Team

- Dott.ssa **Sara Cutini** (Principal Investigator): Dipartimento di Fisica e Geologia - Istituto Nazionale di Fisica Nucleare (INFN)
- Prof. **Paolo Valigi** e Dott. **Gabriele Costante**: Dipartimento di Ingegneria, Intelligent Systems, Automation and Robotics Laboratory (ISARLab) - Università degli Studi di Perugia
- Prof. **Alberto Garinei**: Dipartimento di Ingegneria della Sostenibilità - Università degli Studi "Guglielmo Marconi" Telematica

## Il Progetto

La proposta progettuale è stata presentata nel contesto della call **FISR** (Il Fondo integrativo speciale per la ricerca) 2020 ed è attualmente in fase di valutazione.

### Motivazioni:

- La modellazione delle relazioni tra i parametri clinici e biometrici di un paziente e l'insorgenza della sintomatologia legata al virus SARS-CoV-2 è determinante sia in fase di prevenzione che di terapia.
- Molti risultati scientifici hanno mostrato che esistono delle dipendenze tra l'evoluzione della malattia e le caratteristiche del paziente, a prescindere da eventuali comorbidità (es. il gruppo sanguigno<sup>1</sup>)

<sup>1</sup>*"The ABO blood group locus and a chromosome 3 gene cluster associate with SARS-CoV-2 respiratory failure in an Italian-Spanish genome-wide association analysis"* Ellinghaus et al.



# Il Progetto

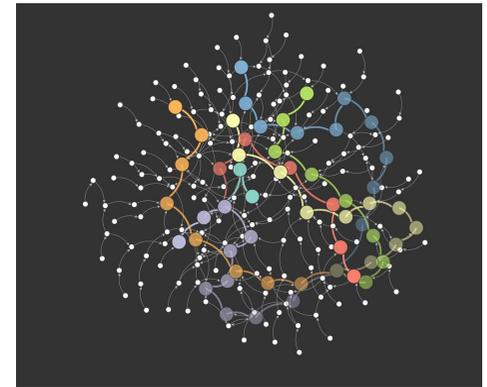
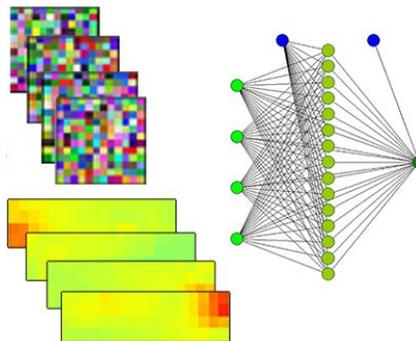
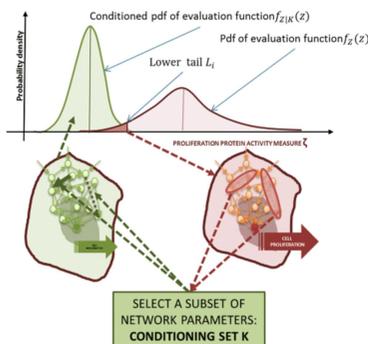
## Motivazioni:

- Tuttavia, molte delle relazioni di interesse dipendono da variabili non misurabili o non note (latenti).
- Inoltre, dati disponibili per le analisi statistiche sono un campione limitato dell'intera popolazione ed elaborare esplicitamente modelli di inferenza robusti ed efficaci può risultare molto complesso e i risultati sono spesso insoddisfacenti.

**Proposta:** Sviluppare modelli previsionali e di analisi basati su tecniche di Machine Learning per l'analisi dei dati dei pazienti positivi a SARS-CoV-2 per individuare eventuali caratteristiche fisiche-biologiche che influenzano lo sviluppo della malattia e della sintomatologia indipendentemente dalla presenza di comorbidità.

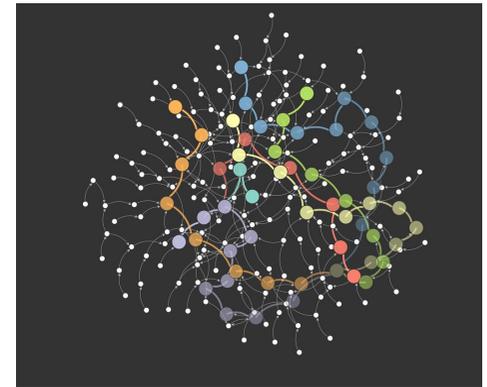
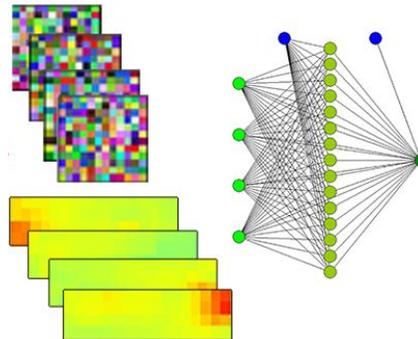
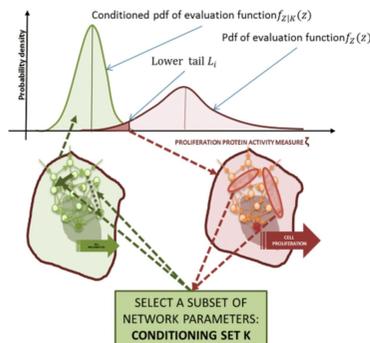
# Metodologie

- Utilizzo di approcci ibridi: model based e data-driven.
- Addestramento di modelli basati su Artificial Neural Network (ANN) per la classificazione della gravità della sintomatologia sulla base dei dati biologici e clinici.
- Utilizzo di strategie basate su Random Forest, Decision Tree ed altre tecniche per l'individuazione dei parametri del paziente più rilevanti rispetto all'insorgenza della malattia.
- Valutazione di approcci basati su Deep Learning.



# Attività Previste

- Acquisizione dei dati dei pazienti.
- Modellazione delle incertezze sulle informazioni acquisite.
- Costruzione del Database e dell'infrastruttura software necessaria.
- Sviluppo e valutazione delle strategie basate sugli algoritmi di Machine Learning.



# Applicazione di metodi di machine learning per la identificazione delle correlazioni tra le variabili biometriche di un individuo e la sintomatologia associata a SARS-CoV-2

*Metodi di machine learning per analisi dati*

WORKSHOP Dipartimento di Ingegneria – INFN Perugia

Sara Cutini, Paolo Valigi, Gabriele Costante e Alberto Garinei

11 Gennaio 2021



UNIVERSITÀ DEGLI STUDI  
DI PERUGIA