

# Dallo sviluppo al deployment di sistemi di IA: Il caso studio di uno strumento a supporto della diagnosi clinica dell'Alzheimer

Marco Moscatelli e Fabio Massimo D'Amore

## Chi sono

- Più di 20 anni di esperienza nel trasporto aereo (IT project manager, Revenue Management per Cargo)
- Ricercatore in ambito IA su tematiche health care in collaborazione ISTC/CNR, IRCCS S. Lucia, La Sapienza
- Docente AS-AI



<https://www.linkedin.com/in/fabio-massimo-d-amore-83859b133/>

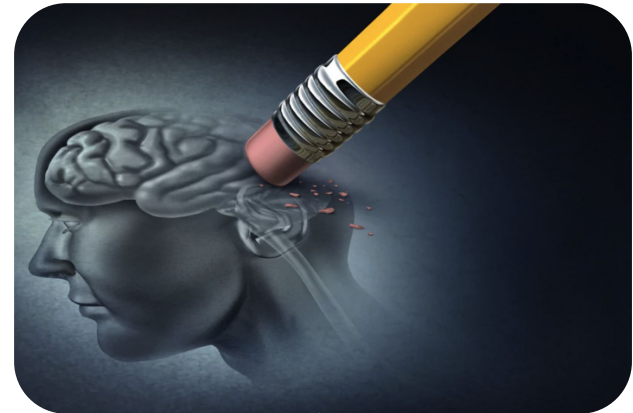
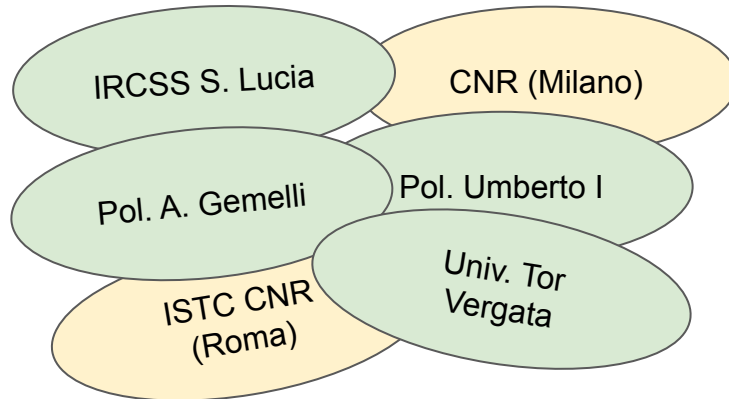
Fabio M. D'Amore

# Agenda

- Introduzione al problema
- Raccolta dati e costruzione Dataset
- La pipeline di Machine Learning
- Modello finale e spiegabilità
- Risultati e conclusioni

## Introduzione - Contesto Clinico

- Alzheimer: principale malattia neurodegenerativa
- Progressione graduale da MCI (Mild Cognitive Impairment)
- Solo 20–40% degli MCI converte in AD (Alzheimer Disease)
- Diagnosi precoce → fondamentale per trattamento



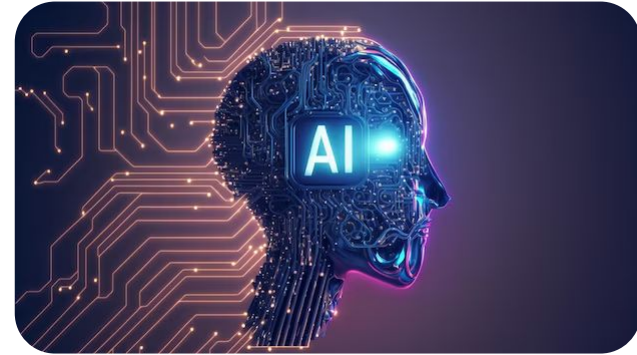
# Introduzione - Gli approcci tradizionali

- Dipendenza da:
  - Imaging (PET, MRI)
  - Biomarcatori invasivi (prelievo del liquido cerebrospinale)
- Costosi, poco scalabili
- Limitata applicabilità clinica reale



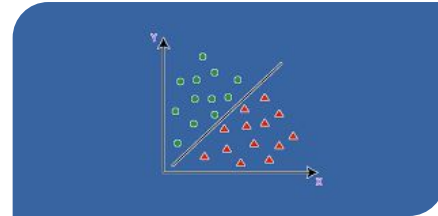
# Introduzione - Obiettivo del progetto

- Predire conversione da MCI  $\rightarrow$  AD entro 18 mesi
- Usando solo:
  - dati clinici non invasivi
  - test neuropsicologici
  - dati demografici
- Costruire modello:
  - robusto
  - interpretabile
  - generalizzabile multi-centro
  - differenze di genere



# Introduzione - Formulazione del problema

- Dati: test visite neuropsicologiche periodiche
- Task: classificazione binaria
- Target:
  - Classe 0  $\rightarrow$  conversione precoce ( $< 1.6$  anni)
  - Classe 1  $\rightarrow$  conversione tardiva o assente
- Setting: supervised learning



## Dataset - Le fonti

- Centri clinici:
  - Santa Lucia (33 pazienti)
  - P. Umberto I (86 pazienti)
  - Gemelli (97 pazienti)
- Dataset longitudinale
- 685 osservazioni etichettate

### Criteri di arruolamento (MCI amnestico)

- Disturbi soggettivi di memoria (paziente + caregiver)
- Deficit oggettivo di memoria episodica ( $\geq 1$  test patologico)
- Cognizione globale preservata ( $MMSE > 23.8$ )
- Funzionamento quotidiano preservato ( $CDR \leq 0.5$ )
- Imaging cerebrale negativo per lesioni rilevanti (TC/RM)

Etica

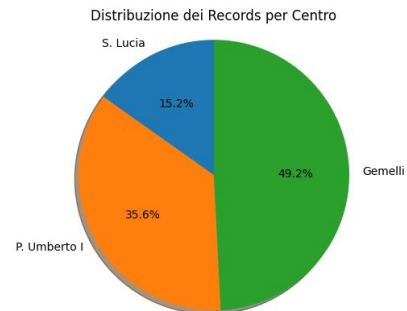
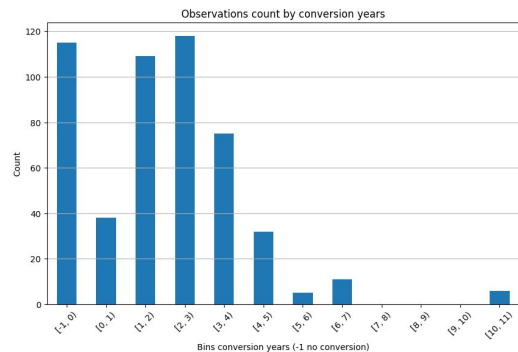
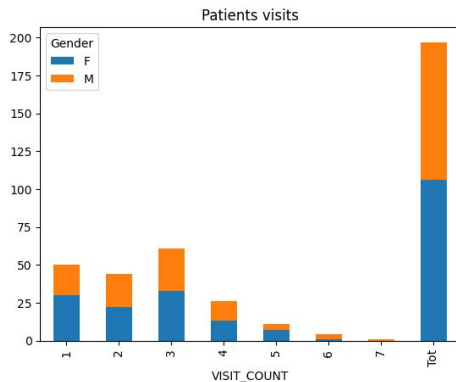




## Dataset - Le features

- Test cognitivi:
  - memoria
  - linguaggio
  - attenzione
- Variabili demografiche
- Misure longitudinali

	S. Lucia	P. Umberto I	Gemelli
features	28	36	29
records	104	244	337
n. medio visite p.p.	3,1	2,8	3,4



## Dataset - Le Sfide

Dataset di modeste dimensioni

Valori mancanti ( $> 40\%$ )

Data entry manuale

Eterogeneità tra centri

## Pipeline - Approccio

- Pipeline:
  - ripetibilità degli esperimenti
  - clinician-constrained
- Coinvolgimento esperti:
  - selezione feature
  - validazione risultati
- Iterazioni continue ML ↔ clinici

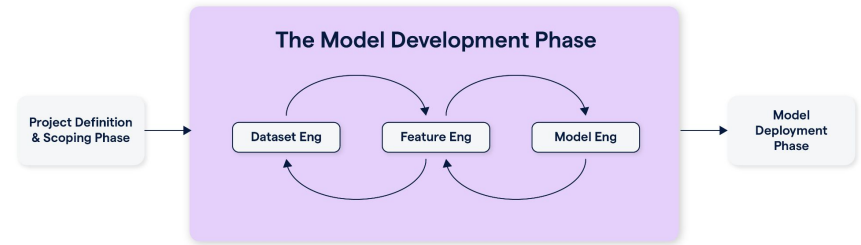
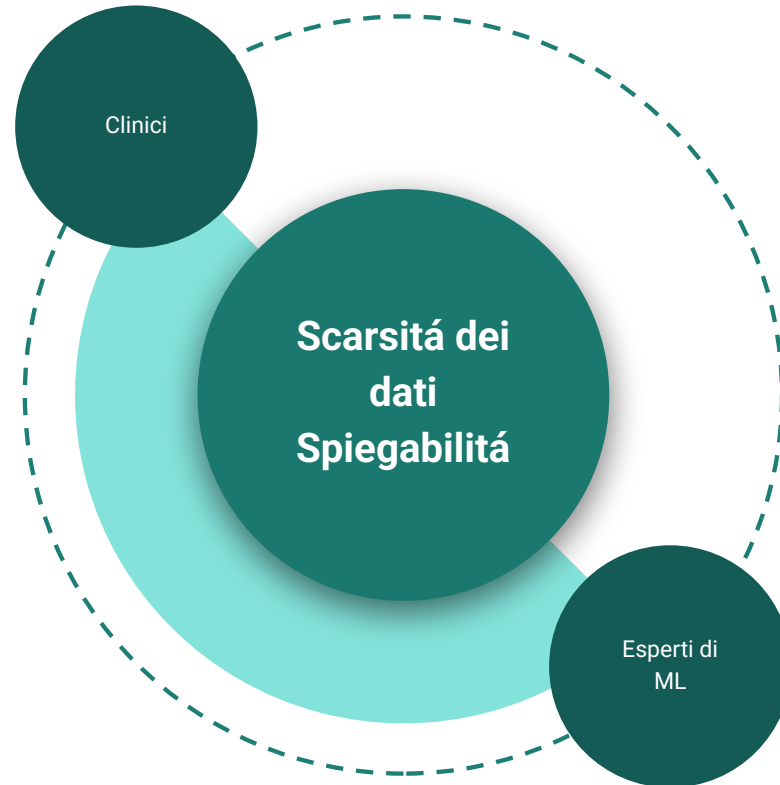
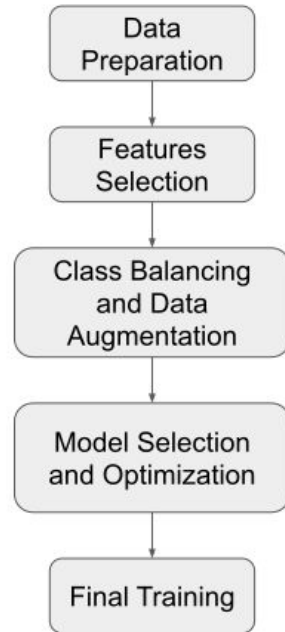


Immagine: <https://www.featureform.com/post/feature-engineering-guide>

## Pipeline - Overview



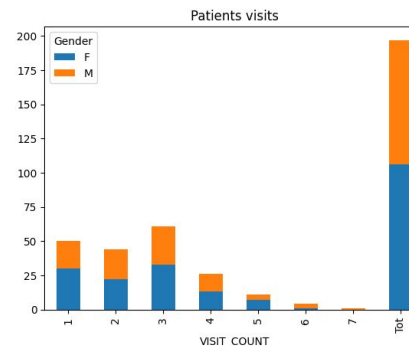
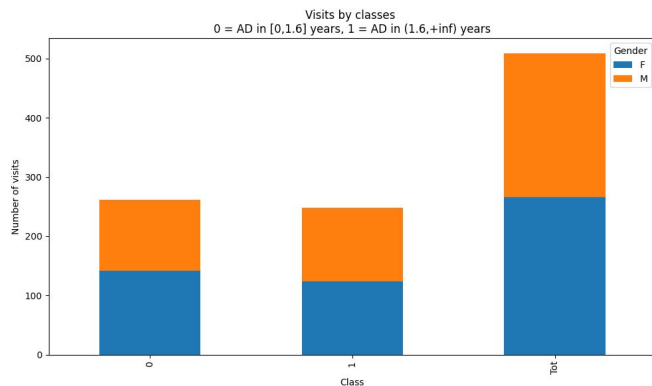
## Pipeline - Data preparation

### Operazioni preliminari

- Armonizzazione dataset multi-centro
  - standardizzazione nomi e formati
- Validazione valori clinici
  - verifica dei range
- Rimozione feature poco informative

### Filtering dei dati

- Rimozione dei dati non etichettabili
  - ultime visite di pazienti non convertiti
- Recupero per training test
  - decision tree ad alta confidenza ( $p > 0.9$ )



## Pipeline - Features engineering

### Feature Selection

- Approccio ibrido:
  - clinico + data-driven
- Metodi:
  - Automatico - Recursive Feature Elimination Cross Validation (Random Forest)
  - Manuale - mediante Features importance clinica

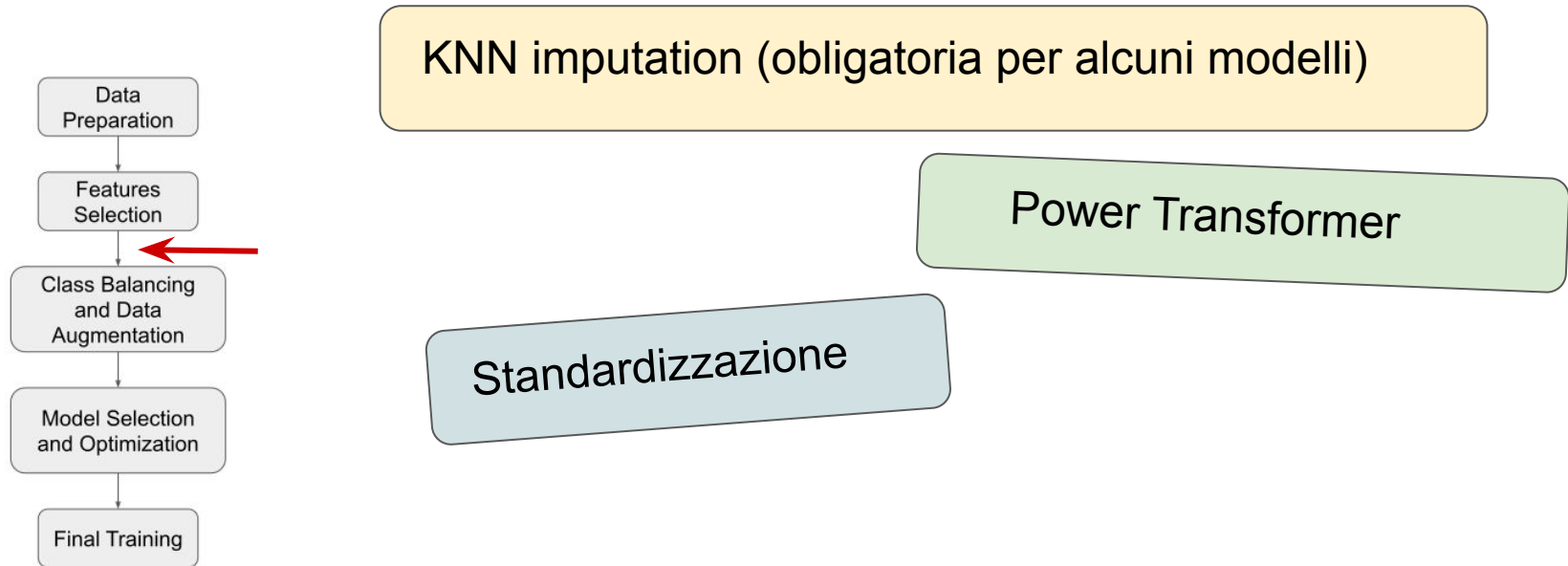
Obiettivo: riduzione overfitting + interpretabilità

### Feature Engineering

- Introduzione di delta features (longitudinali)
  - $\Delta$  tra visite successive /  $\Delta$  tempo
- Selezione basata su correlazione con target

Obiettivo: cattura della dinamica della malattia

## Pipeline - step opzionali



## Pipeline - Augmentation e bilanciamento delle classi

Focalizzato sui dati di training (fold nel caso di cross validation)

- leggero sbilanciamento medio delle classi (0,52% classe 1)
- augmentation mediante Gaussian Copula con realismo spinto ad una similarità  $> 89,8\%$

Con una numerosità media di circa 350 campioni di training vengono generati circa 200 campioni sintetici.



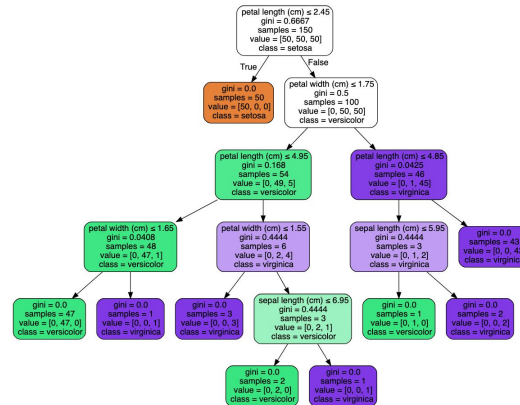
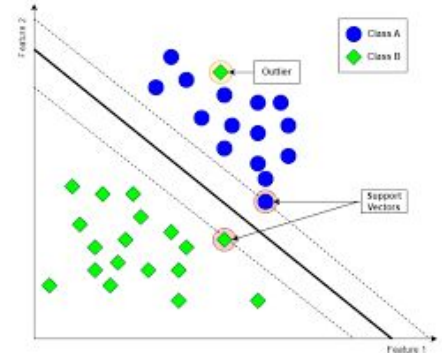
## Pipeline - Modelli

Sono stati selezionati i seguenti modelli

- SVC
- Random Forest
- LightGBM
- XGBoost

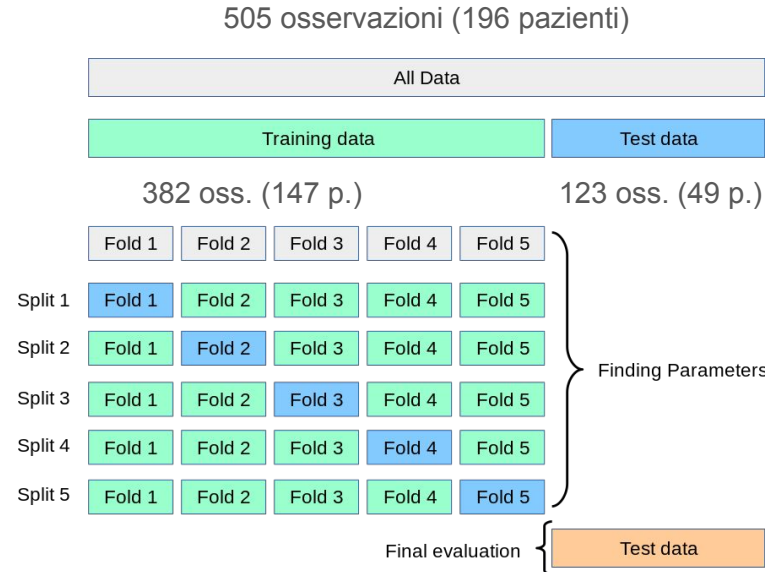
Criteri:

- robustezza su small data
- interpretabilità



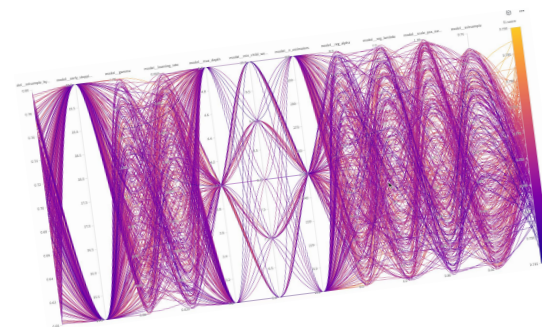
## Pipeline - cross validation

Recupero delle osservazioni non classificate, normalizzazione, bilanciamento delle classi e data augmentation su **fold di training generati a run time**.



## Pipeline - Ottimizzazione

- Ottimizzazione bayesiana ricorsiva
  - supervisione e modifica manuale dei range
- K-fold cross-validation
- Ottimizzazione:
  - iperparametri
  - componenti pipeline



Config parameter	Importance	Correlation
model__scale_pos_weight	High	Low
model__learning_rate	High	Low
model__reg_alpha	High	High
model__min_child_weight	High	High
model__colsample_bytree	High	High
model__gamma	High	High
model__subsample	High	High
model__reg_lambda	High	High
Runtime	Low	Low
model__n_estimators	High	Low



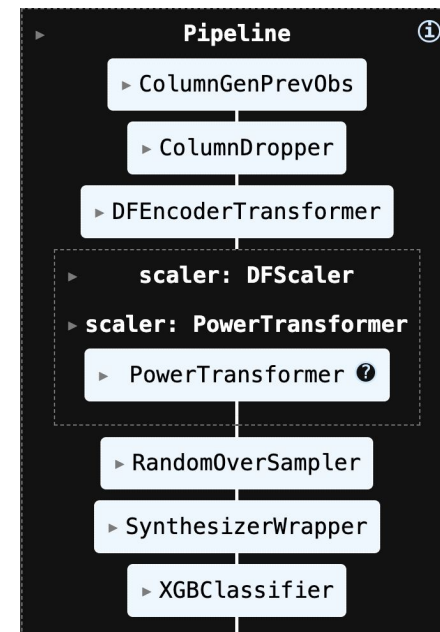
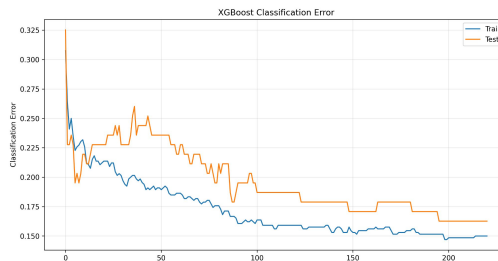
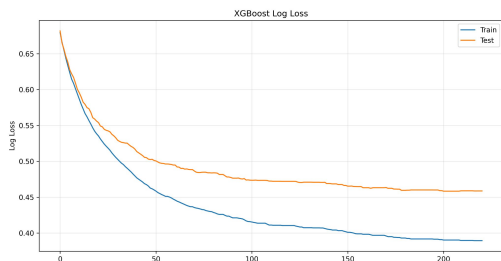
W&B piattaforma di supporto che implementa il tuning automatico degli iperparametri.

# Pipeline - Infrastruttura

- Script Python
- Google Colab
  - Uso di GPU per i modelli che la supportano
- Stima del tempo di ottimizzazione
  - 1 run di ottimizzazione (valutazione di 200 differenti set di parametri, ognuno su 4 fold) impiega circa 1h
  - ogni modello sottoposto a circa 30 run di ottimizzazione
  - totale di circa 120h dedicate all'ottimizzazione della pipeline
  - lo spazio degli iperparametri varia da modello a modello

## Modello finale - Pipeline di machine learning

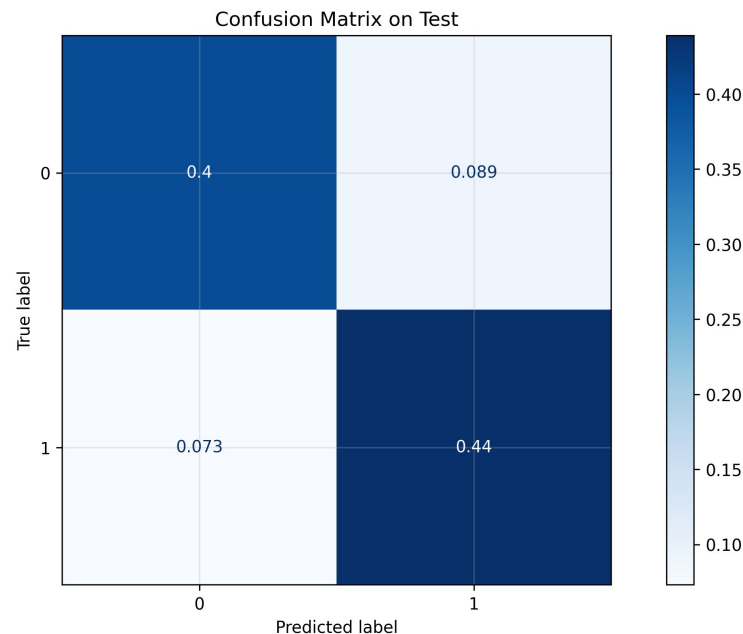
Model	Female&Male		Male		Female	
	f1-score	accuracy	f1-score	accuracy	f1-score	accuracy
SVC	0.724	0.724	0.712	0.713	0.737	0.737
RandomForest	0.797	0.797	0.848	0.848	0.720	0.719
LGBM	0.780	0.780	0.818	0.818	0.737	0.737
<b>XGBoost</b>	<b>0.813</b>	<b>0.813</b>	<b>0.863</b>	<b>0.864</b>	<b>0.755</b>	<b>0.754</b>



## Risultati - performance

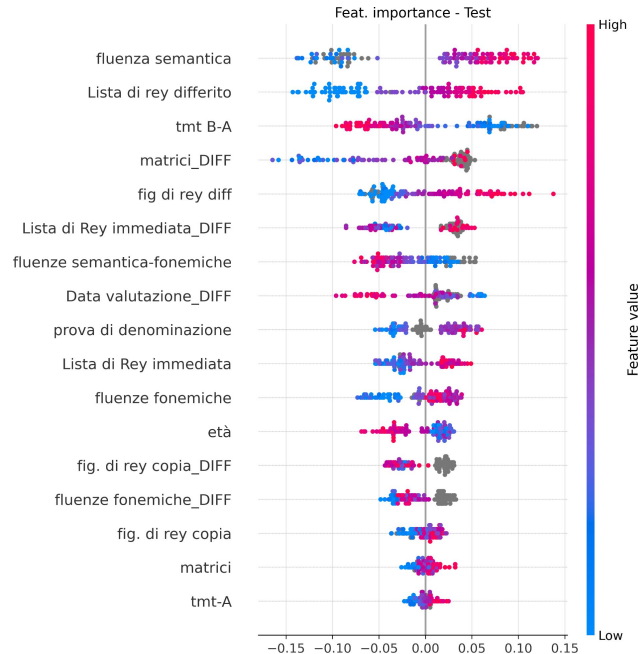
### Report di Classificazione

Class	Precision	Recall	F1-score	Support
0	0.80	0.82	0.81	60
1	0.82	0.81	0.82	63
Accuracy			0.81	123
Macro avg	0.81	0.81	0.81	123
Weighted avg	0.81	0.81	0.81	123



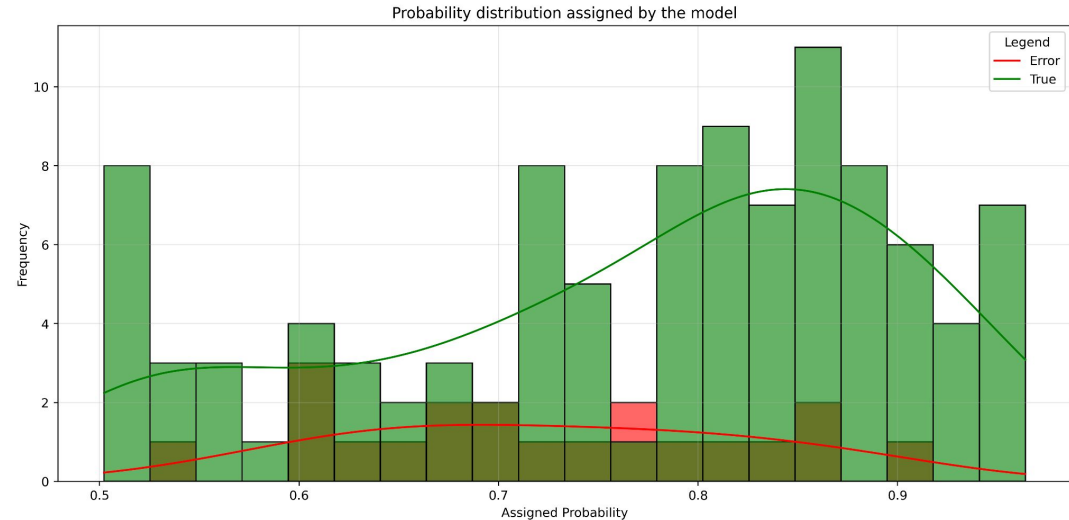
Conf. matrix normalizzata

## Risultati - spiegabilità



Le spiegazioni SHAP sono state integrate nel processo di validazione e utilizzate attivamente dai clinici durante cicli iterativi di feedback.

## Risultati - analisi degli output



- Output informativi
- Supporto alle decisioni cliniche



## Conclusioni - Takeaways

- Pipeline ML completa e modulare
- Uso di dati non invasivi
- Forte integrazione clinica
- Explainability centrale

## Conclusioni - Sviluppi futuri

- Dataset limitato
- Necessità di:
  - più dati
  - validazioni esterne
- Possibile estensione:
  - modelli longitudinali più avanzati