

CORRIERE DELLA SERA



UNIVERSITÀ
DI SIENA
1240



CINECA

DISCLAIMER

ULTIME AVVERTENZE PRIMA DELLA RIVOLUZIONE

BIOMIND

INTELLIGENZA ARTIFICIALE PER L'INNOVAZIONE
IN FARMACIA E BIOINFORMATICA

KNOWLEDGE PARTNER



INTRODUZIONE

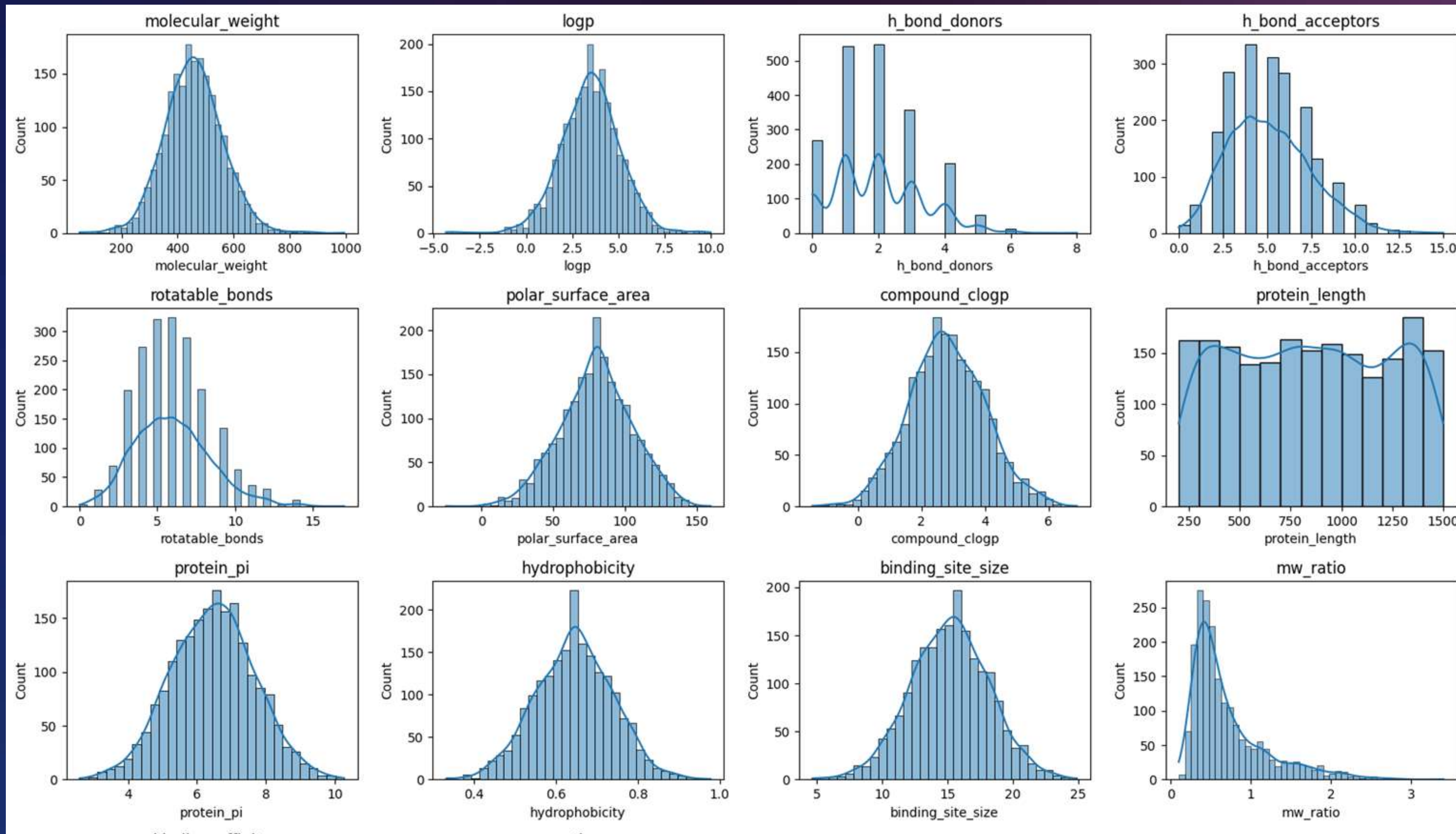
Immaginiamo di avere una proteina specifica, ma di non sapere ancora se e come i composti chimici della nostra libreria interagiscono con essa.

Per affrontare questa sfida utilizziamo l'intelligenza artificiale:

- **Fase di apprendimento:** il sistema viene addestrato su esempi in cui gli esiti delle interazioni sono già noti.
- **Fase di verifica:** lo mettiamo poi alla prova con nuovi esempi mai visti prima, per verificare se ha davvero compreso i modelli e non sta solo ripetendo ciò che ha memorizzato.
- **Applicazione:** se il test viene superato, possiamo affidarci alle sue previsioni per scoprire possibili interazioni tra la proteina e i composti non ancora conosciuti.

INFORMAZIONI SUL DATASET

Per ogni proteina del
dataset sono fornite
12 caratteristiche



PREPROCESSING DATI

I dati a disposizione vengono divisi in due set:

- Training set → 80%
- Validation set → 20%

Successivamente abbiamo utilizzato MinMaxScaler per scalare i valori di tutte le features nell'intervallo $[0,1]$

NOTA: dopo lo scaling le distribuzioni delle features rimangono invariate

ARCHITETTURA

MLP → feed-forward neural network

- Configurazioni testate:

- 1 input layer con 12 neuroni, 2 hidden layers composti da 100 neuroni e 1 neurone di output
- 1 input layer con 12 neuroni, 2 hidden layers composti da 50 neuroni e 1 neurone di output
- 1 input layer con 12 neuroni, 3 hidden layers composti da 100 neuroni e 1 neurone di output
- 1 input layer con 12 neuroni, 3 hidden layers composti da 50 neuroni e 1 neurone di output



COME FUNZIONA UN MLP

Processo di addestramento:

- fase forward → calcolo dell'output della rete neurale
- calcolo dell'errore utilizzando la Binary Cross-Entropy come funzione di loss
- aggiornamento dei pesi della rete sulla base dei valori della funzione di loss tramite il processo di Backpropagation

VANTAGGI E LIMITI DELL'MLP

✓ Vantaggi:

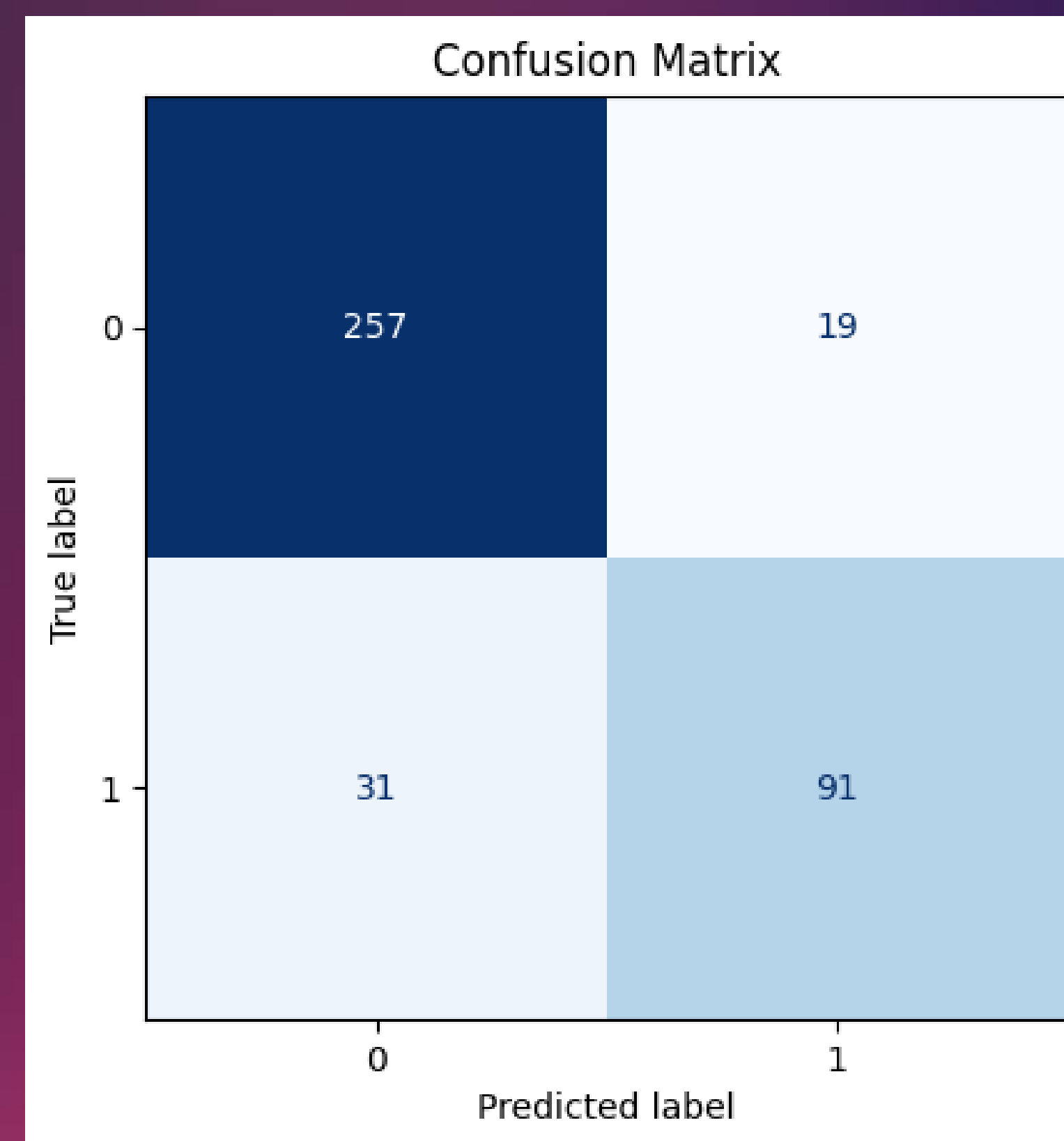
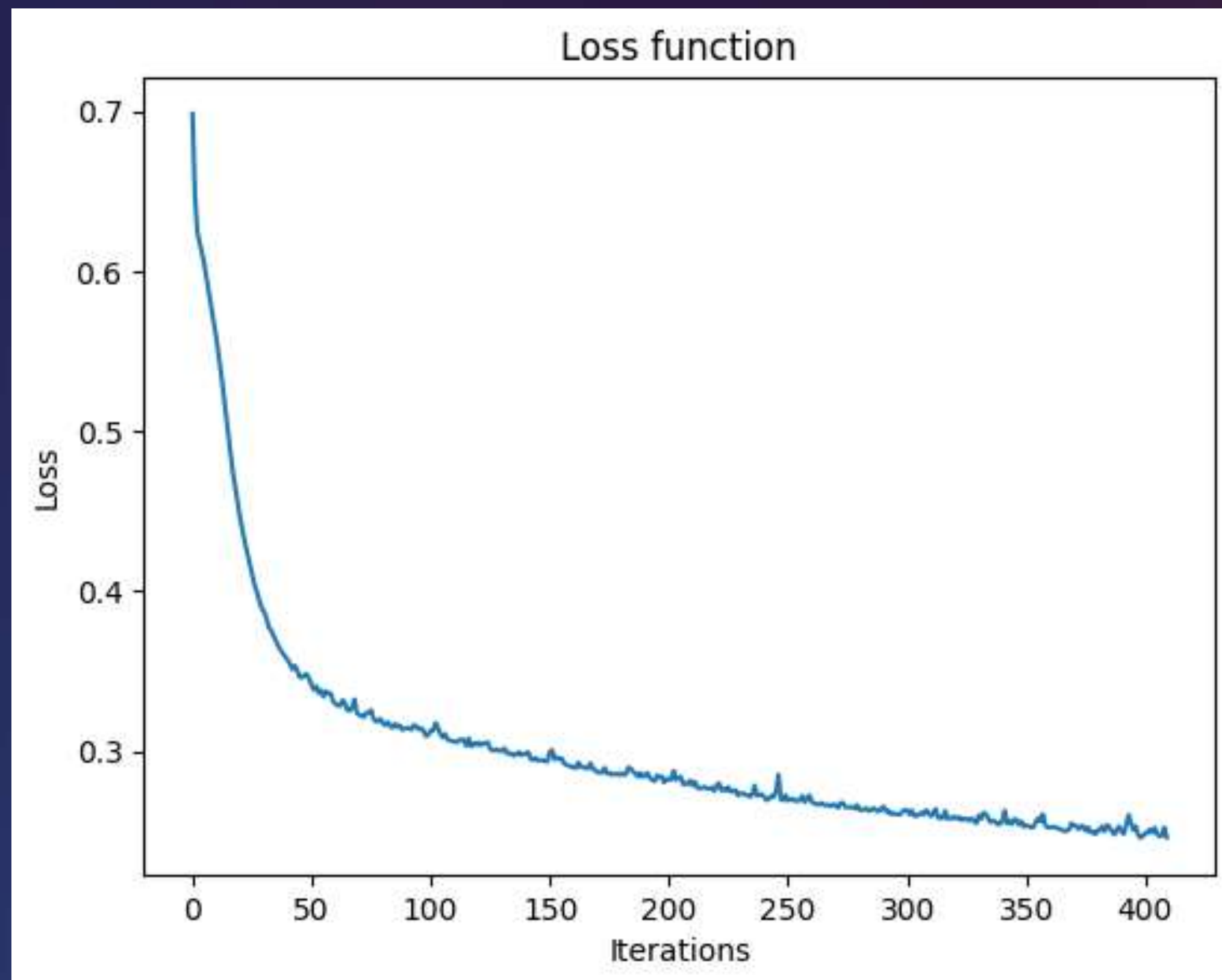
- Capace di modellare relazioni **non lineari complesse**
- Grande **flessibilità**: può adattarsi a diversi tipi di dati (classificazione, regressione, ecc.)
- **Generalizzazione** buona se addestrato correttamente
- Può approssimare qualunque funzione continua (teorema dell'approssimazione universale)

✗ Limiti:

- Richiede un **elevato numero di dati** per addestrarsi efficacemente
- **Sensibile** a overfitting e parametri (learning rate, numero di neuroni, ecc.)
- Addestramento spesso **costoso in termini computazionali**
- Meno interpretabile rispetto a modelli più semplici (scatola nera)

RISULTATI #1

- Dopo aver testato il modello su più configurazioni risulta che il modello migliore è con 2 hidden layer da 50 neuroni ciascuno. Di seguito si riportano i grafici:



RISULTATI #2

- Le metriche del modello:

Accuracy: 0.87437

F1 Score: 0.78448

ROC AUC Score: 0.83853

Classification Report:

	precision	recall	f1-score	support
0	0.89	0.93	0.91	276
1	0.83	0.75	0.78	122

CORRIERE DELLA SERA



UNIVERSITÀ
DI SIENA
1240



CINECA

GRAZIE DELL'ATTENZIONE!

DISCLAIMER

ULTIME AVVERTENZE PRIMA DELLA RIVOLUZIONE

BIOMIND

INTELLIGENZA ARTIFICIALE PER L'INNOVAZIONE
IN FARMACIA E BIOINFORMATICA

KNOWLEDGE PARTNER

