


Protein Classification Via Graph Neural Networks

Candidato:
Riccardo Raciti

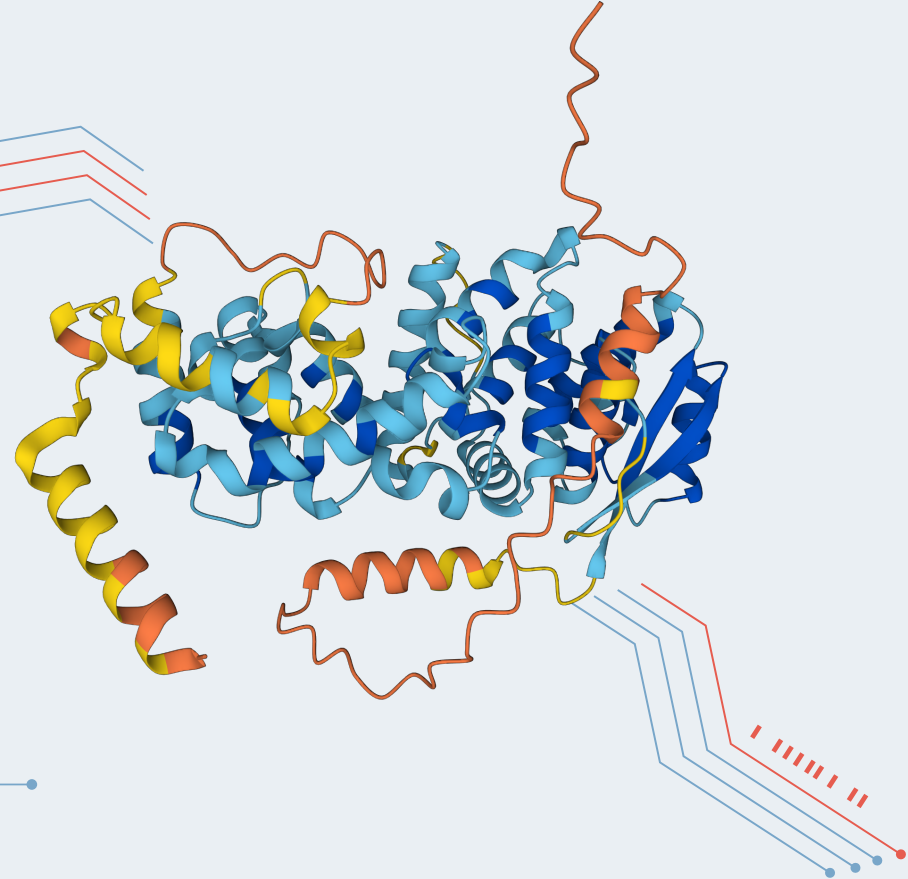
Relatore:
Giovanni Micale





Introduzione

Lo scopo di questa tesi è quello di classificare le proteine basandosi sulla loro morfologia interna, quindi dalla loro sequenza di aminoacidi.



BACKGROUND



Aminoacidi

Sono una vasta categoria di molecole organiche che hanno sia il gruppo funzionale amminico sia quello carbossilico



Catene di aminoacidi

Le catene di aminoacidi sono indicate col termine di peptidi e possono contenere un numero molto variabile di unità



Proteine

Le proteine sono costituite da da diversi aminoacidi collegati tra loro, tipicamente si parla di 300 o più aminoacidi



Grafi

I Grafi sono modelli matematici che permettono di codificare le interazioni tra le componenti di un sistema complesso



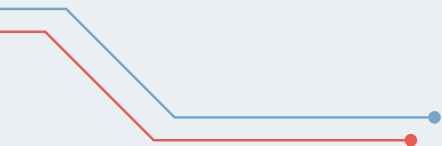
Reti Neurali

Sono un modello di comunicazione composto da neuroni artificiali, utilizzati per la risoluzione di problemi di intelligenza artificiale

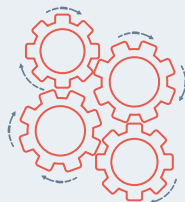


GNN

Una classe di reti neurali che processa dati rappresentabili sotto forma di grafi

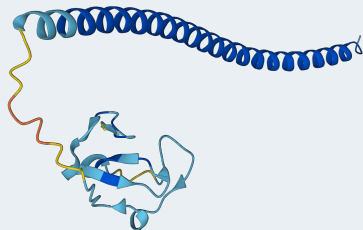


Obiettivo



Classificazione

L'obiettivo è quello di classificare le proteine basandosi sulla loro forma e valutare l'accuratezza del modello creato



FASI

Classificazione Proteine

È stato scaricato il file
contenente tutte le proteine e la
loro relativa classificazione

Creazione Grafi

Sono stati adattati le
strutture 3D delle proteine
in grafi 2D

Classificazione

Utilizzo del modello per la
classificazione delle
proteine

Ottenimento Proteine

Sono state scaricate le
informazioni inerenti la
morfologia delle proteine

Creazione modello

Creazione modello per la
classificazione, tramite
GNN

Valutazione

Valutazione della bontà del
modello di classificazione



Modelli creati

- Modello per la classificazione di un dataset binario
- Modello per la classificazione di un dataset quaternario
- Modello per la classificazione di un dataset senario
- Modello per la classificazione di un dataset ottonario

Per ogni modello sono stati variati dei parametri interni per valutarne il cambiamento di accuratezza

ACCURATEZZA

Numero di nodi	Tipi di Dataset			
	Binario	Quaternario	Senario	Ottinario
12	0.87	0.55	0.53	0.36
32	0.87	0.62	0.43	0.44
128	0.87	0.56	0.54	0.44
256	0.90	0.24	0.54	0.36





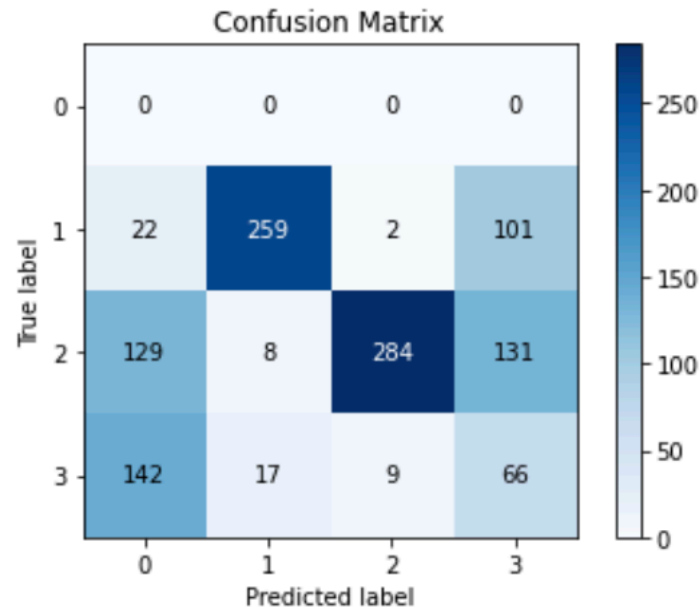
One Vs All

Questo approccio permette di trasformare il problema di classificazione multiclasse in un insieme di problemi di classificazione binario.



Risultati

	Precision	Recall	F1-score	Support
0	0.00	0.00	0.00	0
1	0.91	0.67	0.78	384
2	0.96	0.51	0.67	552
3	0.22	0.28	0.25	234
Accuracy			0.52	1170
Macro avg	0.52	0.37	0.42	1170
Weighted avg	0.80	0.52	0.62	1170





GRAZIE PER L'ATTENZIONE

