



Enhanced *Gene fusion* detection through Graph learning-based strategies

Relatore:

Prof. Rocco Zaccagnino

Candidato:

Gisolfi Andrea

Matr. 0512114162



Il Problema





Il Problema



Un gene di fusione, chiamato anche **gene chimerico** o **gene ibrido**, è la giustapposizione di due geni altrimenti separati.





Il Problema



Un gene di fusione, chiamato anche **gene chimerico** o **gene ibrido**, è la giustapposizione di due geni altrimenti separati.

Questo può avvenire attraverso:





Il Problema



Un gene di fusione, chiamato anche **gene chimerico** o **gene ibrido**, è la giustapposizione di due geni altrimenti separati.

Questo può avvenire attraverso:



Traslocazione





Il Problema



Un gene di fusione, chiamato anche **gene chimerico** o **gene ibrido**, è la giustapposizione di due geni altrimenti separati.

Questo può avvenire attraverso:



Traslocazione



Inversione





Il Problema



Un gene di fusione, chiamato anche **gene chimerico** o **gene ibrido**, è la giustapposizione di due geni altrimenti separati.

Questo può avvenire attraverso:



Traslocazione



Inversione



Duplicazione in tandem





Il Problema

Un gene di fusione, chiamato anche **gene chimerico** o **gene ibrido**, è la giustapposizione di due geni altrimenti separati.

Questo può avvenire attraverso:



Traslocazione



Inversione



Duplicazione in tandem

Possibili implicazioni nel **cancro**





Motivazioni e Obiettivi



Il rilevamento tradizionale utilizza
strumenti di allineamento

...





Motivazioni e Obiettivi



Il rilevamento tradizionale utilizza
strumenti di allineamento

Limiti:

...





Motivazioni e Obiettivi



Il rilevamento tradizionale utilizza
strumenti di allineamento

Limiti:



Bias

...





Motivazioni e Obiettivi

Il rilevamento tradizionale utilizza
strumenti di allineamento

Limiti:



Bias



Falsi positivi

...





Motivazioni e Obiettivi

Il rilevamento tradizionale utilizza
strumenti di allineamento

Limiti:



Bias



Falsi positivi



Carico computazionale

...





Motivazioni e Obiettivi

Il rilevamento tradizionale utilizza
strumenti di allineamento

Limiti:



Bias



Falsi positivi



Carico computazionale

Nuova soluzione basta sulle **Graph Neural
Network** che sia:





Motivazioni e Obiettivi

Il rilevamento tradizionale utilizza
strumenti di allineamento

Limiti:



Bias



Falsi positivi



Carico computazionale

Nuova soluzione basta sulle **Graph Neural
Network** che sia:



Affidabile





Motivazioni e Obiettivi

Il rilevamento tradizionale utilizza
strumenti di allineamento

Limiti:

 Bias

 Falsi positivi

 Carico computazionale

Nuova soluzione basata sulle **Graph Neural
Network** che sia:

 Affidabile

 Innovativa





Motivazioni e Obiettivi

Il rilevamento tradizionale utilizza
strumenti di allineamento

Limiti:



Bias



Falsi positivi



Carico computazionale

Nuova soluzione basta sulle **Graph Neural Network** che sia:



Affidabile

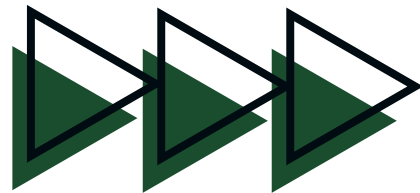


Innovativa



Robusta

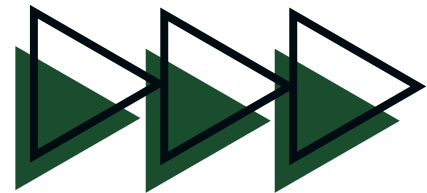




Background metodologico

Grafo di De Bruijn





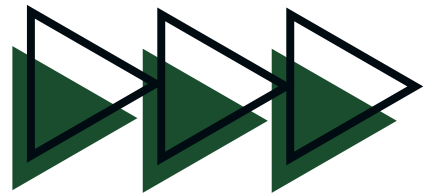
Background metodologico



Grafo di De Bruijn

Un **grafo di De Bruijn** è una struttura utilizzata per rappresentare sequenze genetiche o altri dati lineari in modo compatto.





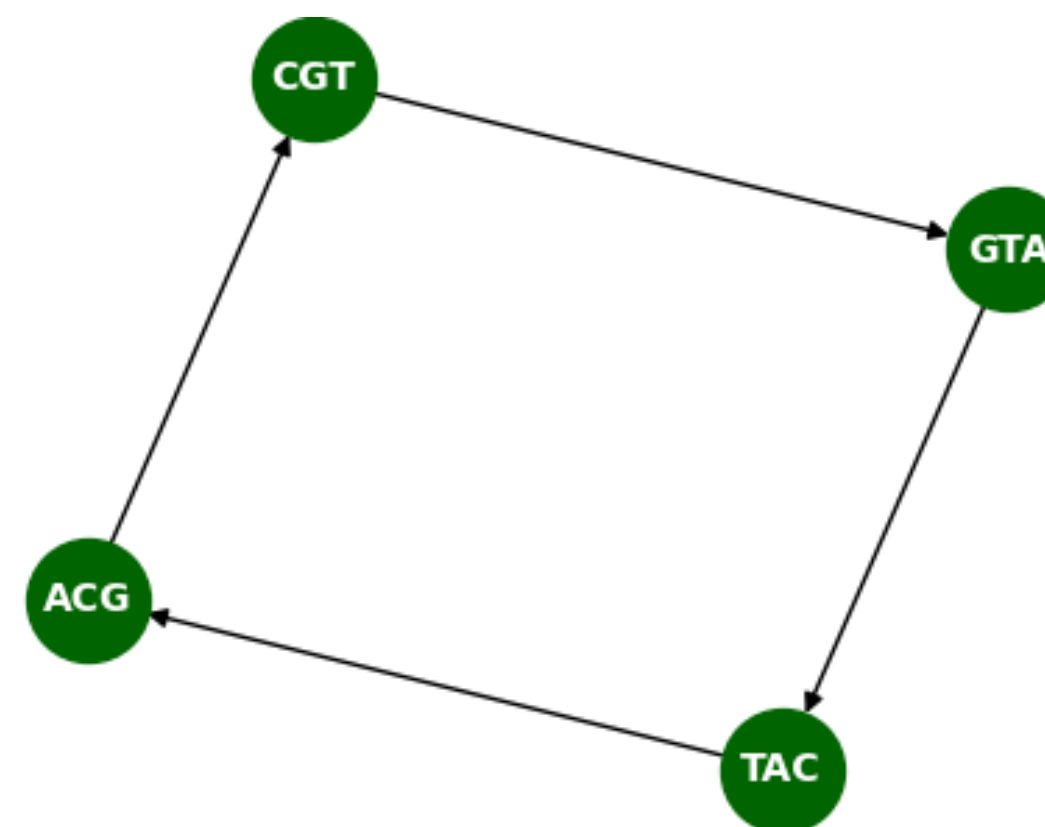
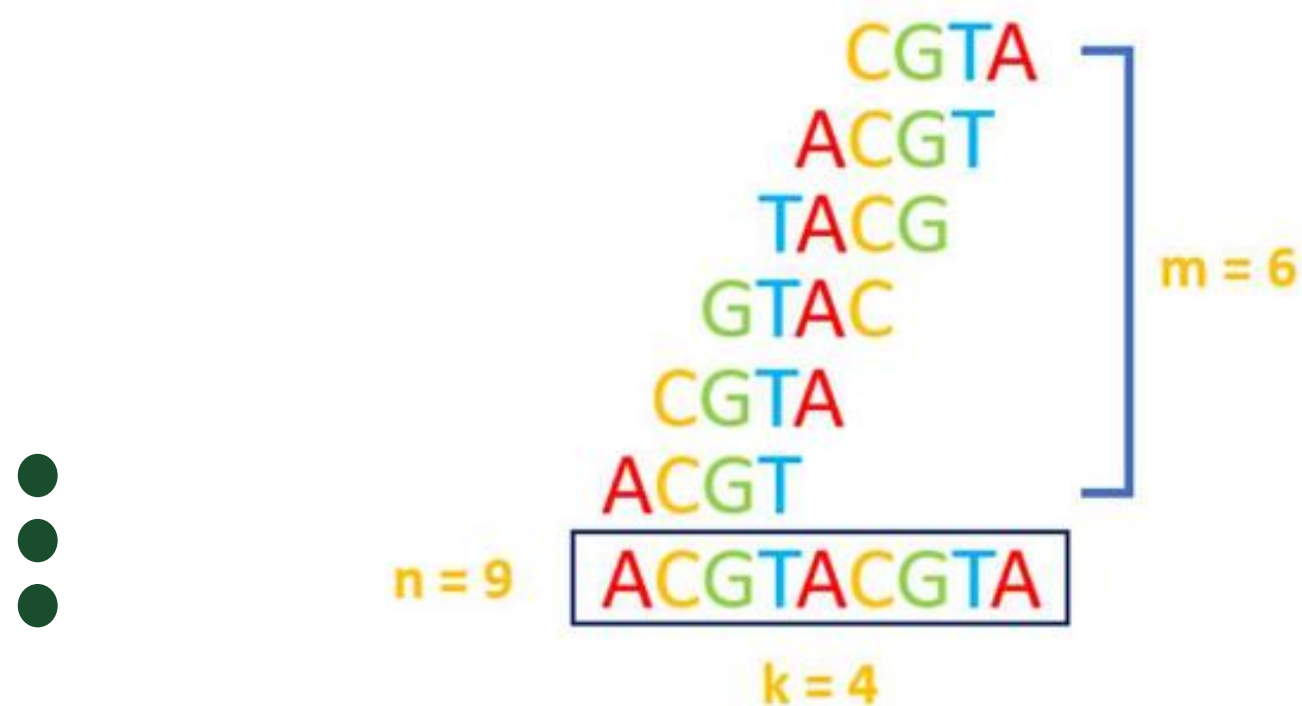
Background metodologico

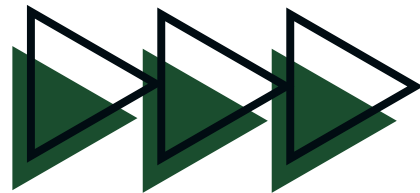


Grafo di De Bruijn

Un **grafo di De Bruijn** è una struttura utilizzata per rappresentare sequenze genetiche o altri dati lineari in modo compatto.

Viene costruito utilizzando i **k-mers** di una sequenza.

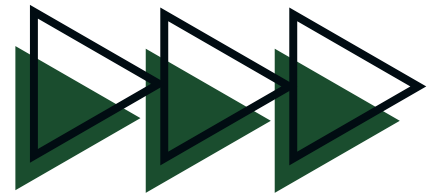




Background metodologico

Graph Neural Network





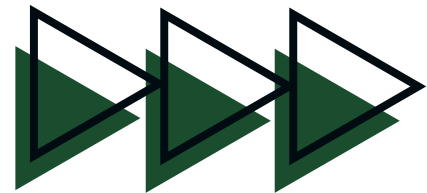
Background metodologico



Graph Neural Network

Una **Graph Neural Network (GNN)** è un modello di apprendimento automatico progettato per lavorare su dati rappresentati come grafi.





Background metodologico

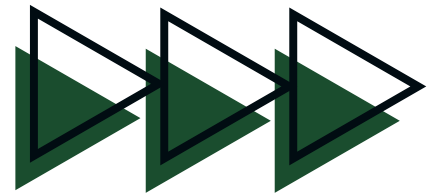


Graph Neural Network

Una **Graph Neural Network (GNN)** è un modello di apprendimento automatico progettato per lavorare su dati rappresentati come grafi.

Sono stati testati tre diversi modelli:





Background metodologico



Graph Neural Network

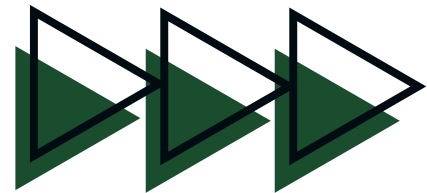
Una **Graph Neural Network (GNN)** è un modello di apprendimento automatico progettato per lavorare su dati rappresentati come grafi.

Sono stati testati tre diversi modelli:



Graph Convolution Network





Background metodologico



Graph Neural Network

Una **Graph Neural Network (GNN)** è un modello di apprendimento automatico progettato per lavorare su dati rappresentati come grafi.

Sono stati testati tre diversi modelli:

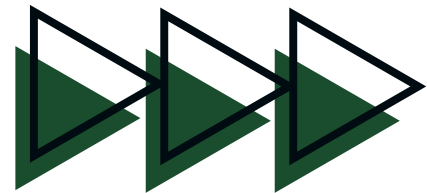


Graph Convolution Network



Graph Attention Network





Background metodologico



Graph Neural Network

Una **Graph Neural Network (GNN)** è un modello di apprendimento automatico progettato per lavorare su dati rappresentati come grafi.

Sono stati testati tre diversi modelli:



Graph Convolution Network

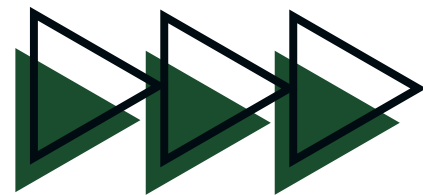


Graph Attention Network



Graph Sample and Aggregate

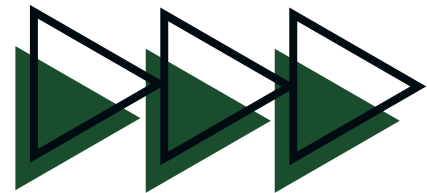




Background metodologico

Tecniche di Encoding





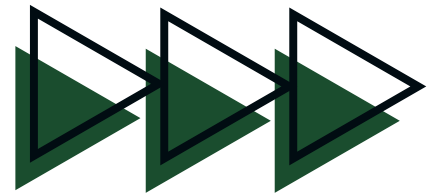
Background metodologico



Tecniche di Encoding

Sono state adottate due tecniche principali di encoding.





Background metodologico



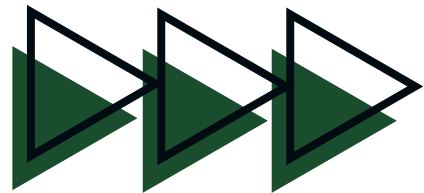
Tecniche di Encoding

Sono state adottate due tecniche principali di encoding.

DNABERT

Modello basato su BERT,
addestrato su k-mers per generare
rappresentazioni avanzate.





Background metodologico



Tecniche di Encoding

Sono state adottate due tecniche principali di encoding.

DNABERT

Modello basato su BERT,
addestrato su k-mers per generare
rappresentazioni avanzate.

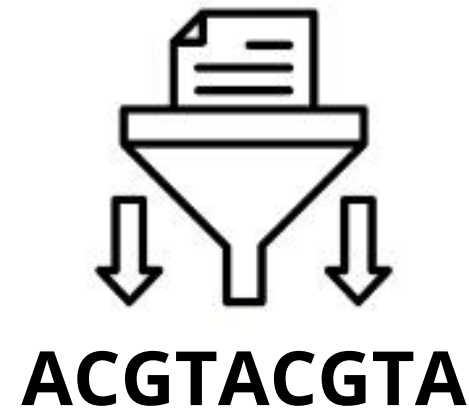
One-Hot Encoding

Ogni nucleotide (A, C, G, T) è
trasformato in un vettore binario
unico.

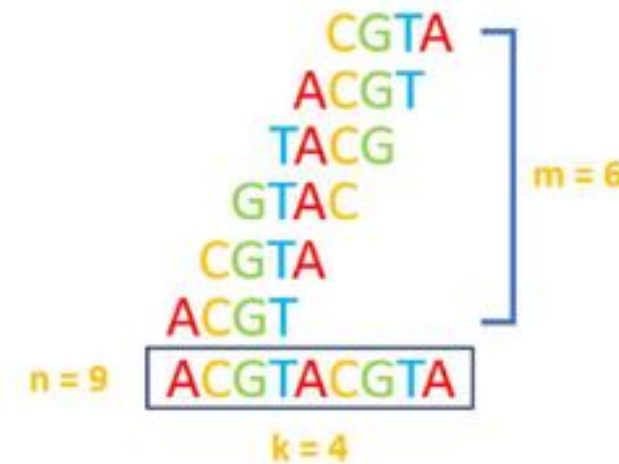


Implementazione

1 sequence extraction



2 k-mers extraction



3 De Bruijn Graph



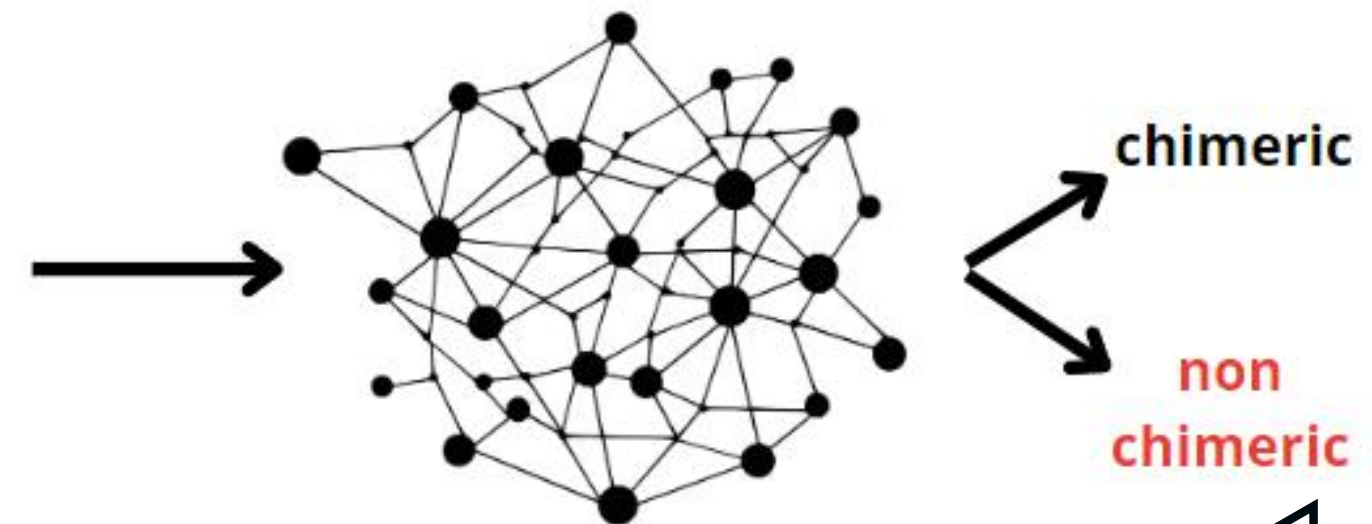
4 k-mers encoding

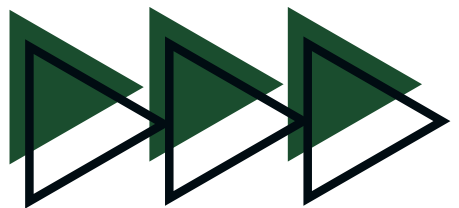


5 graph dataset creation



6 GNN classifier

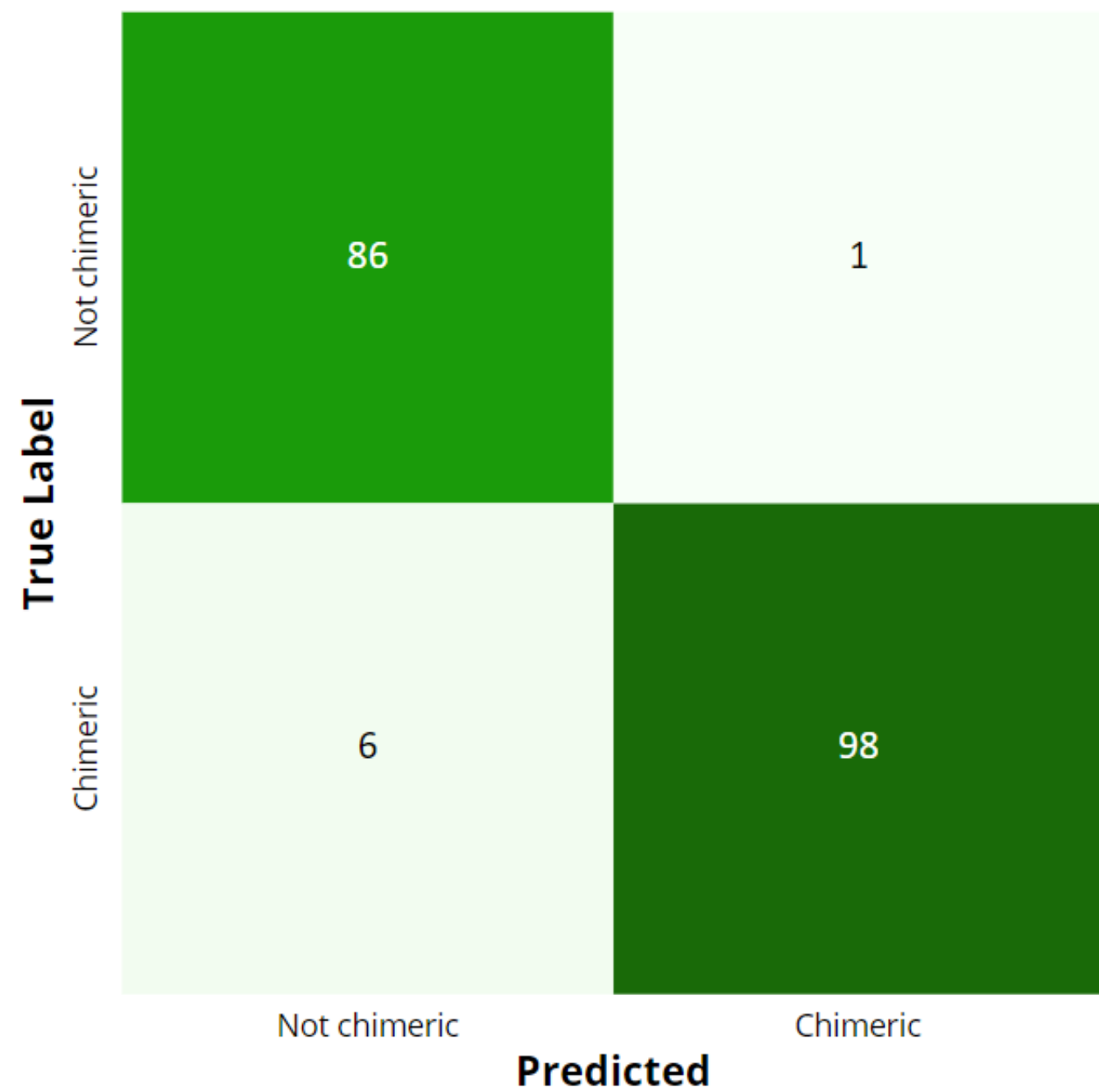




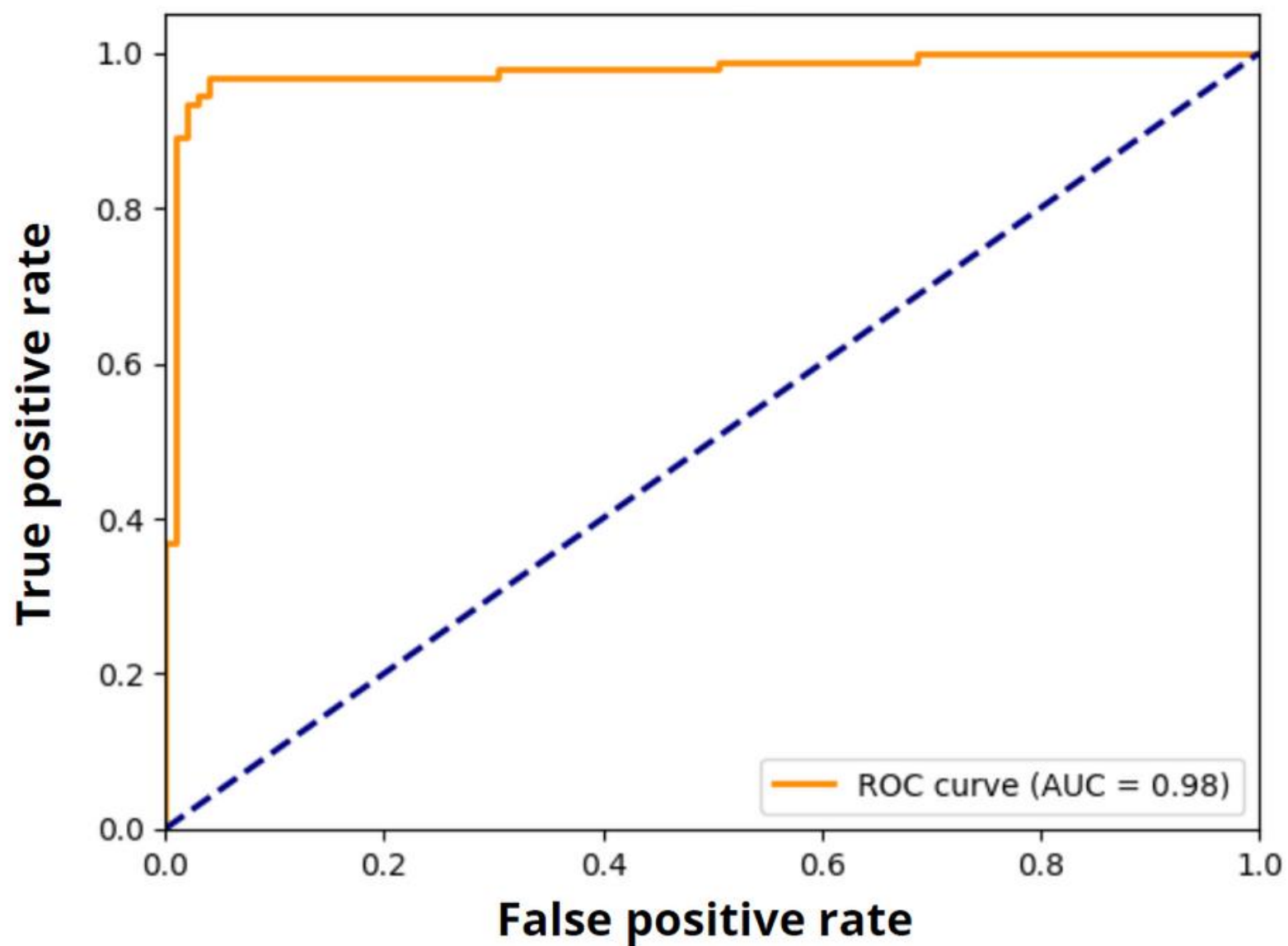
Risultati



GCN - Confusion Matrix



GCN - ROC Curve





Conclusioni

L'analisi sperimentale ha dimostrato che i modelli proposti ottengono delle prestazioni notevoli.



Conclusioni

L'analisi sperimentale ha dimostrato che i modelli proposti ottengono delle prestazioni notevoli.

Sviluppi futuri:



Conclusioni

L'analisi sperimentale ha dimostrato che i modelli proposti ottengono delle prestazioni notevoli.

Sviluppi futuri:



Ottimizzazione



Conclusioni

L'analisi sperimentale ha dimostrato che i modelli proposti ottengono delle prestazioni notevoli.

Sviluppi futuri:



Ottimizzazione



Valutazione su campioni reali



Conclusioni

L'analisi sperimentale ha dimostrato che i modelli proposti ottengono delle prestazioni notevoli.

Sviluppi futuri:



Ottimizzazione



Valutazione su campioni reali



Tecniche di explainability

A faint, light gray background illustration of a graduation cap (mortarboard) with a tassel, and a rolled-up diploma tied with a ribbon, positioned behind the main text.

**GRAZIE PER
L'ATTENZIONE**

