



UNIVERSITÀ DEGLI STUDI DI FIRENZE
SCUOLA DI INGEGNERIA - DIPARTIMENTO DI INGEGNERIA
DELL'INFORMAZIONE

Tesi di Laurea Triennale in Ingegneria Informatica

**IDENTIFICAZIONE DI SOTTOGRAFI PER
SEGMENTAZIONE DI DOCUMENTI SCIENTIFICI**

Candidato
Pietro Siliani

Relatore
Prof. Simone Marinai

Correlatore
Andrea Gemelli

Anno Accademico 2021/2022

Indice

Sommario	i
1 Introduzione	1
1.1 Contesto	1
1.2 Task	2
2 Creazione ed organizzazione dei dati	4
2.1 Dataset	4
2.1.1 PubLayNet	4
2.1.2 PubTables-1M	5
2.1.3 Dataset finale	5
2.2 PyMuPDF	6
2.3 DGL	6
2.4 Costruzione dei grafi	8
2.4.1 Individuazione e Labeling dei Nodi	8
2.4.2 Embeddings dei Nodi	9
2.4.3 Connattività	12
2.4.4 Edge Labeling	13
2.4.5 Bilanciamento	14

3 Struttura del modello	17
3.1 Link Prediction	17
3.2 Node Classification	18
3.3 Modello	18
3.3.1 Proiezione Input	19
3.3.2 GNN	21
3.3.3 Predittore	24
3.3.4 Classificatore	24
4 Addestramento ed esperimenti	26
4.1 Addestramento	26
4.1.1 Split dei dati e K-Fold	26
4.1.2 Early Stopping	27
4.1.3 Training loop	28
4.2 Inferenza	30
4.2.1 Metriche	33
5 Risultati	40
5.1 Risultati quantitativi	40
5.2 Commenti	52
5.2.1 mAP	52
5.2.2 Node Classification e Label Pruning	53
5.2.3 Altre considerazioni	54
5.3 Risultati qualitativi - Alcuni esempi	55
6 Conclusioni	63
Bibliografia	65

Sommario

Negli ultimi decenni l'avvento del World Wide Web ha permesso di condividere e recuperare enormi quantità di dati in modo veloce ed accessibile, un tale strumento consente alle diverse branche della comunità scientifica mondiale di distribuire, confrontare ed integrare i risultati dei propri studi in modo mai visto prima nella storia, a supporto di questa collaborazione sono stati sviluppati archivi come arXiv [1] e PubMed [10], i quali offrono accesso immediato ad un elevato numero di documenti scientifici riguardanti gli ambiti più disparati, come Matematica, Fisica, Economia e Biologia. La crescita ininterrotta della letteratura scientifica, per quanto vantaggiosa, richiede tuttavia di sviluppare delle tecniche in grado di facilitare ed alleggerire sia il processo di revisione delle pubblicazioni, sia l'estrazione di informazioni di interesse, l'analisi manuale di un documento scientifico non è infatti in grado di stare al passo con la frequenza di pubblicazione degli studi più recenti, per superare questo limite sono state sviluppate una serie di metodologie di analisi automatiche che sfruttano, tra le altre cose, anche tecniche pertinenti all'ambito del Deep Learning [8].

Tra gli obiettivi perseguiti dall'analisi automatica di documenti scientifici vi è l'operazione di *Document Layout Analysis*, che consiste nell'individuare, e talvolta classificare, le diverse componenti di una pubblicazione, come l'abstract, la bibliografia, le tavole, le immagini ed i paragrafi, una delle difficoltà principali di questo compito è dovuta all'eterogeneità delle strutture di tali elementi, che rende com-

plicato lo sviluppo di un unico strumento in grado di eseguire *Document Layout Analysis* in modo uniforme [17].

In questo lavoro si presenta un approccio all'operazione di *Document Layout Analysis* su documenti scientifici, basato sull'utilizzo di *Graph Neural Networks* (GNN). La tecnica sviluppata prevede di rappresentare ciascuna pagina di un documento con un grafo, individuando un nodo per ogni parola od immagine. I grafi ottenuti costituiscono l'input di una GNN, il cui output è utilizzato per effettuare *link prediction* e *node classification* in modo da ricavare una molteplicità di sottografi, ognuno corrispondente ad un particolare elemento della pagina.

Sono presenti 6 capitoli, di seguito descritti:

- Nel **Capitolo 1 - Introduzione** si presenta il contesto in cui si va ad operare, riassumendo l'obiettivo del lavoro effettuato.
- Nel **Capitolo 2 - Creazione ed organizzazione dei dati** si introducono i dataset utilizzati, la fase di *pre-processing* dei dati ed i criteri utilizzati per la creazione dei grafi.
- Nel **Capitolo 3 - Struttura del modello** si definiscono le componenti dei modelli addestrati ed il loro funzionamento.
- Nel **Capitolo 4 - Addestramento ed esperimenti** è riportata in dettaglio l'organizzazione del ciclo di apprendimento e la procedura di inferenza.
- Nel **Capitolo 5 - Risultati** sono presentati, confrontati e commentati i principali risultati ottenuti, sia a livello quantitativo che qualitativo.
- Nel **Capitolo 6 - Conclusioni** si analizzano i traguardi ed i limiti evidenziati dalla procedura impiegata, offrendo al contempo spunti di riflessione per il proseguimento dello studio.

Capitolo 1

Introduzione

1.1 Contesto

Il lavoro qui presentato si inserisce all'interno di un contesto di ricerca più ampio, portato avanti da Gemelli et al. e presentato in [3]. L'obiettivo ultimo è quello di riuscire ad eseguire congiuntamente le operazioni di *Document Layout Analysis*, e *Table Understanding*, la prima prevede di individuare e classificare le diverse componenti di un documento scientifico, come titoli, paragrafi, immagini e tabelle, la seconda consiste invece nell'identificazione delle tabelle, nella loro segmentazione strutturale e nell'analisi e comprensione dei dati contenuti in esse. Molti studi in questo ambito trattano i documenti convertendoli in immagini ed impiegando delle tecniche di *Optical Character Recognition* (OCR) per l'individuazione del testo. Un approccio simile è spesso incline ad errori e causa la perdita di informazioni durante la conversione da PDF ad immagine [3]; una delle sfide di questo studio consiste proprio nell'abbandono di strumenti di OCR per lavorare direttamente su documenti in formato PDF, ad esempio tramite l'impiego di PyMuPDF [7] (presentato nel paragrafo **2.2**).

Le pagine che compongono i documenti scientifici sono convertite in grafi, i cui nodi sono individuati in parole ed immagini, e le informazioni estratte da esse sono utilizzate per assegnare un carico informativo ad ogni nodo sotto forma di vettori di *features* numeriche, detti *embeddings*. La struttura così ottenuta è fornita in input ad una Graph Neural Network il cui output è utilizzato in operazioni di *post-processing* per effettuare Document Layout Analysis e Table Understanding.

1.2 Task

Lo scopo di questo lavoro è quello di riuscire a ricavare, a partire dal grafo relativo ad una generica pagina di un documento scientifico, un insieme di sottografi ciascuno corrispondente ad un diverso elemento della pagina. A tale scopo si addestra un modello basato su una Graph Neural Network all'operazione di *link prediction* che consiste nel determinare, per ogni potenziale arco del grafo, se questo appartiene o meno al grafo obiettivo. Nel secondo caso si procede a rimuovere l'arco in questione dal grafo di partenza al fine di generare i sottografi desiderati. Si dicono *positivi* gli archi che fanno parte del grafo obiettivo, mentre quelli che non vi appartengono, e devono quindi essere rimossi, sono indicati come *negativi*. Preso un generico grafo $G = (V, E)$ corrispondente ad una pagina di un documento, sono stati considerati positivi tutti quegli archi $(u, v) \in E$ tali che sia u che v facciano parte del medesimo elemento della pagina. Ad esempio, se u e v sono parole contenute nello stesso paragrafo, o nello stesso titolo, l'arco (u, v) è considerato positivo. Viceversa se u e v fanno parte l'uno di un paragrafo, l'altro di un titolo, o se fanno parte di due paragrafi separati, il relativo arco sarà considerato negativo. In alcuni esperimenti il modello è stato addestrato ad eseguire anche l'operazione di *node classification*, nel tentativo di fornire un ulteriore strumento per eseguire il *pruning* dei grafi durante la fase di inferenza. La classificazione dei nodi consiste

nell’assegnare ad ognuno di essi un’etichetta (o *label*) che ne indichi l’elemento di appartenenza. Ad esempio, se si considerano tre nodi diversi u , v , w tali che sia u che v appartengano ad un paragrafo e w appartenga ad un titolo, il modello sarà addestrato ad assegnare la stessa etichetta ad u e v , indipendentemente dal fatto che essi appartengano allo stesso paragrafo, ed un’etichetta diversa a w .

Capitolo 2

Creazione ed organizzazione dei dati

2.1 Dataset

Il dataset utilizzato per l’addestramento del modello è stato ottenuto unendo due dataset distinti: PubLayNet [19] e PubTables-1M [13]. Entrambi sono composti da articoli pubblicati su PubMed Central.

2.1.1 PubLayNet

PubLayNet è composto da 358353 pagine in formato PDF. Il dataset è orientato a task di *Document Layout Analysis* e *Table Detection*, presenta due file di annotazioni: *train.json* e *test.json* associati rispettivamente ai documenti appartenenti allo *split* di *train* e a quello di *test*. Tali file riportano, per ogni pagina, le *bounding boxes* dei diversi oggetti presenti all’interno della pagina, per ciascun oggetto così delimitato è anche riportato un id relativo alla categoria di appartenenza di tale regione, sono identificate 5 diverse categorie: *title*, *text*, *list*, *table* ed *image*.

2.1.2 PubTables-1M

PubTables-1M è un dataset orientato a task di *Table Detection*, *Table Structure Recognition* e *Table Functional Analysis*. Prevede 947642 tabelle corredate da annotazioni che, similmente a quanto avviene con PubLayNet, delimitano ed identificano le diverse regioni di una tabella come ad esempio *row*, *column*, *table header* e *table cell*.

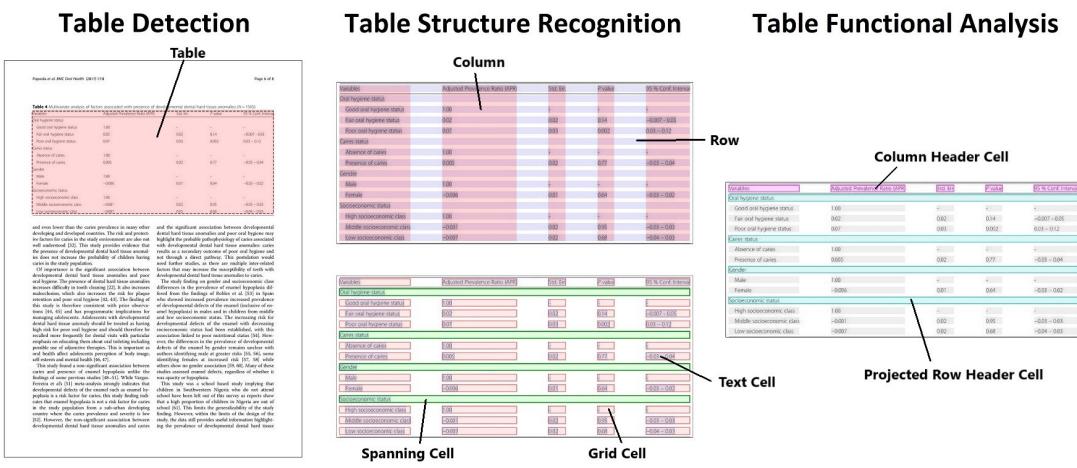


Figura 2.1: Le categorie e le relative aree di una tabella identificate dalle annotazioni di PubTables-1M, come riportate in [12].

2.1.3 Dataset finale

Il dataset utilizzato è formato da pagine provenienti da articoli appartenenti ad entrambe le collezioni precedentemente descritte. Le annotazioni associate distinguono tra 13 classi di oggetti: *other*, *text*, *title*, *list*, *table*, *figure*, *caption*, *column header*, *spanning cell*, *grid cell*, *table cell*, *column* e *row*, di queste le classi *other* e *caption* non appartengono alle annotazioni di nessuno dei due dataset originali e sono state aggiunte dopo l'unione dei due insiemi di documenti. La classe *caption* identifica le didascalie di immagini e tabelle, *other* si riferisce invece a tutti quegli

elementi che non appartengono a nessuna delle altre categorie, come ad esempio l’indice della pagina. Tutte le altre classi appartengono alle categorie identificate dalle annotazioni di PubLayNet o di PubTables-1M.

2.2 PyMuPDF

Al fine di costruire un grafo per ogni pagina del dataset è necessario riuscire ad estrarre dai singoli PDF due tipologie di informazioni fondamentali: il contenuto testuale vero e proprio e le *bounding boxes* delle singole parole. Per fare questo è stata utilizzata PyMuPDF: una libreria Python che fornisce accesso alle API MuPDF tramite le quali è possibile renderizzare, manipolare ed estrarre dati a partire da documenti con estensioni del tipo: “*.pdf*”, “*.xps*”, “*.oxps*”, “*.cbz*”, “*.fb2*” o “*.epub*” [7]. In particolare, per ogni pagina del dataset, è possibile utilizzare la funzione *open()* offerta dalla libreria per generare un’istanza della classe **Document** a partire dal percorso del relativo PDF, ogni istanza di **Document** così ottenuta contiene un’istanza della classe **Page**. Quest’ultima mette a disposizione la funzione *get_text()*, a cui si può passare il parametro “*words*” per ottenere una lista che riporta informazioni quali *bounding box* e stringa testuale relative ad ogni parola della pagina. Tali informazioni sono state utilizzate, come descritto nel paragrafo 2.4, per identificare i nodi e generare per ciascuno di essi sia gli *embeddings*, sia una *label* numerica che identifica la classe dell’oggetto.

2.3 DGL

Per creare e manipolare i grafi del dataset si è utilizzata la libreria DGL [15], mentre il *framework* per *machine learning* adottato è stato PyTorch [9]. In questa configurazione un grafo $G = (V, E)$, dove V è l’insieme dei nodi, E è l’insieme degli

archi, è rappresentato da un’istanza della classe **DGLGraph** mediante l’impiego dei seguenti tensori:

- Un tensore $N \in \mathbb{N}^{|V|}$ contenente indici che si riferiscono ai nodi del grafo, ogni indice individua in modo univoco un singolo nodo.
- Una coppia di tensori (X, Y) tali che $X, Y \in \mathbb{N}^{|E|}$, che individuano gli archi del grafo. Gli elementi di X e Y sono indici dei nodi, in particolare: la coppia (X_i, Y_i) indica che è presente un arco uscente dal nodo indicizzato da X_i e diretto verso quello indicizzato da Y_i . Non è possibile definire grafi non orientati in modo immediato, per trattare un grafo come se fosse non orientato è necessario che per ogni coppia (X_i, Y_i) esista una coppia (X_j, Y_j) tale che: $X_i = Y_j, Y_i = X_j$.

Ad esempio il grafo in Figura 2.2 si ottiene con $N = [0, 1, 2, 3]$, $X = [0, 0, 1, 2]$ e $Y = [1, 2, 2, 3]$.

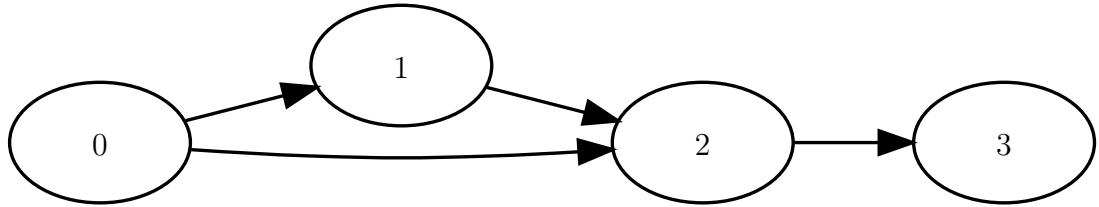


Figura 2.2: Esempio di un grafo

Sempre attraverso l’impiego di tensori possono essere assegnate delle *features* ad ogni nodo e ad ogni arco. Ad esempio, se volessimo associare ad ogni nodo un vettore di *features* $F \in \mathbb{R}^n$ sarebbe sufficiente definire un tensore $\bar{F} \in \mathbb{R}^{|V| \times n}$, in questo modo \bar{F}_i indicherà il vettore di n features corrispondente al nodo i -esimo. Per poter associare tale tensore al grafo è sufficiente aggiungere una voce al dizionario *ndata*, contenuto nella classe **DGLGraph**, utilizzando come chiave una stringa arbitraria e come valore il tensore stesso. DGL fornisce inoltre una

serie di primitive per eseguire *message passing* tra nodi ed archi del grafo, mentre PyTorch mette a disposizione strumenti utili per la realizzazione e l’addestramento di modelli, come algoritmi di ottimizzazione, funzioni per il calcolo della *loss* ed anche alcune strutture più semplici, come la classe **Linear**¹, a partire dalle quali costruire modelli più complessi.

2.4 Costruzione dei grafi

Data una pagina in formato PDF non è immediato definire il grafo ad essa corrispondente. Come prima cosa è necessario individuare quali elementi far corrispondere ai nodi del grafo, alcune opzioni possono essere le singole parole, le righe di testo, le immagini o le tabelle. Una volta stabilito come identificare i nodi si deve definire un criterio in base al quale determinare gli archi tra di essi. Il processo di costruzione dei grafi è descritto di seguito.

2.4.1 Individuazione e Labeling dei Nodi

È stato deciso di identificare i nodi con parole ed immagini. Al fine di costruire un grafo partendo da questi elementi è necessario riuscire ad individuare, per ciascuno di essi, la relativa *bounding box*. Le *bounding boxes* di parole ed immagini sono state estratte tramite l’utilizzo di PyMuPDF [7], introdotto nel paragrafo 2.2, sfruttando la funzione *get_text()*, i nodi del grafo sono individuati dal baricentro delle *bounding boxes* così estratte.

Al fine di generare degli *embeddings* sono state memorizzate sia le *bounding boxes* che le stringhe di testo corrispondenti ai nodi così individuati, alle immagini è stata associata la stringa: “IMAGE!”. Ad ogni nodo è stata poi attribuita un’etichetta (o

¹la classe **Linear** corrisponde ad un layer *fully connected*, ovvero implementa la trasformazione $y = W^T x + b$

label) che ne indica la classe di appartenenza, per farlo si è recuperato il baricentro della relativa *bounding box*, si è poi cercata l'annotazione che lo contenesse ed infine si è assegnata al nodo l'etichetta indicata da tale annotazione. Ai nodi che hanno baricentro compreso entro un'annotazione di classe *table* è stata associata l'etichetta corrispondente a *table*, senza considerare altre potenziali classi come *table cell*, *column* o *row*, le etichette possibili sono dunque: *text*, *title*, *list*, *image*, *table*, *caption* e *other*.

La scala delle *bounding boxes* utilizzata da PyMuPDF non corrisponde a quella impiegata nelle annotazioni, per far combaciare le due è stato necessario dividere i valori riportati da PyMuPDF per un fattore di scala pari a 0.36, e considerare la parte intera inferiore del risultato, ovvero:

$$\text{bbox annotazioni} = \lfloor (\text{bbox PyMuPDF}) \div 0.36 \rfloor$$

2.4.2 Embeddings dei Nodi

Anche se per noi la comprensione di un testo risulta il più delle volte spontanea ed immediata, una parola è in realtà in grado di trasportare una grande quantità di informazioni. Tra queste se ne annoverano una geometrica, relativa ad elementi come dimensione e posizione, ed una testuale, ricavata a partire dai simboli che la compongono, dal suo significato e, talvolta, dal contesto in cui è inserita. Queste due tipologie di dati sono state utilizzate per generare dei vettori di *features* da associare ai nodi, e che costituiranno le unità di input della rete. Suddette *features* sono brevemente descritte di seguito.

Features Geometriche

Le *features* geometriche sono state estratte dal carico informativo offerto dalle *bounding boxes*. Si distinguono 9 *features* ed il vettore è così strutturato: $\langle x_1, y_1, x_2, y_2, w, h, x_c, y_c, A \rangle$ dove, relativamente alla *bounding box* di un nodo:

- (x_1, y_1) e (x_2, y_2) sono le coordinate rispettivamente del vertice in alto a sinistra e di quello in basso a destra. Il sistema di riferimento utilizzato ha l'origine nel vertice in alto a sinistra della pagina e le unità di misura sono i *pixels*.
- w e h sono larghezza ed altezza.
- (x_c, y_c) sono le coordinate del baricentro.
- A è l'area.

SpaCy

Una possibilità per ricavare un vettore di *features* a partire dall'informazione testuale è fornita da SpaCy, una libreria Python *open-source* utilizzata in contesti di *Natural Language Processing* (NLP) [14]. SpaCy mette a disposizione dei modelli pre-addestrati che, data una parola od una frase, vi associano un vettore di *features* ricercandolo in una *look-up table*. Il modello utilizzato per generare gli *embeddings* è stato “en_core_web_lg”, il più grande offerto da SpaCy tra quelli addestrati sulla lingua inglese.

Scibert

Un'alternativa è costituita da Scibert [2], che, similmente a SpaCy, permette di ricavare un vettore di *features* a partire dalle singole parole. Un aspetto vantaggioso di Scibert è il fatto di essere addestrato su documenti scientifici, e dunque su un lessico affine a quello impiegato nei documenti del dataset utilizzato².

²(Si veda il paragrafo **2.1**)

Altre features

Sono stati testati anche i seguenti vettori di *features*:

- Istogramma
- Stile
- Intersezione

Il vettore “istogramma” è strutturato in questo modo: <% di lettere, % di cifre, % di simboli, immagine> dove per ogni nodo:

- I primi 3 valori sono uguali rispettivamente a: $\frac{(\# \text{ di lettere})}{(\# \text{ di caratteri})}$, $\frac{(\# \text{ di cifre})}{(\# \text{ di caratteri})}$ e $\frac{(\# \text{ di simboli})}{(\# \text{ di caratteri})}$.
- L’ultimo valore è un booleano che indica se il nodo in questione è un’immagine o meno. In caso affermativo i primi 3 valori del vettore saranno tutti uguali a 0.

Il vettore “stile” è composto da tre valori binari: <testo semplice, corsivo, grassetto>, le combinazioni utilizzate sono:

- <0, 0, 0> Per le immagini.
- <1, 0, 0> Per le parole che non sono scritte né in corsivo né in grassetto.
- <0, 1, 0> Per le parole scritte in corsivo.
- <0, 0, 1> Per le parole scritte in grassetto.
- <0, 1, 1> Per le parole scritte sia in corsivo che in grassetto.

Infine il vettore “intersezione” è generato, per un nodo v , allargandone leggermente la *bounding box* e controllando se i lati della *bounding box* risultante intersecano le

bounding boxes di altre parole od immagini. È formato da quattro booleani: <sopra, destra, sotto, sinistra>, questi valori sono impostati ad 1 se rispettivamente il lato superiore, destro, inferiore e sinistro della *bounding box* di u intersecano almeno un’altra *bounding box*.

Embeddings finali

Gli *embeddings* effettivamente utilizzati per i nodi sono formati da una concatenazione delle *features* geometriche, testuali ed istogramma, a cui talvolta si possono aggiungere i vettori stile ed intersezione. Alcune possibili combinazioni sono:

- Bounding boxes + Istogramma + SpaCy.
- Bounding boxes + Istogramma + Scibert.
- Bounding boxes + Istogramma + Stile + SpaCy.
- Bounding boxes + Istogramma + Intersezione + Scibert.
- Bounding boxes + Istogramma + Stile + Intersezione + Scibert.

2.4.3 Connettività

Un’altra problematica riguarda il criterio di inserimento di archi, le soluzioni adottate sono descritte di seguito.

Visibility

La connettività di tipo *visibility* prevede che due nodi u e v , identificabili con due punti P e Q siano collegati da un arco se e solo se il segmento (P, Q) non interseca alcun “ostacolo”. Per implementare questa topologia ogni nodo è stato identificato col baricentro della propria *bounding box*, e gli ostacoli sono stati individuati nelle *bounding boxes* degli altri nodi e negli altri archi già presenti.

Knn

La connettività *k-nearest neighbours* (knn) prevede di inserire un arco diretto da un nodo u verso un nodo v se, detti P e Q i punti corrispondenti rispettivamente ad u e v , la distanza euclidea tra P e Q è tra le k distanze più piccole di P da altri nodi del grafo. Poichè non è necessariamente vero che la distanza tra P e Q sia tra le k più piccole anche per Q , un grafo *k-nearest neighbours* è di per sè orientato.

Altre considerazioni

In entrambi i casi gli archi introdotti sono archi non orientati. Nel caso di connettività knn, una volta ottenuti gli archi applicando la definizione, questi sono stati sostituiti da archi non orientati. È stata inoltre fissata una lunghezza massima degli archi pari a 500 *pixels*, per evitare di collegare due elementi eccessivamente lontani così da impedire uno scambio di messaggi tra essi in fase di addestramento. Infine ad ogni arco è stata assegnata una singola *feature* corrispondente alla distanza euclidea tra i baricentri dei nodi collegati.

2.4.4 Edge Labeling

Per poter addestrare un modello di *link prediction* è necessario avere, per ogni arco, una *label* binaria che indichi se l'arco è positivo e quindi deve essere mantenuto, oppure se è negativo e deve essere rimosso. In particolare si considerano positivi tutti gli archi (u, v) tali che u e v abbiano il baricentro contenuto nella *bounding box* dello stesso oggetto. Si presentano però due eccezioni:

- Esistono più annotazioni che si riferiscono alle tabelle, come *table*, *column* o *row*. Un generico nodo di una tabella potrebbe avere il proprio baricentro contenuto sia nell'annotazione *table*, che delimita la tabella nella sua totalità, sia nell'annotazione *column* che nell'annotazione *row*, che definiscono

rispettivamente una colonna od una riga. È quindi necessario stabilire se separare una tabella in più oggetti, ad esempio considerando negativi gli archi che collegano due nodi appartenenti a colonne diverse, oppure se considerare la tabella come un singolo oggetto. È stata adottata la seconda opzione, dunque gli archi che collegano due elementi qualsiasi di una tabella sono sempre considerati positivi, purchè entrambi i nodi collegati abbiano il proprio baricentro contenuto nella *stessa* annotazione con id *table*.

- Gli archi tra nodi appartenenti alla categoria *other* sono sempre considerati positivi, indipendentemente dal fatto che i due nodi facciano parte della stessa annotazione.

2.4.5 Bilanciamento

Il dataset utilizzato³ risulta fortemente sbilanciato: ogni grafo presenta circa il 90% di archi etichettati come positivi. È quindi necessario bilanciare i dati utilizzati per l'apprendimento al fine di addestrare il modello a riconoscere correttamente sia le istanze positive che quelle negative. Un primo espediente adottato per effettuare il bilanciamento è stata la rimozione delle “isole”, ovvero di tutti quei nodi u appartenenti alla categoria *text* tali che esista un percorso con più di k archi da u ad un qualsiasi altro nodo v appartenente ad una classe differente [3]. Tuttavia prima di eliminare un nodo u si controlla se esiste almeno un arco negativo che abbia un vertice in quel nodo, in tal caso u non viene eliminato, questo controllo viene eseguito per evitare di ridurre ulteriormente il numero di archi negativi. Il valore di k adottato è stato $k = 5$ ed un esempio del risultato ottenuto applicando questo metodo ad un grafo *visibility* può essere visionato in figura 2.3.

La rimozione delle isole si dimostra particolarmente efficace per bilanciare la di-

³Si veda il paragrafo 2.1

strubuzione delle classi dei nodi, tuttavia non risolve del tutto il problema di sbilanciamento degli archi. A livello generale si sono comunque riscontrate prestazioni migliori nei modelli addestrati su grafi in cui erano state rimosse le isole, pertanto tale pratica è stata mantenuta anche quando il modello non è stato addestrato ad effettuare *node classification*. Per ottenere un bilanciamento definitivo è stato eseguito un *upsampling* degli esempi negativi: in ogni grafo è stata aggiunta una quantità di archi negativi generati casualmente tale da controbilanciare il numero di quelli positivi. L'aggiunta di archi, positivi o negativi che siano, modifica inevitabilmente la topologia del grafo, addestrare una Graph Neural Network su un grafo con una struttura irregolare può rivelarsi inconcludente in quanto tali reti si basano prevalentemente su operazioni di *message passing* per le quali si presuppone di lavorare con una connettività significativa. Pertanto per ogni grafo nel dataset di *train* sono state generate due versioni: una prima versione che, a meno della rimozione delle isole, risulta non bilanciata, ed una seconda versione bilanciata. La prima è utilizzata per realizzare la fase di *message passing* e calcolare gli *embeddings* finali dei nodi, la seconda è utilizzata durante la fase di *link prediction*, questa soluzione permette di avere a disposizione una quantità bilanciata di archi positivi e negativi senza compromettere la connettività dei grafi utilizzati durante la fase di *message passing*.

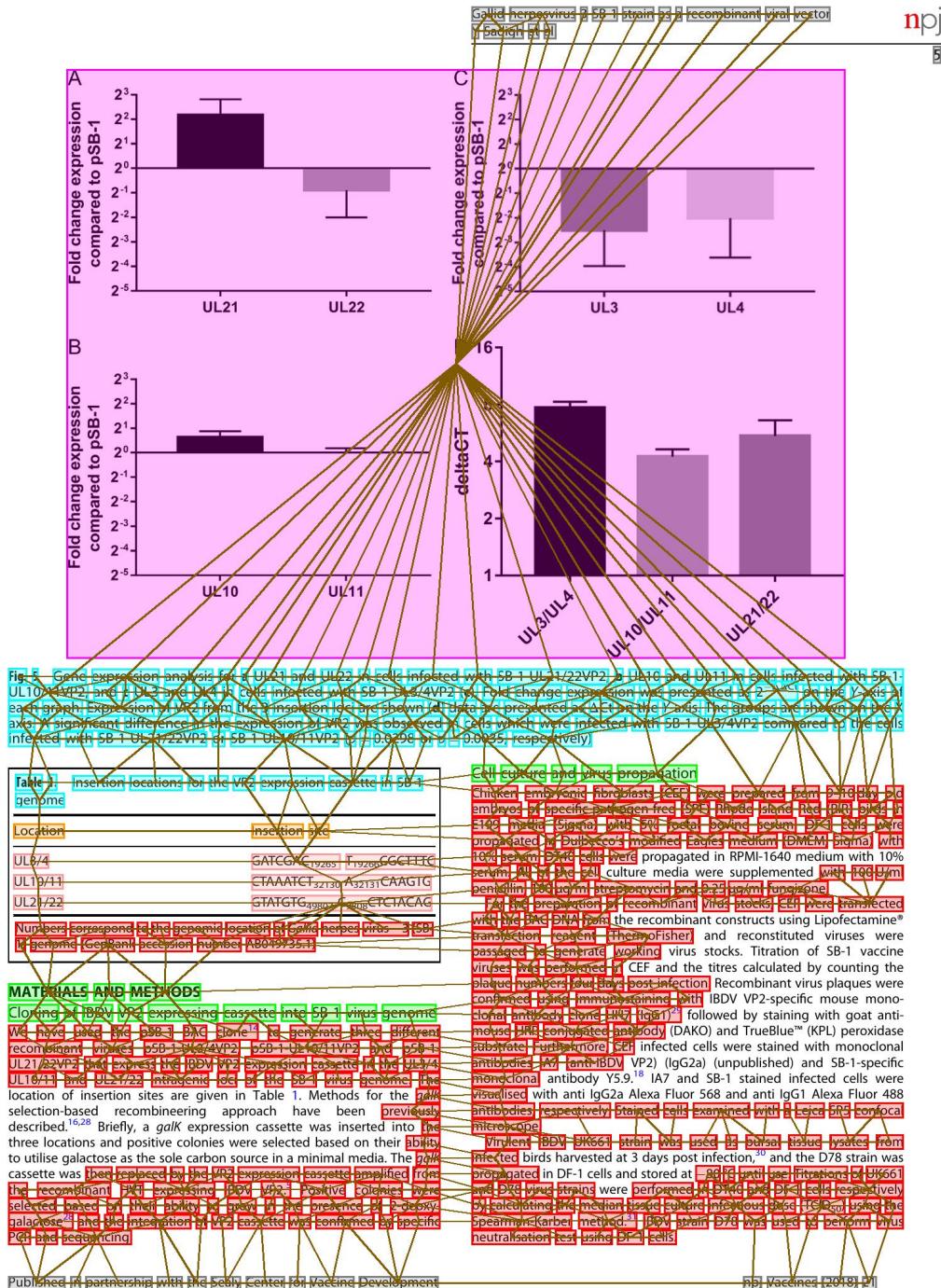


Figura 2.3: Un esempio di grafo visibility in seguito alla rimozione delle isole.

Capitolo 3

Struttura del modello

Durante gli esperimenti sono state testate diverse strutture sia per il modello che per i grafi. Di seguito viene data una presentazione generale del modello, del suo funzionamento e delle tecniche utilizzate. Dei riferimenti ad alcune delle architetture impiegate si possono invece trovare nel capitolo 5.

3.1 Link Prediction

Dal punto di vista procedurale l'operazione di *link prediction* può essere descritta nel seguente modo: inizialmente si sfrutta una Graph Neural Network che riceve in input sia il grafo che i vettori di *features* dei nodi¹ e restituisce in output gli *embeddings* finali di ogni nodo. Successivamente per ogni arco (u, v) si calcola un *embedding* a partire da quelli di u e v ad esempio tramite concatenazione o somma, nel primo caso si avrebbe:

$$h_{(u,v)} = h_u || h_v$$

Dove $h_{(u,v)}$ indica l'*embedding* finale dell'arco (u, v) e h_u, h_v indicano rispettivamente gli *embeddings* finali di u e v . In seguito per ogni arco (u, v) si calcola uno

¹Per le features utilizzate si veda 2.4.2

$score \in \mathbb{R}$ a partire da $h_{(u,v)}$. Da ultimo si ricava la probabilità che l'arco sia positivo applicando la funzione sigmoide logistica σ allo $score$, dunque:

$$p_{(u,v)} = \sigma(s_{(u,v)})$$

Dove $p_{(u,v)}$ è la predizione finale sull'arco (u, v) mentre $s_{(u,v)}$ è lo $score$ del medesimo arco.

3.2 Node Classification

In alcuni esperimenti il modello è stato addestrato ad eseguire in concomitanza sia l'operazione di *link prediction* che quella di *node classification*. Quest'ultima prevede di ricavare dall'*embedding* h_v di un nodo $v \in V$, calcolato dalla Graph Neural Network, un vettore di *scores* $s_v \in \mathbb{R}^n$, dove n è il numero di possibili classi a disposizione. Applicando la funzione *softmax* a s_v si ricava un vettore di n probabilità p_v , le classi sono identificate da dei valori numerici che vanno da 1 a n , al nodo v si assegna la classe $i \in [1, n]$ tale che: $i = \underset{j=1,2,\dots,n}{\operatorname{argmax}}(p_{v,j})$.

3.3 Modello

Il modello addestrato negli esperimenti può essere concettualmente separato in quattro parti:

- Un insieme di layer *fully connected* che processano l'input in modo da assegnare pari importanza alle diverse *features* utilizzate.
- Una Graph Neural Network che calcola gli *embeddings* finali di ogni nodo attraverso una fase di *message passing*.

- Un predittore che, a partire dagli *embeddings* di due nodi u e v , calcola la probabilità che l'arco (u, v) sia positivo.
- Un classificatore che, a partire dall'*embedding* di un nodo u assegna una classe a tale nodo. Il classificatore è stato impiegato solo nei modelli utilizzati per fare anche *node classification*.

I dettagli implementativi di sudette componenti sono forniti nelle sezioni successive.

3.3.1 Proiezione Input

Come affermato nel paragrafo 2.4.2 gli *embeddings* dei nodi sono composti da *features* con significati diversi che possono essere combinate in più modi, in ogni caso il numero di *features* testuali risulta di gran lunga superiore rispetto a quello delle altre categorie. Ad esempio, utilizzando SpaCy si ottengo vettori di *features* testuali di 300 valori, mentre concatenando i vettori di *features* geometriche, istogramma, stile ed intersezione si ottiene un vettore di soli 20 elementi. Ciò potrebbe indurre il modello ad assegnare maggiore importanza alle *features* testuali, trascurando le altre. Per evitare questo comportamento si possono innanzitutto dividere gli *embeddings* dei nodi in dei sotto vettori di *features*, tali che ognuno di essi contenga una specifica tipologia di informazione (ad esempio geometrica) o una concatenazione di diverse tipologie (ad esempio istogramma + stile), dopodichè è possibile utilizzare dei layer *fully connected* per proiettare ciascuno di questi sotto vettori in uno spazio di dimensione n . Ad esempio, assumendo per semplicità di utilizzare solo *features* geometriche e testuali e di proiettarle in due spazi separati, detti rispettivamente g_v , t_v i vettori di *features* geometriche e testuali di un nodo $v \in V$, n_g il numero di *features* geometriche, n_t il numero di quelle testuali ed

indicati con g'_v e t'_v gli output dei layer *fully connected* si ha:

$$g'_v = W_g^T g_v + b_g \quad (3.1)$$

$$t'_v = W_t^T t_v + b_t \quad (3.2)$$

Con:

$$g_v \in \mathbb{R}^{n_g}$$

$$t_v \in \mathbb{R}^{n_t}$$

$$g'_v, t'_v \in \mathbb{R}^n$$

$$W_g^T \in \mathbb{R}^{n \times n_g}, b_g \in \mathbb{R}^n$$

$$W_t^T \in \mathbb{R}^{n \times n_t}, b_t \in \mathbb{R}^n$$

Una volta eseguita la proiezione si applica una funzione di attivazione e si concatenano i risultati, in particolare la funzione di attivazione utilizzata è stata:

$$\text{LeakyReLU}(x) = \begin{cases} x & \text{se } x > 0 \\ \alpha x & \text{se } x \leq 0 \end{cases} \quad (3.3)$$

Dove è stato fissato $\alpha = 0.01$.

Dunque detto c_v il risultato della concatenazione delle *features* del nodo v ed f il numero di diversi sotto vettori di *features* utilizzati, riferendosi al caso in cui si hanno solo *features* geometriche e testuali si ha:

$$c_v = (\text{LeakyReLU}(g'_v) \parallel \text{LeakyReLU}(t'_v)) \quad (3.4)$$

Con:

$$c_v \in \mathbb{R}^{nf}$$

Da ultimo si utilizza un ulteriore layer *fully connected*, l'embedding finale h_v è ottenuto in questo modo:

$$h_v = \text{LeakyReLU}(W^T c_v + b) \quad (3.5)$$

Con:

$$W \in \mathbb{R}^{(nf) \times (nf)}, b \in R^{nf}$$

La figura 3.1 schematizza il procedimento appena descritto.

3.3.2 GNN

La Graph Neural Network utilizzata è stata costruita a partire da una versione leggermente modificata del layer proposto in [4], denominata GraphSAGE-GCN e presentata in [3]. La rete prevede che ogni nodo $v \in V$ del grafo $G = (V, E)$ possieda un vettore di *features* h_v , ottenuto in seguito alla trasformazione descritta nel paragrafo **3.3.1**, ed una *mailbox* m_v . Ad ogni passo dell'algoritmo implementato da GraphSAGE-GCN (schematizzato in figura 3.2) ciascun nodo v riceve dei messaggi dai propri vicini $N(v)$, il messaggio inviato dal nodo u verso il nodo v ha la seguente forma:

$$m_{u \sim v} = w_e h_u \quad (3.6)$$

Dove h_u è il vettore di *features* di u e w_e è la distanza normalizzata di u da v . Al termine di un generico passo k dell'algoritmo, la *mailbox* di ciascun nodo v sarà:

$$m_v^k = \{w_e h_u^{k-1} | \forall u \in N(v)\} \quad (3.7)$$

I vettori nella *mailbox* sono poi aggregati calcolandone una media in questo modo:

$$h_{N(v)}^k = \frac{\sum_{m \in m_v^k} m}{|N(v)|} \quad (3.8)$$

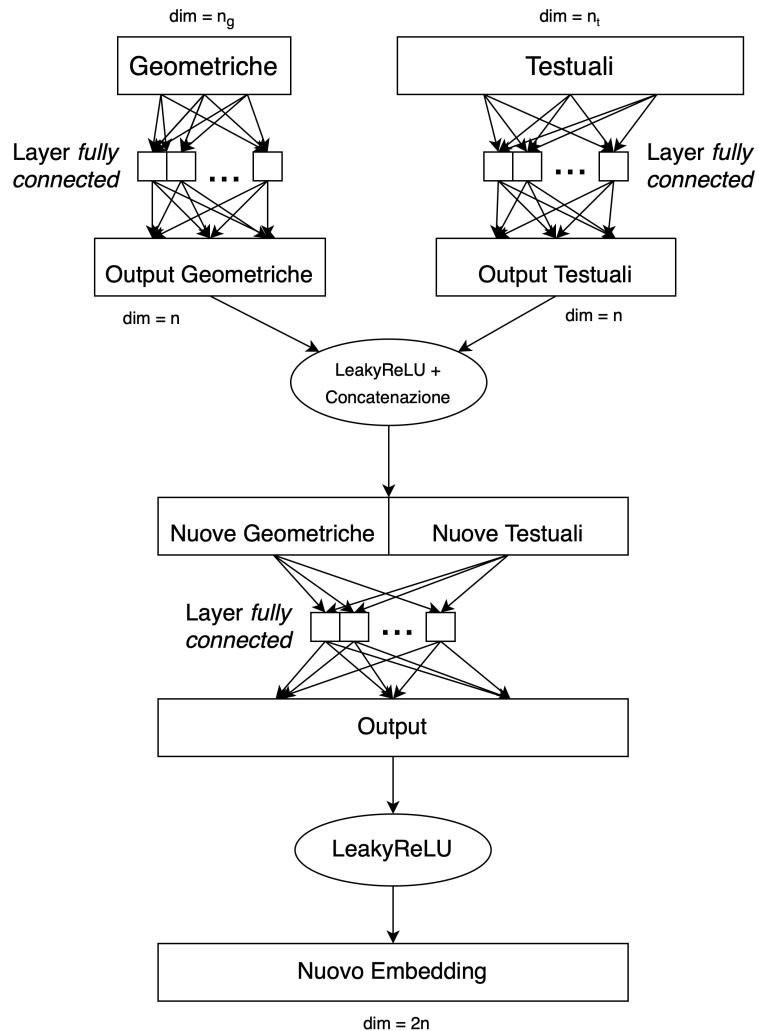


Figura 3.1: Schema di proiezione con *features* geometriche e testuali.

Successivamente si concatena h_v^{k-1} con $h_{N(v)}^k$ ed il vettore risultante è dato in input ad un layer *fully connected*. Si utilizza infine una funzione di attivazione per ottenere l'*embedding* aggiornato di v , ovvero:

$$h_v^k = \theta(W^T \times (h_v^{k-1} \| h_{N(v)}^k)) \quad (3.9)$$

Dove è stata utilizzata $\theta(x) = \text{ReLU}(x)$ con:

$$\text{ReLU}(x) = \begin{cases} x & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases} \quad (3.10)$$

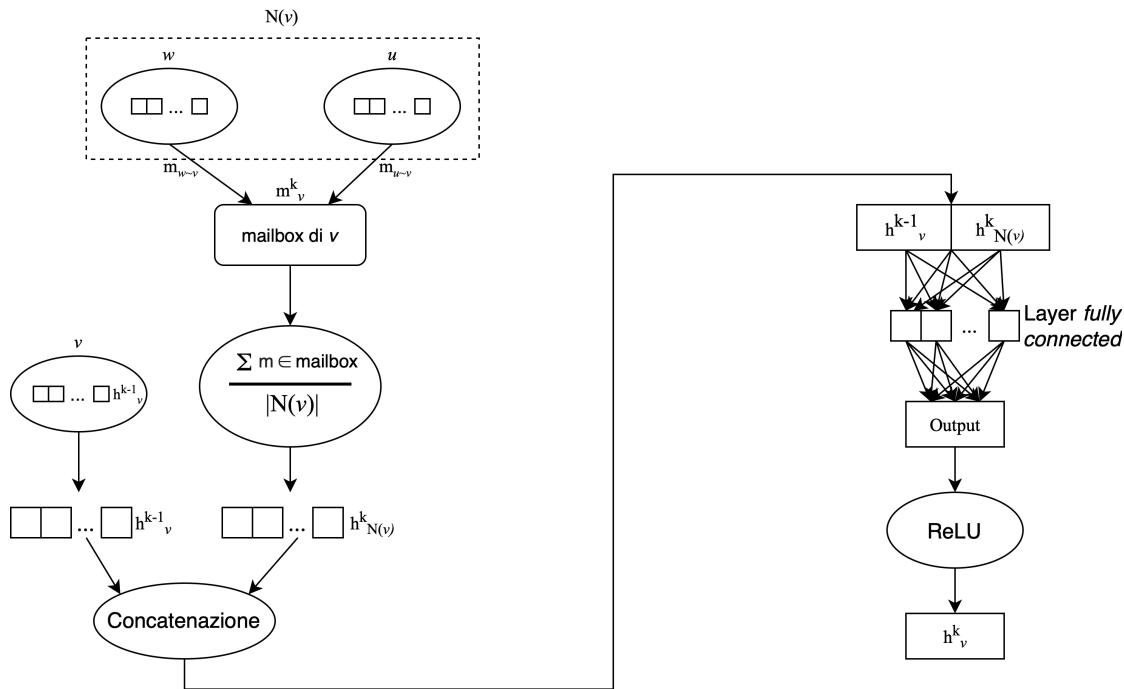


Figura 3.2: Schema di un passo dell'algoritmo applicato al nodo v con vicini u e w .

3.3.3 Predittore

Il ruolo del predittore è quello di calcolare uno *score* per ogni arco (u, v) a partire dagli *embeddings* dei nodi u e v . A tale scopo inizialmente si ottiene l'*embedding* dell'arco a partire da quello dei nodi, ad esempio tramite concatenazione, in tal caso:

$$h_{(u,v)} = h_u \parallel h_v \quad (3.11)$$

Dopodichè si utilizza un Multilayer Perceptron (MLP) per calcolare lo *score* $y_{(u,v)}$ partendo da $h_{(u,v)}$. Il MLP utilizzato è composto da un singolo *hidden layer* di dimensione n_h mentre la dimensione dell'*input layer* è n_i . Se $h'_{(u,v)}$ è l'*embedding* di (u, v) calcolato dall'*hidden layer* del MLP, il comportamento del predittore si può riassumere in questo modo:

$$h'_{(u,v)} = \text{ReLU}(W_0^T h_{(u,v)} + b_0) \quad (3.12)$$

$$s_{(u,v)} = W_1^T h'_{(u,v)} + b_1 \quad (3.13)$$

dove:

$$W_0^T \in \mathbb{R}^{n_h \times n_i}, b_0 \in \mathbb{R}^{n_h}$$

$$W_1^T \in \mathbb{R}^{1 \times n_h}, b_1 \in \mathbb{R}$$

Per calcolare la probabilità $p_{(u,v)}$ che l'arco (u, v) sia positivo si utilizza la funzione sigmoide logistica σ , pertanto:

$$p_{(u,v)} = \sigma(s_{(u,v)}) = \frac{1}{1 + e^{-s_{(u,v)}}} \quad (3.14)$$

3.3.4 Classificatore

Nei casi in cui si esegue anche *node classification* è stato utilizzato un classificatore per assegnare ad ogni nodo $v \in V$ una classe. Tale classificatore è formato da un

layer *fully connected* che proietta l'*embedding* h_v in un vettore $s_v \in \mathbb{R}^n$, dove n è il numero di possibili classi. Da s_v si ricava un vettore di probabilità p_v utilizzando la funzione softmax, ovvero:

$$p_{v,j} = \text{softmax}(s_v)_j = \frac{e^{s_{v,j}}}{\sum_{k=1}^n e^{s_{v,k}}} \quad (3.15)$$

la classe i assegnata al nodo v sarà quindi $i = \underset{j=1,2,\dots,n}{\text{argmax}}(p_{v,j})$.

Capitolo 4

Addestramento ed esperimenti

4.1 Addestramento

4.1.1 Split dei dati e K-Fold

Il dataset utilizzato¹ è composto da un *training set* di circa 24000 pagine mentre il *test set* ne contiene 1724. La quantità di dati effettivamente utilizzati per l’addestramento del modello è di 12000 grafi, questo limite è stato imposto, in modo indiretto, dalla necessità di eseguire più volte le operazioni di *pre-processing* al fine di costruire e testare diverse combinazioni di *features*, connettività e tecniche di bilanciamento. Limitare il numero di pagine da processare ha permesso di ottenere un buon compromesso tra abbondanza di dati e tempo impiegato per la fase di preparazione dei grafi.

Per scongiurare il fenomeno dell’*overfitting*, per il quale un modello si adatta completamente ai dati su cui è addestrato e perde la capacità di generalizzare, è stata

¹Si veda il paragrafo **2.1.3**

utilizzata la tecnica di *K-Fold cross-validation* scegliendo $K = 10$. Il funzionamento del meccanismo di *K-Fold* è il seguente: una volta fissato un intero K , l'insieme di addestramento viene prima permutato in modo casuale e poi diviso in K sottinsiemi con intersezione vuota e con lo stesso numero di elementi. Ad ogni epoca di apprendimento uno solo di questi sottinsiemi è utilizzato come *validation set*, mentre l'unione di tutti gli altri costituisce il *training set* effettivo. Il *validation set* non è utilizzato per addestrare il modello, ma per verificarne il comportamento alla fine di ogni epoca e quindi individuare il fenomeno dell'*overfitting*. L'utilizzo di *K-Fold cross-validation* permette invece di ridurre la tendenza all'*overfitting* del modello, presentando ad ogni epoca un *training set* leggermente differente da quello utilizzato nell'epoca precedente. La tecnica di *K-Fold* tuttavia non è sufficiente a tutelare in modo affidabile dall'*overfitting*, ed ha più che altro lo scopo di rilevare il fenomeno quando esso avviene, per un'effettiva prevenzione è stata impiegata la strategia di *early stopping*, descritta nel paragrafo **4.1.2**.

4.1.2 Early Stopping

Il meccanismo di *early stopping* prevede di monitorare la *loss* calcolata sul *validation set* (*validation loss*) e di fermare l'addestramento una volta che questa smette di diminuire in modo consistente. Alla fine di ogni epoca la *validation loss* v calcolata viene comparata con la miglior *validation loss* v' tra quelle registrate nelle epoche precedenti: se $v < v' - \Delta$, dove Δ è un parametro fissato, il valore di v' viene aggiornato con v , quando ciò avviene lo stato attuale del modello è memorizzato in un *checkpoint*. È inoltre utilizzato un ulteriore parametro, chiamato *patience*, per fissare un limite al numero di epoche terminate senza aggiornamento di v' che si è disposti ad aspettare prima di arrestare l'addestramento. In altre parole se il valore di v' rimane costante per un numero di epoche maggiore di *patience* l'addestramento termina ed il modello migliore si può recuperare dall'ultimo

checkpoint creato.

4.1.3 Training loop

Di seguito sono descritti la struttura del ciclo di apprendimento sviluppato ed alcuni accorgimenti utilizzati per far fronte ai limiti imposti dalla memoria RAM e VRAM a disposizione. All'inizio di ogni epoca di addestramento si determina quale tra i K sottoinsiemi del *training set* utilizzare come *validation set*. Inoltre poichè la quantità di dati utilizzata è tale da non rendere possibile il mantenimento contemporaneo di ogni grafo nè in memoria RAM nè tantomeno in VRAM, l'insieme di addestramento viene diviso in dei *batch* di 100 grafi ciascuno. Il modello viene quindi addestrato in modo iterativo su ogni singolo *batch* caricando i grafi che lo compongono direttamente dalla memoria principale e spostandoli in VRAM, fortunatamente l'*overhead* dovuto al caricamento continuo di dati dalla memoria principale si è rivelato quasi del tutto trascurabile.

La fase di addestramento vera e propria prevede come prima cosa di calcolare gli *embeddings* finali dei nodi impiegando i layer di proiezione delle *features* e la rete GraphSAGE-GCN. Una volta ottenuti gli *embeddings* finali dei nodi si esegue l'operazione di *link prediction* utilizzando il predittore per ricavare una predizione per ogni arco. A questo punto si calcola la *loss* del modello, come funzione di *loss* l è stata utilizzata la *Binary Cross Entropy Loss*:

$$l = \frac{1}{n_e} \sum_{i=1}^{n_e} y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \quad (4.1)$$

Dove p_i e y_i sono rispettivamente la predizione ed il valore di *ground-truth* sull'arco i -esimo e n_e è il numero complessivo di archi nel *batch*.

Se si sceglie di eseguire anche *node classification* si utilizzerà il classificatore per ottenere la classe di ciascun nodo e si calcolerà la relativa *loss* usando la *Multiclass*

Cross Entropy Loss ridotta tramite media:

$$l_{ce} = \{l_1, \dots, l_i, \dots l_{n_v}\}, l_i = \sum_{k=1}^n 1\{y_i = k\} \log(p_{i,k}) \quad (4.2)$$

$$l = \frac{1}{n_v} \sum_{i=1}^{n_v} l_{ce,i} \quad (4.3)$$

Dove n è il numero di classi, $p_{i,k}$ la probabilità che il nodo i -esimo abbia classe k , y_i la classe del nodo i -esimo e n_v il numero di nodi nel *batch*.

Una volta calcolata la *loss* si aggiustano i parametri del modello e si seleziona il *batch* successivo su cui effettuare l'addestramento. Se si esegue *node classification* la *loss* da minimizzare aggiustando i parametri è ottenuta come somma delle *loss* calcolate per *link prediction* e *node classification*. Terminato l'addestramento sui *batch* che compongono il *training set* di una data epoca si passa alla fase di validazione, si procede cioè ad applicare il modello attuale al *validation set*, anch'esso diviso in *batch* per alleggerire il carico sulla memoria.

In generale è desiderabile calcolare una *loss* complessiva L sia al termine della fase di addestramento che della fase di validazione, in modo da rendere evidente il comportamento del modello indipendentemente dai *batch*, L è calcolata in questo modo:

$$L = \frac{1}{N} \sum_{i=1}^B l_i n_i \quad (4.4)$$

Dove l_i e n_i sono rispettivamente la *loss* e la quantità di esempi² relativi al *batch* i -esimo, N è la somma del numero di esempi di tutti i *batch* e B è il numero complessivo di *batch*.

La *loss* calcolata sui grafici di validazione viene utilizzata per implementare l'*early stopping* descritto nel paragrafo 4.1.2. Conclusa la validazione si passa all'epoca

²i. e. archi se ci si riferisce all'operazione di *link prediction*, nodi per il *task* di *node classification*

successiva ed il procedimento viene ripetuto fino a quando l'*early stopping* non arresta l'addestramento.

4.2 Inferenza

Si noti che a scopo sintetico tutte le volte che in questo paragrafo ci si vorrà riferire al vertice in alto a sinistra di una *bounding box* esso sarà indicato con **TL**, mentre quello in basso a destra sarà denominato **BR**.

Al termine dell'addestramento, per verificare l'efficacia del modello ottenuto, lo si è utilizzato per fare inferenza sui grafi dell'insieme di test. Il procedimento prevede come prima cosa di tagliare gli archi che il modello individua come negativi, per fare questo si utilizza la rete per calcolare, per ogni arco (u, v) , la probabilità $p_{(u,v)}$ che l'arco sia positivo, in seguito si eliminano tutti gli archi la cui $p_{(u,v)}$ è minore di una soglia (o *threshold*) fissata.

Successivamente le diverse entità della pagina sono individuate calcolando le componenti connesse del grafo, poi per la *bounding box* di ciascun nodo v nella componente connessa C si definiscono:

- $x_{0,v}$ = coordinata x di **TL**
- $y_{0,v}$ = coordinata y di **TL**
- $x_{1,v}$ = coordinata x di **BR**
- $y_{1,v}$ = coordinata y di **BR**

ed i seguenti insiemi:

- $X_0 = \{x_{0,v} | \forall v \in C\}$
- $Y_0 = \{y_{0,v} | \forall v \in C\}$

- $X_1 = \{x_{1,v} | \forall v \in C\}$

- $Y_1 = \{y_{1,v} | \forall v \in C\}$

a questo punto **TL** della *bounding box* di C avrà coordinate: $(\min(X_0), \min(Y_0))$, **BR** sarà invece individuato da: $(\max(X_1), \max(Y_1))$. Può capitare che, in genere a causa di errori nella classificazione degli archi, le *bounding boxes* delle componenti connesse individuate si intersechino, in questi casi è possibile unirle utilizzando uno tra due possibili criteri:

- **intersezione:** tutte le *bounding boxes* che si intersecano sono unite in un'unica *bounding box*.
- **inclusione:** se la *bounding box* $B1$ include la *bounding box* $B2$ allora $B2$ è unita a $B1$. Le *bounding boxes* che si intersecano e basta sono lasciate intatte.

Se $(x_{0,B1}, y_{0,B1})$ e $(x_{1,B1}, y_{1,B1})$ sono rispettivamente le coordinate di **TL** e **BR** di una *bounding box* $B1$ e $(x_{0,B2}, y_{0,B2})$ e $(x_{1,B2}, y_{1,B2})$ sono le coordinate dei medesimi vertici di una *bounding box* $B2$, la *bounding box* risultante dall'unione di $B1$ e $B2$ avrà coordinate di **TL** e **BR** corrispondenti a: $(\min(x_{0,B1}, x_{0,B2}), \min(y_{0,B1}, y_{0,B2}))$, $(\max(x_{1,B1}, x_{1,B2}), \max(y_{1,B1}, y_{1,B2}))$.

of recombinant insulin could be completely avoided or the quantity reduced.

Although the variables that clinically influence the efficiency of each transplant are high in numbers so that a direct relationship between IEQ/kg and C-peptide production is difficult to demonstrate, we can agree that the previously determined threshold of 2500–3000 IEQ/kg is associated with most successful clinical results [18].

From the islet allotransplantation setting, we also learned that even minimal quantities of insulin and C-peptide can reduce, respectively, hypoglycemia unawareness and its frequently tragic consequences, as well as the blood vessel atherosclerotic deterioration, principal cause of all typical diabetes complications [25, 26]. Also, because it is an auto-transplant, the absence of a need for immunosuppression allows not only for a better engraftment of the islets, but also for a potentially longer graft survival time than is observed for the islets used in allotransplantation. The improved survival is due in part to the absence of rejection, and in part to the fact that the most used immunosuppressive regimens include drugs that are toxic for the insulin-producing beta cells [27].

The frequently less than lifelong survival of the graft might not be due entirely to the stress that the isolation protocol itself imposes on the islets, but also to the less than ideal site of the infusion. The sinusoids of the liver, while optimal to guarantee blood exposure to the islets and a proper geographical location—being a part of the portal system in which insulin is physiologically secreted—might not be the most suited site for properly lodging them. It is demonstrated that the islets transplanted into the liver accumulate amyloid deposits that, not only damage the liver's

of recombinant insulin could be completely avoided or the quantity reduced.

Although the variables that clinically influence the efficiency of each transplant are high in numbers so that a direct relationship between IEQ/kg and C-peptide production is difficult to demonstrate, we can agree that the previously determined threshold of 2500–3000 IEQ/kg is associated with most successful clinical results [18].

From the islet allotransplantation setting, we also learned that even minimal quantities of insulin and C-peptide can reduce, respectively, hypoglycemia unawareness and its frequently tragic consequences, as well as the blood vessel atherosclerotic deterioration, principal cause of all typical diabetes complications [25, 26]. Also, because it is an auto-transplant, the absence of a need for immunosuppression allows not only for a better engraftment of the islets, but also for a potentially longer graft survival time than is observed for the islets used in allotransplantation. The improved survival is due in part to the absence of rejection, and in part to the fact that the most used immunosuppressive regimens include drugs that are toxic for the insulin-producing beta cells [27].

The frequently less than lifelong survival of the graft might not be due entirely to the stress that the isolation protocol itself imposes on the islets, but also to the less than ideal site of the infusion. The sinusoids of the liver, while optimal to guarantee blood exposure to the islets and a proper geographical location—being a part of the portal system in which insulin is physiologically secreted—might not be the most suited site for properly lodging them. It is demonstrated that the islets transplanted into the liver accumulate amyloid deposits that, not only damage the liver's

(a) *ground truth*

(b) nessun post-processing

of recombinant insulin could be completely avoided or the quantity reduced.

Although the variables that clinically influence the efficiency of each transplant are high in numbers so that a direct relationship between IEQ/kg and C-peptide production is difficult to demonstrate, we can agree that the previously determined threshold of 2500–3000 IEQ/kg is associated with most successful clinical results [18].

From the islet allotransplantation setting, we also learned that even minimal quantities of insulin and C-peptide can reduce, respectively, hypoglycemia unawareness and its frequently tragic consequences, as well as the blood vessel atherosclerotic deterioration, principal cause of all typical diabetes complications [25, 26]. Also, because it is an auto-transplant, the absence of a need for immunosuppression allows not only for a better engraftment of the islets, but also for a potentially longer graft survival time than is observed for the islets used in allotransplantation. The improved survival is due in part to the absence of rejection, and in part to the fact that the most used immunosuppressive regimens include drugs that are toxic for the insulin-producing beta cells [27].

The frequently less than lifelong survival of the graft might not be due entirely to the stress that the isolation protocol itself imposes on the islets, but also to the less than ideal site of the infusion. The sinusoids of the liver, while optimal to guarantee blood exposure to the islets and a proper geographical location—being a part of the portal system in which insulin is physiologically secreted—might not be the most suited site for properly lodging them. It is demonstrated that the islets transplanted into the liver accumulate amyloid deposits that, not only damage the liver's

of recombinant insulin could be completely avoided or the quantity reduced.

Although the variables that clinically influence the efficiency of each transplant are high in numbers so that a direct relationship between IEQ/kg and C-peptide production is difficult to demonstrate, we can agree that the previously determined threshold of 2500–3000 IEQ/kg is associated with most successful clinical results [18].

From the islet allotransplantation setting, we also learned that even minimal quantities of insulin and C-peptide can reduce, respectively, hypoglycemia unawareness and its frequently tragic consequences, as well as the blood vessel atherosclerotic deterioration, principal cause of all typical diabetes complications [25, 26]. Also, because it is an auto-transplant, the absence of a need for immunosuppression allows not only for a better engraftment of the islets, but also for a potentially longer graft survival time than is observed for the islets used in allotransplantation. The improved survival is due in part to the absence of rejection, and in part to the fact that the most used immunosuppressive regimens include drugs that are toxic for the insulin-producing beta cells [27].

The frequently less than lifelong survival of the graft might not be due entirely to the stress that the isolation protocol itself imposes on the islets, but also to the less than ideal site of the infusion. The sinusoids of the liver, while optimal to guarantee blood exposure to the islets and a proper geographical location—being a part of the portal system in which insulin is physiologically secreted—might not be the most suited site for properly lodging them. It is demonstrated that the islets transplanted into the liver accumulate amyloid deposits that, not only damage the liver's

(c) inclusione

(d) intersezione

Figura 4.1: Differenza tra inclusione ed intersezione eseguite sullo stesso paragrafo.

Le regioni così individuate delimitano le diverse entità che fanno parte della pagina analizzata. In genere è stata osservata una tendenza dei modelli ad eliminare un buon numero di archi positivi, soprattutto all'aumentare del valore di *threshold*,

ciò può portare al fenomeno di *oversplitting*, per cui invece di identificare un elemento nella sua totalità, questo viene divisa in delle regioni separate più piccole. Il fenomeno è più comune nelle zone a bassa densità di connessioni, come nei titoli e nelle tabelle, ma persiste talvolta anche nei paragrafi (Figura 4.2).

Quando il modello è addestrato ad eseguire anche *node classification* le etichette assegnate ai nodi possono essere utilizzate sia per classificare le componenti connesse individuate, sia per eseguire un *pruning* aggiuntivo del grafo. La classificazione di una componente connessa avviene assegnandovi l'etichetta più frequente nei nodi che la compongono mentre l'utilizzo delle *labels* per il taglio degli archi prevede di rimuovere tutti gli archi (u, v) in cui l'etichetta di u è diversa da quella di v .

¹ Institute of Plastic and Craniofacial Surgery, SOBRAPAR, Campinas, SP, Brazil

² Plastic Surgery Division, Department of Surgery, Universidade de São Paulo (USP), São Paulo, SP, Brazil

Correspondence should be addressed to Cassio Eduardo Raposo-Amaral, cassioraposo@hotmail.com

Received 22 July 2012 | Accepted 1 December 2012

Academic Editor: Renato Da Silva Freitas

Figura 4.2: Esempio di *oversplitting*.

4.2.1 Metriche

Le metriche utilizzate per valutare il modello sono state:

- **Area Under The Receiver Operating Characteristics Curve o ROC AUC** per l'operazione di *link prediction*.
- **Balanced Accuracy** per il compito di *node classification*.
- **Mean Average Precision o mAP** per valutare l'inferenza sul *test set*.

Le prime due metriche sono definite a partire dalla **Sensitività**, che esprime la frazione di esempi positivi correttamente individuati dal modello, e **Specificità**,

che equivale invece alla frazione di esempi negativi classificati correttamente. Indicando con **TP** il numero di Veri Positivi, con **FP** i Falsi Positivi, con **TN** i Veri Negativi e con **FN** i Falsi Negativi si ha:

$$\text{Sensitività} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FN}} \quad (4.5)$$

$$\text{Specificità} = \frac{\mathbf{TN}}{\mathbf{TN} + \mathbf{FP}} \quad (4.6)$$

è utile inoltre definire la frazione di falsi positivi FPR come:

$$\text{FPR} = 1 - \text{Specificità} = \frac{\mathbf{FP}}{\mathbf{TN} + \mathbf{FP}} \quad (4.7)$$

ROC AUC

Per calcolare la ROC AUC (o semplicemente AUC) è necessario tracciare la curva ROC (Figura 4.3). Per farlo si considera innanzitutto un vettore p tale che p_i esprima la probabilità calcolata dal modello che l'istanza i -esima sia positiva e si fissa una soglia (o *threshold*) t . Tutti gli esempi i con $p_i \geq t$ saranno classificati come positivi, viceversa a tutti quelli per i quali $p_i < t$ sarà assegnata la classe negativa. Ordinando in modo decrescente gli elementi del vettore p è possibile individuare tutte le possibili soglie che separano due elementi consecutivi nel vettore ordinato, tracciando i valori corrispondenti di FPR e Sensitività rispettivamente sull'asse delle ascisse e sull'asse delle ordinate si ottiene la curva ROC, l'AUC non è altro che l'area sottesa da tale curva.

L'AUC aiuta a rendere l'idea di quanto il modello in questione sia in grado di distinguere tra le due classi, il minimo valore possibile è 0 mentre il massimo è 1. Pur essendo una metrica adatta a problemi sbilanciati, è stato difficile valutare le prestazioni effettive dei modelli esclusivamente tramite AUC, molto probabilmente perché nel problema affrontato lo sbilanciamento tra classi positive e negative è

eccessivo anche per una metrica del genere. Per essere più corretti: un AUC maggiore riflette solitamente una capacità di inferenza migliore, ma bastano pochi punti percentuali di differenza per variare notevolmente i risultati ottenuti³.

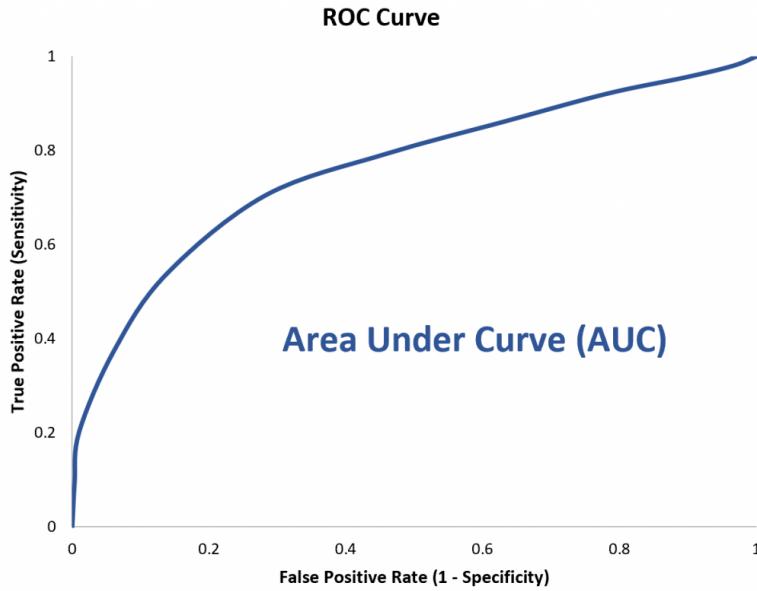


Figura 4.3: Un esempio di curva ROC [18].

Balanced Accuracy

La Balanced Accuracy è calcolata in questo modo:

$$\text{Balanced Accuracy} = \frac{\text{Sensibilità} + \text{Specificità}}{2} \quad (4.8)$$

è una metrica adatta a problemi di classificazione multiclassse sbilanciati, il minimo valore possibile è 0 ed il massimo è 1. Tale metrica permette di ottenere dei valori maggiormente rappresentativi dello stato del modello rispetto alla semplice **Accuracy**, che è invece calcolata come:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4.9)$$

³Per confronti tra risultati quantitativi si consulti il paragrafo 5.1.

Mean Average Precision

Per comparare le prestazioni del modello con altri sistemi già esistenti è stato inoltre calcolato un valore di **Mean Average Precision** o mAP. A differenza delle metriche precedenti questa è stata utilizzata per valutare le prestazioni a livello di inferenza, quindi solo dopo aver individuato le *bounding boxes* dei diversi elementi di una pagina. Prima di definire in cosa consiste la metrica di mAP è necessario introdurre i concetti di **Precision**, **Recall** ed **Intersection over Union** (o **IoU**).

Precision e Recall

Se assumiamo di possedere un insieme di predizioni ed un insieme di *ground truths*, ovvero di valori corretti per ogni esempio, possiamo definire:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4.10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4.11)$$

Il valore di Precision indica la proporzione di esempi positivi correttamente individuati su tutti gli esempi che sono stati classificati come positivi. Quello di Recall equivale invece alla proporzione di positivi individuati su tutti gli esempi che sono effettivamente positivi nell'insieme di *ground truth*.

IoU

Per applicare le nozioni di Precision e Recall a problemi di Object Detection, in cui è necessario individuare le *bounding boxes* dei diversi oggetti di interesse presenti in un'immagine, è possibile utilizzare il concetto di Intersection over Union. In particolare considerate due *bounding boxes*, una estratta dall'insieme di predizioni, l'altra dall'insieme di *ground truth*, si può calcolare il valore di IoU come:

$$\text{Intersection over Union} = \frac{\text{Area dell'Intersezione}}{\text{Area dell'Unione}} \quad (4.12)$$

A questo punto è possibile calcolare i punteggi di Precision e Recall fissando una soglia t : se le due *bounding boxes* hanno un valore di IoU superiore a t si conteggia un Vero Positivo. Se invece si registra un valore di IoU inferiore a t si conteggia un Falso Positivo, lo stesso avviene se esiste una *bounding box* nell'insieme di predizioni che non interseca alcuna *bounding box* nell'insieme di *ground truth*. Infine se per una *bounding box* dell'insieme di *ground truth* non esiste alcuna *bounding box* nell'insieme di predizioni che la interseca si conteggia un Falso Negativo.

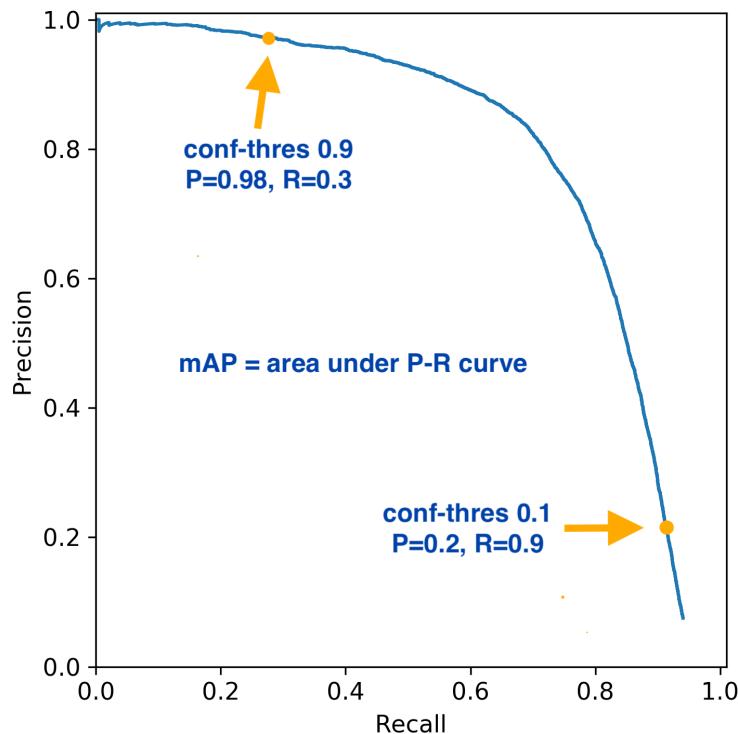


Figura 4.4: Un esempio di curva Precision-Recall [5]

mAP

I valori di Precision e Recall dipendono quindi dalla soglia t , pertanto è possibile tracciare una curva Precision-Recall rappresentando i valori di Recall sull'asse delle ascisse e i valori di Precision sull'asse delle ordinate per diversi valori di t . Per le valutazioni condotte sono state utilizzate le seguenti soglie: <0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95>. L'mAP equivale all'area sottesa dalla curva Precision-Recall costruita in questo modo (Figura 4.4).

Considerazioni

La scelta della metrica mAP per la valutazione della capacità di inferenza può essere messa in discussione, se non altro perché i modelli addestrati non eseguono *Object Detection* e ricavano le *bounding boxes* degli elementi di una pagina a partire dai nodi che formano le componenti connesse. Inoltre le *bounding boxes* di *ground truth* utilizzate per la valutazione corrispondono a quelle individuate dalle annotazioni di PubLayoutNet e PubTables-1M, le quali sono ottenute tramite euristiche e pertanto tendono a frammentare alcuni elementi, come i paragrafi, distinguendoli in più regioni separate, ulteriori riflessioni a riguardo sono esposte nel paragrafo **5.2.1**. Tuttavia non è stato individuato alcun modello che si comportasse in modo simile a quelli sviluppati nel corso degli esperimenti, è stato quindi necessario confrontarsi con sistemi che eseguono operazioni di *Object Detection* su documenti scientifici e che si basano su tecniche di OCR, per i quali la metrica mAP è uno standard adeguato.

Come metro di paragone è stato utilizzato il modello denominato “faster_rcnn_R_50_FPN_3x” pre-addestrato su circa 7000 pagine provenienti dalla collezione di PubLayoutNet e proposto da Layout Parser⁴: un *toolkit* che mette a disposizione una varietà di strumenti per facilitare le operazioni di *Document*

⁴Il modello è recuperabile qui: <https://layout-parser.readthedocs.io/en/latest/notes/modelzoo.html>

Image Analysis, tra cui un *repository* di modelli pre-addestrati a compiti di *Document Layout Analysis* [11]. Infine poichè l'insieme di *ground truth* utilizzato non contiene le regioni relative alle annotazioni *other*, esse sono state rimosse dall'insieme di predizioni al fine di effettuare la valutazione tramite mAP.

Capitolo 5

Risultati

5.1 Risultati quantitativi

Di seguito sono riportati i risultati ottenuti dagli esperimenti condotti. La soglia t^* utilizzata in fase di inferenza per distinguere archi postivi da negativi è stata ottenuta tracciando, per ogni modello, la curva ROC (introdotta nel paragrafo 4.2.1) relativa al *training set* e scegliendo t^* in modo da soddisfare¹:

$$t^* = \min_t |\text{Sensibilità}(t) + FPR(t) - 1| \quad (5.1)$$

Ogni tabella si riferisce ad una specifica architettura della rete e connettività dei grafi, in particolare per ogni tabella sono specificati:

- **Conn:** il tipo di connettività usata per i grafi.
- **GCN L:** il numero di layer nascosti usati nella GCN.
- **HU:** il numero di unità utilizzate per ogni layer nascosto della GCN e dell'MLP usato come predittore.

¹L'equazione utilizzata corrisponde ad un metodo piuttosto comune per tarare la soglia t^* [6].

- **Dim P:** la dimensione degli spazi in cui sono proiettate le *features* dei nodi prima di essere mandate in input alla GCN.
- **Aggr:** indica come il MLP aggrega gli *embeddings* dei nodi prima della predizione, “cat” indica concatenazione, “sum” indica somma.
- **NC:** indica se si esegue anche *node classification* o meno.

Di seguito si offre una spiegazione delle colonne utilizzate:

- **Nome:** è l’identificatore specifico assegnato ad ogni riga.
- **mAP:** riporta il punteggio di mAP ottenuto con la configurazione specificata dalla riga. L’mAP è espresso in percentuale.
- **AUC:** riporta il punteggio di ROC AUC ottenuto dal modello.
- **bal.:** se “sì” indica che l’addestramento è stato effettuato sui grafi bilanciati, altrimenti significa che l’unica tecnica di bilanciamento adottata è stata la rimozione delle isole.
- **merge:** indica la tipologia di risoluzione dei conflitti utilizzata (intersezione o inclusione).
- **features:** indica le *features* utilizzate.
- **label prun.:** presente solo in caso di *node classification*, se “sì” indica che sono stati tagliati tutti gli archi che collegavano due nodi classificati in modo diverso.
- **acc:** presente solo in caso di *node classification*, riporta il punteggio di *balanced accuracy* ottenuto dal modello.

N.B.: Tutti i risultati qui riportati sono stati ottenuti dividendo le *features* nei sottovettori: <geometriche>, <istogramma + stile + intersezione>, <testuali> e proiettando ciascuno di essi in uno spazio di dimensione **Dim P** secondo la procedura descritta nel paragrafo **3.3.1**.

Tabella 5.1: **Conn:** visibility, **GCN L:** 4, **HU:** 256, **Dim P:** 100,**Aggr:** cat, **NC:** no

Nome	mAP	AUC	bal.	merge	features
faster rcnn	81.86				
A	47.39	0.9832	sì	inclusion	bbox + <u>scibert</u> + hist + style + inters
B	45.73	0.9832	sì	intersection	bbox + <u>scibert</u> + hist + style + inters
C	26.76	0.9688	no	inclusion	bbox + <u>scibert</u> + hist + style + inters
D	26.18	0.9688	no	intersection	bbox + <u>scibert</u> + hist + style + inters

Tabella 5.2: **Conn:** knn, **GCN L:** 4, **HU:** 256, **Dim P:** 100,**Aggr:** sum, **NC:** no

Nome	mAP	AUC	bal.	merge	features
faster rcnn	81.86				
E	28.55	0.9580	sì	inclusion	bbox + <u>scibert</u> + hist + style + inters
F	30.60	0.9580	sì	intersection	bbox + <u>scibert</u> + hist + style + inters
G	31.62	0.9527	no	inclusion	bbox + <u>scibert</u> + hist + style + inters
H	32.66	0.9527	no	intersection	bbox + <u>scibert</u> + hist + style + inters
I	31.69	0.9513	sì	inclusion	bbox + <u>spacy</u> + hist + style + inters
J	31.99	0.9513	sì	intersection	bbox + <u>spacy</u> + hist + style + inters
K	34.76	0.9470	no	inclusion	bbox + <u>spacy</u> + hist + style + inters
L	31.82	0.9470	no	intersection	bbox + <u>spacy</u> + hist + style + inters

Tabella 5.3: **Conn:** visibility, **GCN L:** 4, **HU:** 256, **Dim P:** 100,**Aggr:** sum, **NC:** no

Nome	mAP	AUC	bal.	merge	features
faster rcnn	81.86				
M	39.61	0.9789	sì	inclusion	bbox + <u>scibert</u> + hist + style + inters
N	38.59	0.9789	sì	intersection	bbox + <u>scibert</u> + hist + style + inters
O	43.55	0.9807	no	inclusion	bbox + <u>scibert</u> + hist + style + inters
P	42.61	0.9807	no	intersection	bbox + <u>scibert</u> + hist + style + inters
Q	40.99	0.9700	sì	inclusion	bbox + <u>spacy</u> + hist + style + inters
R	38.89	0.9700	sì	intersection	bbox + <u>spacy</u> + hist + style + inters

Tabella 5.4: **Conn:** knn, **GCN L:** 4, **HU:** 256, **Dim P:** 100,**Aggr:** sum, **NC:** sì

Nome	mAP	AUC	bal.	merge	features	label prun.	acc
faster rcnn	81.86						
S	18.60	0.9291	sì	inclusion	bbox + <u>scibert</u> + hist + style + inters	no	0.8641
T	26.24	0.9291	sì	intersection	bbox + <u>scibert</u> + hist + style + inters	no	0.8641
U	11.15	0.9291	sì	inclusion	bbox + <u>scibert</u> + hist + style + inters	sì	0.8641
V	25.14	0.9291	sì	intersection	bbox + <u>scibert</u> + hist + style + inters	sì	0.8641
W	16.00	0.9265	no	inclusion	bbox + <u>scibert</u> + hist + style + inters	no	0.8491
X	24.30	0.9265	no	intersection	bbox + <u>scibert</u> + hist + style + inters	no	0.8491
Y	10.72	0.9265	no	inclusion	bbox + <u>scibert</u> + hist + style + inters	sì	0.8491
Z	23.93	0.9265	no	intersection	bbox + <u>scibert</u> + hist + style + inters	sì	0.8491

Tabella 5.5: **Conn:** knn, **GCN L:** 4, **HU:** 256, **Dim P:** 100,**Aggr:** sum, **NC:** sì

Nome	mAP	AUC	bal.	merge	features	label prun.	acc
faster rcnn	81.86						
AA	19.44	0.9299	sì	inclusion	bbox + <u>spacy</u> + hist + style + inters	no	0.8217
AB	25.67	0.9299	sì	intersection	bbox + <u>spacy</u> + hist + style + inters	no	0.8217
AC	13.56	0.9299	sì	inclusion	bbox + <u>spacy</u> + hist + style + inters	sì	0.8217
AD	23.98	0.9299	sì	intersection	bbox + <u>spacy</u> + hist + style + inters	sì	0.8217
AE	18.24	0.9271	no	inclusion	bbox + <u>spacy</u> + hist + style + inters	no	0.8521
AF	23.94	0.9271	no	intersection	bbox + <u>spacy</u> + hist + style + inters	no	0.8521
AG	11.67	0.9271	no	inclusion	bbox + <u>spacy</u> + hist + style + inters	sì	0.8521
AH	22.67	0.9271	no	intersection	bbox + <u>spacy</u> + hist + style + inters	sì	0.8521

Tabella 5.6: **Conn:** visibility, **GCN L:** 4, **HU:** 256, **Dim P:** 100,**Aggr:** sum, **NC:** sì

Nome	mAP	AUC	bal.	merge	features	label prun.	acc
faster rcnn	81.86						
AI	36.24	0.9642	sì	inclusion	bbox + <u>scibert</u> + hist + style + inters	no	0.8298
AJ	37.66	0.9642	sì	intersection	bbox + <u>scibert</u> + hist + style + inters	no	0.8298
AK	26.56	0.9642	sì	inclusion	bbox + <u>scibert</u> + hist + style + inters	sì	0.8298
AL	34.85	0.9642	sì	intersection	bbox + <u>scibert</u> + hist + style + inters	sì	0.8298
AM	37.36	0.9650	no	inclusion	bbox + <u>scibert</u> + hist + style + inters	no	0.8556
AN	36.59	0.9650	no	intersection	bbox + <u>scibert</u> + hist + style + inters	no	0.8556
AO	20.21	0.9650	no	inclusion	bbox + <u>scibert</u> + hist + style + inters	sì	0.8556
AP	33.20	0.9650	no	intersection	bbox + <u>scibert</u> + hist + style + inters	sì	0.8556

Tabella 5.7: **Conn:** visibility, **GCN L:** 4, **HU:** 256, **Dim P:** 100,**Aggr:** sum, **NC:** sì

Nome	mAP	AUC	bal.	merge	features	label prun.	acc
faster rcnn	81.86						
AQ	29.39	0.9605	sì	inclusion	bbox + <u>spacy</u> + hist + style + inters	no	0.8429
AR	31.27	0.9605	sì	intersection	bbox + <u>spacy</u> + hist + style + inters	no	0.8429
AS	20.80	0.9605	sì	inclusion	bbox + <u>spacy</u> + hist + style + inters	sì	0.8429
AT	28.59	0.9605	sì	intersection	bbox + <u>spacy</u> + hist + style + inters	sì	0.8429
AU	31.90	0.9714	no	inclusion	bbox + <u>spacy</u> + hist + style + inters	no	0.8566
AV	32.70	0.9714	no	intersection	bbox + <u>spacy</u> + hist + style + inters	no	0.8566
AW	23.55	0.9714	no	inclusion	bbox + <u>spacy</u> + hist + style + inters	sì	0.8566
AX	30.30	0.9714	no	intersection	bbox + <u>spacy</u> + hist + style + inters	sì	0.8566

Tabella 5.8: **Conn:** visibility, **GCN L:** 4, **HU:** 256, **Dim P:** 100,**Aggr:** cat, **NC:** sì

Nome	mAP	AUC	bal.	merge	features	label prun.	acc
faster rcnn	81.86						
AY	36.27	0.9679	sì	inclusion	bbox + <u>spacy</u> + hist + style + inters	no	0.8145
AZ	40.86	0.9679	sì	intersection	bbox + <u>spacy</u> + hist + style + inters	no	0.8145
AAA	23.35	0.9679	sì	inclusion	bbox + <u>spacy</u> + hist + style + inters	sì	0.8145
AAB	36.10	0.9679	sì	intersection	bbox + <u>spacy</u> + hist + style + inters	sì	0.8145
AAC	30.51	0.9746	no	inclusion	bbox + <u>spacy</u> + hist + style + inters	no	0.8178
AAD	33.26	0.9746	no	intersection	bbox + <u>spacy</u> + hist + style + inters	no	0.8178
AAE	22.05	0.9746	no	inclusion	bbox + <u>spacy</u> + hist + style + inters	sì	0.8178
AAF	29.59	0.9746	no	intersection	bbox + <u>spacy</u> + hist + style + inters	sì	0.8178

Tabella 5.9: **Conn:** visibility, **GCN L:** 4, **HU:** 256, **Dim P:** 100,**Aggr:** cat, **NC:** sì

Nome	mAP	AUC	bal.	merge	features	label prun.	acc
faster rcnn	81.86						
AAG	31.24	0.9695	sì	inclusion	bbox + <u>scibert</u> + hist + style + inters	no	0.8333
AAH	33.73	0.9695	sì	intersection	bbox + <u>scibert</u> + hist + style + inters	no	0.8333
AAI	19.64	0.9695	sì	inclusion	bbox + <u>scibert</u> + hist + style + inters	sì	0.8333
AAJ	30.23	0.9695	sì	intersection	bbox + <u>scibert</u> + hist + style + inters	sì	0.8333
AAK	41.26	0.9613	no	inclusion	bbox + <u>scibert</u> + hist + style + inters	no	0.8466
AAL	41.73	0.9613	no	intersection	bbox + <u>scibert</u> + hist + style + inters	no	0.8466
AAM	24.79	0.9613	no	inclusion	bbox + <u>scibert</u> + hist + style + inters	sì	0.8466
AAN	37.22	0.9613	no	intersection	bbox + <u>scibert</u> + hist + style + inters	sì	0.8466

5.2 Commenti

5.2.1 mAP

È immediato notare che, nonostante gli elevati valori di AUC, i modelli presentati non riescono a raggiungere un mAP che si avvicini a quello ottenuto con il modello “faster_rcnn_R_50_FPN_3x” (per semplicità *faster rcnn*). Ciò è imputabile ai seguenti fattori:

- Una possibile inadeguatezza della misura di mAP per valutare le prestazioni del modello, dovuta principalmente alla sostanziale differenza tra *faster rcnn* e i modelli addestrati. Il primo infatti esegue operazioni di *Object Detection*, pertanto è difficile che le *bounding boxes* individuate da esso corrispondano esattamente a quelle di *ground truth*, tuttavia anche con una modesta quantità di errori esse otteranno comunque dei valori di Intersection over Union sufficienti a considerare i diversi elementi della pagina come correttamente identificati. Al contrario, nel caso dei nostri modelli, le *bounding boxes* sono ricavate a partire da quelle dei nodi, ottenute tramite PyMuPDF, e si riesce quindi a raggiungere una precisione maggiore nella delineazione dei confini di una regione (figura 5.3). Ciò fa sì che quando una regione individuata da un nostro modello viene conteggiata come un Vero Positivo il suo valore di IoU sarà prossimo ad 1, cosa di cui la misura di mAP non tiene conto, in quanto è sufficiente che l’IoU sia superiore ad una certa soglia t affinché si conteggi un Vero Positivo. Inoltre, poichè basta un singolo arco mal classificato per variare notevolmente il numero di nodi in una componente connessa, un unico errore può causare l’individuazione di *bounding boxes*

molto più grandi o molto più piccole del dovuto (figura 5.1), che ottengono valori di IoU particolarmente bassi e aumentano il numero di Falsi Negativi. Anche se quest'ultima problematica può essere indubbiamente imputata ad una generica fragilità dei modelli costruiti, che sono molto sensibili agli errori di classificazione, si deve comunque considerare che, a giudicare dai risultati qualitativi, la maggior parte delle volte gli archi negativi mal classificati uniscono due parole appartenenti allo *stesso* paragrafo, ciò si ricollega alla problematica successiva.

- Le regioni di *ground truth* sono ricavate dalle annotazioni di PubLayNet, le quali tendono a frammentare i paragrafi in più elementi, individuando il confine di una *bounding box* ogni volta l'autore vada a capo. Per modelli basati su OCR è più facile individuare le regioni in cui ciò avviene, mentre senza l'utilizzo di informazioni visive risulta assai più complicato, si veda ad esempio la figura 5.2. Inoltre se da una parte i nostri modelli sono addestrati a partire dalle annotazioni di PubLayNet, e quindi in linea teorica ci si aspetterebbe che fossero in grado di ricalcarle, dall'altra si può discutere sul fatto che un singolo paragrafo debba essere effettivamente diviso in più regioni o meno. Se si assume come obiettivo quello di individuare i paragrafi, senza necessariamente fare caso a quando si va a capo, i modelli presentati otterrebbero ragionevolmente delle valutazioni migliori, anche se, a causa dell'assenza di *ground truths* che riflettano questo obiettivo, non abbiamo ancora sufficienti risultati quantitativi per giustificare tale ipotesi.

5.2.2 Node Classification e Label Pruning

Si può osservare come i risultati ottenuti rimuovendo gli archi in base alle etichette dei nodi (i. e. voce *label pruning* nelle tabelle) siano consistentemente peggiori

di quelli raggiunti senza questa operazione. Ciò è dovuto principalmente al fatto che spesso il classificatore compie errori su nodi vicini tra loro. Tendono quindi a formarsi dei *cluster* di vertici mal classificati che generano frammentazioni eccessive (Figura 5.7). In questi casi l'utilizzo dell'intersezione per la risoluzione dei conflitti si dimostra leggermente migliore dell'inclusione, ma è necessario approntare ulteriori migliorie affinchè le *labels* dei nodi diventino un criterio valido per la potatura degli archi.

5.2.3 Altre considerazioni

Dai risultati quantitativi emerge che la connettività *visibility* risulta una scelta migliore rispetto a quella knn. Inoltre le *features* testuali Scibert forniscono in genere risultati maggiormente apprezzabili rispetto a quelle SpaCy, ed infatti l'mAP più alto è attribuito proprio al modello A, che sfrutta grafi *visibility* con *features* Scibert.

È interessante notare come, nonostante il forte sbilanciamento del problema, addestrare su grafi non bilanciati fornisca spesso risultati solo leggermente inferiori rispetto a quando si bilancia il *training set*. Rimane da indagare se ciò è dovuto ad una generale inefficienza della metodologia di bilanciamento adottata o, al contrario, ad una buona capacità di apprendimento e generalizzazione anche in caso di addestramento su esempi non bilanciati.

5.3 Risultati qualitativi - Alcuni esempi

Si riportano infine alcuni dei risultati qualitativi ottenuti. Si noti che quando viene eseguita *node classification* le regioni del documento individuate vengono anche classificate. Al momento della visualizzazione del risultato ad ogni classe viene fatto corrispondere un colore come indicato di seguito:

- **Classe:** *text*, **Colore:** rosso.
- **Classe:** *title*, **Colore:** verde.
- **Classe:** *list*, **Colore:** blu.
- **Classe:** *table*, **Colore:** giallo.
- **Classe:** *caption*, **Colore:** celeste.
- **Classe:** *image*, **Colore:** magenta.
- **Classe:** *other*, **Colore:** grigio.

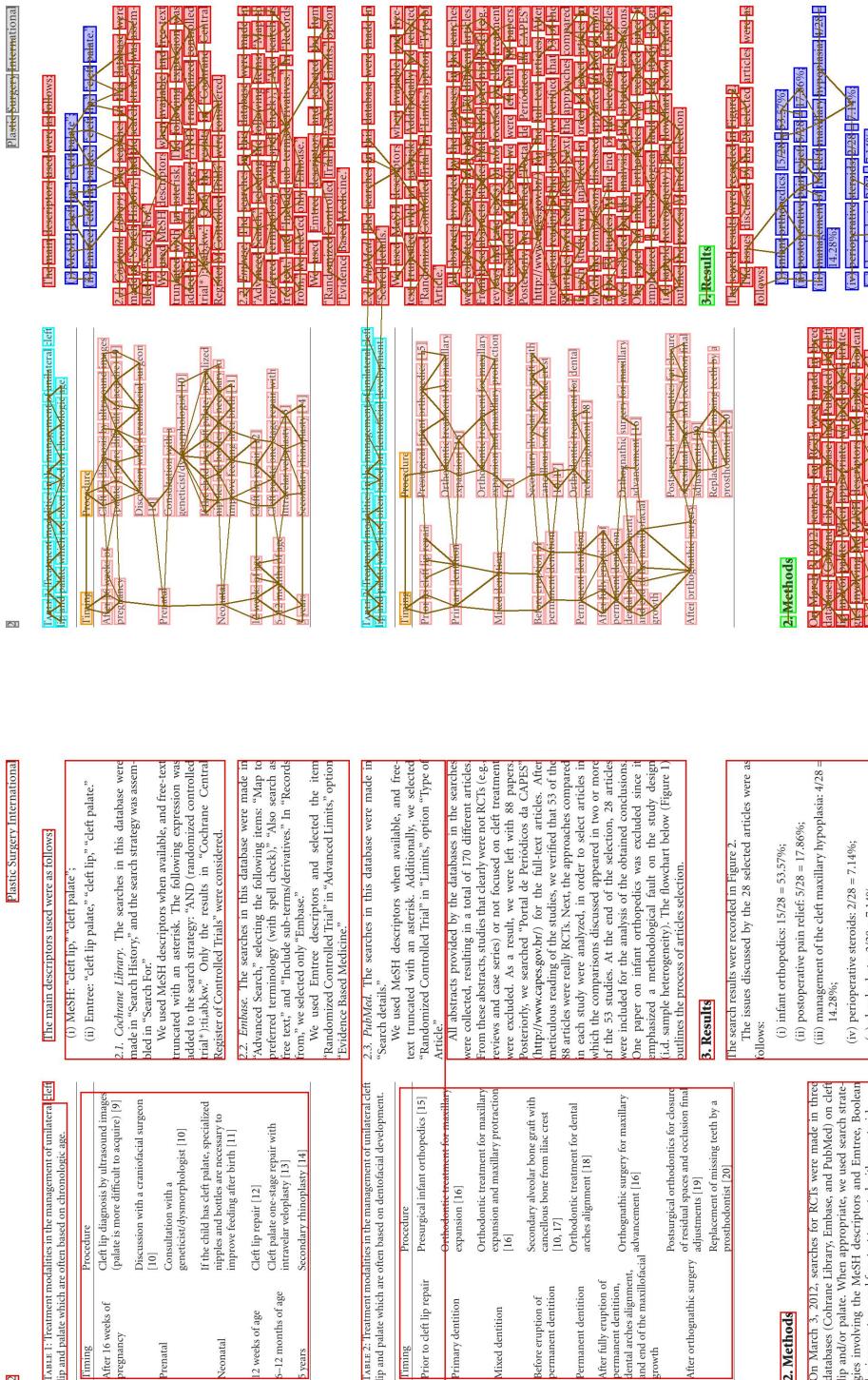


Figura 5.1: Esempio di errore dovuto a pochi archi mal classificati: La *caption* della tabella 2 viene unita al paragrafo a destra.

N.B.: Il modello A non esegue *node classification*, i nodi sono stati colorati solo per una migliore visualizzazione.

(a) *Bounding boxes* identificate.

(b) Grafo relativo alla pagina.

TABLE 1: Targeting MAbS for TH1 and target tissue

Targeting MAb	Target receptor	Experimental model and target tissue	Target receptor	Experimental model and target tissue
Murine OX26 [53]	Rat TIR	(a) Rat Ce-6/RC2 glomeruli in culture (b) Rat Ce-390 in cultures (c) In vivo transport via rat BBB and rat brain cells (neuron and glial). Gene delivery	(a) Rat Ce-6/RC2 glomeruli (b) In vivo transport via rat BBB and rat brain cells (neuron and glial). Gene delivery	
Rat 8D3 [54]	Mouse TIR	In vivo transport via rat BBB and rat brain cells (neuron and glial). Gene delivery		
Murine 8314 [60]	Human IR	(a) Human U87 glioma cultures (b) In vivo transport via primate/human BBB and brain cells (neuron and glial). Gene delivery	(a) Human U87 glioma cultures (b) In vivo transport via primate/human BBB and brain cells (neuron and glial). Gene delivery	
8D3 + 8314 [23]	Mouse TIR + human IR	(a) Experimental human brain tumors in scid mice (b) In vivo transport via mouse BBB and brain primate/human cells (neuron and glial). Gene delivery	(a) Experimental human brain tumors in scid mice (b) In vivo transport via mouse BBB and brain primate/human cells (neuron and glial). Gene delivery	
Chimeric anti-TIR [55]	Mouse TIR			
Humanized anti-TIR [26]	Human IR			

MAbs has been developed (Table 1), and their efficacy in delivering TH1s to brain has been demonstrated in experimental animal model *in vivo* [1, 3, 4, 19–25, 27]. The 83-14 murine MAb to the TIR and the OX26 murine MAb to the rat TIR are used to target human and rat tissues, respectively (Table 1). The OX26 TRIMAb is active only in rats, so the rat 8D3 TIR against the mouse TIR is used in mice (Table 1) [20, 21, 27–33]. The 83-14 HIRMAb does not cross-react with the insulin receptor in rodents or even New World primates such as the squirrel monkey. There are 2 potential mechanisms for the decline in the expression of this transgene, that is, promoter inactivation and plasmid degradation. The levels of both luciferase enzyme activity and plasmid DNA decay in the primate brain and liver were measured, and both processes decayed with a half-life of approximately 2 days, which indicates that the transfection of the luciferase gene expression is mainly due to plasmid degradation [33].

The organ distribution of the lacZ transgene was also investigated at the cellular level with histochemistry following *in vivo* applications of TH1s. The SV40 promoter enables the sequential receptor-mediated transcytosis of the TH1 across the BBB followed by the receptor-mediated endocytosis of the TH1 into the brain cell (Figure 1c). TH1s have also been successfully constructed to target human tumor cells in a scid mouse model wherein dual targeting MAbs were directed to the mouse TIR and HIR (Table 1 and Figure 1d), i.e., MAb1 and MAb2 [23].

3. Brain Expression of Reporter Genes

In *in vivo* applications of TH1s were initially investigated with luciferase and lacZ reporter genes *in vivo* [3, 4, 20, 21, 34]. TH1s were constructed with the expression plasmid of interest (i.e., luciferase reporter gene) and engineered with either the TRIMAb for rodents or the HIRMAb for Rhesus monkeys, respectively. The doses of DNA encapsulated in TH1s and administered IV was 5 or 70 μ g per rat or primate respectively, which are equivalent 20 and 12 μ g/kg body weight, respectively. When the transgene is driven by the widely read SV40 promoter, the levels of luciferase were ~10 ng luciferase/mg protein in the monkey brain. High levels of expression were also seen in peripheral tissues that are rich in the target receptor, including liver, spleen, and lung [27, 34]. A 50-fold increase in the tissue levels

3

TABLE 1: Targeting MAbs for TH1 and target tissue

Targeting MAb	Target receptor	Experimental model and target tissue	Target receptor	Experimental model and target tissue
Murine OX26 [54]	Rat TIR	(a) Rat Ce-6/RC2 glomeruli in culture (b) Rat Ce-390 in cultures (c) In vivo transport via rat BBB and rat brain cells (neuron and glial). Gene delivery	(a) Rat Ce-6/RC2 glomeruli (b) In vivo transport via rat BBB and rat brain cells (neuron and glial). Gene delivery	
Rat 8D3 [54]	Mouse TIR	In vivo transport via rat BBB and rat brain cells (neuron and glial). Gene delivery		
Murine 8314 [60]	Human IR	(a) Human U87 glioma cultures (b) In vivo transport via primate/human BBB and brain cells (neuron and glial). Gene delivery	(a) Human U87 glioma cultures (b) In vivo transport via primate/human BBB and brain cells (neuron and glial). Gene delivery	
8D3 + 8314 [23]	Mouse TIR + human IR	(a) Experimental human brain tumors in scid mice (b) In vivo transport via mouse BBB and brain primate/human cells (neuron and glial). Gene delivery	(a) Experimental human brain tumors in scid mice (b) In vivo transport via mouse BBB and brain primate/human cells (neuron and glial). Gene delivery	
Chimeric anti-TIR [55]	Mouse TIR			
Humanized anti-TIR [26]	Human IR			

of luciferase was reported in primates, as compared to rat and mouse tissues. The high levels of expression in non-primate animal models *in vivo* [1, 3, 4, 19–23, 27]. The 83-14 murine MAb to the HIR and the OX26 murine MAb to the rat TIR are used to target human and rat tissues, respectively (Table 1). The OX26 TRIMAb is active only in rats, so the rat 8D3 TIR against the mouse TIR is used in mice (Table 1) [20, 21, 27–33]. The 83-14 HIRMAb does not cross-react with the insulin receptor in rodents or even New World primates such as the squirrel monkey. However, this HIRMAb does cross-react with the insulin receptor of Old World primates such as the Rhesus monkey. Since the plasmid DNA must be delivered to the nuclear compartment of brain cells, the TH1 must traverse both the BBB and the brain cell plasma membrane (BCM) behind the BBB (Figure 1c). Owing to high expression of the TIR on R on both the BBB and BCM, the targeting MAb enables the sequential receptor-mediated transcytosis of the TH1 across the BBB followed by the receptor-mediated endocytosis of the TH1 into the brain cell (Figure 1c). TH1s have also been successfully constructed to target human tumor cells in a scid mouse model wherein dual targeting MAbs were directed to the mouse TIR and HIR (Table 1 and Figure 1d), i.e., MAb1 and MAb2 [23].

If luciferase was reported in primates, as compared to rat and mouse tissues, the high levels of expression in non-primate animal models *in vivo* [1, 3, 4, 19–23, 27]. The 83-14 murine MAb to the HIR and the OX26 murine MAb to the rat TIR are used to target human and rat tissues, respectively (Table 1). The OX26 TRIMAb is active only in rats, so the rat 8D3 TIR against the mouse TIR is used in mice (Table 1) [20, 21, 27–33]. The 83-14 HIRMAb does not cross-react with the insulin receptor in rodents or even New World primates such as the squirrel monkey. However, this HIRMAb does cross-react with the insulin receptor of Old World primates such as the Rhesus monkey. Since the plasmid DNA must be delivered to the nuclear compartment of brain cells, the TH1 must traverse both the BBB and the brain cell plasma membrane (BCM) behind the BBB (Figure 1c). Owing to high expression of the TIR on R on both the BBB and BCM, the targeting MAb enables the sequential receptor-mediated transcytosis of the TH1 across the BBB followed by the receptor-mediated endocytosis of the TH1 into the brain cell (Figure 1c). The gene expression was also successfully constructed to target human tumor cells in a scid mouse model wherein dual targeting MAbs were directed to the mouse TIR and HIR (Table 1 and Figure 1d), i.e., MAb1 and MAb2 [23].

If luciferase was reported in primates, as compared to rat and mouse tissues, the high levels of expression in non-primate animal models *in vivo* [1, 3, 4, 19–23, 27]. The 83-14 murine MAb to the HIR and the OX26 murine MAb to the rat TIR are used to target human and rat tissues, respectively (Table 1). The OX26 TRIMAb is active only in rats, so the rat 8D3 TIR against the mouse TIR is used in mice (Table 1) [20, 21, 27–33]. The 83-14 HIRMAb does not cross-react with the insulin receptor in rodents or even New World primates such as the squirrel monkey. However, this HIRMAb does cross-react with the insulin receptor of Old World primates such as the Rhesus monkey. Since the plasmid DNA must be delivered to the nuclear compartment of brain cells, the TH1 must traverse both the BBB and the brain cell plasma membrane (BCM) behind the BBB (Figure 1c). Owing to high expression of the TIR on R on both the BBB and BCM, the targeting MAb enables the sequential receptor-mediated transcytosis of the TH1 across the BBB followed by the receptor-mediated endocytosis of the TH1 into the brain cell (Figure 1c). The gene expression was also successfully constructed to target human tumor cells in a scid mouse model wherein dual targeting MAbs were directed to the mouse TIR and HIR (Table 1 and Figure 1d), i.e., MAb1 and MAb2 [23].

(b) ground truth

Figura 5.3: Confronto tra *ground truth* e modello A. N.B.: Le regioni *other* non sono presenti nell'insieme di *ground truth*.

(a) modello A

Review Article

Protocols in Cleft Lip and Palate Treatment: Systematic Review

Pedro Ribeiro Soares de Ladeira¹ and Nivaldo Alonso^{2,3}

¹ School of Medicine, University of São Paulo, São Paulo, SP, Brazil
² Division of Burns and Plastic Surgery, School of Medicine, University of São Paulo, São Paulo, SP, Brazil
³ Rua Afonso Braga, 473, cj 65, Vila Nova Conceição, 0451-1000, São Paulo, SP, Brazil

Correspondence should be addressed to Nivaldo Alonso, nivaldo.alonso@gmail.com

Received 27 June 2012; Accepted 24 September 2012

Academic Editor: Renato Da Silveira Freitas

Copyright © 2012 P. R. de Ladeira and N. Alonso. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objectives. To find clinical decisions on cleft treatment based on randomized controlled trials (RCTs). Methods. Searches were made in PubMed, Embase, and Cochrane Library on cleft lip and palate. From the 170 articles found in the searches, 28 were considered adequate to guide clinical practice. Results. A scarce number of RCTs were found approaching cleft treatment. The experimental clinical approaches analyzed in the 28 articles were infant orthopedics, rectal acetaminophen, palatal block with bipuvacaine, infraorbital nerve block with bupivacaine, osteogenesis distraction, intravenous dexamethasone, sodium phosphate and allopurinol with bone morphogenic protein-2 (BMP-2). Conclusions. Few randomized controlled trials were found approaching cleft treatment, and fewer related to surgical repair of this deformity. So there is a need for more multicenter collaborations, mainly on surgical area, to reduce the variety of treatment modalities and to ensure that the cleft patient receives an evidence-based clinical practice.

The most broadcast treatment modalities in the management of unilateral cleft lip and palate are listed in Table 1 (chronologic age) and Table 2 (developmental development).

The presented management of cleft lip and palate is not accepted exactly by all cleft centers, and there is a striking diversity of clinical practice in area [4]. Evidence-based medicine should be the answer to the uncertainties in the treatment; however there is a paucity of high level of evidence (i.e., systematic reviews and randomized controlled trials) on cleft lip and palate [6]. Therefore, many clinical decisions are made based on biased evidence, such as the decision of when to perform secondary bone graft, which is answered by many surgeons with the information of a retrospective study [7].

Aiming to find clinical decisions based on randomized controlled trials (RCTs), searches on cleft lip and/or palatal were done in three main scientific databases: Cochrane Library, Embase, and PubMed [8]. Posteriorly, we selected articles that could validate or change the presented management.

1. Introduction

Clefts are the most prevalent craniofacial birth defects and the second most common birth anomaly, second only to clubfoot [1]. In the United States of America, it is estimated that \$100,000 are spent to rehabilitate a child born with oral cleft [2].

The approach of the patient with cleft lip and palate is multidisciplinary, and the cleft team should be ideally composed by craniofacial surgeons, otolaryngologists, geneticists, anaesthesiologists, speech-language pathologists, nutritionists, orthodontists, prosthodontists, and psychologists, and to be capable of treating even rare facial clefts with excellence, neurosurgeons, and ophthalmologists. In this manner, it is possible to provide long-term followup through the entire child's development and achieve all of the following treatment goals: normalized facial aesthetic, integrity of the primary and secondary palate, normal speech and hearing, airway patency, class I occlusion with normal masticatory function, good dental and periodontal health, and normal psychosocial development [3].

The most broadcast treatment modalities in the management of unilateral cleft lip and palate are listed in Table 1 (chronologic age) and Table 2 (developmental development).

The presented management of cleft lip and palate is not accepted exactly by all cleft centers, and there is a striking diversity of clinical practice in area [4]. Evidence-based medicine should be the answer to the uncertainties in the treatment; however there is a paucity of high level of evidence (i.d., systematic reviews and randomized controlled trials) on cleft lip and palate [6]. Therefore, many clinical decisions are made based on biased evidence, such as the decision of when to perform secondary bone graft, which is answered by many surgeons with the information of a retrospective study [7].

Aiming to find clinical decisions based on randomized controlled trials (RCTs), searches on cleft lip and/or palatal were done in three main scientific databases: Cochrane Library, Embase, and PubMed [8]. Posteriorly, we selected articles that could validate or change the presented management.

(a) modello K

(b) ground truth

Figura 5.4: Confronto tra *ground truth* e modello K. N.B.: Le regioni *other* non sono presenti nell'insieme di *ground truth*.

Clinical Study

Lip Height Improvement during the First Year of Unilateral Complete Cleft Lip Repair Using Cutting Extended Mohler Technique

Cassio Eduardo Raposo-Amaral,¹ André Pecchi Giancoli,¹
Rafael Denadai,¹ Frederico Figueiredo Marques,¹ Renato Salazar Somensi,¹
Cesar Augusto Raposo-Amaral,¹ and Nivaldo Alonso²

¹Institute of Plastic and Craniofacial Surgery, SOBRAL/UFSCar, Campinas, SP, Brazil
²Plastic Surgery Division, Department of Surgery, Universidade de São Paulo (USP), São Paulo, SP, Brazil

Correspondence should be addressed to Cassio Eduardo Raposo-Amaral, cassiopra@uolmail.com.

Received 23 July 2012; Accepted 1 December 2012

Academic Editor: Renato Da Silva Freitas

Copyright © 2013 Cassio Eduardo Raposo-Amaral et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective: To compare the cutaneous lip height at early and late postoperative periods and to objectively determine the average amount of lip height improvement during the first year of unilateral complete cleft lip repair using Cutting extended Mohler technique. **Methods:** In this prospective cohort study, 26 unilateral complete cleft patients and 50 nondelt subjects were included. Photographs were taken between 12 and 16 weeks (T1) and also taken between 12 and 13 months after surgery (T2). The cutaneous lip height distance (photogrammetric lip analysis) obtained in these two periods of time were measured and statistically analyzed. **Results:** The average lip heights were $24\% \pm 9\%$ in T1 and $8\% \pm 6\%$ in T2 ($P < 0.01$). The average lip height asymmetry in the nondelt individuals was $4.52\% \pm 1.89\%$. **Conclusion:** Since all principles to obtain a symmetrical Cupid's bow were performed, the postoperative pull-up of Cupid's bow is probably owed to the scar contracture, which improves by 2 times during the first year after surgery.

1. Introduction

Ralph Millard revolutionized the treatment of cleft lip by describing the innovative principles to repair a unilateral cleft lip, that allows surgeons around the globe to treat patients with different racial characteristics [1–4]. Consequently, his principles remain as a foundation to the development of surgical strategies and tactics to improve the results in the cleft lip repair worldwide [5–7]. Millard [8] used Millard's cheloplasty, using his own technique, that adds a more vertical incision in the cleft philtrum column, creating a faint scar that represent a mirror image of the contralateral philtrum column. In the Mohler technique [8], the Millard's C-flap is used to fill the gap created by the downward rotation of the cleft lip segment, instead of the lateral advancement segment as proposed by Millard [5]. Thus, a short lip height may be produced as a consequence of these maneuvers.

straight-line scar and absence of the lateral advancement segment fulfilling the medial gap after the back-cut incision)

Cutting's modifications of the Mohler technique are the following: (1) an extension of the medial incision toward the columella, (2) the Millard's back-cut incision never passes the nondelt philtrum column, (3) a more vertical incision than that described by Millard that creates a, and (4) wider C-flap that fills the medial rotation defect [9]. Cutting and Dayan [9] subsequently analyzed cleft patients who underwent a cheloplasty, using the Cutting extend Mohler technique, in order to respond whether this technique produces a short lip height and lip width in two different postoperative periods of time. Cleft and nondelt distances in each patient were measured preoperatively and postoperatively at two different periods of time (1 to 13 months and at 2 years or more) and statistically compared [9]. Their data did not show statistically significant changes in lip height over time and

Figura 5.5: Confronto che evidenzia il fenomeno di oversplitting (regione centrale della pagina).

(a) modello D

(b) ground truth

TABELLA 1: Prediction of respiratory weaning outcomes

Author	Number	Outcome	Study	Author	Number	Outcome	Study
Zappa et al. [41]	100	RNP or pro-BNP & ECHO predict WO	Prospective	Zappa et al. [41]	100	RNP or pro-BNP & ECHO predict WO	Prospective
Gerbaut et al. [38]	44	BNP or pro-BNP & ECHO did not predict WO	Prospective	Gerbaut et al. [38]	44	BNP or pro-BNP & ECHO did not predict WO	Prospective
Papiaonamou et al. [7]	42	Cardiorespiratory dynamics predict WO	Prospective	Papiaonamou et al. [7]	42	Cardiorespiratory dynamics predict WO	Prospective
Chien et al. [36]	52	Percent change of less than 20% in RNP Predicts WO	Prospective	Chien et al. [36]	52	Percent change of less than 20% in BNP predicts WO	Prospective
Melkonian-Dessal et al. [37]	102	Lower levels before SFT may predicts WO	Prospective	Melkonian-Dessal et al. [37]	102	Lower BNP levels before SFT may predicts WO	Prospective
Moschietto et al. [42]	68	ECHO predicts WO	Prospective	Moschietto et al. [42]	68	ECHO predicts WO	Prospective
Schifflinan et al. [43]	24	ECHO did not predicts WO	Prospective	Schifflinan et al. [43]	24	ECHO did not predicts WO	Prospective
Calie et al. [46]	117	ECHO predicts WO	Prospective	Calie et al. [46]	117	ECHO predicts WO	Prospective
Grasso et al. [40]	19	BNP predicts WO	Prospective	Grasso et al. [40]	19	BNP predicts WO	Prospective

RNP: rechtocardioritme; WO: weaning outcome; BNP: weaning outcome; BNP: B-type natriuretic peptide.

Author	Number	Outcome	Study	Author	Number	Outcome	Study
Zappa et al. [41]	100	RNP or pro-BNP & ECHO predict WO	Prospective	Zappa et al. [41]	100	RNP or pro-BNP & ECHO predict WO	Prospective
Gerbaut et al. [38]	44	BNP or pro-BNP & ECHO did not predict WO	Prospective	Gerbaut et al. [38]	44	BNP or pro-BNP & ECHO did not predict WO	Prospective
Papiaonamou et al. [7]	42	Cardiorespiratory dynamics predict WO	Prospective	Papiaonamou et al. [7]	42	Cardiorespiratory dynamics predict WO	Prospective
Chien et al. [36]	52	Percent change of less than 20% in RNP Predicts WO	Prospective	Chien et al. [36]	52	Percent change of less than 20% in BNP predicts WO	Prospective
Melkonian-Dessal et al. [37]	102	Lower levels before SFT may predicts WO	Prospective	Melkonian-Dessal et al. [37]	102	Lower BNP levels before SFT may predicts WO	Prospective
Moschietto et al. [42]	68	ECHO predicts WO	Prospective	Moschietto et al. [42]	68	ECHO predicts WO	Prospective
Schifflinan et al. [43]	24	ECHO did not predicts WO	Prospective	Schifflinan et al. [43]	24	ECHO did not predicts WO	Prospective
Calie et al. [46]	117	ECHO predicts WO	Prospective	Calie et al. [46]	117	ECHO predicts WO	Prospective
Grasso et al. [40]	19	BNP predicts WO	Prospective	Grasso et al. [40]	19	BNP predicts WO	Prospective

RNP: rechtocardioritme; WO: weaning outcome; BNP: B-type natriuretic peptide.

Etiology of weaning failure. Weaning failure due to excessive preload should be treated with diuretic. It is important to rule out extra cardiac causes of weaning failure in such cases. Vasodilator therapy is indicated for weaning failure due to excessive afterload or myocardial ischemia. Additionally, alteration in ITP can be prevented by the use of CPAP/BIPAP (Figure 2). Noninvasive ventilation decreases cardiac stress load and should be utilized in weaning patient with poor cardiac reserves [47, 48]. In fact, positive pressure therapy is now the standard of care for treating patient with acute pulmonary edema and decreases afterload [3, 19]. Using the same physiological concepts, Marino and Langille et al. and others have introduced the concept of resistive loaded breathing to augment cardiac output during cardiopulmonary resuscitation [49-52].

4 Conclusion

Evaluation and prediction of weaning failure from cardiac origin remain complicated. Current prediction models are difficult to implement clinically at bedside. Echocardiography remains a valuable tool to monitor respiratory weaning process and requires expertise in image interpretation. Additionally, the need for multiple assessments makes it difficult to implement echocardiography as a routine monitor in the intensive care setting. Serum BNP and NT-pro-BNP appear promising to identify patients with heart failure during weaning process. However, laboratory turnover time and the cut-off values for HF pose a clinical challenge for data interpretation in the intensive care arena.

5 Key Messages for Practicing Intensivists

(i) Ischemic heart disease, valvular heart disease, systolic or diastolic dysfunction contributes to increase in cardiac load and weaning failure.
(ii) Extra demand on cardiac working load imposed by SFT may become apparent when transferring patient in the ICU. However, due to certain limitation relating to patient, it cannot be used in every patient [22, 45, 46] (Table 1).
(iii) Diuretic therapy may be considered for excessive pre-load.
(iv) Noninvasive positive pressure ventilation is beneficial for weaning-induced pulmonary edema.

(a) modello AJ

(b) ground truth

Figura 5.6: Confronto tra modello AJ, che esegue node classification, e ground truth.

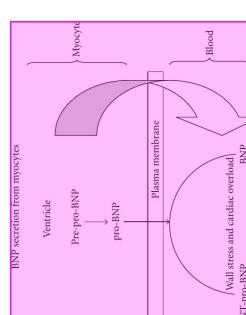


Figura 3: BNP: B-type natriuretic peptide; NT-pro-BNP: N-terminal pro-B-type natriuretic peptide.

Evaluation and prediction of weaning failure from cardiac origin remain complicated. Current prediction models are difficult to implement clinically at bedside. Echocardiography remains a valuable tool to monitor respiratory weaning process and requires expertise in image interpretation. Additionally, the need for multiple assessments makes it difficult to implement echocardiography as a routine monitor in the intensive care setting. Serum BNP and NT-pro-BNP appear promising to identify patients with heart failure during weaning process. However, laboratory turnover time and the cut-off values for HF pose a clinical challenge for data interpretation in the intensive care arena.

4 Conclusion

Evaluation and prediction of weaning failure from cardiac origin remain complicated. Current prediction models are difficult to implement clinically at bedside. Echocardiography remains a valuable tool to monitor respiratory weaning process and requires expertise in image interpretation. Additionally, the need for multiple assessments makes it difficult to implement echocardiography as a routine monitor in the intensive care setting. Serum BNP and NT-pro-BNP appear promising to identify patients with heart failure during weaning process. However, laboratory turnover time and the cut-off values for HF pose a clinical challenge for data interpretation in the intensive care arena.

5 Key Messages for Practicing Intensivists

(i) Ischemic heart disease, valvular heart disease, systolic or diastolic dysfunction contributes to increase in cardiac load and weaning failure.
(ii) Extra demand on cardiac working load imposed by SFT may become apparent when transferring patient in the ICU. However, due to certain limitation relating to patient, it cannot be used in every patient [22, 45, 46] (Table 1).
(iii) Diuretic therapy may be considered for excessive pre-load.
(iv) Noninvasive positive pressure ventilation is beneficial for weaning-induced pulmonary edema.

(a) modello AJ

(b) ground truth

Figura 5.6: Confronto tra modello AJ, che esegue node classification, e ground truth.

Research Article

Esthetic Composition of Smile in Individuals with Cleft Lip, Alveolus, and Palate: Visibility of the Periodontium and the Esthetics of Smile

Luis Augusto Esper,¹ Michelye Cristiane Sbrana,¹ Mércia Jussara da Silva Cunha,¹ Guilherme Santos Moreira,² and Ana Lúcia Pompeia Fraga de Almeida^{2,3}

¹Oral Rehabilitation, Bauru School of Dentistry, University of São Paulo, 17012-901 Bauru, Brazil
²Searc de Periodoncia, Hospital de Reabilitação de Anomalias Craniofaciais da Universidade da S. Paulo, Rua Santa Marchinha, 3-20 Vila Universitária, 17012-900 Bauru, SP, Brazil
³Department of Prosthetic Dentistry at Bauru School of Dentistry, Universidade de São Paulo, 17012-901 Bauru, SP, Brazil

Correspondence: lucia.almeida@usp.br

Received 5 August 2012; Accepted 15 September 2012

Academic Editor: Nicarla Mosso

Copyright © 2012 Luis Augusto Esper et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective: To evaluate characteristics of smile related to visibility in individuals with cleft lip, alveolus, and palate. **Design, Cross-sectional.** Setting: HRAC/USP, Brazil. **Patients:** Individuals with repaired complete unilateral cleft lip and palate ($n = 45$), aged 15–30 years. **Interventions:** Frontal facial photographs were obtained in natural and forced smiles ($n = 135$). Six specialists in periodontics evaluated the photographs as the smile line, thickness, and curve of the upper lip. **Main Outcome Measures:** The smile area was compared with the contralateral region. Results: Results were expressed as percentages and means. The findings were compared between groups of periodontists. Results: Statistically significant relationship was observed in the smile line and curve of the upper lip at rest and thinner in the forced smile, also evaluated by the group nonopposed with cleft care. The curve of the upper lip in natural and forced smiles was considered as close to straight by both groups, regardless the side. Conclusion: The smile in individuals with cleft was regarded as average for both nonopposed sides. The thickness was characterized as average to thin, being thinner in forced smile and when analyzed by the group not experienced with cleft care. In the average, the curve of the upper lip was considered as straight. The present study elucidates some characteristics related to the smile in individuals with repaired unilateral cleft lip, alveolus, and palate.

Research Article

Esthetic Composition of Smile in Individuals with Cleft Lip, Alveolus, and Palate: Visibility of the Periodontium and the Esthetics of Smile

Luis Augusto Esper,¹ Michelye Cristiane Sbrana,¹ Mércia Jussara da Silva Cunha,¹ Guilherme Santos Moreira,² and Ana Lúcia Pompeia Fraga de Almeida^{2,3}

¹Oral Rehabilitation, Bauru School of Dentistry, University of São Paulo, 17012-901 Bauru, Brazil
²Searc de Periodoncia, Hospital de Reabilitação de Anomalias Craniofaciais da Universidade da S. Paulo, Rua Santa Marchinha, 3-20 Vila Universitária, 17012-900 Bauru, SP, Brazil
³Department of Prosthetic Dentistry at Bauru School of Dentistry, Universidade de São Paulo, 17012-901 Bauru, SP, Brazil

Correspondence: Ana.L.Fraga@usp.br

Received 5 August 2012; Accepted 15 September 2012

Academic Editor: Nicarla Mosso

Copyright © 2012 Luis Augusto Esper et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective: To evaluate characteristics of smile related to visibility in individuals with cleft lip, alveolus, and palate. **Design, Cross-sectional.** Setting: HRAC/USP, Brazil. **Patients:** Individuals with repaired complete unilateral cleft lip and palate ($n = 45$), aged 15–30 years. **Interventions:** Frontal facial photographs were obtained in natural and forced smiles ($n = 135$). Six specialists in periodontics evaluated the photographs as the smile line, thickness, and curve of the upper lip. **Main Outcome Measures:** The smile area was compared with the contralateral region. Results: Results were expressed as percentages and means. The findings were compared between groups of periodontists. Results: Statistically significant relationship was observed in the smile line and curve of the upper lip at rest and thinner in the forced smile, also evaluated by the group nonopposed with cleft care. The curve of the upper lip in natural and forced smiles was considered as close to straight by both groups, regardless the side. Conclusion: The smile in individuals with cleft was regarded as average for both nonopposed sides. The thickness was characterized as average to thin, being thinner in forced smile and when analyzed by the group not experienced with cleft care. In the average, the curve of the upper lip was considered as straight. The present study elucidates some characteristics related to the smile in individuals with repaired unilateral cleft lip, alveolus, and palate.

Luis Augusto Esper,¹ Michelye Cristiane Sbrana,¹ Mércia Jussara da Silva Cunha,¹ Guilherme Santos Moreira,² and Ana Lúcia Pompeia Fraga de Almeida^{2,3}

¹Oral Rehabilitation, Bauru School of Dentistry, University of São Paulo, 17012-901 Bauru, Brazil
²Searc de Periodoncia, Hospital de Reabilitação de Anomalias Craniofaciais da Universidade da S. Paulo, Rua Santa Marchinha, 3-20 Vila Universitária, 17012-900 Bauru, SP, Brazil
³Department of Prosthetic Dentistry at Bauru School of Dentistry, Universidade de São Paulo, 17012-901 Bauru, SP, Brazil

Correspondence: Ana.L.Fraga@usp.br

Received 5 August 2012; Accepted 15 September 2012

Academic Editor: Nicarla Mosso

Copyright © 2012 Luis Augusto Esper et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective: To evaluate characteristics of smile related to visibility in individuals with cleft lip, alveolus, and palate. **Design, Cross-sectional.** Setting: HRAC/USP, Brazil. **Patients:** Individuals with repaired complete unilateral cleft lip and palate ($n = 45$), aged 15–30 years. **Interventions:** Frontal facial photographs were obtained in natural and forced smiles ($n = 135$). Six specialists in periodontics evaluated the photographs as the smile line, thickness, and curve of the upper lip. **Main Outcome Measures:** The smile area was compared with the contralateral region. Results: Results were expressed as percentages and means. The findings were compared between groups of periodontists. Results: Statistically significant relationship was observed in the smile line and curve of the upper lip at rest and thinner in the forced smile, also evaluated by the group nonopposed with cleft care. The curve of the upper lip in natural and forced smiles was considered as close to straight by both groups, regardless the side. Conclusion: The smile in individuals with cleft was regarded as average for both nonopposed sides. The thickness was characterized as average to thin, being thinner in forced smile and when analyzed by the group not experienced with cleft care. In the average, the curve of the upper lip was considered as straight. The present study elucidates some characteristics related to the smile in individuals with repaired unilateral cleft lip, alveolus, and palate.

I. Introduction and Literature Review

I.1. Introduction and Literature Review

The facial appearance is a key element in the psychological development of individuals with cleft lip, alveolus, and palate, who usually present characteristics of social introversion [1,2]. Their rehabilitation should be conducted with a view to enhance the functional and esthetic characteristics and consequently the quality of life.

The smile is one of the most important facial expressions

and should be carefully analyzed as a whole before the accomplishment of dental treatment, aiming to establish the harmony between the teeth and adjacent soft tissues

achieving an esthetic and pleasant smile [3]. Evaluation of the face should be conducted together with the intraoral

examination for the establishment of treatment goals that may meet the patient's expectations and reestablish the most esthetic conditions possible [4].

The esthetic analysis of smile by the visual perception [5] should consider it as a unique composition, in which some elements act in combination to make the smile harmonious and pleasant for the observer. Significant disruptions in the composition deviate the attention to the undesired element alterations in the gingival component, lip shape, and filling of the gingival papilla, call the attention due to their distinguished role in the composition of smile [5].

Concerning the position of the upper lip, the smile may be classified as high, average, and low [6], which is very

(a) modello AU

Figura 5.7: Differenza di comportamento con *label pruning* (modello AW) e senza (modello AU).

(b) modello AW

Capitolo 6

Conclusioni

In questo elaborato è stato proposto un metodo per eseguire *Document Layout Analysis* estraendo le informazioni necessarie direttamente dai documenti in formato PDF, senza appoggiarsi ad OCR, e sviluppando un sistema basato sulle GNN. Sono stati presentati gli strumenti di *pre-processing* adottati e la procedura di costruzione dei grafi, sono stati poi descritti il funzionamento della rete sviluppata, del processo di addestramento e di quello di inferenza, valutando e commentando i risultati ottenuti.

Il lavoro svolto è molto promettente, anche se è necessario implementare delle migliorie ulteriori per ovviare alla mancanza di informazioni visive, che come osservato nel paragrafo **5.2.1** possono fare la differenza nella valutazione attraverso mAP. Un altro limite è costituito dalla fragilità dovuta all'utilizzo delle componenti connesse per l'identificazione delle diverse regioni di un documento: è sufficiente che il predittore sbagli a classificare un singolo arco negativo per ottenere delle *bounding boxes* notevolmente più ampie del dovuto. Per affrontare questo problema potrebbe essere necessario appoggiarsi ad una fase di *post-processing* più impegnativa, che sia in grado di perfezionare il lavoro svolto dal predittore. Un'altra possibile soluzione potrebbe trovarsi nella riduzione del numero di archi che compongono

il grafo, adottando una connettività più semplice, come la β -skeleton proposta in [16]. Altre considerazioni riguardano gli *embeddings* dei nodi: si potrebbe ad esempio rifinire la codifica di alcuni vettori di *features*, come quelli “intersezione”, per identificare in modo più preciso le entità che circondano un nodo e dunque riuscire a trasmettere informazioni relative alla struttura di un paragrafo durante la fase di *message passing*. Infine sarebbe necessario chiarire la questione posta nel paragrafo **5.2.1** riguardante la segmentazione dei paragrafi, compito che risulta particolarmente difficile senza l’utilizzo di OCR e talvolta non strettamente necessario per la corretta identificazione delle porzioni di testo. Si potrebbero dunque generare delle regioni di *ground truth* che racchiudano i diversi paragrafi nella loro totalità, e provare ad addestrare e valutare un modello a partire da esse. Si deve però considerare che ciò corrisponderebbe ad un ingente investimento in termini di tempo, in quanto per la costruzione dell’insieme di *ground truth* non sarebbe possibile utilizzare le tecniche euristiche che invece sono state adottate per ottenere le annotazioni di PubLayNet e PubTables-1M.

Bibliografia

- [1] About arxiv. <https://arxiv.org/about>.
- [2] Iz Beltagy, Kyle Lo, and Arman Cohan. Scibert: A pretrained language model for scientific text. 2019.
- [3] Andrea Gemelli, Emanuele Vivoli, and Simone Marinai. Graph neural networks and representation embedding for table extraction in pdf documents. *Conference Proceedings ICPR*, 2022.
- [4] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 1025–1035, Red Hook, NY, USA, 2017. Curran Associates Inc.
- [5] Glenn Jocher. Precision-recall curve. <https://github.com/ultralytics/yolov3/issues/898>.
- [6] Gianluca Malato. Are you still using 0.5 as a threshold? <https://www.yourdatateacher.com/2021/06/14/are-you-still-using-0-5-as-a-threshold/>, Jun 2021.
- [7] Jorj X. McKie. Pymupdf. <https://github.com/pymupdf/PyMuPDF>.

- [8] Zara Nasar, Syed Waqar Jaffry, and Muhammad Malik. Information extraction from scientific articles: a survey. *Scientometrics*, 117, 09 2018.
- [9] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019.
- [10] About - pubmed. <https://pubmed.ncbi.nlm.nih.gov/about/>.
- [11] Zejiang Shen, Ruochen Zhang, Melissa Dell, Benjamin Charles Germain Lee, Jacob Carlson, and Weining Li. Layoutparser: A unified toolkit for deep learning based document image analysis. *arXiv preprint arXiv:2103.15348*, 2021.
- [12] Brandon Smock. Pubtables-1m. <https://github.com/microsoft/table-transformer>.
- [13] Brandon Smock, Rohith Pesala, and Robin Abraham. Pubtables-1m: Towards comprehensive table extraction from unstructured documents, 2021.
- [14] spacy 101: Everything you need to know. <https://spacy.io/usage/spacy-101>.
- [15] Minjie Wang, Da Zheng, Zihao Ye, Quan Gan, Mufei Li, Xiang Song, Jinjing Zhou, Chao Ma, Lingfan Yu, Yu Gai, Tianjun Xiao, Tong He, George Karypis, Jinyang Li, and Zheng Zhang. Deep graph library: A graph-centric, highly-performant package for graph neural networks, 2019.

- [16] Renshen Wang, Yasuhisa Fujii, and Ashok C. Popat. Post-ocr paragraph recognition by graph convolutional networks, 2021.
- [17] Huichen Yang and William H. Hsu. Vision-based layout detection from scientific literature using recurrent convolutional neural networks. 2020.
- [18] ZACH. How to interpret a roc curve (with examples). <https://www.statology.org/interpret-roc-curve/>.
- [19] Xu Zhong, Jianbin Tang, and Antonio Jimeno-Yepes. Publaynet: largest dataset ever for document layout analysis. *CoRR*, abs/1908.07836, 2019.