



Dipartimento di Ingegneria dell'Informazione
Corso di Laurea in Ingegneria Biomedica

Tesi di Laurea Magistrale

Analisi di immagini tomografiche ad alta risoluzione attraverso
reti neurali convoluzionali per lo studio delle interstiziopatie
polmonari

Relatori:

Prof. Vincenzo Positano

Dott.ssa Sara Colantonio

Dott.ssa Danila Germanese

Candidata:

Rossana Buongiorno

12 giugno 2020

Anno Accademico 2019-2020

Introduzione

Le interstiziopatie polmonari (*Interstitial Lung Disease*, ILD) sono un gruppo numeroso ed eterogeneo di patologie croniche non neoplastiche che causano la cicatrizzazione del tessuto connettivo del parenchima polmonare e dell'interstizio alveolare e la compromissione della normale funzionalità respiratoria. Sebbene siano poco conosciute e rare, nel loro insieme rappresentano la causa più frequente di malattie polmonari croniche non ostruttive. Attualmente, si annoverano più di 200 forme diverse di ILD con cause, prognosi e terapia che possono variare notevolmente; per questo, è stata standardizzata la classificazione delle patologie in idiopatiche, secondarie, sarcoidosi e non classificabili. L'identificazione precisa della pneumopatia e la collocazione in uno dei gruppi è fondamentale per individuare la terapia migliore e formulare una prognosi. Una diagnosi corretta consiste in una valutazione completa del quadro clinico da parte di pneumologo, anatomo-patologo e radiologo. Nello specifico, l'esame radiologico di riferimento è la tomografia computerizzata del torace ad alta risoluzione (*High Resolution Computed Tomography*, HRCT) che permette di ottenere immagini ad alta definizione rispetto ad una radiografia tradizionale. Nell'analizzare le tomografie, il radiologo deve innanzitutto stabilire se vi è *Usual Interstitial Pneumonia* (UIP) ovvero presenza di pattern istopatologici tipici della malattia; una volta accertata la presenza dei pattern, se ne valuta l'estensione, correlata sia con il grado di coinvolgimento patologico che con la gravità delle alterazioni fisiologiche. Tuttavia, l'incidenza rara delle interstiziopatie e la numerosità dei pattern fanno sì che non tutti i radiologi abbiano un grado di esperienza adatto a discriminare visivamente le piccole differenze in attenuazione tra tessuto sano e tessuto malato e a caratterizzare correttamente l'anomalia a fronte della sovrapposizione tra diversi tipi di strutture. Inoltre, la malattia si diffonde lungo tutti i polmoni e la segmentazione manuale da parte di un operatore clinico, passaggio indispensabile per una valutazione semi-quantitativa, risulta tediosa e faticosa.

Per far fronte a queste problematiche, quindi nel tentativo di rimediare alla variabilità intra- ed inter-osservatore, la ricerca si sta impegnando nello sviluppo di tecniche nuove per il riconoscimento automatico dei pattern UIP. Allo stato attuale, vi sono approcci basati sull'analisi della distribuzione dei livelli di grigio nell'istogramma e sull'analisi della texture dell'immagine, che però mancano di riproducibilità, spesso sotto-stimano la presenza dei pattern e non considerano la stadiazione della malattia. *Computer-Aided Lung Informatics for Pathology Evaluation and Ratings* (CALIPER), software sviluppato dal Biomedical Imaging Resource Laboratory della Mayo Clinic, è lo strumento di riferimento per la quantificazione dei pattern UIP in pazienti affetti da ILD; è validato su test di funzionalità respiratoria e ottenuto dall'elaborazione di una quantità consistente di dati ma non è una risorsa *open-source* e, al momento, è destinato ai soli scopi di ricerca. Inoltre, essendo il risultato di un processo di *features engineering*, partito da classificazioni fatte da operatori clinici diversi, presenta un bias, che è causa di identificazioni errate, o mancate, dei pattern.

Il presente lavoro di tesi mira a indagare la capacità del deep learning di fornire un risultato accurato e affidabile, che prescindano dall'operatore, dal macchinario, dal protocollo di acquisizione dell'immagine e dall'algoritmo di ricostruzione. Il deep learning, infatti, si distingue dalle tecniche tradizionali perché fornisce strumenti che imparano autonomamente a classificare i dati, indipendentemente dal contesto in cui vengono acquisiti.

Verrà proposto un approccio basato su una rete neurale convoluzionale (*Convolutional Neural Network*, CNN), la UIP-net, per l'automatizzazione del processo di segmentazione e, quindi, la discriminazione dei pattern UIP dal tessuto sano. Per progettare correttamente la UIP-net, è stato, innanzitutto, condotto uno studio della letteratura sulle applicazioni delle CNN alla segmentazione delle immagini mediche; l'obiettivo era ricercare l'architettura più adatta ad analizzare immagini risolte e limitare la perdita di informazioni che si verifica man mano che i dati vengono elaborati da strati consecutivi della rete. Le informazioni rilevanti, per il riconoscimento automatico dei pattern UIP, risiedono nella texture delle immagini, quindi nell'arrangiamento spaziale e nella differenza d'intensità tra pixels adiacenti. La struttura scelta per la UIP-net è, quindi, l'*Encoder-Decoder*, sviluppata, in origine, per segmentare la cartilagine di ginocchio; in seguito, è stata opportunamente modificata dopo aver identificato le differenze sostanziali tra le due applicazioni. Infatti, se da un lato la versione originale doveva discrimi-

nare i pixels sulla base delle *intensità* dei livelli di grigio, dall'altro, la UIP-net deve classificare le regioni delle immagini (*Region Of Interest*, ROI) sulla base della *differenza* delle intensità dei livelli di grigio dei pixels che vi appartengono. Alla luce di questo, sono stati eliminati gli strati di *batch-normalization* e *pooling* perché sotto-campionano i dati, causando una forte perdita di informazione, ed è stato ridotto il numero di strati. L'architettura finale della UIP-net si compone, quindi, di uno strato di input, che riceve un'immagine di dimensione 492×492 , tre strati di codifica e tre strati di decodifica convolutivi, ad ognuno dei quali segue uno strato di attivazione ReLU, a eccezione dell'ultimo in cui si applica una funzione di attivazione tangente iperbolica. Gli strati di codifica riducono la dimensione delle immagini in input, eseguendo un'operazione di convoluzione con un kernel (o filtro) 5×5 , nel primo strato, e 3×3 negli strati successivi; gli output degli strati di codifica sono delle features maps, ovvero le risposte delle ROI delle immagini ispezionate, ai filtri applicati. Gli strati di decodifica, invece, eseguono un'operazione, detta di *up-convoluzione*, e, al contrario, sovra-campionano i dati per riportare le immagini alle dimensioni di partenza. Ad ogni step di codifica, il numero di features maps raddoppia; al contrario, si dimezza durante la decodifica. Infine, l'ultimo strato è di tipo convolutivo e assembla tutte le features maps in un'immagine con le stesse dimensioni dell'input. Dopo averne definito l'architettura, sono stati scelti la funzione di costo, la funzione di ottimizzazione e le metriche di accuratezza per l'addestramento della rete: rispettivamente sono l'entropia incrociata binaria, l'*Adaptive Moment Estimation* e gli indici di *Dice* e *Jaccard*.

Il dataset per l'addestramento e il test della UIP-net contava 20 HRCT di pazienti affetti da Fibrosi Idiopatica Polmonare (IPF, sotto-categoria delle ILD), messe a disposizione dall'Unità Operativa Radiodiagnostica 2 dell'Azienda Ospedaliera Universitaria Pisana di Cisanello. Da queste, è stata ottenuta la *ground truth* attraverso il software CALIPER, che, prendendo in input le intere scansioni volumetriche, fornisce in output le singole slices segmentate, sottoforma di immagini RGB, quindi a colori, per la differenziazione dei pattern UIP. Tuttavia, per gli scopi del lavoro di tesi, non è stata fatta una classificazione multi-classe ma binaria, quindi dalle immagini RGB sono state ottenute delle maschere che avessero pixels uguali a 1 se appartenenti ad uno dei pattern UIP e pixels uguali a 0 se appartenenti al tessuto sano. Dopo aver ottenuto la *ground truth*, le immagini sono state pre-processate per abbattere il tempo computazionale necessario per l'addestramento, riducendo sia le dimensioni del campo visivo dell'immagine,

che il numero di pixels da analizzare. Per questo, è stato implementato un algoritmo di clustering non esclusivo, il *fuzzy C-mean*, che associasse ogni pixel dell'immagine ad una classe di appartenenza: campi polmonari e fondo o regione addominale. A seguito del pre-processing, le immagini pronte per l'addestramento avevano pixels non nulli solo se appartenenti ai polmoni.

L'addestramento della rete è stato eseguito con 13 delle 20 TC ma, dal momento che la rete è stata progettata per lavorare con immagini 2-D proprio per aumentare il numero di dati a disposizione, il training set intero contava 3200 immagini. Per individuare la strategia di addestramento migliore che limitasse il fenomeno dell'overfitting, sono stati eseguiti 20 addestramenti, variando la numerosità del dataset mantenendo fisso il numero di epoche e variando il numero di epoche mantenendo fisso il numero di esempi. Il modello migliore è risultato quello addestrato con 800 campioni per 50 epoche, in quanto la funzione di perdita e la metrica di accuratezza, in termini di indice di Dice, calcolati sul training e validation set si assestavano pressappoco sugli stessi valori ad ogni epoca. Una volta scelto il numero di campioni e di epoche coi quali addestrare la rete, è stata eseguita una convalida incrociata per valutarne la capacità di generalizzazione, attraverso una *k-fold cross validation*, imponendo k uguale a 5. Quindi, il training set è stato diviso in 5 gruppi uguali e per 5 volte la rete è stata addestrata su 640 immagini e testata su 160. Gli indici di Dice e Jaccard medi, calcolati per ogni split della convalida incrociata, sono, rispettivamente, 76,1% e 61,7%.

Una volta addestrato il modello con convalida incrociata, è stato eseguito il test su 7 delle 20 TC attraverso un'analisi quantitativa e un'analisi qualitativa. L'analisi quantitativa è consistita nel calcolo di cinque indici di confronto con la ground truth: indici di Dice e Jaccard, BF-score, sensibilità e specificità; l'analisi qualitativa, invece, si è basata sul giudizio attribuito da un radiologo esperto. Proprio in seguito all'ispezione visiva dei risultati, sono state migliorate le segmentazioni con un'operazione di post-processing: rispetto a CALIPER, infatti, la UIP-net classifica i vasi sanguigni e le vie aeree come appartenenti ai polmoni o al parenchima polmonare includendo, spesso, anche organi esterni come l'intestino. Per far fronte alla problematica, attraverso il software 3D-Slicer che implementa un algoritmo di segmentazione automatica per l'isolamento dei campi polmonari, dei vasi sanguigni e delle vie aeree, sono state ottenute delle maschere binarie, a partire dalle quali sono state eliminate le strutture erro-

neamente classificate dalla UIP-net. Gli indici di qualità della performance sono stati calcolati sulle immagini prima e dopo le operazioni di post-processing. In entrambi i casi, il BF-score e la sensibilità hanno presentato i valori massimi, arrivando rispettivamente a 83,73% e 84,6%, senza post-processing e 96,7% e 85,9%, con il post-processing.

Per concludere, alla luce dei valori degli indici di prestazione della UIP-net a confronto con CALIPER, il *gold standard*, e anche a seguito della valutazione qualitativa di un radiologo esperto, emerge che il deep learning e, nello specifico, le reti convoluzionali sono strumenti che promettono di superare i limiti delle tecniche allo stato dell'arte, primo fra tutti il bias dovuto al giudizio degli operatori clinici necessario per classificare le immagini.

L'elaborato si articolerà come segue: nel capitolo 1, verrà inquadrato il contesto clinico del lavoro, per cui verranno descritte l'anatomia e la fisiologia dell'apparato respiratorio e approfonditi gli aspetti principali dell'IPF. Si illustreranno anche i principi di funzionamento della Tomografia Computerizzata (TC), con particolare riguardo all'HRCT. Nel capitolo 2, verrà riportato lo stato dell'arte dei metodi automatici per l'analisi quantitativa delle immagini HRCT delle interstiziopatie polmonari e le applicazioni principali, come l'analisi dell'istogramma e della texture dell'immagine. Nel capitolo 3, verrà fornita una panoramica generale sulle reti neurali convoluzionali; poi, il focus si sposterà sull'applicazione delle CNN all'imaging medico, soffermandosi su due reti ampiamente utilizzate, la U-net e l'Encoder-Decoder. Nel capitolo 4, verranno illustrati i metodi coi quali è stato svolto il lavoro, descrivendo l'architettura della CNN progettata e il pre-processing delle immagini che compongono il dataset utilizzato per l'addestramento e il test della rete. Nel capitolo 5, verranno illustrati e commentati i risultati dopo aver descritto la strategia di addestramento della CNN; seguiranno le analisi quantitative dei valori degli indici di performance della rete e le analisi qualitative sulle immagini segmentate dalla UIP-net da parte di un operatore clinico specializzato in radiologia toracica. Entrambe le analisi sono state eseguite sia prima che dopo una correzione fatta in post-processing. Nel capitolo 6, si concluderà la discussione e si illustreranno gli eventuali sviluppi futuri.

Indice

1	Apparato respiratorio e IPF	9
1.1	Anatomia dell'apparato respiratorio	9
1.1.1	Le vie aeree	10
1.1.2	I polmoni	11
1.2	Fisiologia dell'apparato respiratorio	13
1.2.1	Funzioni respiratorie e volumi polmonari	13
1.2.2	Esami funzionali	15
1.2.2.1	Volumetria polmonare e valutazione della diffusione alveolo- capillare	16
1.3	La fibrosi idiopatica polmonare (IPF)	17
1.3.1	Diagnosi e pattern UIP	18
1.4	Tomografia computerizzata e HRCT	21
1.4.1	Protocollo di acquisizione delle immagini	26
2	AI per il supporto alla diagnosi di interstiziopatia polmonare: stato dell'arte	29
2.1	Quantitative Computed Tomography (QCT)	29
2.1.1	Analisi dell'istogramma dell'immagine	32
2.1.2	Analisi della texture dell'immagine	35
2.1.2.1	Adaptive Multiple Feature Method (AMFM)	36
2.1.2.2	CALIPER	38
2.2	IPF e Deep Learning	40

3	Le reti neurali convoluzionali (CNN)	44
3.1	Introduzione alle CNN	44
3.1.1	Architettura	45
3.1.1.1	Strato convolutivo	46
3.1.1.2	Strati di attivazione e di <i>pooling</i>	48
3.1.1.3	Strato <i>fully-connected</i>	51
3.1.2	Addestramento	51
3.1.2.1	<i>Forward-propagation</i>	53
3.1.2.2	Aggiornamento dei pesi e <i>backward-propagation</i>	53
3.1.3	CNN per l'imaging medico	54
3.1.3.1	U-net	55
3.1.3.2	Encoder-Decoder	57
4	Metodi	59
4.1	Obiettivi del lavoro	59
4.2	UIP-net: rete convoluzionale per la segmentazione delle interstiziopatie	61
4.2.1	Architettura	63
4.2.1.1	Scelta degli iper-parametri	66
4.2.2	Addestramento	67
4.2.2.1	Funzione costo e di ottimizzazione	67
4.2.2.2	Metriche di accuratezza	70
4.3	Dataset	71
4.3.1	Ottenimento della ground truth	72
4.3.2	Pre-processing delle immagini	75
4.3.2.1	Clustering non esclusivo: algoritmo fuzzy C-means	76
4.3.2.2	Isolamento dei polmoni e ridimensionamento dell'immagine	78
5	Risultati	81
5.1	Training	81
5.1.1	Convalida incrociata	84
5.2	Test e valutazione della performance	85

5.2.1	Analisi quantitativa	86
5.2.2	Analisi qualitativa e correzioni in post-processing	89
5.2.2.1	Eliminazione delle strutture esterne ai polmoni e dei vasi . .	91
5.3	Discussione	94
6	Conclusioni	97

Capitolo 1

Apparato respiratorio e IPF

Nel primo capitolo, verranno descritti l'apparato respiratorio dal punto di vista anatomico e fisiologico e la Fibrosi Idiopatica Polmonare (IPF) nel tentativo di giustificare la metodologia sia dei processi di segmentazione allo stato dell'arte che di quello sviluppato durante il lavoro di tesi. Verranno citate e descritte le vie aeree superiori e inferiori per poi riportare la suddivisione dei polmoni in lobi e lobuli. Poi, passando alla fisiologia, verranno definiti i volumi polmonari e descritte le più importanti funzionalità respiratorie. Per quanto riguarda la parte dedicata all'IPF, invece, essa si articolerà in una descrizione della malattia, del suo decorso e del processo di diagnosi per il quale è necessario definire e riportare le linee guida stabilite dall'American Thoracic Society e dall'European Respiratory Society (ATS/ERS). Infine, verranno richiamati i punti salienti delle metodiche di riferimento per l'imaging diagnostico della malattia, ovvero la tomografia computerizzata tradizionale e ad alta risoluzione (High Resolution Computed Tomography, HRCT), con un accenno conciso al protocollo di acquisizione adottato nella pratica clinica.

1.1 Anatomia dell'apparato respiratorio

La respirazione consiste nello scambio di gas tra l'organismo e l'ambiente che lo circonda. Nell'essere umano, la respirazione consente l'assorbimento di ossigeno da parte di cellule, tessuti e organi e l'eliminazione di anidride carbonica, prodotto di scarto dei principali processi metabolici cellulari. Lo scambio gassoso avviene all'interno dei polmoni, posti nella cavità

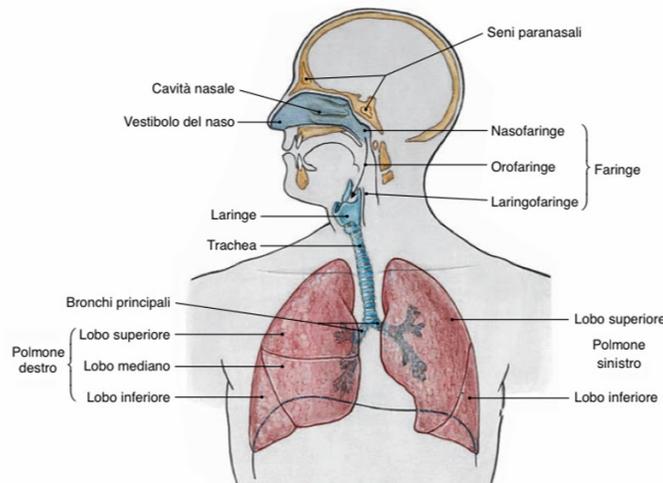


Fig 1.1: organi dell'apparato respiratorio

toracica, che sono in comunicazione con l'esterno attraverso le vie aeree. L'apparato respiratorio, quindi, è un sistema anatomicamente complesso che comprende tutti gli organi deputati alla respirazione (Fig 1.1); esso è costituito dalle vie aeree superiori e inferiori, dai polmoni e dalle pleure, membrane sierose che facilitano lo scorrimento dei polmoni sulla parete toracica durante gli atti respiratori.

1.1.1 Le vie aeree

Le vie aeree, o vie respiratorie, si dividono in *superiori* e *inferiori*. Le vie aeree superiori comprendono il naso, la cavità orale, la faringe e la laringe; le vie aeree inferiori comprendono la trachea e i bronchi.

Il tratto superiore inizia con la sezione *superficiale*, costituita dalle fosse nasali che sono a contatto diretto con l'esterno e che, col naso, formano l'apertura principale esterna. Il **naso** ha una porzione esterna visibile fatta di osso e cartilagine con due aperture, dette narici, e una interna inserita all'interno della scatola cranica. Lo spazio all'interno del naso è detta cavità nasale ed è divisa in due porzioni da un setto verticale, il setto nasale.

La **faringe** è la parte della gola situata dietro la cavità orale e nasale e rappresenta una via di transito per l'aria che, dall'esterno, confluisce verso i polmoni. Dalla faringe, poi, l'aria passa alla **laringe**, anch'esso organo cavo che contiene le corde vocali. La laringe ha il compito di

convogliare l'aria alla trachea, ai bronchi e ai polmoni, di impedire l'ingresso dei cibi nelle vie aeree inferiori e di emettere i suoni.

Il tratto inferiore inizia con la **trachea**: condotto a parete fibromuscolare lungo circa 12 cm, che percorre la parte inferiore del collo e la parte superiore del torace in senso verticale. Ha un diametro di circa 2 cm ed è costituita da circa 20 anelli cartilaginei connessi tra di loro. All'altezza della IV vertebra toracica, si divide nei due bronchi principali, destro e sinistro[1]. I **bronchi** principali originano dalla biforcazione inferiore della trachea e costituiscono l'ultimo tratto delle vie respiratorie. I due bronchi si dirigono verso l'ilo polmonare dove si ramificano ulteriormente a formare l'albero bronchiale. Così, i bronchi principali e le prime ramificazioni rappresentano il segmento *extrapolmonare*; la parte dell'albero che si trova tutto all'interno del polmone è il segmento *intrapolmonare*. Le ramificazioni che originano dalla trachea fanno sì che ogni lobo abbia il proprio bronco, detto *bronco lobare* che si suddivide ancora in *bronco segmentario* e via via sino ai *bronchioli terminali* e ai *bronchioli alveolari*.

1.1.2 I polmoni

I **polmoni** sono due e si trovano nella cavità toracica che riempiono quasi completamente, lasciando un vano nel centro, il mediastino, occupato da cuore e altre strutture tra le quali esofago e trachea. Ciascun polmone ha la forma di cono irregolare ed è alto circa 23 cm. Il polmone destro è diviso da due profonde incisive in tre parti o *lobi*: superiore, medio e inferiore; invece, il polmone sinistro è diviso in due soli lobi: superiore e inferiore (Fig 1.2). Le incisive prendono il nome di *scissure*: la grande scissura separa i lobi superiore ed inferiore di entrambi i polmoni, la piccola scissura si trova solo a destra e separa il lobo superiore dal medio. La faccia dei polmoni che si affaccia sul mediastino presenta una regione, il cosiddetto *ilo polmonare*, attraverso il quale penetrano i bronchi e le arterie polmonari e fuoriescono le vene e i vasi linfatici.

La struttura interna del polmone dipende dalle molteplici ramificazioni dei bronchi che, dividendosi e suddividendosi, terminano in tubuli sottili detti *bronchioli polmonari* che, a loro volta, terminano negli *alveoli polmonari*. Gli alveoli polmonari assieme ai bronchioli respiratori dai quali dipendono, ai dotti e ai sacchi alveolari rappresentano l'unità fondamentale del polmone o *lobulo primario* in quanto attraverso esso avviene lo scambio di gas tra aria alveo-

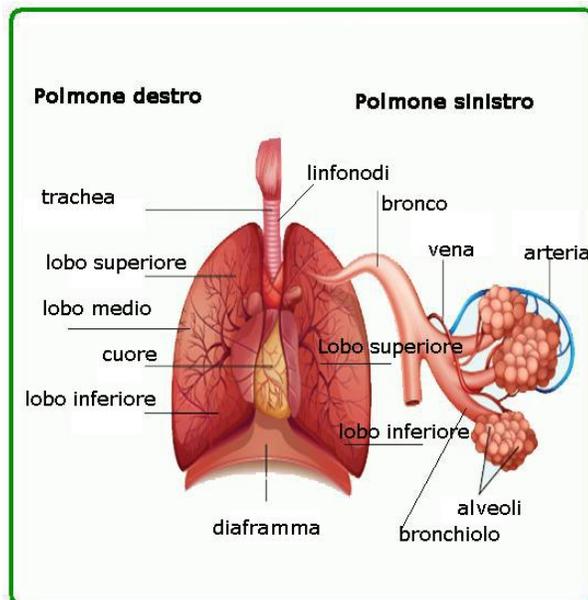


Fig 1.2: suddivisione dei polmoni e dei bronchi

lare e sangue capillare alveolare.

La superficie dei polmoni è suddivisa in moltissimi poligoni detti **lobuli polmonari**. Ogni lobulo è un'entità indipendente ed è in contatto con altri lobuli attraverso il connettivo interstiziale che diventa via via più visibile con l'età a causa del deposito progressivo di granuli di carbone, trasportati con l'aria negli alveoli. I bronchi principali, penetrati nell'ilo polmonare, si suddividono in rami sempre più numerosi e sottili, gli ultimi dei quali, i più fini, finiscono ciascuno in un lobulo polmonare e prendono il nome di *bronco lobulare*. Ogni bronco lobulare si divide, poi, in *bronchioli intra-lobulari* ed ognuno di questi dà origine a *bronchioli terminali* dai quali si dipartono i *bronchioli respiratori*. Ogni bronchiolo respiratorio si suddivide infine in *condotti alveolari* che hanno una parete tutta occupata da estroflessioni emisferiche contigue, gli *alveoli*, che terminano, dilatandosi, negli *infundiboli*. Il lobulo polmonare, in definitiva, è formato dalle ramificazioni del bronco lobulare e dei condotti alveolari; gli alveoli sono tanto numerosi che vengono a contatto reciprocamente con la loro parete. Bronchioli terminali, bronchioli respiratori, condotti alveolari e alveoli formano un *acino polmonare* (Fig. 3).

L'insieme di tutte le ramificazioni bronchiali all'interno dei lobuli prende il nome di **parenchima** polmonare. L'unità fondamentale del parenchima è proprio l'acino polmonare con diametro di 1 cm e costituito da un numero di alveoli compreso tra 60 e 120.

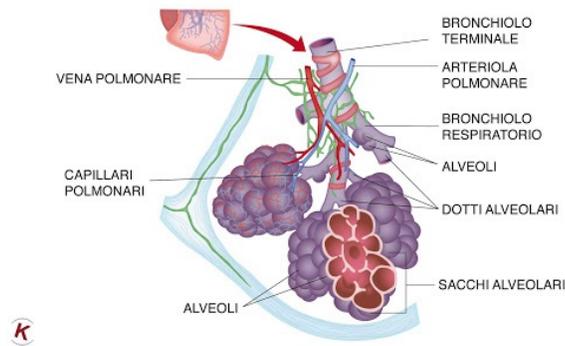


Fig 1.3: acino polmonare.

1.2 Fisiologia dell'apparato respiratorio

1.2.1 Funzioni respiratorie e volumi polmonari

Come già affermato nel paragrafo 1.1, la respirazione consiste nello scambio di gas tra l'organismo e l'ambiente esterno. Nel caso dell'uomo, si può dire che consiste nell'assorbimento di ossigeno e nell'eliminazione di anidride carbonica: l'ossigeno viene portato negli alveoli polmonari grazie alla funzione ventilatoria polmonare e da qui trasferito ai tessuti per mezzo della funzione cardiocircolatoria ed ematica; a livello cellulare, si forma anidride carbonica che viene portata nell'ambiente esterno ad opera della circolazione e della ventilazione (Fig 1.4).

Della funzione respiratoria, quindi, si possono distinguere le seguenti componenti[2]:

- *componente ventilatoria*: ha il compito di trasportare ossigeno dall'ambiente esterno agli alveoli polmonari e anidride carbonica in senso inverso;
- *componente alveolocapillare*: per il passaggio dei gas respiratori tra alveoli e sangue;
- *componente cardiocircolatoria*: per il trasporto di ossigeno ai tessuti in attività e dell'anidride carbonica da questi agli alveoli;
- *componente tessutale*: ha il compito di utilizzare l'ossigeno pervenuto nel metabolismo cellulare e di produrre anidride carbonica.

Il processo mediante il quale l'aria contenuta negli alveoli viene rinnovata è detto *ventilazione polmonare* e avviene con i movimenti respiratori di inspirazione, che consiste nell'introduzione

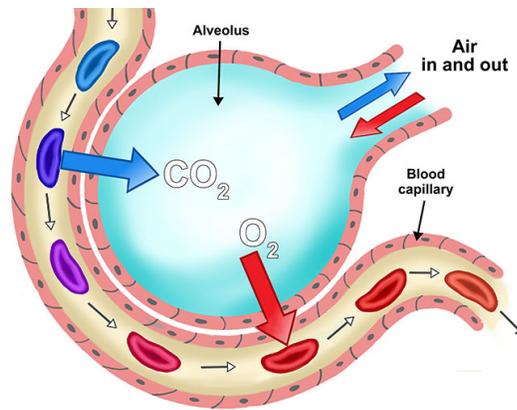


Fig 1.4: scambio di gas tra aria alveolare e capillare alveolare.

ne di aria nei polmoni, ed espirazione che permette l'uscita della stessa dai polmoni. Entrambi i movimenti vengono compiuti ad opera dei muscoli intercostali e del diaframma. Infatti, durante l'inspirazione i muscoli intercostali si contraggono sollevando le costole e favorendo l'espansione della gabbia toracica, mentre il diaframma si contrae comprimendo verso il basso gli organi addominali e garantendo ai polmoni un maggiore spazio di espansione. Durante l'espirazione, invece, i muscoli intercostali si rilassano richiamando verso il basso le costole e riducendo il volume della gabbia toracica e il diaframma si rilassa. I polmoni, durante gli atti respiratori, seguono passivamente i movimenti della gabbia toracica sebbene le loro insite proprietà elastiche fanno sì che si oppongano alle deformazioni. Un'altra resistenza all'espansione dei polmoni è data dagli ostacoli che il flusso gassoso incontra nelle vie aeree.

Ad ogni atto respiratorio, in un uomo che respira regolarmente, entra una quantità di aria di circa 500 ml detta *volume ventilatorio*. Mediante un'inspirazione volontaria massimale (eseguita, cioè, ampliando quanto più possibile la gabbia toracica), un uomo è in grado di immettere nei polmoni circa 3000 ml di aria: questa rappresenta la *riserva inspiratoria*. Nei polmoni, poi, dopo un'espirazione normale permane sempre una certa quantità di aria e questa viene detta *riserva espiratoria*; pur facendo uno sforzo massimale espiratorio, non si sarà mai in grado di espellere tutta l'aria contenuta nei polmoni, a cui si dà il nome di *volume residuo*. La somma del volume ventilatorio, della riserva inspiratoria e della riserva espiratoria prende il nome di *capacità vitale* che sommata al volume residuo definisce la *capacità polmonare totale* che esprime il volume di aria massimale che può essere contenuto nei polmoni. Il suo valore, in un soggetto sano, è di circa 6 litri.

Una volta espletata la funzione ventilatoria, l'ossigeno trasportato negli alveoli viene in contatto con il sangue che scorre nei capillari alveolari attraverso la *membrana alveolo-capillare* costituita dalla parete alveolare e da quella capillare. Il sangue proveniente dai tessuti in attività è povero di ossigeno e ricco di anidride carbonica; a livello alveolo-capillare si riscontra, quindi, uno scambio di ossigeno verso il sangue e di anidride carbonica in senso inverso. Il sangue giunto alla fine del capillare alveolare ha perso le sue caratteristiche venose e ha assunto quelle arteriose. Questo processo di *diffusione gassosa* avviene anche a livello dei tessuti dove il sangue proveniente dai polmoni cede ossigeno e si arricchisce di anidride carbonica. Lo scambio di anidride carbonica e di ossigeno tra le cellule dei tessuti e il sangue viene detta *respirazione interna* mentre lo scambio tra il sangue dei capillari polmonari e l'aria dei polmoni prende il nome di *respirazione esterna*. In quest'ultimo caso, a livello degli alveoli, il sangue circola nella rete finissima e fittissima dei capillari e viene a trovarsi quasi a contatto diretto con l'aria. In questa situazione, si deve stabilire un equilibrio in cui ciascun gas presente o nel liquido capillare o nell'aria alveolare deve passare da un punto di maggior pressione a uno di minor pressione. Per cui, l'ossigeno passerà dall'aria alveolare al sangue e l'anidride carbonica dal sangue all'aria alveolare. A livello dei tessuti, invece, succede il fenomeno contrario: il sangue ricco di ossigeno viene in contatto coi liquidi che bagnano le cellule che sono saturi di anidride carbonica continuamente prodotta da esse. Di nuovo, si stabilisce un equilibrio ma sarà il sangue a cedere ossigeno e l'anidride carbonica ad abbandonare i tessuti, per trasferirsi nel sangue. L'aumento di attività metabolica dei tessuti determina sempre un aumento del suo fabbisogno di ossigeno che può essere apportato ai tessuti o attraverso un maggiore afflusso di sangue nei capillari tissutali o mediante un aumento di quantità di ossigeno prelevata da un dato volume di sangue. In definitiva, il sangue è indispensabile per il trasporto dei gas respiratori dai polmoni ai tessuti e viceversa e il cuore ha il compito di mantenere e regolare la circolazione ematica in base alle richieste energetiche.

1.2.2 Esami funzionali

Nelle malattie polmonari, la funzione respiratoria può presentarsi alterata, quindi insufficiente. Si definisce *insufficiente respiratorio* quel soggetto con un'alterazione di uno degli organi che partecipano alla funzione respiratoria, anche in assenza di sintomi evidenti[2]. Gli esami

di funzionalità respiratoria forniscono un aiuto indispensabile ai fini di una diagnosi funzionale corretta o per trarre deduzioni prognostiche. Come detto nel paragrafo precedente, la funzione respiratoria si esplica attraverso quattro componenti; le varie componenti possono essere studiate indipendentemente le une dalle altre oppure contemporaneamente.

Lo studio globale, invece, viene eseguito sottoponendo il soggetto a un esercizio fisico di diversa entità e con diverse modalità.

1.2.2.1 Volumetria polmonare e valutazione della diffusione alveolo-capillare

La volumetria polmonare viene svolta attraverso uno spirometro attraverso cui è possibile misurare i volumi di gas espirati e valutare il volume della capacità vitale ovvero il volume di aria che il soggetto espira quando esegue un'espiazione massima preceduta da un'inspirazione massima. Un esame più complesso, invece, consiste nella registrazione del volume espiratorio massimo al secondo, ovvero il volume di aria che un soggetto può espirare nel primo secondo di espiazione, quando essa viene eseguita più rapidamente e profondamente possibile. Questa grandezza è fondamentale per valutare la *capacità di rapida espiazione* che è strettamente collegata all'elasticità della gabbia toracica, alla forza di contrazione dei muscoli, alla canalizzazione bronchiale e all'elasticità del parenchima polmonare.

La misurazione del volume residuo, che è la quantità di aria presente alla fine di un'espiazione forzata, è importante per inquadrare un'eventuale disfunzione respiratoria. In base alla conoscenza di questi volumi e dei relativi rapporti, le disfunzioni possono essere distinte in due diversi gruppi:

- di forma *restrittiva*: si presenta nei casi di riduzione della superficie polmonare ventilata in cui vi è una diminuzione dei volumi che, però, conservano i rapporti funzionali;
- di forma *ostruttiva*: tipica di quei casi che presentano un comportamento diverso delle singole grandezze volumetriche e in cui la ventilazione polmonare è disomogenea.

Non è detto che l'aria che penetra nell'apparato respiratorio durante l'atto respiratorio partecipi agli scambi gassosi; il volume di aria che non si equilibra con il sangue alveolare capillare prende il nome di *spazio morto respiratorio funzionale*. Esso aumenta nel caso di fibrosi, a causa dell'ispessimento della membrana alveolare e nei processi ostruttivi, a causa dell'aumento

di volume degli alveoli che non permette a tutta l'aria di equilibrarsi. Le alterazioni degli scambi gassosi sono studiate prevalentemente attraverso la ricerca della tensione parziale dell'ossigeno e dell'anidride carbonica, presenti in una piccola quantità di sangue prelevato dalle arterie. Nella sindrome restrittiva l'ispessimento della membrana alveolo-capillare impedisce all'ossigeno di equilibrarsi con quello capillare. Nella sindrome ostruttiva vi è un'alterazione del rapporto ventilazione-perfusione dei singoli alveoli a causa della differente canalizzazione bronchiale e alla diversa elasticità della parete alveolare.

1.3 La fibrosi idiopatica polmonare (IPF)

Le pneumopatie dell'interstizio sono un gruppo di patologie che interessano il tessuto connettivo. Il tessuto connettivo, interponendosi tra gli alveoli ed i vasi sanguigni, garantisce il supporto necessario agli scambi gassosi; la compromissione degli scambi gassosi porta all'insorgenza di dispnea (mancanza di respiro) o tosse secca, che sono i primi sintomi più comuni della patologia.

Nel 2002[3] l'American Thoracic Society (ATS) e l'European Respiratory Society (ERS) hanno classificato le interstiziopatie e introdotto il termine di *malattia diffusa del parenchima polmonare (Diffuse Parenchymal Lung Disease, DPLD)* a sottolineare come proprio il parenchima sia il primo sito danneggiato. La classificazione raggruppa le DPLD in DPLD da causa nota, DPLD granulomatose, DPLD idiopatiche (ovvero a causa sconosciuta) ed altre forme. Le DPLD idiopatiche (*Idiopathic Interstitial Pneumonia, IIP*) sono state ulteriormente suddivise in Fibrosi Polmonare Idiopatica (IPF) e IIP diverse dall'IPF.

La **fibrosi idiopatica polmonare**, quindi, è una forma specifica di fibrosi interstiziale polmonare idiopatica, associata alla presenza di un pattern istopatologico o radiologico di *Usual Interstitial Pneumoniae (UIP)* che coinvolge il polmone in modo eterogeneo. L'esame radiologico avrà esito normale nella maggior parte dei pazienti nelle prime fasi della patologia, mentre negli stadi avanzati mostrerà una riduzione notevole dei volumi polmonari e opacità reticolari nelle regioni sub-pleuriche che aumentano dagli apici alle basi dei polmoni (Fig 1.5). Ad ora, l'evoluzione naturale della patologia è di difficile previsione: alcuni pazienti restano stabili per anni fino a quando un evento scatenante determina un netto peggioramento del quadro cli-

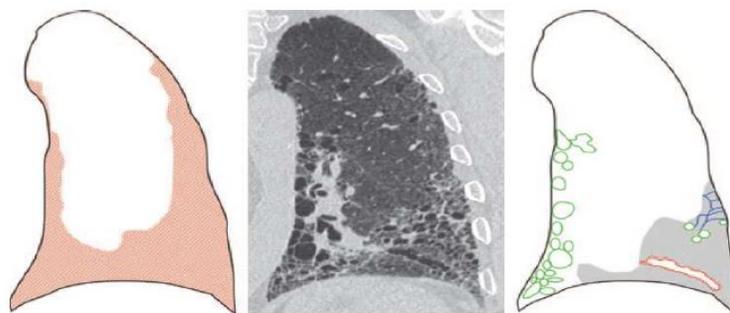


Fig 1.5[4]: distribuzione, immagine e pattern UIP

nico che può anche portare a morte; altri pazienti hanno un'evoluzione rapida dal momento della diagnosi anche senza fattori scatenanti; altri ancora presentano un lento declino che può precipitare quando si verifica un evento scatenante[5].

A fronte dell'imprevedibilità della patologia, allora, sono stati ricercati i fattori che correlasero meglio con la prognosi; tra questi, spiccano le caratteristiche delle immagini ottenute tramite l'HRCT dei polmoni con IPF. Infatti, l'estensione di tessuto malato si è dimostrata essere correlata sia con il grado di coinvolgimento patologico che con la gravità delle alterazioni fisiologiche. Diversi studi hanno focalizzato la loro attenzione nel dimostrare statisticamente come lo *score fibrotico*, ovvero il punteggio soggettivo che il radiologo attribuisce all'estensione della malattia sulle immagini, sia correlato anche alla previsione dell'evoluzione della malattia; infatti, questo si è dimostrato essere, ancor più dei risultati istopatologici, correlato con la prognosi[6].

1.3.1 Diagnosi e pattern UIP

Avere una diagnosi precisa di quale sia la pneumopatia diffusa è fondamentale per individuare la terapia corretta e fare previsione sull'evoluzione dello stato di salute del paziente.

Una diagnosi corretta è il risultato di un processo multidisciplinare che richiede una completa valutazione clinica da parte dello pneumologo, del radiologo, del chirurgo toracico e dell'anatomo-patologo. Generalmente, il processo che conduce ad una diagnosi inizia con l'indagine sulla storia clinica del paziente, con gli esami fisici, con le radiografie tradizionali del torace e con i test funzionali. Sulla base di questi, la pneumopatia viene innanzitutto definita come idiopatica o non idiopatica; se è idiopatica, il paziente viene sottoposto ad una tomogra-

fia ad alta risoluzione (HRCT) il cui compito principale è discriminare pazienti con features tipiche dell'IPF da pazienti che presentano features meno specifiche che possono essere associate ad altre pneumonie[7]. L'esame radiologico di tomografia ad alta risoluzione, quindi, è quasi sempre necessario perché, sebbene sia insufficiente se preso da solo, è fondamentale sia per avere un primo orientamento diagnostico sia per monitorare nel tempo la progressione della patologia.

Le HRCT devono essere attentamente analizzate da radiologi esperti nella valutazione delle features tipiche delle interstiziopatie polmonari, che sono l'*honeycombing*, le *bronchiectasie da trazione*, le *opacità a vetro smerigliato* e il *quadro reticolare*[4].

- L'*honeycombing* si riferisce ad un fitto reticolo a trama spessa che delimita aree trasparenti (quindi radiologicamente nere) di piccole dimensioni;
- Le bronchiectasie/bronchiolectasie da trazione consistono nella dilatazione bronchiale localizzata o diffusa che possono apparire come cisti o microcisti che, giustapposte, possono essere difficilmente distinguibili dall'*honeycombing*;
- Le opacità a vetro smerigliato consistono in una minor trasparenza dei campi polmonari e, quando circondano i bronchi o bronchioli, possono erroneamente spingere a categorizzarli come bronchiectasie/bronchiolectasie;
- Il quadro reticolare si riferisce alla presenza di linee radiopache intrecciate fra loro.

La distribuzione di tutti questi pattern è tipicamente subpleurica e basale: in caso contrario, il pattern TC non è compatibile con la diagnosi di UIP.

Le linee guida per una corretta diagnosi radiologica raccomandano di incorporare i quattro pattern appena citati in tre categorie (Fig 1.7): *pattern UIP*, *possibile pattern UIP*, *pattern non UIP*. Il pattern UIP consiste nella coesistenza di quadro reticolare e *honeycombing* nelle regioni subpleuriche e basali; il possibile pattern UIP nel solo quadro reticolare riscontrato nelle regioni sub-pleuriche e basali mentre il pattern non UIP è caratterizzato da almeno una delle seguenti caratteristiche:

1. interessamento polmonare localizzato soprattutto in campo medio e superiore del polmone;

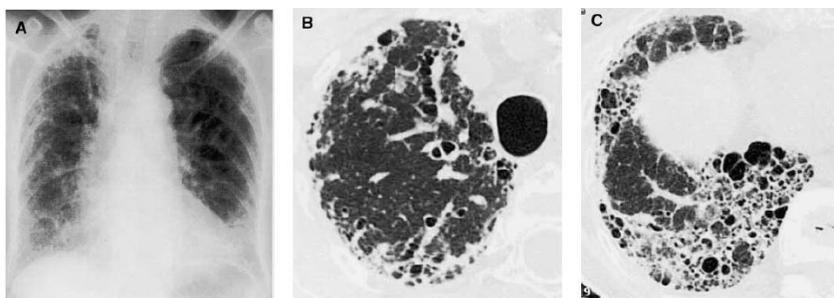


Fig 1.6[7]: IPF. (A) radiografia al torace tradizionale che mostra un'opacità reticolare tipica, marcata alle basi dei polmoni, con honeycombing; la radiografia tradizionale potrebbe addirittura apparire normale in pazienti con IPF. (B-C) mostrano un'immagine più risoluta e precisa in cui appare evidente la predominanza delle alterazioni ai tessuti con bronchiectasie ed honeycombing in stato aggravato.

UIP Pattern (All Four Features)	Possible UIP Pattern (All Three Features)	Inconsistent with UIP Pattern (Any of the Seven Features)
<ul style="list-style-type: none"> • Subpleural, basal predominance • Reticular abnormality • Honeycombing with or without traction bronchiectasis • Absence of features listed as inconsistent with UIP pattern (see third column) 	<ul style="list-style-type: none"> • Subpleural, basal predominance • Reticular abnormality • Absence of features listed as inconsistent with UIP pattern (see third column) 	<ul style="list-style-type: none"> • Upper or mid-lung predominance • Peribronchovascular predominance • Extensive ground glass abnormality (extent > reticular abnormality) • Profuse micronodules (bilateral, predominantly upper lobes) • Discrete cysts (multiple, bilateral, away from areas of honeycombing) • Diffuse mosaic attenuation/air-trapping (bilateral, in three or more lobes) • Consolidation in bronchopulmonary segment(s)/lobe(s)

Definition of abbreviation: UIP = usual interstitial pneumonia.

Fig 1.7[5]: tabella riassuntiva delle categorie raccomandate dalle linee guida ATS/ERS.

2. interessamento peribroncovasale;
3. opacità a vetro smerigliato con estensione maggiore delle alterazioni reticolari;
4. presenza di micronoduli diffusi;
5. presenza di cisti;
6. aree a mosaico, zone di air trapping e consolidamenti

In Fig 1.8, è riassunto schematicamente il processo; il passaggio fondamentale è la corretta analisi delle immagini HRCT, perché è in seguito a questa che si decide con quale operazione, di natura molto invasiva, procedere; possono seguire la biopsia trans-bronchiale, il lavaggio bronco-alveolare o la biopsia chirurgica del polmone affinché vengano prelevati campioni da cui estrapolare pattern istopatologici che, in combinazione con quelli radiologici, confermino o smentiscano il quadro diagnostico.

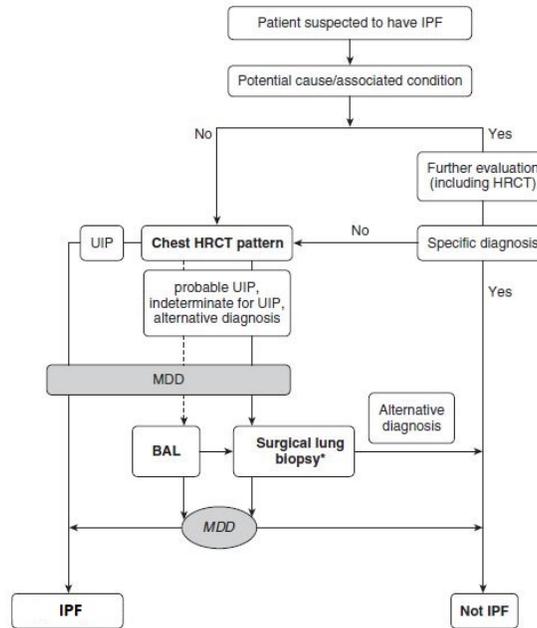


Fig 1.8[8]: schema riassuntivo del processo di diagnosi. Il paziente con sospetta IPF dev'essere attentamente interrogato per identificare una possibile causa di ILD; se non ne viene trovata nessuna, oppure esami successivi non portano ad una diagnosi, allora si esegue un'HRCT. Se risulta sufficiente, i pattern radiologici possono essere valutati da un team multi-disciplinare (MDD = multidisciplinary discussion) che può diagnosticare definitivamente l'IPF oppure prescrivere il lavaggio broncoalveolare (BAL = Bronchoalveolar Lavage) o la biopsia. I risultati di questi ultimi due verranno, a loro volta, valutati per concludere il processo di diagnosi che stabilirà se è IPF o altro.

1.4 Tomografia computerizzata e HRCT

La **tomografia computerizzata** (*Computed Tomography*) è una procedura di imaging diagnostico che impiega raggi X per la ricostruzione di sezioni trasversali del corpo; si basa sul principio per il quale è possibile risalire alla densità del tessuto a partire dall'attenuazione che il fotone subisce quando lo attraversa.

Nella CT, il tubo a raggi X ruota attorno al paziente nel *gantry* (Fig 1.9), emettendo un fascio conico di fotoni con energia tipicamente compresa tra 20 e 150 keV; il fascio attraversa uno strato di materiale biologico di spessore x ed un rivelatore, posto dalla parte diametralmente opposta rispetto all'emettitore, rivela il fotone uscente dal campione. In verità, soltanto negli scanner di prima generazione il rivelatore era singolo; a fronte del tempo di acquisizione elevato che aumentava il tasso di artefatti da movimento nell'immagine e limitava la risoluzione



Fig 1.9: scanner TC con lettino e gantry; il gantry è una struttura circolare rotante, che monta un tubo radiogeno su un lato e uno o più rivelatori, sul lato diametralmente opposto.

temporale, l'innovazione tecnologica ha permesso alle nuove generazioni TC di adottare un sistema *multi-detettore* in cui più rivelatori sono disposti su più file; così è possibile acquisire molti più dati a parità di tempo.

Una volta rivelato il fotone, in accordo con l'equazione (1.1), viene misurata l'attenuazione. I dati noti sono I_0 , I e x , rispettivamente l'intensità del raggio incidente, della radiazione uscente e lo spessore del volume attraversato. L'incognita è il *coefficiente di attenuazione lineare*, μ , che dipende sia dalla densità del materiale attraversato, quindi dal numero atomico, che dall'energia del fascio che lo investe.

$$I = I_0 e^{-\mu x} \quad (1.1)$$

Se il fascio di raggi X ha bassa energia e investe un oggetto con elevato numero atomico e spessore ampio allora subirà un'attenuazione maggiore; viceversa, se attraversa un materiale a bassa densità, percorre uno spessore piccolo e ha energia più alta, allora l'attenuazione sarà minore. Per questo, a parità di energia del fascio e spessore del corpo attraversato, gli oggetti a densità maggiore appariranno chiari perché subiscono massima attenuazione, come le ossa, e gli oggetti a densità minore appariranno più scuri perché subiranno minima attenuazione, come i polmoni, pieni di aria.

L'attenuazione è dovuta principalmente a due fenomeni di assorbimento: l'*effetto fotoelettrico* e l'*effetto Compton* (Fig 1.10). L'effetto fotoelettrico consiste nell'assorbimento completo del

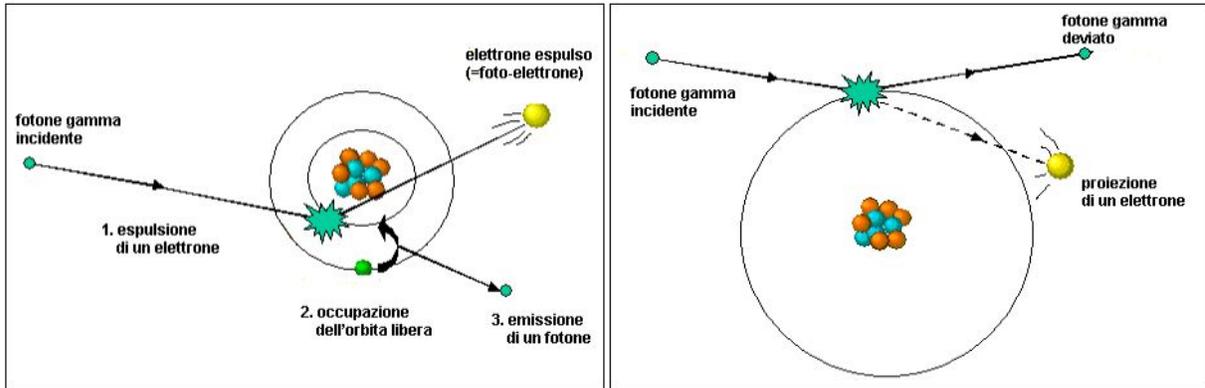


Fig 1.10[9]: Effetto fotoelettrico ed effetto Compton

fotone incidente che scalza un elettrone di un orbitale interno, che viene espulso, e provoca l'emissione di un raggio X caratteristico a seguito del riempimento della vacanza dell'orbitale interno da parte di un elettrone proveniente da un orbitale più esterno. L'effetto Compton, invece, è un fenomeno di scattering dovuto all'interazione tra il fotone incidente e un elettrone dell'orbitale più esterno che assorbe parte dell'energia e devia la traiettoria del fotone incidente che prosegue, provocando altri scattering o arrestandosi definitivamente per assorbimento fotoelettrico. L'effetto Compton è un effetto indesiderato che peggiora il contrasto dell'immagine ed espone ad un maggior rischio anche l'operatore. Una volta misurato il coefficiente di attenuazione, si ricava, applicando una trasformazione lineare, il relativo *numero di Hounsfield*; la scala Hounsfield è una scala di unità di misura usata per descrivere quantitativamente la radiodensità del materiale e si ottiene con la (1.2): μ_{acqua} è la radiodensità dell'acqua distillata a pressione e temperatura in condizioni standard (0 Hounsfield Unit) mentre μ_{aria} quella dell'aria (-1000 HU). Valori tipici dei principali tessuti sono rappresentati in Tabella 1.

$$HU = 1000 \times \frac{\mu - \mu_{\text{acqua}}}{\mu_{\text{acqua}} - \mu_{\text{aria}}} \quad (1.2)$$

Una volta acquisiti i segnali provenienti dai rivelatori, i dati *raw* uscenti dallo scanner vengono processati attraverso algoritmi matematicamente complessi, per ottenere un'immagine tridimensionale di una sezione bidimensionale perpendicolare all'asse del sistema di acquisizione. L'informazione sulla terza dimensione, quindi sulla profondità, viene ottenuta grazie alle molteplici proiezioni dell'oggetto lungo diverse direzioni. Le dimensioni tipiche di un'immagine

Tessuto	HU
Aria	-1000
Polmone	-500
Adipe	da -100 a -50
Acqua	0
Fluido cerebrospinale	15
Rene	30
Sangue	da +30 a +45

Tabella 1: valori HU di alcuni tessuti tipici

tomografica in uno scanner medicale standard sono di 512×512 voxel e profondità di 16 bit per pixel, anche se con la tecnologia attuale si possono ottenere risultati anche migliori sia in termini di numero di voxel che di risoluzione spaziale.

I volumi di dati ottenuti grazie alla TC sono manipolabili attraverso un sistema noto come *windowing* attraverso cui è possibile visualizzare le varie strutture anatomiche in base alla loro capacità di attenuare il fascio di raggi X; è per questo che con la TC si ottengono risultati migliori rispetto alla radiologia tradizionale per quanto riguarda la differenziazione dei tessuti molli e dei muscoli. Attraverso l'operazione di *windowing*, infatti, si equalizza l'istogramma dei livelli di grigio dell'immagine per migliorarne il contrasto. Il processo consiste nella manipolazione dei livelli di grigio dell'immagine attraverso i numeri di Hounsfield per mettere in evidenza le strutture di maggiore interesse: la larghezza della finestra, infatti, determina il range di numeri CT che l'immagine deve contenere. Una finestra più ampia mostrerà una transizione da zone chiare a zone scure meno netta di quella che verrebbe mostrata da una finestra più ristretta[10]. Per cui, l'ampiezza della finestra si riflette sulla possibilità di discriminare coefficienti di attenuazione di tessuti molli diversi. Il centro della finestra, invece, determina la luminosità dell'immagine. Valori tipici di ampiezza e centro della finestra per i polmoni sono, rispettivamente, 1500 e -600.

Per la diagnosi di interstiziopatia polmonare, la metodica di riferimento è la **tomografia computerizzata ad alta risoluzione** del torace (*High Resolution Computed Tomography*) che rive-

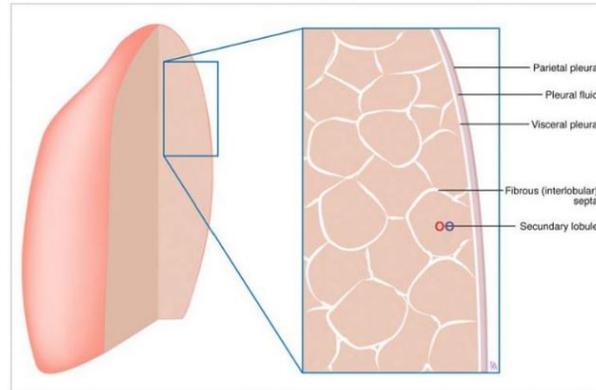


Fig 1.11[11]: lobulo secondario e setti connettivali.

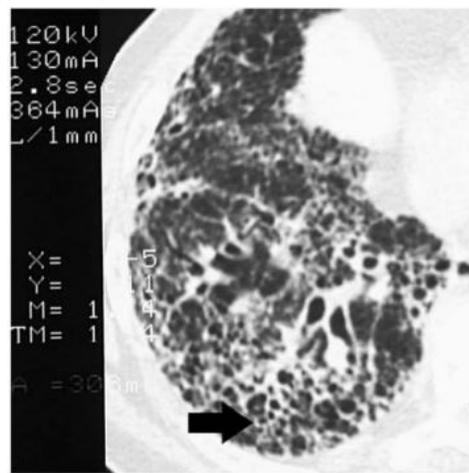


Fig 1.12[11]: le features HRCT dell'IPF comprendono l'ispessimento dei setti intra- e inter-lobulari

ste un ruolo tanto fondamentale da sostituirsi, talvolta, anche all'anatomia patologica. Infatti, le immagini in vivo acquisite risultano paragonabili ad un preparato anatomico. Nelle HRCT, lo spessore di fetta è tipicamente compreso tra i 0.625 mm e 1.5 mm, e la frequenza spaziale usata dall'algoritmo di ricostruzione dell'immagine è elevata o medio-elevata. Per questo, l'HRCT rappresenta l'unica metodica che permette l'identificazione del lobulo secondario, che è la più piccola porzione di polmone delimitata dai setti connettivali di dimensione 0,2 - 0,3 mm (Fig 1.11). A seguito dello sviluppo e dell'impiego di scanner CT multidetector (MDCT) in grado di acquisire, in un unico respiro, dati isotropi lungo tutto il torace, due approcci possono essere utilizzati per l'acquisizione delle immagini: la tecnica *assiale* e la tecnica *volumetrica* (Fig 1.13). La più tradizionale è la tecnica assiale tramite la quale si ottengono immagini di sezioni spesse circa 1 mm acquisite ogni 10-20 mm lungo tutto il torace. La tecnica volumetrica,

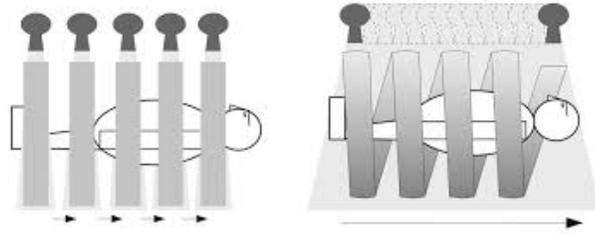


Fig 1.13: tecnica assiale (sinistra) e tecnica volumetrica (destra)

invece, è da preferire nel caso delleILD perché permette di acquisire dinamicamente immagini di strutture sovrapposte e di fare ricostruzioni multi-planari così da facilitare la differenziazione tra honeycombing e bronchiectasie da trazione, features cruciali per la diagnosi di IPF[12]. La tecnica volumetrica si esegue facendo avanzare il lettino durante la rotazione del gantry così che le proiezioni seguano una traiettoria ellittica; però, dev'essere eseguita solo con quegli scanner multi-detettore che hanno almeno 16 file di detettori: se il numero è inferiore, il tempo di acquisizione aumenta eccessivamente rispetto alla capacità del paziente di trattenere il respiro. Un altro punto a favore delle tecniche volumetriche risiede nell'applicazione di algoritmi di post-processing alle immagini 3-D, come la *maximum intensity projection* (MIP), che evidenzia strutture molto dense contro uno sfondo poco denso (utile per la differenziazione dei vasi), oppure la *minimum intensity projection* (minIP) che esplora strutture ipodense contro uno sfondo naturalmente iperdenso (utile per delineare le strutture bronchiali contro il parenchima polmonare); ci sono, per ultimo, algoritmi per l'analisi quantitativa che lavorano su dati volumetrici di polmoni e vie aeree.

Comunque, la scelta di una o dell'altra tecnica dipende dalla strumentazione disponibile, dalle indicazioni cliniche e da considerazioni accurate sulla dose da somministrare al paziente: per la tecnica assiale la dose si aggira attorno agli 0,7 mSv mentre per la tecnica volumetrica rientra in un intervallo che va dai 6,4 a 7,8 mSv.

1.4.1 Protocollo di acquisizione delle immagini

Per ottimizzare la resa dell'esame CT, il protocollo di acquisizione deve seguire un'accurata revisione della storia clinica del paziente e degli eventuali esami radiologici già presenti; la tecnica scelta, poi, deve garantire che la qualità dell'immagine soddisfi il quesito diagnostico

compatibilmente col minor rateo di esposizione alle radiazioni a cui si può sottoporre il paziente.

Sebbene la maggior parte delle operazioni di uno scanner CT siano automatizzate, un certo numero di parametri tecnici restano ancora operatore-dipendente e, dal momento che questi possono influenzare significativamente il valore diagnostico dell'esame, è necessario che il protocollo venga stabilito una volta definiti i seguenti punti:

1. Scelta della tecnica, se assiale o volumetrica.
2. Esposizione alle radiazioni proporzionale alle dimensioni del paziente: se giovane, s'incoraggia l'applicazione di un voltaggio inferiore;
3. Controllo automatico dell'esposizione per la minimizzazione della dose (modulazione angolare e longitudinale della corrente del tubo catodico);
4. Collimazione del fascio;
5. Spessore della sezione per i sistemi multi-detettore. In qualsiasi caso, dev'essere inferiore al millimetro e mezzo;
6. Pitch¹, tempo di rotazione del gantry e velocità con cui viene fatto avanzare il lettino, inferiore al secondo;
7. Dimensione della matrice dell'immagine;
8. Definizione del campo di vista, centrato sui polmoni e stabilito sulla base delle dimensioni del paziente per ottimizzare la risoluzione spaziale;
9. Posizionamento corretto del paziente all'isocentro per minimizzare la dose irradiata e ottimizzare la qualità dell'immagine;

¹È un parametro adimensionale definito come la distanza percorsa dal lettino durante una scansione diviso l'ampiezza della collimazione del fascio radiante; si calcola all'isocentro ed ha come valore di riferimento il valore 1 che significa che lo spostamento del lettino è uguale alla dimensione della collimazione. Quando è inferiore a 1 invece, lo spostamento del lettino è minore rispetto alla collimazione e questo determina maggior irradiazione e sovracampionamento dei dati. Al contrario, per valori superiori a 1 si ha sottocampionamento e risparmio di dose radiante.

10. Istruzioni precise sullo svolgimento dell'atto respiratorio da parte del paziente. Solitamente, le immagini vengono acquisite mentre si trattiene il respiro ma se l'esame viene fatto durante l'espirazione, si acquisisce al termine di essa;
11. Impostazione del *windowing* dei livelli di grigio, quindi definizione del centro e ampiezza della finestra;
12. Algoritmo ed intervallo di ricostruzione dell'immagine;
13. Configurazione del singolo detettore nei sistemi multi-detettore;
14. Modalità di ricostruzione dell'immagine: multi-planare, curvilinea, MIP, o minIP;
15. Piano dell'immagine: assiale, coronale o sagittale.

In conclusione, l'HRCT riveste un ruolo centrale nella diagnosi di pazienti con sospetta IPF. La presenza di features cliniche tipiche e di pattern UIP, se riscontrata da clinici e radiologi esperti, da sola basterebbe a confermare i sospetti; tuttavia, nonostante l'importanza universalmente riconosciuta a tale indagine, non esiste un numero adeguato di specialisti in grado di interpretarla correttamente[13], con il risultato che spesso il paziente riceve un inquadramento diagnostico ed un trattamento terapeutico errato.

Capitolo 2

AI per il supporto alla diagnosi di interstiziopatia polmonare: stato dell'arte

Nel secondo capitolo verrà riportato lo stato dell'arte dei metodi automatici per l'analisi delle immagini tomografiche ad alta risoluzione (HRCT) delle interstiziopatie polmonari. Verrà spiegato in cosa consiste la tomografia computerizzata quantitativa (QCT) e quali sono le principali applicazioni, come l'analisi dell'istogramma e della texture dell'immagine. Verranno, poi, descritte le metodologie di sviluppo di due software ampiamente validati in letteratura: l'Adaptive Multiple Feature Method (AMFM) e il Computer-Aided Lung Informatics for Pathology Evaluation and Rating (CALIPER). Infine, si spiegherà perché, in che modo e attraverso quali applicazioni il deep learning supporta la diagnosi superando i limiti dei metodi precedenti.

2.1 Quantitative Computed Tomography (QCT)

Il rapido sviluppo delle tecniche di imaging ha fornito nuovi strumenti per lo studio morfologico e funzionale dei sistemi biologici umani. Nel caso delle patologie a carico dell'apparato respiratorio, negli ultimi due decenni, la tomografia computerizzata (CT) è stata sfruttata per caratterizzare dettagliatamente le strutture degli organi coinvolti. In particolare, le acquisizioni volumetriche hanno migliorato la qualità dell'analisi dei dati CT grazie ad algoritmi

computerizzati, categorizzati sotto il nome di *Quantitative Computed Tomography* (QCT), che forniscono uno strumento non invasivo per la visualizzazione diretta e la quantificazione delle strutture anatomiche nonché per la valutazione dell'evoluzione dei processi patologici. Molte immagini CT sono state, quindi, analizzate per definire biomarker da correlare su descrittori clinici, funzionali e prognostici.

Le applicazioni di maggior interesse sono: l'individuazione computer-assistita (Computer-Aided Detection, CAD) di noduli, l'analisi della texture delle lesioni, l'analisi densitometrica e la segmentazione delle vie aeree nelle patologie ostruttive e l'analisi della texture nelle interstiziopatie polmonari (*Interstitial Lung Disease*, ILD). Al momento, però, ci sono dei limiti[14]:

- Non è ancora chiaro se, e in che misura, questi metodi siano riproducibili al variare della strumentazione, quindi dello scanner e del protocollo di acquisizione. Questo ostacola l'analisi accurata negli studi multicentrici;
- La scelta dell'algoritmo di ricostruzione influenza la risoluzione dell'immagine la quale, a sua volta, si riflette sulla forma dell'istogramma dei livelli di grigio. Ad esempio, gli algoritmi di ricostruzione con risoluzione spaziale elevata riducono la skewness e la kurtosis[15];
- La densità dei polmoni influenza le features dell'immagine ma non può, ovviamente, essere standardizzata; infatti, oltre a non essere omogenea nello stesso paziente, varia da persona a persona per due fattori principali:
 1. Il livello di inspirazione prima di effettuare l'esame CT;
 2. la comorbidità con altre patologie.
- Molte metriche si limitano alla valutazione dell'estensione della malattia senza darne informazioni circa la stadiazione, che potrebbe essere lieve, moderata o grave;
- L'esigenza di avere immagini ad alta risoluzione, e con poco rumore, rende molti software difficilmente fruibili a causa del carico e del tempo computazionale. Pochi, infatti, si trovano sul mercato.

A causa di tutto questo, la QCT è confinata al solo settore della ricerca e non ancora adottata nella pratica clinica, nonostante molti studi abbiano dimostrato la forte correlazione tra le metriche fornite dagli strumenti automatici e quelle estratte dai tradizionali esami diagnostici, come i test funzionali.

Alcuni dei metodi computerizzati, utilizzati in studi condotti su pazienti con Fibrosi Idiopatica Polmonare (IPF) per l'analisi quantitativa della malattia nelle immagini HRCT, sono: l'analisi dell'istogramma e della texture dell'immagine, l'Adaptive Multiple Feature Method (AMFM) e il Computer-Aided Lung Informatics for Pathology Evaluation and Rating (CALIPER).

L'**analisi dell'istogramma** dell'immagine consiste nell'osservazione combinata di features quali il *valor medio*, la *skewness* e la *kurtosis*. Queste metriche risultano essere sufficientemente riproducibili al variare della dose ma se usati per quantificare i singoli pattern ILD, come l'honeycombing, l'opacità a vetro smerigliato (Ground Glass Opacity, GGO) e la reticolazione, risultano insufficienti. Per questo, sono stati implementati metodi più sofisticati basati sull'analisi della texture.

L'**analisi della texture** parte dalla selezione, da parte di osservatori esperti, di una regione di interesse (*Region-Of-Interest*, ROI) dell'immagine volumetrica che contiene un insieme di pattern. Dalla ROI vengono, poi, estratte le features di texture ed un algoritmo di machine learning sviluppa un modello predittivo.

L'**AMFM** è un software progettato per il riconoscimento dei pattern HRCT attraverso una combinazione di features che ne descrivono la texture; un classificatore bayesiano, poi, riconosce e quantifica il volume occupato dai pattern radiologici. Così, l'enfisema viene distinto dalla fibrosi o la GGO dal polmone normale. L'implementazione in 2-D non riesce ad identificare l'honeycombing a causa della bassa attenuazione in queste regioni (sono cisti aeree); per questo motivo, altri studi hanno esteso l'applicazione di AMFM al caso tridimensionale.

Infine, **CALIPER** è un tool che permette la caratterizzazione automatica del parenchima polmonare; esso usa una tecnica *vision-based* di analisi tridimensionale dell'istogramma e di analisi morfologica per classificare ogni pixel dell'immagine come appartenente ad uno di 7 pattern differenti: normale, GGO, reticolare, honeycombing, oppure aree con leggera, moderata o severa attenuazione. CALIPER è stato dimostrato essere riproducibile e robusto al variare delle tecniche di acquisizione dell'immagine, nonché della dose, su immagini ottenute con diversi

algoritmi di ricostruzione, come la retroproiezione filtrata o gli algoritmi iterativi.

La precisione, la sensibilità e l'ovvia mancanza di variabilità inter-osservatore dei risultati ottenuti attraverso tutti questi metodi hanno reso l'analisi quantitativa una candidata a sostituire la valutazione visiva e soggettiva del radiologo. Non solo: la capacità, dimostrata, di individuare anche i più fini cambiamenti in termini di densità e morfologia gioca un ruolo fondamentale per l'identificazione del trattamento terapeutico migliore e per la predizione del decorso clinico.

Ad esempio, la kurtosis può predire la mortalità[16] oppure, in combinazione col valor medio e la skewness, il tasso di sopravvivenza[17] nei pazienti non trapiantati, alla stregua degli score visivi assegnati da radiologi esperti. Inoltre, Jacob et al. hanno dimostrato come anche quantificare l'estensione fibrotica, o dei singoli pattern, alla baseline (ovvero al primo esame eseguito dopo la diagnosi) possa predire, più accuratamente dei test funzionali, il decorso della malattia[18].

Il futuro della QCT è garantito dal crescente sviluppo tecnologico che suggerisce come l'integrazione con i sistemi automatici per l'analisi quantitativa dei dati possa avere significato prognostico nella pratica clinica per la gestione personalizzata del paziente. Sebbene negli ultimi anni la QCT sia stata provata in vari contesti clinici con risultati interessanti, sono tuttora in corso studi prospettici per confermare che l'informazione sullo stato progressivo della fibrosi, estraibile dalla QCT, possa essere accettata come biomarker[19].

2.1.1 Analisi dell'istogramma dell'immagine

I pattern ILD hanno caratteristiche morfologiche molto eterogenee e l'istogramma dei livelli di grigio dell'immagine manca di un valore soglia di densità che discrimini il tessuto sano dal tessuto malato. Ciononostante, un valido aiuto nella valutazione dell'estensione della malattia proviene dall'osservazione combinata di features, quali il *valor medio*, la *skewness* e la *kurtosis*. La kurtosis descrive la forma della distribuzione dei livelli di grigio nell'istogramma ed è inversamente proporzionale allo spessore delle code: se il valore della kurtosis è maggiore di 0 allora la distribuzione si dice *leptocurtica* e lo spessore delle code è inferiore rispetto ad una distribuzione *platicurtica*, la cui kurtosis assume valori negativi. La skewness, invece, è un indice di asimmetria della distribuzione.

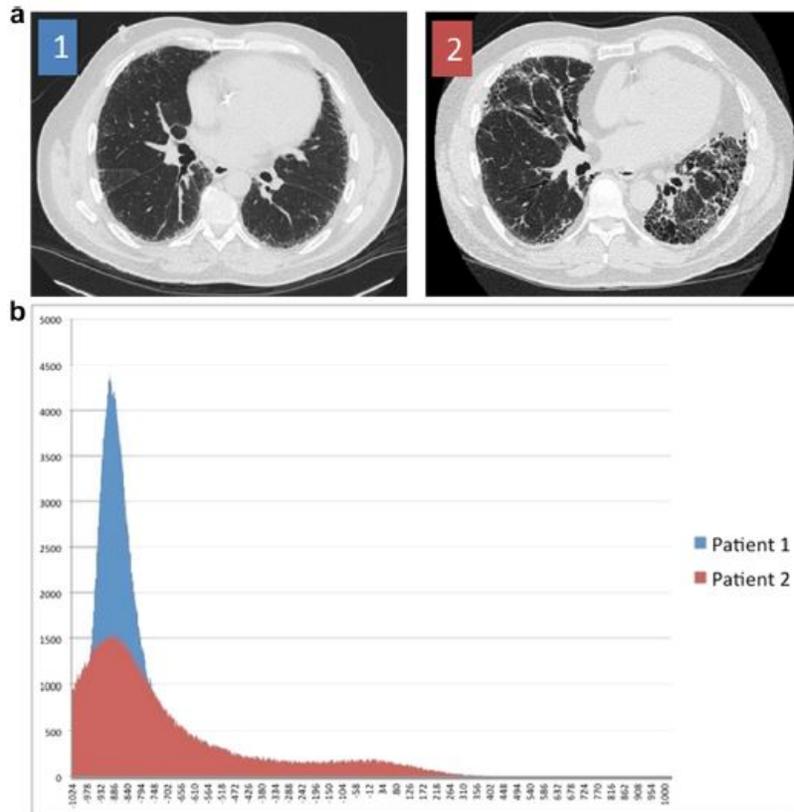


Fig 2.1[20]: istogrammi a confronto; l'immagine 1 è la CT di un paziente sano, l'immagine 2 di un paziente malato. Gli istogrammi relativi differiscono nel picco (valor medio), nella simmetria (skewness) e nell'elongazione della coda (kurtosis).

L'istogramma dell'immagine di un polmone sano ha un picco tipicamente intorno ai -800 HU ed è spostato verso sinistra. Un aumento di tessuto molle, dovuto alla fibrosi, come il collagene, si riflette sull'aumento della densità polmonare media (Mean Lung Density, MLD); diminuiscono di conseguenza la skewness e la kurtosis (Fig 2.1). In più, sia la kurtosis che la skewness si sono rivelate correlate ai test funzionali, come la misura della quantità di aria espirata (*Forced Vital Capacity*, FVC) durante un test di respirazione forzata[21] (*Forced Expiratory Volume*, FEV). Anche Best et al.[22] hanno dimostrato la correlazione tra valor medio, skewness e kurtosis e test funzionali in uno studio condotto su 144 pazienti. In particolare, la kurtosis è risultata ben correlata sia con l'FVC che con la capacità di diffusione alveolo-capillare (*Lung Diffusing capacity for carbon monoxide*, DLco). In un altro studio condotto su 167 pazienti, invece, è stata calcolata la differenza tra gli indici misurati alla baseline e quelli misurati dopo un anno: il risultato si è dimostrato predittivo della mortalità[23]. In aggiunta

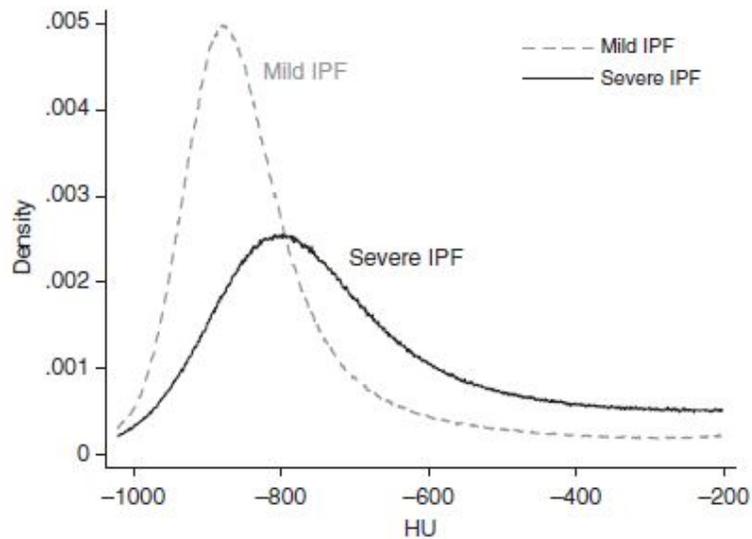


Fig 2.2[19]: istogrammi a confronto. Due pazienti con lieve e grave fibrosi: la kurtosis è, rispettivamente, pari a 5 e 0,41; la skewness, 1,48 e 0,80.

alle informazioni sull'estensione e il progresso della malattia, gli indici di skewness e kurtosis possono essere indicativi anche della stadiazione: generalmente, valori elevati riguardano una fibrosi lieve, valori più bassi una fibrosi moderata o anche grave[19] (Fig 2.2).

Oltre alle metriche appena descritte, anche il quarantesimo e l'ottantesimo percentile calcolati sull'istogramma correlano sia con i risultati dei test funzionali sia con lo score visivo attribuito dai radiologi[24]. Nello specifico, se calcolati alla baseline e ad un anno, le variazioni del quarantesimo percentile correlano con le variazioni di estensione della malattia, soprattutto nel caso dell'opacità a vetro smerigliato; variazioni calcolate sull'ottantesimo percentile, invece, sono indici di variazione delle opacità reticolari.

Riassumendo, il valor medio, la skewness e la kurtosis sono indicativi sia dell'estensione che della stadiazione della malattia, nonché predittivi della mortalità; differenze riscontrate sui valori degli indici equivalgono a cambiamenti delle funzionalità respiratorie, dovute al decorso della malattia. Sebbene da solo non basti a quantificare i singoli pattern UIP, l'istogramma dei livelli di grigio può fornire, in aggiunta, una prima indicazione circa l'evoluzione di opacità a vetro smerigliato e quadri reticolari, attraverso il quarantesimo e l'ottantesimo percentile.

2.1.2 Analisi della texture dell'immagine

Le metriche calcolate sull'istogramma dell'immagine, che includono skewness e kurtosis, sebbene siano facili da ottenere e correlate ai test funzionali quali FVC e DLco, rappresentano indici grossolani adatti ad una prima valutazione sull'estensione dei pattern ma non vantano una precisione sufficiente a descriverne la complessità dei contorni e della morfologia[25]. Le caratteristiche salienti delle *Usual Interstitial Pneumoniae* (UIP) nelle immagini CT, infatti, risiedono sia nei valori di intensità dei pixel che nell'arrangiamento spaziale e sono riassumibili nella *texture* dell'immagine.

Anthimopoulos et al. hanno definito la texture come una ripetizione stocastica di alcune strutture di dimensioni relativamente ridotte rispetto alla regione che le include[26]. L'analisi della texture, simulando la percezione visiva e i processi di apprendimento dell'uomo, valuta sia la densità che la morfologia del pattern in esame, determinandone - nel caso delle ILD - il tipo, la stadiazione e l'estensione, come farebbe l'occhio esperto del radiologo[27].

Nel 2006, Xu et al., integrando statistiche dell'istogramma e features frattali, hanno creato l'Adaptive Multiple Feature Method (AMFM) 3-D, tool con cui vengono discriminate con successo le alterazioni parenchimali come l'enfisema e l'ILD[28] e che si ispira alla versione precedente bidimensionale. L'integrazione dell'analisi di texture con l'analisi delle statistiche descrittive dell'istogramma, in uno studio condotto da Zavaletta et al.[29], ha permesso la collocazione dei pattern HRCT in una di cinque classi tra normale, reticolare, opacità a vetro smerigliato, honeycombing ed enfisema. Sulla base di tale metodo, è stato poi sviluppato Computer-Aided Lung Informatics for Pathology Evaluation and Ratings (CALIPER), software in grado di individuare e quantificare il volume occupato dai pattern in misura correlata ai parametri fisiologici e al tasso di sopravvivenza dei pazienti[30]; inoltre, il volume dei vasi e dell'honeycombing, misurati alla baseline dal CALIPER, si sono rivelati predittivi anche della mortalità[31].

Riassumendo, i metodi di classificazione che si basano sull'analisi della texture dell'immagine superano i limiti dei metodi basati sulla sola analisi dell'istogramma; riescono, infatti, non solo a valutare l'estensione della malattia ma a classificare il tipo di pattern. Tuttavia, richiedono un carico e un tempo computazionale significativamente maggiori; inoltre, essendo stati implementati con algoritmi diversi, resta ancora da chiarire se i risultati che si ottengono con

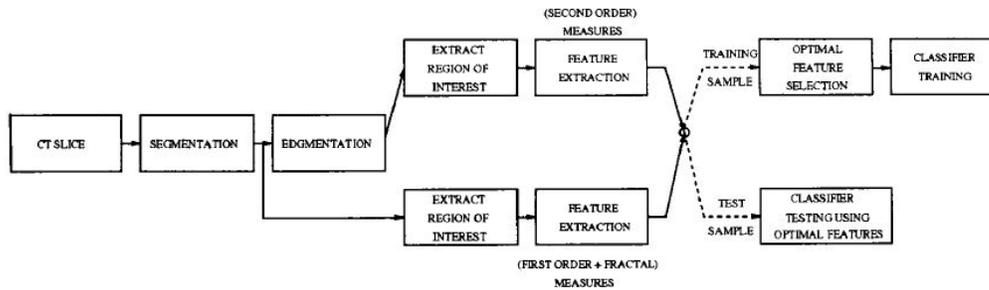


Fig 2.3[32]: diagramma di flusso del metodo AMFM 2-D.

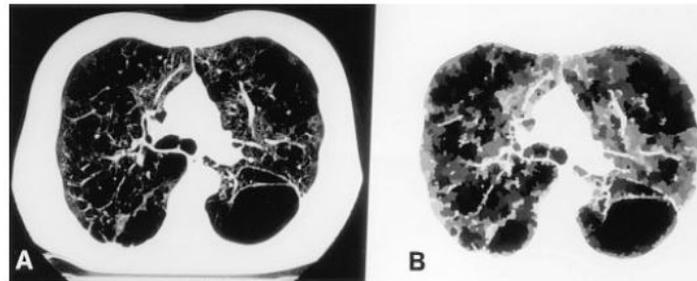


Fig 2.4[32]: immagine CT di un paziente con enfisema: originale (sinistra) e dopo il region growing (destra).

uno piuttosto che con un altro siano consistenti[13].

2.1.2.1 Adaptive Multiple Feature Method (AMFM)

Sull'*Adaptive Multiple Feature Method* (AMFM) si basa il funzionamento di diversi software sviluppati per l'analisi automatica della texture dell'immagine. La Fig 2.3 riporta il diagramma di flusso del metodo applicato al caso 2-D. Il passaggio preliminare è la segmentazione delle slices CT per la rimozione di strutture diverse dai polmoni. Sulle immagini ottenute, poi, viene eseguita un'ulteriore segmentazione con la tecnica del *region growing*: essa unisce pixels adiacenti in regioni in cui la differenza tra i livelli di grigio è trascurabile. Il livello di grigio assegnato ad ogni nuova regione è il valor medio dei livelli di grigio dei pixels inclusi (Fig 2.4). A questo punto, vengono definite delle regioni di interesse (*Region Of Interest*, ROI) sulle immagini ottenute con la prima segmentazione e delle ROI sulle immagini risultate dal *region growing*, entrambe classificate in base al consenso di radiologi esperti che forniscono la *ground truth*; di queste ROIs si calcolano, rispettivamente, le features del primo e del second'ordine. Le cinque features del primo ordine sono valor medio, varianza, skewness, kurtosis ed entro-

pia. Le features del secondo ordine, invece, sono undici: cinque estratte dalla matrice delle lunghezze delle *run*¹ dei livelli di grigio (*Grey-Level Run Length Matrix*, GLRLM) in cui viene riportata la frequenza con cui si osservano, nell'immagine, elementi consecutivi con una certa lunghezza e con una certa intensità, e sei estratte dalla matrice di co-occorrenza (*Grey Level Co-occurrence Matrix*, GLCM) che riporta le frequenze con cui ogni livello di grigio appare accanto ad altri ad una distanza ben definita. Tutte le features vengono, poi, normalizzate per le dimensioni dei pixels e del polmone. Le ROIs disponibili vengono divise in due insiemi: l'insieme di training e l'insieme di test. L'insieme di training fornisce gli esempi che il classificatore impara; in poche parole, durante la fase di training, vengono selezionate le features maggiormente descrittive della texture delle ROIs incluse nell'insieme di training e sulla base di queste vengono definiti i parametri del classificatore. Alle ROIs dell'insieme di test, invece, il classificatore associa una label, da confrontare con la ground truth, sulla base dei parametri definiti durante la fase di training.

Se da un lato, per la discriminazione tra tessuto sano ed enfisema, il metodo 2-D ha fornito risultati migliori rispetto a quelli basati sull'analisi dell'istogramma dei livelli di grigio, che misurano la densità media polmonare (*Mean Lung Density*, MLD) o il quinto percentile (HIST)[32], dall'altro ha presentato dei limiti nell'individuazione dell'honeycombing[33]. Dal momento che l'esame CT fornisce immagini volumetriche e molte delle caratteristiche determinanti dei pattern UIP sono di natura tridimensionale, le features 2-D non sono sufficienti e risultano meno sensitive. Alla luce di questo, Xu et al., in uno studio condotto su 20 pazienti sottoposti ad esame CT con scanner multi-detettore, hanno dimostrato come features volumetriche calcolate sull'istogramma, e di texture, riescano meglio a distinguere le alterazioni parenchimali dell'enfisema e dell'ILD[28]. Le features dell'istogramma sono la media, la varianza, la kurtosis e l'entropia; altre features descrittive della distribuzione dei livelli di grigio nel volume di interesse (*Volume Of Interest*, VOI) sono il numero totale di voxel sotto un soglia di -910 HU (lieve enfisema), il numero totale di voxel sopra una soglia di -864 HU (tessuto sano) o la moda (ovvero il valore che ricorre più frequentemente). Le features del second'ordine, invece, includono quelle calcolate a partire dalle matrici GLRLM e GLCM. Per ultime, sono state calcolate anche le features frattali, adatte alla natura eterogenea del tessuto del polmone[33]. In Fig 2.5

¹Si definisce *run* un gruppo di elementi consecutivi nell'immagine con stessa intensità dei livelli di grigio

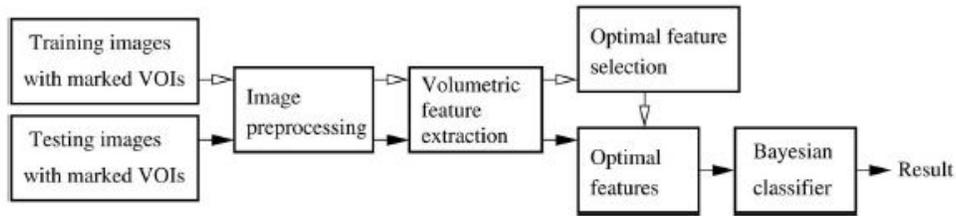


Fig 2.5[28]: diagramma di flusso del metodo AMFM 3-D.

è riportato il diagramma di flusso del metodo applicato al caso 3-D, usato da Xu et al.

Infine, in un altro studio, condotto da Salisbury et al.[34] per valutare l'efficacia di un trattamento terapeutico, l'AMFM è stato in grado non solo di quantificare oggettivamente il volume di polmone impegnato dalla patologia ma anche, in accordo con lo score visivo di tre esperti radiologi, di fornire valutazioni sul rischio di peggioramento dello stato di salute. Così, i risultati ottenuti con l'AMFM 3-D possono rappresentare dei biomarker indicativi sia della severità dell'IPF che del decorso terapeutico.

2.1.2.2 CALIPER

Computer-Aided Lung Informatics for Pathology Evaluation and Rating (CALIPER) è un software che analizza le immagini HRCT per la caratterizzazione del parenchima polmonare, utilizzando sia le features dell'istogramma 3-D che le caratteristiche morfologiche e di texture.

Innanzitutto, CALIPER isola i polmoni dalle strutture circostanti appartenenti al torace, seguendo il metodo di Hu et al.[35] che consiste in tre operazioni principali: una di sogliatura dei livelli di grigio, un'altra che separa il polmone sinistro dal destro e l'ultima di *smoothing* per riempire le vacanze lasciate dalle arterie e dalle vene nei campi polmonari. I polmoni, poi, vengono divisi nei rispettivi lobi (superiore, medio ed inferiore per il destro; superiore e inferiore per il sinistro). Anche le vie aeree vengono segmentate con un'operazione di sogliatura, a cui si aggiungono la tecnica del region growing 3-D e l'analisi delle componenti connesse, ovvero gruppi di pixel che condividono un vertice o uno spigolo. Per ultimi, vengono estratti i vasi sanguigni col metodo proposto da Shikata et al.[36] che applica un filtro multi-scala per la rilevazione di strutture tubulari. Dei dati così pre-processati, poi, CALIPER classifica regioni di interesse volumetriche (*Volume Of Interest, VOI*) di dimensioni 15x15x15. Il dataset di training dei classificatori di features consiste in una serie di VOI opportunamente segmentate da

radiologi esperti e comprovate da esami istopatologici. Dalle VOI vengono estratte features dell'istogramma, frattali e di texture, della matrice GLCM, GLRLM, che il classificatore imparerà ad analizzare su altri input al fine di predirne la label. Infatti, l'istogramma locale delle VOI, centrate sui voxel del parenchima di nuove immagini, viene confrontato con l'istogramma di 34 classi identificate in fase di training. In questo modo, CALIPER rileva pattern quali opacità a vetro smerigliato, reticolari e honeycombing nonché i vasi sanguigni come arterie e vene. In più, CALIPER quantifica tre gradi di bassa attenuazione nel parenchima, due dei quali associabili ad enfisema e il terzo al tessuto morfologicamente sano, mentre non è in grado di quantificare i consolidamenti e le bronchiectasie da trazione[37]. Il numero di voxel appartenenti ad ogni classe identificata viene contato in entrambi i polmoni e viene calcolato anche il numero totale di voxel appartenenti ai polmoni, a prescindere dalla classe. I voxel identificati come vasi vengono considerati come tessuto sano nel calcolo del volume polmonare totale[38]. L'obiettivo è di fornire non solo un valore numerico che rappresenti una misura, assoluta o percentuale, di volume di tessuto malato, ma anche una rappresentazione visiva efficace che riassume più variabili in un *glifo* (Fig 2.6). Il glifo fornisce una panoramica globale delle caratteristiche salienti del parenchima individuate dal CALIPER e del volume polmonare; è diviso da raggi che si dipartono a cerchio da un centro comune che rappresenta il centro dei polmoni e che dividono i polmoni in lobi. All'interno di ogni lobo, viene rappresentata, con colori diversi a seconda della classe, la distribuzione dei pattern con le rispettive percentuali di volume. L'area del glifo rappresenta il volume totale dei polmoni calcolato.

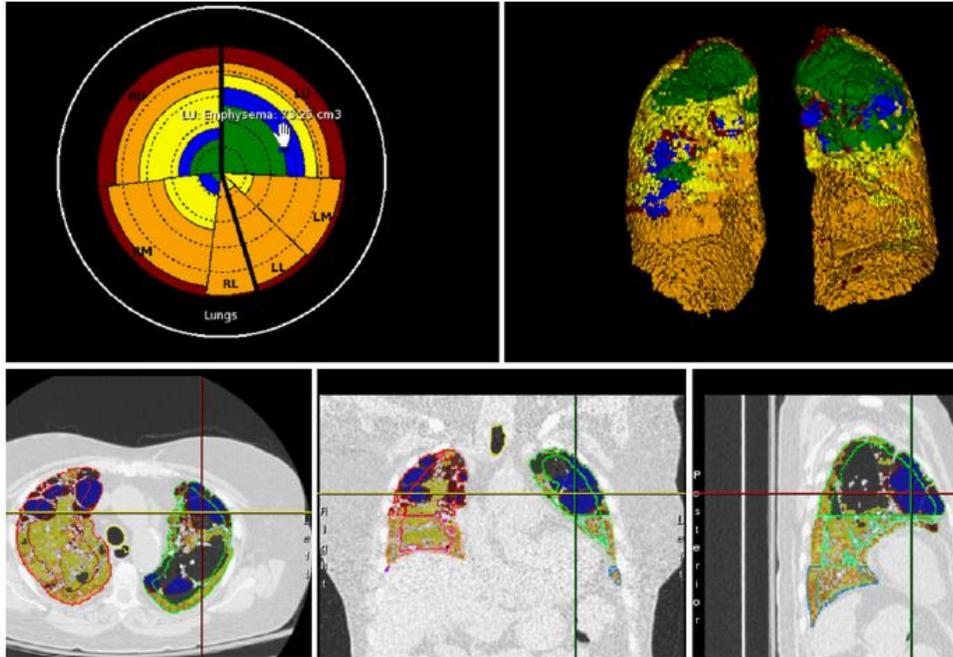


Fig 2.6[38]: in alto a sinistra, glifo con distribuzione di 5 pattern HRCT (rosso: honeycombing; arancione: quadro reticolare; giallo: opacità a vetro smerigliato; blu: enfisema; verde: tessuto sano). R/L = right/left lung; U/M/L = upper/middle/lower lung. Il cerchio bianco esterno rappresenta la capacità polmonare totale predetta del paziente. In questo esempio, cliccando sull'enfisema che si trova nella parte superiore sinistra del glifo, appare la relativa regione, localizzata sulle scansioni 2-D triplanari (fondo della figura). Vi è infine, un rendering 3-D dell'intero polmone coi pattern evidenziati a colori

CALIPER è stato utilizzato per analisi retrospettive su migliaia di dataset provenienti da diversi centri ed utilizzato per analisi prospettiche di circa 3000 HRCT del *Lung Tissue Research Consortium*: i risultati si sono mostrati fortemente correlati, più dello score visivo attribuito dai radiologi [39, 40], con i test funzionali, con il tasso di sopravvivenza dei pazienti e con il peggioramento delle funzionalità respiratorie[39, 41, 42].

2.2 IPF e Deep Learning

I metodi di analisi delle immagini più affidabili nell'ambito delle interstiziopatie polmonari, come quelli descritti finora, sono basati sul processo di *features engineering*, che consiste nella selezione di features significative, estratte da dati pre-esistenti e conosciuti, per definirne altre con un maggior potere predittivo. Nell'imaging diagnostico, lo svantaggio del features engi-

neering risiede nella necessità di disporre di label, o segmentazioni, fornite da medici esperti e nella conseguente variabilità inter-osservatore universalmente riconosciuta alle interpretazioni delle immagini, che costituisce un ostacolo alla robustezza e alla standardizzazione dei risultati[43, 44].

Gli strumenti sviluppati attraverso algoritmi di apprendimento automatico, invece, sono in grado di imparare, in maniera autonoma, le features rappresentative senza che ci sia il bisogno di esplicitarle: questo è il vantaggio principale del *deep learning*. Gli algoritmi di deep learning forniscono risultati riproducibili in poco tempo, sono low-cost, hanno un livello di accuratezza simile a quello dell'uomo e possono tornare utili in quei centri in cui il numero di clinici esperti nell'analisi delle immagini al torace è esiguo[45]. Recentemente, è cresciuto l'interesse da parte di studiosi nello sviluppo di algoritmi di deep learning che fornissero un supporto alla diagnosi, più che alla prognosi, **classificando** o **segmentando** i pattern nei campi polmonari.

Infatti, Walsh et al.[45], nel 2018, hanno usato una rete neurale pre-addestrata per classificare i pattern HRCT in *Usual Interstitial Pneumoniae* (UIP) o non-UIP, sulla base delle linee guida dell'American Thoracic Society e dell'European Respiratory Society. La rete è stata validata sul giudizio di clinici, non coinvolti in fase di sviluppo, e sia la sensibilità che la specificità sono risultati maggiori rispetto a quelle dei radiologi. In un altro studio, condotto da Anthimopoulos et al.[46], una nuova rete neurale convoluzionale (*Convolutional Neural Network*, CNN) è stata prima progettata e poi testata per la classificazione di *patches* di immagini 2-D in 7 categorie diverse (Fig 2.7): tessuto sano, opacità a vetro smerigliato (Ground Glass Opacity, GGO), quadro reticolare, una combinazione di GGO e quadro reticolare, honeycombing, micronoduli e consolidamenti. I risultati hanno mostrato un'accuratezza dell'85% dimostrando il potenziale delle CNN nell'analisi dei pattern HRCT. Anche Kim et al.[47] hanno progettato una rete neurale convoluzionale che, lavorando sulle immagini 2-D, classificasse i pixel come tessuto sano, GGO, honeycombing, quadro reticolare, enfisema o consolidamento. In più, i risultati sono stati messi a confronto con una *Support Vector Machine*² e hanno dimostrato come

²Per Support Vector Machine si intende un modello di apprendimento supervisionato che analizza i dati per associarli ad una di due classi, con una classificazione o una regressione; una SVM definisce un iperpiano, o un insieme di più iperpiani, che abbia la massima distanza dai punti più vicini di ogni classe nell'insieme di training.

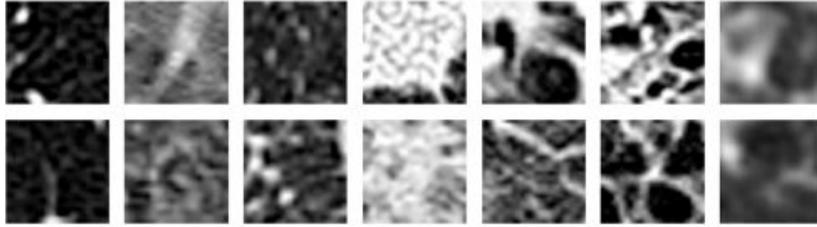


Fig 2.7[46]: 7 pattern classificati dalla CNN. Da sinistra a destra: tessuto sano, GGO, micronoduli, consolidamento, quadro reticolare, honeycombing, e GGO/quadro reticolare.

la performance della rete fosse migliore.

È solo negli ultimi due anni, invece, che sono state sviluppate reti neurali fatte di soli strati convolutivi (*fully convolutional*) per la segmentazione automatica delle immagini. Nel 2019, Anthimopoulos et al.[26] hanno progettato ad-hoc una rete che ha superato le prestazioni di metodi già performanti e precedentemente testati[46, 48, 49], risparmiando carico e tempo computazionale, sebbene siano stati riconosciuti alcuni errori di segmentazione. Ad esempio, il tessuto sano è stato spesso classificato o come quadro reticolare o come enfisema, in quest'ultimo caso a causa delle densità simile dei due pattern.

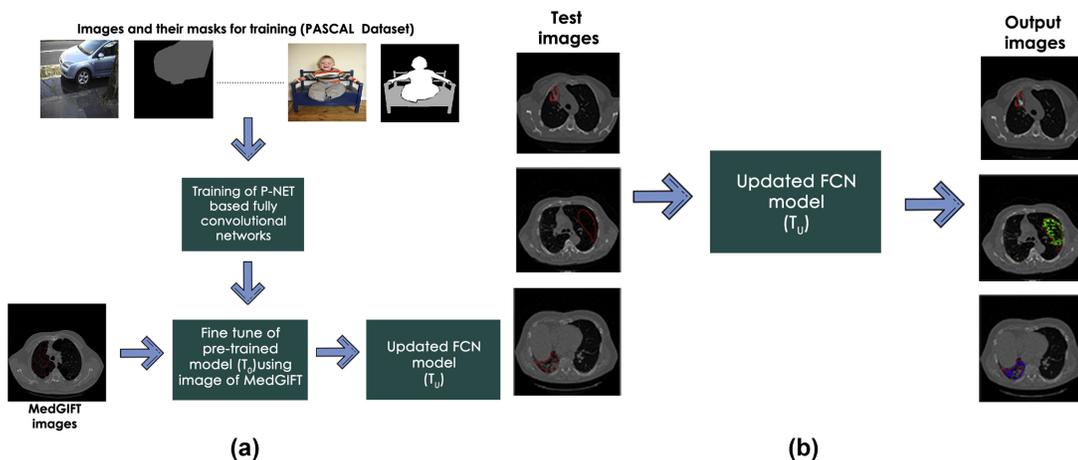


Fig 2.8[50]: workflow sviluppato da Agarwala et al. In rosso la *ground truth*, in celeste, verde e blu, rispettivamente, i consolidamenti, l'enfisema e la fibrosi individuati dalla rete.

Infine, nel 2020, Agarwala et al.[50] hanno pre-addestrato una rete P-net[51] con immagini fotografiche di natura quotidiana, i cui parametri sono stati poi modificati, una volta addestrata, per ottimizzare i risultati delle performance sulle HRCT del torace prese da un database con sole immagini di interstiziopatia. Il workflow è in Fig 2.8. I risultati sono soddisfacenti per quanto riguarda l'identificazione di consolidamenti, fibrosi ed enfisema anche a fronte dello scarso numero di patches etichettate a disposizione per il training.

Capitolo 3

Le reti neurali convoluzionali (CNN)

Nel terzo capitolo, dopo una breve introduzione, verranno descritti architettura e addestramento delle reti neurali convoluzionali (*Convolutional Neural Network*, CNN). Nella sezione dedicata all'architettura, si fornirà una spiegazione del funzionamento degli strati principali di cui è composta una rete; nella parte dedicata all'addestramento, si descriverà in cosa consiste, quali sono le strategie utilizzate e come la scelta degli iper-parametri possa influenzarne la resa. Infine, il focus si sposterà verso l'applicazione delle CNN all'imaging medico con un riguardo particolare verso le due reti che hanno ispirato il lavoro, la *U-net* e l'*Encoder-Decoder*. L'obiettivo è delineare un quadro che aiuti nella comprensione di ogni scelta progettuale fatta per la UIP-net, rete convoluzionale costruita ad-hoc per la segmentazione delle interstiziopatie polmonari durante il lavoro di tesi.

3.1 Introduzione alle CNN

Le **reti neurali convoluzionali** (*Convolutional Neural Network*, CNN o CNN) sono reti neurali artificiali esplicitamente progettate per l'analisi delle immagini. I neuroni che compongono una CNN hanno un'organizzazione che si ispira all'organizzazione della corteccia visiva animale, in cui ogni neurone risponde singolarmente a regioni del campo visivo. Hubel e Wiesel, nel 1968, individuavano due cellule principali nel sistema visivo animale[52]:

1. **Cellule semplici** che rilevano features in una regione locale del campo visivo;

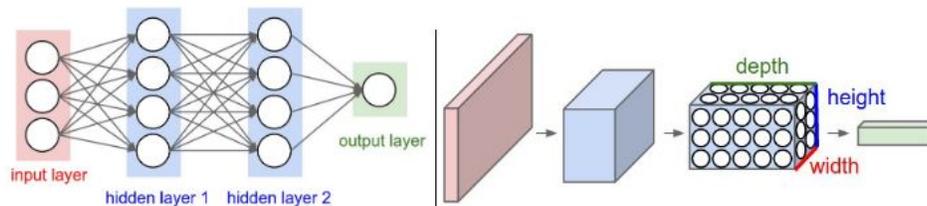


Fig 3.1[53]: a sinistra, una rete neurale feed-forward; a destra, una rete convoluzionale con neuroni definiti in tre dimensioni (larghezza, altezza e profondità).

2. **Cellule complesse** che ricevono i segnali delle cellule semplici e, combinandoli, li rielaborano.

Allo stesso modo del cervello umano, in una rete neurale convoluzionale, ogni neurone analizza una feature dell'immagine, così l'oggetto viene semplificato prima di essere riconosciuto su scala globale.

A differenza delle reti neurali *feed-forward*¹, che ricevono in input un vettore, i neuroni degli strati di una rete convoluzionale sono arrangiati in tre dimensioni: larghezza, altezza e profondità (Fig 3.1); la larghezza e l'altezza tengono conto dell'organizzazione spaziale retinotipica dell'immagine in input, la profondità, invece, tiene conto del numero di features rilevate e riassunte nelle *features maps*.

3.1.1 Architettura

L'architettura di una rete neurale convoluzionale consiste in una sequenza di strati (o *layers*), i cui tipi principali sono: convoluzionali, di attivazione, di *pooling* e *fully-connected*. A monte, c'è lo strato di input, con i valori dei pixel dell'immagine originale. Lo strato successivo esegue un'operazione di convoluzione tra i neuroni, che sono piccoli kernel (o filtri), e regioni circoscritte dell'immagine; in output fornisce una matrice di dimensione $n \times m \times p$, dove p è il numero di filtri applicati, quindi il numero di features viste dalla rete. Lo strato di attivazione applica una funzione ad ogni elemento dell'immagine e lascia inalterata la dimensione dell'input; diversamente accade per lo strato di pooling che, eseguendo un sotto-campionamento in 2-D (quindi sulla lunghezza e sulla larghezza), ne dimezza le dimensioni. Infine, lo strato

¹Per la descrizione della rete feed-forward si rimanda alla sezione 3.1.2, paragrafo 3.1.2.1

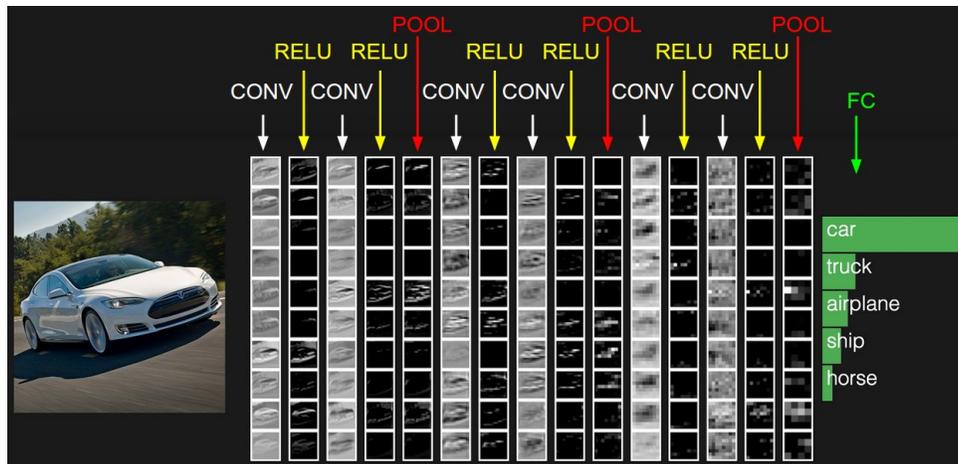


Fig 3.2[53]: esempio di CNN per la classificazione di un'immagine. Sulla sinistra lo strato di input conserva i pixel originali, sulla destra lo strato di output fornisce i punteggi di ogni classe; nel mezzo, gli strati tipici.

fully-connected è fatto da neuroni connessi a tutti i neuroni dello strato precedente e, di fatto, è un vettore la cui profondità è pari al numero di classi in cui si vuole categorizzare l'immagine: ogni elemento del vettore contiene un punteggio associato alla probabilità che l'immagine appartenga a quella classe (Fig 3.2).

Non tutti i layers sono definiti da parametri da ottimizzare durante il training: ad esempio, sia lo strato di attivazione che di pooling implementano una funzione ben precisa; lo strato convolutivo, invece, necessita di parametri per attribuire correttamente i pesi agli elementi dei filtri in base alle caratteristiche dell'immagine che devono evidenziare.

3.1.1.1 Strato convolutivo

I neuroni degli strati convolutivi sono dei kernel (o filtri) di lunghezza e larghezza ridotte rispetto alle dimensioni dell'immagine originale, ma di uguale profondità. Ogni neurone, piuttosto che a ogni pixel dell'immagine, è connesso a una regione che prende il nome di *campo recettivo*. In analogia coi sistemi visivi animali, il campo recettivo rappresenta la porzione di spazio nella quale lo stimolo visivo dev'essere localizzato affinché il singolo neurone possa rispondere. Ogni filtro viene, quindi, *convoluto* con il campo recettivo, ovvero vengono moltiplicati i pesi degli elementi che compongono il kernel con i pixel dell'immagine che ricopre, e poi sommati (Fig 3.3). Il filtro viene fatto scorrere in lunghezza e in larghezza e, una volta percorsa l'intera immagine, produce una matrice, l'*activation map*, che rappresenta la risposta

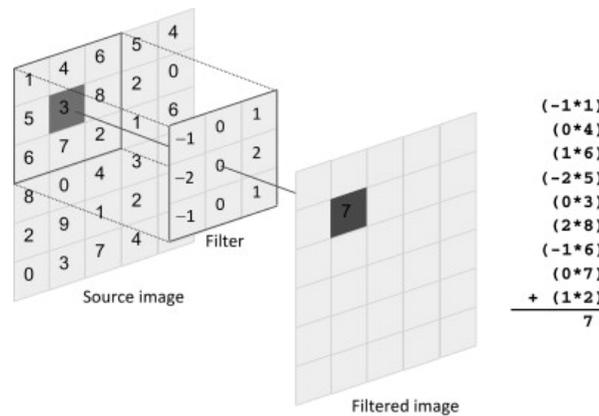


Fig 3.3[54]: operazione di convoluzione.

dell'input al filtro per ogni posizione. Le activation maps di ogni filtro vengono, poi, impilate a formare l'output che, quindi, consiste in uno stack di tante matrici quanti sono i filtri applicati. La dimensione dell'output dipende da tre **iper-parametri**: la *profondità*, lo *stride* e infine lo *zero padding*, che concorrono a definirne lunghezza e larghezza.

La **profondità** coincide con il numero di filtri utilizzati, uno per ogni feature dell'immagine (ad esempio, l'orientazione dei contorni o la texture).

Lo **stride** è il passo con il quale il filtro scorre sull'immagine: se pari a 1, allora il filtro viene applicato a campi recettivi sfalsati di un pixel. Aumentando lo stride, la dimensione dell'output diminuisce.

Lo **zero padding** rappresenta lo spessore con cui viene espansa l'immagine; si può aggiungere una cornice di 0 oppure degli stessi valori della cornice esterna dell'immagine. Lo zero padding permette di attribuire all'output le dimensioni desiderate.

L'ultimo iper-parametro è la **dilatazione** che rappresenta il passo con cui si esegue l'operazione di convoluzione. Ogni elemento del kernel può essere moltiplicato per pixels dell'immagine sfalsati di una quantità pari alla dilatazione; la dilatazione è utile quando si vogliono cogliere più informazioni spaziali a fronte di un numero ridotto di strati.

Il *tuning* degli iper-parametri è un passaggio cruciale per il buon funzionamento della rete e per la correttezza dei risultati. La scelta degli iper-parametri influenza, infatti, due fattori molto importanti:

1. La dimensione spaziale dell'output, che si può calcolare come funzione della dimensione

dell'input (W), del filtro (F), dello stride (S) e dello zero padding (P), secondo la (3.1).

$$OutputSize = \frac{(W - F + 2P)}{S} + 1 \quad (3.1)$$

2. Il controllo sui dati utilizzati dalla rete; se lo stride e lo zero padding sono tali che la dimensione dell'output, calcolata seguendo la (3.1), sia un numero non intero, il singolo neurone non analizza l'input in maniera ordinata e simmetrica. La rete potrebbe arbitrariamente forzare i parametri e il comportamento, a training concluso, potrebbe essere imprevedibile e inspiegabile.

Infine, l'aspetto che rende le reti neurali convolutive preferibili alle reti neurali feed-forward nell'analizzare le immagini è l'**invarianza spaziale**; ovvero, se è lecito fare l'assunzione per la quale una feature è significativa indipendentemente dalla posizione nel piano dell'immagine, allora i pesi attribuiti ai neuroni possono essere mantenuti costanti sul piano bidimensionale, quindi variare solo lungo la profondità. In questo modo, si riducono drasticamente i parametri da calcolare in fase di training e il carico computazionale si alleggerisce.

Non sempre ha senso fare quest'assunzione, però; infatti, un'immagine può presentare features completamente diverse in una regione dell'immagine piuttosto che in un'altra, come accade nei task di *face recognition*.

3.1.1.2 Strati di attivazione e di *pooling*

Nella corteccia cerebrale, la trasmissione dell'informazione fra un nodo di *Ranvier* di un neurone e il successivo avviene se la differenza di potenziale tra interno ed esterno della membrana mielinica supera un valore soglia e si innesca il potenziale di azione. Allo stesso modo, nelle reti neurali artificiali, l'informazione si propaga grazie ad una funzione di attivazione la cui scelta influenza la convergenza della rete. Infatti, i pesi dei neuroni, in fase di training, vengono aggiornati ad ogni iterazione per ridurre la differenza tra l'output predetto e l'output reale, seguendo la direzione della derivata della funzione di attivazione.

Nel caso specifico delle CNN, gli **strati di attivazione** operano sugli output degli strati convolutivi e le funzioni maggiormente utilizzate sono:

1. La funzione sigmoide: è stabile e, anche a fronte di brusche variazioni dell'input, differenziabile; comprime tutti i valori in un range tra 0 e 1 ed è definita dalla (3.2).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3.2)$$

Lo svantaggio principale è che è lenta a convergere perché per valori di x elevati la funzione satura, la sua derivata diminuisce notevolmente e i pesi vengono modificati di poco (ha luogo il *vanishing gradient*).

2. La funzione tangente iperbolica: simile alla funzione sigmoide ma estende il range di valori tra -1 e +1 (3.3).

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3.3)$$

3. La funzione *Rectifier Linear Unit* (ReLU): è una funzione molto semplice da calcolare che pone a zero tutti i valori negativi mentre lascia inalterati tutti i valori maggiori o uguali zero (3.4).

$$f(x) = \max(0, x) \quad (3.4)$$

La semplicità della funzione la rende particolarmente utile nei layers intermedi nei quali la quantità di passaggi richiede calcoli notevoli. In più, rispetto alla funzione sigmoide, non satura e riduce l'effetto del *vanishing gradient*.

Generalmente, negli strati intermedi si sceglie come funzione di attivazione la ReLU, per diverse ragioni:

- Riduce l'effetto *vanishing gradient* per cui aiuta la rete a convergere più velocemente; infatti, la sua dinamica non presenta regioni di saturazione e a variazioni degli input sono associate le giuste variazioni degli output;
- Incoraggia la sparsità della rete ponendo a zero metà dei neuroni (la funzione è nulla su metà del suo dominio). Questo se da un lato aiuta a ridurre l'overfitting, dall'altro velocizza il meccanismo di retro-propagazione dell'errore² in fase di training e quindi il raggiungimento della convergenza;

²Per una spiegazione si rimanda alla sezione 3.1.2, paragrafo 3.1.2.2

- È efficiente da calcolare dal punto di vista computazionale: la funzione da implementare richiede solo un controllo sul segno;
- È non lineare: questo rende la rete un approssimatore universale più preciso.

Gli strati di attivazione di output, invece, prediligono la funzione sigmoide, o tangente iperbolica. Come gli strati di attivazione, anche gli **strati di pooling** implementano funzioni fisse su tutti gli input indistintamente, per ogni iterazione; perciò i parametri che li definiscono non devono essere aggiornati in fase di training.

Gli strati di pooling operano su porzioni delle feature maps in uscita dallo strato convolutivo e di attivazione, come succedeva con la convoluzione sul campo recettivo; riducono la dimensione spaziale dell'immagine e il numero di parametri da calcolare, progressivamente, con la profondità della rete (Fig 3.4).

L'elemento costitutivo di ogni strato di pooling è un kernel di dimensioni F che scorre con uno stride S sull'immagine. Il volume di uscita ha profondità pari alla profondità delle feature maps e lunghezza e larghezza di dimensioni:

$$\frac{W - F}{S} + 1$$

dove W è la lunghezza o la larghezza della feature map.

Le funzioni di pooling più utilizzate sono il *max pooling*, che calcola il valore massimo tra gli elementi inclusi nella sotto-regione, e l'*average pooling* che ne calcola la radice quadrata della somma dei quadrati. Il max pooling causa la perdita di informazione maggiore perché più pixel vengono scartati; l'average pooling, invece, causa la perdita dell'informazione sui valori puntuali dei pixel perché tutti vengono condensati in un unico pixel. In entrambi i casi, comunque, sia le dimensioni del kernel che lo stride si scelgono inferiori a 3 per limitare le conseguenze della compressione. La perdita di informazione, infatti, ha portato gli strati di pooling a cadere in disuso nelle reti convoluzionali più moderne, come le *variational autoencoders* (VAEs) e le *generative adversarial networks* (GANs)[53].

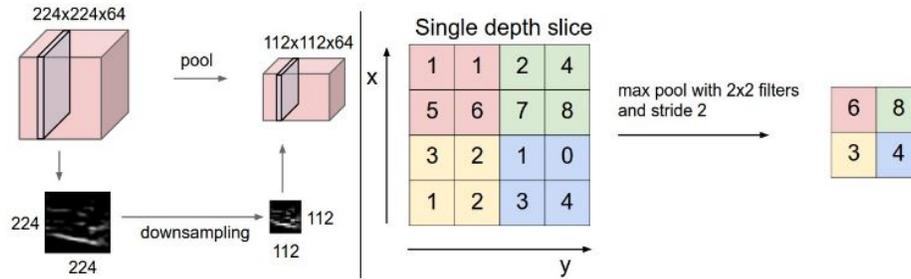


Fig 3.4[53]: operazione di pooling.

3.1.1.3 Strato *fully-connected*

Ogni neurone dello strato *fully-connected* (FC) è connesso a tutti i neuroni dello strato precedente; per questo, viene generalmente utilizzato come strato di output perché qui è necessario analizzare, senza distinzioni, tutta l'informazione contenuta nelle feature maps. Tuttavia, a fronte del numero elevato di connessioni, aumenta il numero di pesi da apprendere e, di conseguenza, il rischio che la rete interpoli i dati di ingresso, senza generalizzare (si ha il cosiddetto *overfitting*). Si rimedia applicando, come metodo di regolarizzazione, la tecnica del *drop-out* che consiste nella disattivazione casuale di un certo numero di neuroni ad ogni epoca, così da impedire alla rete di adattarsi ai dati di training. In uscita, lo strato *fully-connected* fornisce un vettore con profondità pari al numero delle possibili classi di appartenenza dell'immagine, e i valori contenuti sono punteggi associati alla probabilità che i dati vi appartengano.

3.1.2 Addestramento

L'addestramento di una rete neurale artificiale consiste in un problema di ottimizzazione globale di una funzione che aggiorna i pesi dei neuroni nei vari strati; si minimizza una funzione che, tipicamente, è:

$$\min_{x \in R^m} E(w) = \sum_{p=1}^P E_p(w)$$

in cui w sono i pesi, x gli input, P il numero totale di campioni e E_p è l'errore relativo al p -mo campione, calcolato come la distanza tra l'uscita desiderata e l'uscita fornita dalla rete. È possibile utilizzare diverse metriche di errore; la più utilizzata è l'errore quadratico medio[55] che viene calcolato su un sottoinsieme dei dati disponibili, il **training set**.

Dal momento che lo scopo dell'addestramento è modellare il processo di generazione dei dati, più che farne un'interpolazione, la scelta dell'architettura, dei dati, la definizione della metrica di errore e la strategia di addestramento devono assicurare buone capacità di generalizzazione[56, 57]. Per questo, è opportuno che la complessità del modello, in termini di numero di parametri liberi, sia commisurata alla numerosità del dataset. Solitamente, si ricorre a tre strategie di addestramento:

1. La *stabilizzazione strutturale*: consiste nello scegliere l'architettura addestrando una serie di reti a cui si aggiungono, o tolgono, strati, quindi numero di neuroni[58]. Per ciascuna di queste reti, vengono determinati i pesi minimizzando l'errore sul training set; le prestazioni vengono, poi, confrontate attraverso una *cross-validation* valutando, cioè, l'errore che ogni rete commette su un altro insieme di dati, il **validation set**, non inclusi nel training set. Alla fine, si sceglie la rete che fornisce l'errore minimo sul validation set e, una volta addestrata, la rete viene testata su un terzo insieme di dati, il **test set**, non utilizzato né per la scelta dell'architettura né per la determinazione dei parametri.
2. La *regolarizzazione*: consiste nell'aggiungere alla funzione di errore un termine di penalità sul singolo peso, così da restringere l'insieme entro il quale i parametri vengono scelti. La nuova funzione obiettivo diventa:

$$E(w) = \sum_{p=1}^P E_p(w) + \gamma \|w\|$$

3. L'*early-stopping*: consiste nell'interruzione prematura del processo di minimizzazione della funzione d'errore. Si basa sull'idea di valutare periodicamente l'errore che la rete commette sul validation set, perché può aumentare se l'errore sul training set diventa troppo piccolo. In questo caso, potrebbe iniziare la fase di overfitting in cui la rete interpola i dati di training a discapito della capacità di generalizzazione.

Qualunque sia la strategia di addestramento scelta, talvolta è necessario ripetere il processo di minimizzazione dell'errore con varie architetture o funzioni di errore; per questo, disporre di funzioni di ottimizzazione efficienti è fondamentale per costruire una rete neurale. Le difficoltà computazionali, infatti, potrebbero risiedere in:

- Forti non linearità della funzione di errore;
- Elevata dimensionalità del vettore dei pesi ed elevato numero di campioni nel training set;
- Presenza di minimi locali e non globali;
- Impossibilità di convergenza globale a partire da punti iniziali arbitrari, a cui si rimedia in parte introducendo il termine di regolarizzazione.

L'addestramento di una rete neurale si svolge tipicamente in due fasi: una di *forward propagation* e una di *backward propagation*.

3.1.2.1 *Forward-propagation*

Una rete neurale artificiale organizzata a strati, in cui ogni neurone riceve input solo dai neuroni del livello precedente e propaga l'output solo ai neuroni degli strati successivi, si dice rete *feed-forward*. In un rete feed-forward, non vi sono auto-collegamenti o connessioni con neuroni di uno stesso strato e la direzione del flusso di informazione va dall'ingresso all'uscita; si dice che la rete implementa una *forward propagation*. La forward-propagation riduce il rischio di generare loop che ostacolano la rete nel produrre l'output.

L'elaborazione dell'informazione di ogni neurone degli strati compresi tra l'ingresso e l'uscita, chiamati strati nascosti (*hidden layers*), avviene in due fasi:

1. Una fase di **pre-attivazione**, in cui si esegue la somma pesata degli input;
2. Una fase di **attivazione**, nella quale la somma pesata viene passata ad una funzione di attivazione del neurone per trasmettere l'informazione allo strato successivo.

3.1.2.2 *Aggiornamento dei pesi e backward-propagation*

L'aggiornamento dei pesi, alla k -ma iterazione, avviene a partire dal calcolo del gradiente della funzione di errore (E), detta anche funzione costo o di perdita. Ad esempio, il metodo iterativo più diffuso per l'ottimizzazione delle funzioni differenziabili è il metodo di *discesa stocastica*

del gradiente che aggiorna i pesi seguendo la (3.5).

$$w^{(k+1)} = w^k - \eta \nabla E(w^k) \quad (3.5)$$

Lo scalare η rappresenta la velocità di apprendimento, il *learning rate*, la cui scelta può influenzare il training della rete; infatti, valori elevati spingono il gradiente in un minimo locale piuttosto che globale, valori bassi, invece, rallentano il raggiungimento della convergenza. Oltre al metodo di discesa del gradiente, esistono molte altre funzioni di ottimizzazione dei pesi della rete, come l'*Adaptive Gradient* (AdaGrad) oppure la *Root Mean Square Propagation* (RMSProp).

Comunque, tutti i metodi iterativi per l'aggiornamento dei pesi rientrano in due classi:

1. Metodi *batch* in cui ad ogni passo i pesi vengono aggiornati utilizzando informazioni relative o a tutti i campioni dell'insieme di training o a sotto-insiemi di esso.
2. Metodi *on-line* in cui ad ogni passo i pesi vengono aggiornati tenendo conto soltanto di un singolo campione.

Al crescere del numero di pesi da aggiornare, il calcolo della derivata parziale della funzione gradiente, sebbene in sé non richieda un onere computazionale eccessivo, potrebbe rappresentare il collo di bottiglia del processo di ottimizzazione globale. Per questo motivo, l'algoritmo di *backpropagation* calcola solo l'errore E allo stadio uscita della rete e lo retro-propaga ai precedenti, fino a quello di input. Ad ogni strato, poi, solo i pesi dei neuroni attivati, quindi quelli non abbattuti dallo strato di attivazione precedente, vengono aggiornati dall'errore propagato; è evidente il risparmio computazionale che ne consegue.

3.1.3 CNN per l'imaging medico

Le reti neurali convoluzionali vengono utilizzate tipicamente per classificare le immagini in input con una *label* fornita in output. Soprattutto nell'imaging medico, però, molti tasks visivi richiedono un grado di dettaglio maggiore per localizzare oggetti diversi all'interno della stessa immagine, che è anche l'obiettivo principale della **segmentazione**. Anche per la segmentazione delle immagini si può far ricorso alle reti convoluzionali: si tratta, infatti, di un processo

che consiste principalmente nell'associare ad ogni pixel dell'immagine la classe dell'oggetto di appartenenza.

Nell'eseguire la segmentazione, la CNN non analizza l'immagine per intera ma la esamina sotto-regione per sotto-regione. Di ogni sotto-regione, che corrisponde al campo recettivo dello strato convolutivo, poi, la rete impara le caratteristiche rilevanti; non solo: impara anche ad associare, alle caratteristiche, la classe di appartenenza del pixel, o dell'intorno del pixel. La potenzialità delle reti convoluzionali in ambito medico è stata dimostrata in varie applicazioni: dalla segmentazione degli organi addominali[59] a quella del fegato policistico[60], dalla classificazione di tumori, alla mammella o alla prostata, in benigni o maligni[61, 62], al neuro-imaging per la rilevazione di aneurismi, melanomi e iper-intensità della materia bianca[63, 64, 65, 66]. Di conseguenza, in letteratura, molte sono le architetture proposte, ma due tra queste hanno particolarmente ispirato la progettazione della rete convoluzionale per la segmentazione delle interstiziopatie polmonari del lavoro di tesi: la **U-net** e l'**auto-encoder**.

3.1.3.1 U-net

Il limite principale delle reti neurali artificiali applicate all'imaging medico è la quantità di dati di cui necessitano per essere addestrate. Infatti, le reti *deep*, composte, cioè, da molti strati, hanno molti parametri da stimare in fase di training e aumentando il numero di parametri, aumenta anche il numero di dati necessari con cui modellarli. A fronte dell'impossibilità di disporre di un gran numero di immagini, Ciresan et al.[67] hanno progettato una rete a cui viene fornito in input solo una *patch* dell'immagine intera. Se da un lato, questo amplia l'insieme di dati disponibili per il training che, seppur molto lentamente, viene portato a termine con successo, dall'altra aumenta la ridondanza delle informazioni a causa dell'inevitabile *overlapping* delle patches.

Per far fronte al problema, allora, Ronneberg et al. hanno proposto la *U-net*, una rete la cui architettura (cosiddetta *fully-convolutional*) necessita di un numero ridotto di immagini per il training pur garantendo risultati precisi[68].

Alla porzione di rete fatta da strati convolutivi che riducono la dimensione dell'input, si concatena una seconda sezione che, sovra-campionando i dati, ne aumenta la dimensione in larghezza e lunghezza, riducendone la profondità. Il percorso convolutivo in cui si sotto-campiona e

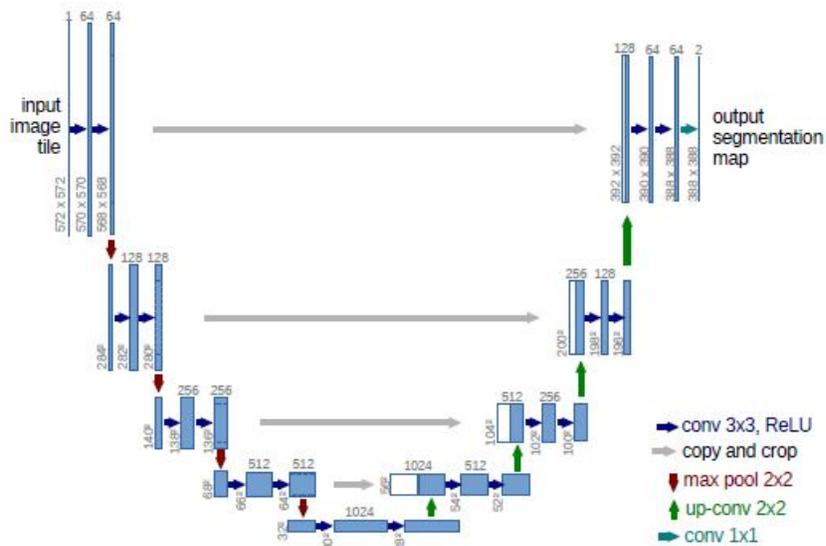


Fig 3.5[68]: Architettura della U-net.

quello concatenato, in cui si sovra-campiona, sono simmetrici per non perdere informazioni. Non si perdono, cioè, le features analizzate dalla rete negli step convolutivi. Così, la rete intera assume una forma a U, da cui prende il nome.

L'architettura è in Fig 3.5 e consiste in una parte convolutiva (a sinistra) che riduce la dimensione dell'ingresso ed una parte espansiva che l'aumenta (a destra). La parte convolutiva è quella tipica di una rete convoluzionale: consiste nell'applicazione ripetuta di un'operazione di convoluzione con kernel di dimensioni 3x3, seguita dall'applicazione di una funzione di attivazione ReLU ed un max pooling con kernel 2x2 e stride 2 per il sotto-campionamento. Ad ogni sotto-campionamento, il numero di features maps viene raddoppiato. La parte che sovra-campiona riporta progressivamente le dimensioni delle immagini a quelle di partenza, dimezzando ad ogni passo il numero di features maps; per rimediare, l'altra metà delle features maps viene copiata dal corrispettivo strato della parte convolutiva. All'interno dello stesso stadio, viene eseguita la solita convoluzione con kernel di dimensioni 3x3 e la funzione di attivazione ReLU. Nello strato finale, invece, la convoluzione viene eseguita con filtro 1x1.

La rete è stata utilizzata in due applicazioni: per la segmentazione delle strutture neuronali su immagini microscopiche e per la segmentazione delle cellule; in entrambi i casi ha raggiunto buoni risultati anche a fronte dei pochi dati forniti alla rete.

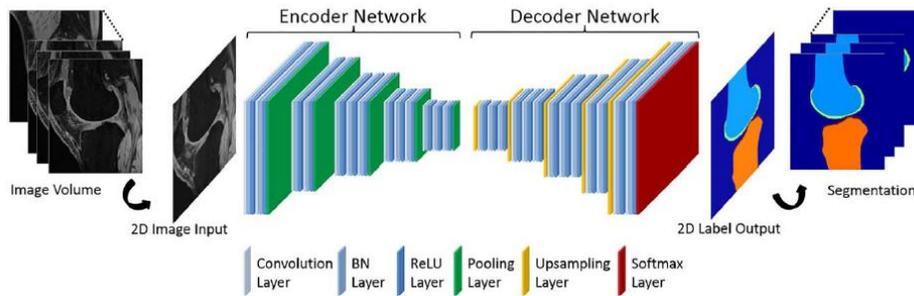


Fig 3.6[69]: Architettura Encoder-Decoder.

3.1.3.2 Encoder-Decoder

Una progettazione più accurata e attenta è richiesta se la rete neurale convoluzionale deve segmentare immagini mediche con un'alta risoluzione e con dettagli fini da preservare. Liu et al.[69] hanno proposto una rete che:

- Classificasse ogni pixel dell'immagine ad alta risoluzione;
- Usasse uno schema *up-convolutivo* efficiente dal punto di vista dei calcoli e della memoria computazionale richiesta, così da essere rapida nell'analisi dell'immagine;
- Supportasse il calcolo parallelo dei framework più accessibili per migliorare le performances e allo stesso tempo fosse facile da implementare.

L'architettura scelta è un'architettura *Encoder-Decoder* (Fig 3.6) che è composta, appunto, da una rete *encoder* seguita da una rete *decoder*. L'encoder è composta da strati che applicano una convoluzione in 2-D, una *batch-normalization*³, una funzione di attivazione ReLU e un max-pooling. Ogni gruppo encoder viene ripetuto cinque volte in modo da comprimere bene l'informazione ma, allo stesso tempo, rilevare quelle features dell'immagine che sono robuste e invarianti nello spazio. All'encoder segue subito il decoder, senza strati fully-connected nel mezzo. La rete decoder ha lo stesso numero di strati dell'encoder ma ne esegue le operazioni inverse quindi sovra-campiona, applicando una convoluzione, una batch-normalization e

³La batch-normalization consiste nella normalizzazione delle features maps di ogni strato e ad ogni iterazione; i vantaggi sono principalmente due: si velocizza la convergenza e l'apprendimento di ogni strato è indipendente dagli strati che lo precedono

una funzione di attivazione ReLU. Il processo di sovra-campionamento riesce a preservare in maniera efficiente sia i contorni dell'immagine che le features con un basso consumo di memoria in fase di test. Infatti, l'operazione di sovra-campionamento viene eseguita solo sui valori mantenuti dopo il max-pooling, di cui si conservano gli indici; mantenere in memoria gli indici, piuttosto che le intere features maps, alleggerisce il carico. Lo strato finale del decoder, poi, è un classificatore che produce, per ogni pixel, una mappa di probabilità che ha la stessa risoluzione dell'immagine di input.

Capitolo 4

Metodi

Nel quarto capitolo, dopo averne chiarito l'obiettivo, verranno illustrati i metodi coi quali è stato condotto il lavoro. Verrà prima fornita una descrizione della UIP-net, rete neurale convoluzionale progettata per la segmentazione delle interstiziopatie polmonari; poi, si parlerà del dataset, descrivendo il pre-processing delle immagini di cui si compone, propedeutico all'addestramento e al test della rete.

4.1 Obiettivi del lavoro

Nel capitolo 2, sono stati messi in luce i limiti dei metodi attuali per l'analisi quantitativa delle immagini tomografiche ad alta risoluzione che impediscono ai radiologi di servirsene nella pratica clinica.

1. La **riproducibilità**: la correttezza dei risultati non è garantita al variare della strumentazione e dal protocollo di acquisizione. Le case produttrici di scanner tomografici incorporano tecnologie differenti per l'acquisizione delle immagini e sono i tecnici, o i fisici medici, che valutano quando, e in che modo, eseguire l'esame radiologico sul paziente. Perciò, manca una standardizzazione che garantisca la consistenza dei risultati al variare dei contesti.
2. La **variabilità degli algoritmi di ricostruzione dell'immagine**: ogni sistema ricostruisce le immagini con algoritmi a risoluzione spaziale diversa. Algoritmi a bassa ri-

soluzione generano immagini più *smooth* di quelle generate con algoritmi a risoluzione maggiore; un metodo di analisi automatica dell'immagine, a maggior ragione se sviluppato con approccio *data-driven* e ottenuto dopo un processo di *features engineering*, non garantirà risultati robusti perché i valori delle features sono influenzati dal filtro spaziale.

3. La **variabilità inter- e intra-osservatore**, che rappresenta un bias intrinseco ai classificatori addestrati su dati diversi classificati da radiologi diversi.
4. L'**accessibilità** dei software già disponibili. Sia CALIPER, sviluppato nel Biomedical Imaging Resource Laboratory della Mayo Clinic, che AMFM, di cui si è parlato nel capitolo 2, sono tools sicuramente validati su esami funzionali e prognostici, ma non sono risorse alle quali è facile accedere, perché sono destinate ai soli scopi di ricerca, e non sono codici *open-source*. Inoltre, sono in vendita sul mercato a prezzi elevati.

Il bisogno di uno strumento di apprendimento automatico indipendente dallo scanner di acquisizione, dall'algoritmo di ricostruzione e dal giudizio degli operatori clinici e che sia anche open-source ha spostato l'interesse verso il *deep learning*. Richiedendo un minimo pre-processing sui dati, il deep learning riesce a identificare autonomamente le features rappresentative dell'immagine[70, 71]. Il rapido sviluppo di CPUs e GPUs in grado di sopportare carichi computazionali notevoli, poi, ha portato il deep learning ad un successo senza precedenti; così, la capacità delle reti neurali profonde di rilevare features gerarchiche, ovvero features con grado di dettaglio maggiore a partire da features meno dettagliate, ha incitato l'impiego delle reti neurali convoluzionali (*Convolutional Neural Network*, CNN) nella segmentazione automatica delle immagini mediche. La segmentazione, infatti, è il punto di partenza indispensabile per un'analisi quantitativa a supporto della diagnosi, ma richiede tempo, energia ed esperienza da parte del radiologo e automatizzarla può rappresentare un valido aiuto.

Per far fronte a tutte le problematiche, nella fase sperimentale del lavoro di tesi è stata progettata la **UIP-net** (*Usual Interstitial Pneumoniae*), una rete neurale convoluzionale per la segmentazione automatica del tessuto malato nei polmoni di pazienti affetti da Fibrosi Idiopatica Polmonare (IPF).

L'architettura della rete neurale è stata progettata nel rispetto di diverse esigenze, tra le quali:

1. Preservare i dettagli dell'immagine, perché il contenuto informativo maggiore è nella texture.
2. Assicurare un training efficace a fronte della numerosità limitata dei dati; è un problema comune del deep learning, infatti, non disporre di un dataset adeguatamente ampio all'addestramento di reti neurali con molti strati.
3. Ottimizzare tempo e costo computazionali, per rendere la rete uno strumento pratico e accessibile.

4.2 UIP-net: rete convoluzionale per la segmentazione delle interstiziopatie

Nel capitolo 1, la descrizione dell'apparato respiratorio e dei polmoni ha messo in evidenza la complessità delle strutture anatomiche e le problematiche dovute alla sovrapposizione di più tessuti, come quello del parenchima e dei vasi sanguigni o delle vie aeree; queste caratteristiche possono fuorviare l'occhio del medico nell'analisi visiva dei pattern nelle immagini radiologiche, così come possono confondere la rete convoluzionale che le segmenta. È per questo che, come è stato anche spiegato nel paragrafo 1.4, la metodica di riferimento per l'imaging diagnostico delle interstiziopatie polmonari è la tomografia computerizzata ad alta risoluzione (*High Resolution Computed Tomography*, HRCT). L'HRCT, infatti, fornisce immagini volumetriche e con un elevato grado di dettaglio, che aiuta a rimediare ai problemi dovuti alla complessità anatomica dell'apparato.

Nel capitolo 2, invece, gli studi riportati suggeriscono che, nel caso specifico delle interstiziopatie polmonari, le informazioni salienti, per l'identificazione corretta dei pattern UIP¹ nel polmone, sono nella texture.

In definitiva, lo studio preliminare è stato condotto per progettare una rete convoluzionale che fornisca risultati accurati e affidabili, tenendo conto delle caratteristiche specifiche del problema e ha fatto luce su varie esigenze, tra le quali:

¹Si veda il Paragrafo 1.3.1

1. Preservare l'elevata risoluzione delle immagini HRCT. L'elevata risoluzione di cui godono le immagini, infatti, non è solo quella spaziale ma anche quella associata alla profondità del pixel, ovvero al numero di bit con cui viene codificato il livello di grigio, che è più elevato rispetto alle comuni metodiche di imaging;
2. Limitare il più possibile la perdita di informazione: anche la più piccola variazione d'intensità tra un pixel e quello adiacente può essere un fattore discriminante. È questa, infatti, la texture;
3. Progettare una rete addestrabile:
 - Con un numero limitato di dati, a fronte dell'impossibilità di disporre di un dataset troppo ampio;
 - Con immagini 2-D e in scala di grigio: così che, prendendo le singole slices, si potesse aumentare il numero di dati del training set;
 - Senza disporre di segmentazioni manuali;
 - Con risorse computazionali limitate, anche a fronte dell'elevato contenuto informativo da processare, per fornire risultati rapidi ma precisi allo stesso tempo.

Alla luce di tutto questo, la scelta dell'architettura da cui partire è ricaduta sulla rete Encoder-Decoder, descritta nel Paragrafo 3.1.3.2. e utilizzata da Liu et al., per due motivi:

1. Perché preserva i dettagli più fini dell'immagine. Applicandola alle immagini morfologiche di ginocchio, la rete di Liu et al. segmentava correttamente la cartilagine;
2. Per risparmiare tempo e memoria. Rispetto alla U-net di Ronnenberg et al., l'Encoder-Decoder recupera le informazioni più risolte, compresse dall'encoder, conservando solo gli indici dei neuroni attivati dagli strati convolutivi precedenti, e trasmessi da quelli di pooling, non le intere feature maps.

Sono due le differenze sostanziali rispetto al problema risolto da Liu et al., però:

1. Le immagini da segmentare erano immagini di risonanza magnetica che opera sulla base di principi fisici completamente diversi da quelli della CT e, di conseguenza, i valori dei livelli di grigio dei pixel sono diversi;

2. L'obiettivo di Liu et al. era discriminare il tessuto cartilagineo (che è un tessuto molle) dall'osso. Queste sono strutture la cui differenza sostanziale sta nelle intensità dei livelli di grigio dei pixels, non nella texture, come nel caso delle interstiziopatie polmonari.

Fatte queste due precisazioni, all'architettura Encoder-Decoder sono state apportate delle modifiche:

1. Sono stati eliminati gli strati di batch-normalization e di pooling per minimizzare il più possibile la perdita di informazione. Dal momento che ogni pixel è significativo ed è bene preservare anche la più piccola differenza d'intensità tra un pixel e il suo vicino, anche cambiare l'operazione di max pooling in average pooling non funziona: essa combina più pixels in uno solo;
2. Gli strati dell'Encoder e del Decoder, sono 3 e non 5;
3. È stata modificata la funzione di attivazione dell'ultimo strato: la softmax è stata sostituita dalla tangente iperbolica. In questo modo, il range di valori in uscita è aumentato, da $[0; 1]$ a $[-1; 1]$, e, dal momento che si tratta di una funzione simmetrica rispetto all'origine, la convergenza si è velocizzata. Inoltre, data la sua dinamica, la funzione tangente iperbolica è stabile per brusche variazioni degli input; questo rappresenta un vantaggio quando si retro-propaga l'errore.

4.2.1 Architettura

L'architettura della UIP-net è in Fig 4.1. È composta da tre strati encoder e tre strati decoder convolutivi; ad ogni strato convolutivo, segue uno strato di attivazione con funzione ReLU, ad eccezione dell'ultimo strato in cui si applica la funzione tangente iperbolica. Ad ogni step convolutivo nella fase di codifica, il numero di features maps raddoppia; per cui, il primo fornirà in uscita uno stack di 32 features maps, il secondo 64 ed il terzo 128. Dei tre strati del decoder, invece, solo i primi due seguono al contrario l'encoder e producono in uscita, rispettivamente, 64 e 32 features maps; l'ultimo strato mantiene lo stesso numero di features maps del precedente ma aumentano larghezza e lunghezza delle immagini. L'ultimo strato convolutivo assembla tutte le features maps in un'immagine che ha le stesse dimensioni dell'immagine in input. In

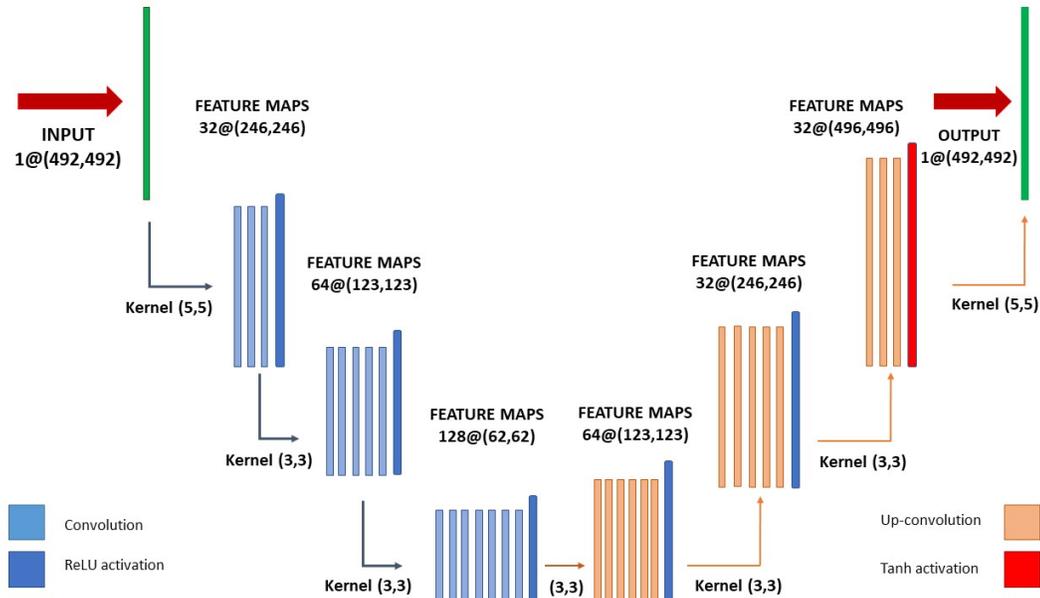


Fig 4.1: Architettura della UIP-net.

Fig 4.2, a titolo d'esempio, sono riportati i 32 kernel dello strato convolutivo.

Gli strati *up-convolutivi* eseguono un sovra-campionamento per generare delle features maps di dimensioni maggiori rispetto a quelle fornite in input. Come gli strati convolutivi, anche questi sono definiti da parametri quali lo stride (s) e il padding (p); l'operazione che eseguono può essere riassunta in quattro passi (Fig 4.3):

1. Calcolo di nuovi parametri z e p' a partire da s e p ;
2. Inserire z numeri di zeri tra ogni riga e colonna: questo aumenta la dimensione dell'input, i , portandola a $(2 \times i - 1)(2 \times i - 1)$;
3. Fare il padding sulla matrice ottenuta a seconda del valore di p ;
4. Eseguire una convoluzione sull'immagine generata al passo precedente con uno stride pari a 1.

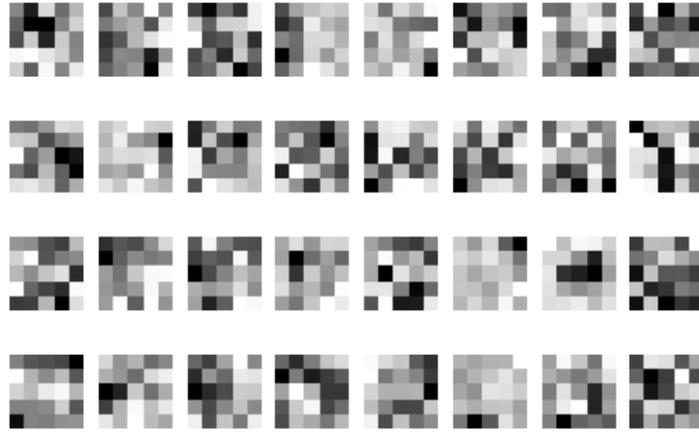


Fig 4.2: 32 kernel del primo strato convolutivo.

In Tabella 2, vi è una stima della memoria richiesta da ogni strato, sia per il mantenimento delle features maps che dei parametri, solo in fase di forward-propagation.

1. Per le features maps si richiedono 58 MB per ogni immagine;
2. Per i parametri si richiedono 782 KB per ogni immagine.

Il numero totale di parametri è 195.489, tutti addestrabili.

Layer	Dimensione	Memoria	Pesi
Input	$492 \times 492 \times 1$	242K	0
Conv1	$246 \times 246 \times 32$	1.9M	832
Conv2	$123 \times 123 \times 64$	968K	18496
Conv3	$62 \times 62 \times 128$	492K	73856
Upconv1	$124 \times 124 \times 64$	984K	73792
Upconv2	$248 \times 248 \times 32$	2M	18464
Upconv3	$496 \times 496 \times 32$	7.9M	9248
Conv4	$492 \times 492 \times 1$	242K	801

Tabella 2: riassunto degli strati della UIP-net.

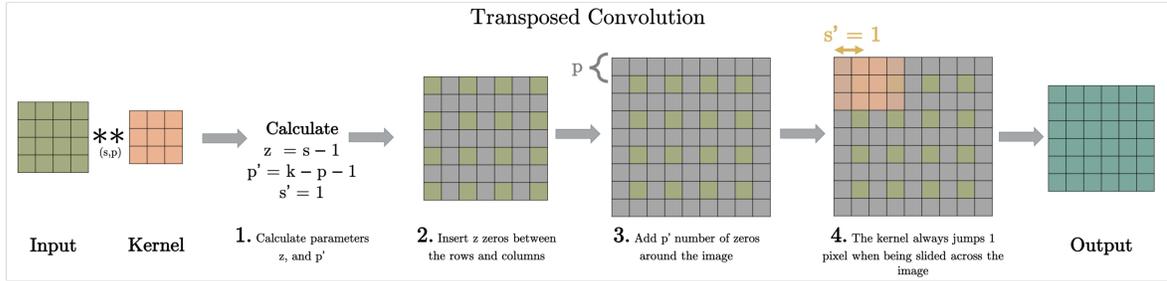


Fig 4.3: *up-convoluzione*

4.2.1.1 Scelta degli iper-parametri

Tipicamente, sono tre le regole da seguire per la scelta degli iper-parametri:

1. Lo strato di input, che contiene l'immagine, deve avere dimensioni pari;
2. Gli strati convolutivi devono operare con kernel piccoli (3x3 o 5x5) e stride pari a 1; l'uso di kernel di dimensioni maggiori (ad esempio 7x7) è consigliato solo nel primo strato, che vede l'immagine in ingresso;
3. Il padding deve mantenere inalterata la dimensione dell'input.

Contestualizzate alla CNN progettata, sono state applicate come di seguito:

1. L'input ha dimensione 492x492 così da rendere la UIP-net simmetrica e avere in output un'immagine con le stesse dimensioni dell'input;
2. Per ridurre il rischio che la memoria trabocchi, solo nel primo e nell'ultimo strato sono stati utilizzati kernel più grandi, cioè di dimensioni 5x5. Negli strati interni, invece, sono stati utilizzati kernel 3x3 così da diminuire il numero di parametri da calcolare, a fronte dell'aumento del numero di features maps, ma allo stesso tempo cogliere tutti i dettagli utili;
3. Non essendoci strati di pooling, che tipicamente sono i principali responsabili del sottocampionamento delle immagini, lo stride è stato scelto pari a 2. Tenuto conto del valore dello stride, il valore del padding (p) di ogni strato è stato scelto così che la dimensione dell'immagine dimezzasse; p , quindi, è pari a $\frac{k-1}{2}$, con k che rappresenta la dimensione del kernel. Questo è il valore che solitamente si attribuisce all'iper-parametro quando si

desidera che l'output abbia la stessa dimensione dell'input, se lo stride è pari a 1. Visto che lo stride è stato fissato a 2, ad ogni strato la dimensione dell'immagine si dimezza.

4.2.2 Addestramento

Nel paragrafo 3.1.2, è stato spiegato come viene svolto l'addestramento di una rete convoluzionale, a partire dalla scelta della **funzione costo**, che misura l'errore che la rete compie nella predizione del risultato ad ogni passo, e della **funzione di ottimizzazione**, con la quale aggiornare i pesi di ogni strato. In più, a seconda dell'applicazione, si scelgono delle **metriche di accuratezza**, ovvero dei punteggi attribuiti ai risultati, da valutare durante la fase di training e di validation.

Per l'addestramento della UIP-net, presentata al paragrafo precedente, sono state scelte:

1. L'*entropia incrociata binaria* come funzione costo;
2. L'*Adaptive Moment Estimation* (ADAM) come funzione di ottimizzazione;
3. Gli indici di *Dice* e *Jaccard* come metriche di accuratezza, per controllare l'andamento del training.

4.2.2.1 Funzione costo e di ottimizzazione

La **funzione di costo** scelta è la funzione di entropia incrociata binaria (*binary cross-entropy*). L'*entropia*, in statistica, è una funzione che misura la casualità di una variabile aleatoria e quanto più una variabile è casuale, tanto minore sarà il suo contenuto informativo e tanto maggiore sarà l'entropia. L'*entropia incrociata* quantifica la differenza tra distribuzioni di probabilità diverse sullo stesso insieme di dati, solitamente tra la distribuzione dei dati veri e la distribuzione dei dati predetti. Di conseguenza, quanto minore sarà l'entropia incrociata tanto più le due distribuzioni saranno simili. La formula matematica dell'entropia incrociata è:

$$H_p(q) = - \sum_{c=1}^C q(y_c) \cdot \log(p(y_c))$$

Solitamente, si utilizza l'*entropia incrociata binaria* per problemi di classificazione binaria, in cui il classificatore deve associare ad ogni dato la probabilità che appartenga ad una di due classi. Nel caso specifico della rete progettata ad-hoc per segmentare le interstiziopatie polmonari,

ogni pixel dell'immagine può appartenere al tessuto sano o al tessuto malato e la funzione di costo valuta la bontà della predizione conoscendo a priori la classe di ogni pixel.

La formula matematica dell'entropia incrociata binaria è:

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))$$

dove N è il numero totale di pixel nell'immagine, y è la label da predire e $p(y)$ è la probabilità predetta dalla rete che il pixel abbia valore 1, ovvero che sia appartenente ad un tessuto malato. Quindi, alla funzione di costo si aggiunge un termine pari al prodotto tra il valore vero della label del pixel e la probabilità che il pixel abbia quella label secondo la rete; in questo modo, si riescono a penalizzare le predizioni sbagliate; infatti, la funzione entropia incrociata binaria:

1. Aumenta se il pixel è di un tessuto malato ma la predizione della rete lo associa al tessuto sano;
2. Diminuisce se il pixel è di un tessuto malato ed effettivamente la rete fornisce un valore di probabilità che sia malato più elevato rispetto al valore di probabilità che sia sano.

Per ogni pixel e ad ogni passo, la rete calcola la cross-entropia tra la distribuzione vera dei dati e quella predetta; di tutti i valori di cross-entropia così ottenuti, poi, ne fa una media che rappresenta il valore della funzione di costo definitivo per quel passo di training.

La **funzione di ottimizzazione** scelta è l'*Adaptive Moment Estimation* (ADAM). L'ADAM è un algoritmo di discesa stocastica del gradiente che combina i vantaggi dell'*Adaptive Gradient* (AdaGrad) e del *Root Mean Square Propagation* (RMSProp). L'AdaGrad è un metodo che adatta i valori del learning rate alla distribuzione dei parametri: per distribuzioni sparse, il learning rate sarà maggiore; per distribuzioni dense, il learning rate sarà più basso. Per questo motivo, l'AdaGrad converge più rapidamente, soprattutto quando la distribuzione dei dati è prevalentemente sparsa. L'RMSProp, come l'AdaGrad, è un metodo con learning rate adattivo, ma, a differenza dell'AdaGrad, modifica il valore del learning rate sulla base della media mobile, pesata per un fattore di decadimento esponenziale, dei valori assunti dal gradiente in una finestra di iterazioni. In questo modo, viene data più importanza ai valori dei gradienti delle iterazioni più recenti e per valori elevati, il learning rate viene cambiato di poco. Questo è un vantaggio se i dati sono rumorosi. Superando l'AdaGrad e il RMSProp, l'ADAM valuta sia la media che

la varianza dei valori assunti dal gradiente in una finestra di iterazioni, pesati per due fattori di decadimento. L'aggiornamento dei pesi avviene in misura proporzionale al rapporto tra la media e la varianza dei valori assunti dal gradiente nella finestra di iterazioni nella quale vengono calcolati; in più rispetto all'RMSProp, l'ADAM corregge i valori di media (momento del primo ordine) e varianza (momento del secondo ordine) per i due fattori di decadimento esponenziale, perché rappresentano un bias che allontana il valor vero della media e della varianza dal valore atteso. Riassumendo, l'aggiornamento dei pesi, ad ogni iterazione, richiede i seguenti passaggi[72]:

1. Calcolare il gradiente coi parametri correnti;
2. Aggiornare la media mobile esponenziale della media e della varianza del gradiente alle iterazioni passate;
3. Correggere per il bias introdotto dal fattore di decadimento esponenziale;
4. Aggiornare i pesi in misura proporzionale al rapporto tra la media e la varianza, corretti e scalati per il learning rate.

La velocità di convergenza maggiore, dovuta alla correzione dei bias dei decadimenti esponenziali, unita al fatto che ADAM funziona bene sia con dati sparsi che non stazionari, quindi rumorosi, sono stati i motivi principali per cui è stata scelta come funzione di ottimizzazione per la fase di addestramento della rete. Gli iper-parametri della funzione ADAM, con i valori numerici scelti, sono:

- Learning rate: 0,001;
- Decadimento esponenziale β_1 della finestra per il calcolo del momento del prim'ordine: 0,9;
- Decadimento esponenziale β_2 della finestra per il calcolo del momento del second'ordine: 0,999;

4.2.2.2 Metriche di accuratezza

Per controllare l'andamento della performance della rete in fase di addestramento, sono state utilizzate due metriche: l'indice di *Dice* e l'indice di *Jaccard*. Si tratta di indici che confrontano il risultato di un algoritmo automatico per la segmentazione di immagini e un'immagine che fa da *ground truth* ovvero l'immagine vera, che la rete dev'essere in grado di predire quanto più accuratamente possibile a training concluso. Date S , la segmentazione predetta dall'algoritmo e G , la ground truth, l'indice di Jaccard si definisce come:

$$J = \frac{|S \cap G|}{|S \cup G|} \quad (4.1)$$

dove $|S \cap G|$ è uguale al numero di pixel in comune tra le due maschere, detta anche *overlapping area* e $|S \cup G|$ è il numero di pixel ottenuto dall'unione delle due maschere. I pixel che si trovano nell'*overlapping area* si dicono **veri positivi** (TP). I veri positivi vengono sottratti dal numero di pixel dell'area dell'unione al denominatore della (4.1) per evitare che vengano contati due volte. Oltre i veri positivi, ci sono i **falsi positivi** (FP) che sono quei pixel erroneamente segmentati, e i **falsi negativi**, ovvero quei pixel mancati che dovevano essere segmentati. In termini di veri positivi, falsi positivi e falsi negativi, l'indice di Jaccard può essere riscritto come:

$$J = \frac{TP}{(TP + FP + FN)}$$

L'indice di Dice, invece, è definito come:

$$D = \frac{2|S \cap G|}{|S| + |G|}$$

quindi è uguale al rapporto tra il numero di pixel in comune tra le due maschere e la media della numerosità delle due maschere. Anche l'indice di Dice può essere espresso attraverso TP, FP e FN:

$$D = \frac{2 \times TP}{(TP + FP) + (TP + FN)}$$

Rispetto all'indice di Jaccard, l'indice di Dice attribuisce un valore doppio ai veri positivi; per questo è una misura di similarità meno severa. Sia l'indice di Dice che l'indice di Jaccard saranno uguali a 1 se le due maschere sono perfettamente coincidenti, a 0 se sono disgiunte (4.4). L'indice di Jaccard può essere definito in funzione dell'indice di Dice e viceversa:

$$J = \frac{D}{2 - D}, D = \frac{2J}{1 + J} \quad (4.2)$$

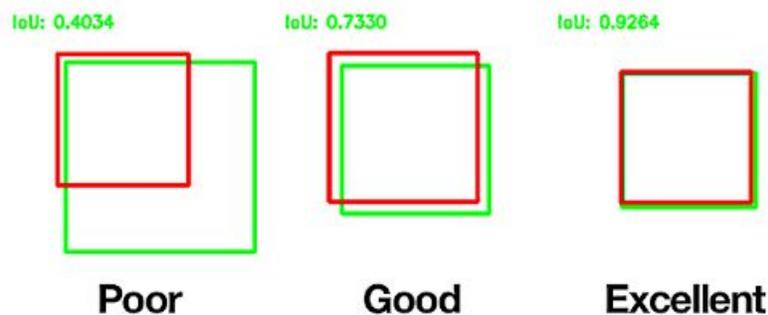


Fig 4.4: esempi di valori dell'indice di Jaccard per diversi matching delle maschere.

Se l'addestramento procede correttamente, gli indici di Jaccard e di Dice, a partire da un valore molto basso crescono progressivamente ad ogni iterazione, fino ad arrivare ad un valore massimo. Raggiunto il valore massimo, gli indici non aumentano né diminuiscono: a questo punto è bene stoppare il training per evitare l'overfitting. Quindi, durante il training, guardando sia valori della funzione di costo (che deve decrescere) che gli indici di accuratezza (che devono aumentare), si valuta se l'addestramento sta procedendo correttamente: se la funzione di costo mantiene valori elevati e le metriche valori bassi allora la rete non sta imparando; se la funzione di costo decresce progressivamente mentre le metriche crescono progressivamente, allora la rete sta imparando; se, infine, sia la funzione di costo che le metriche dopo aver raggiunto il plateau, rispettivamente, aumentano e diminuiscono se valutati sull'insieme di validation, allora la rete si sta adattando ai dati in ingresso e perde la capacità di generalizzare.

4.3 Dataset

Le immagini tomografiche ad alta risoluzione che compongono il dataset utilizzato per l'addestramento e il test della rete convoluzionale descritta nella sezione 4.2, sono immagini di torace di pazienti affetti da Fibrosi Idiopatica Polmonare contenute nel database dell'Unità Operativa di Radiodiagnostica 2 dell'Azienda Ospedaliera Universitaria Pisana (AOUP) dell'Ospedale di Cisanello di Pisa. Sono immagini DICOM in scala di grigio e di dimensioni 512x512, tutte ottenute dallo stesso scanner (SIEMENS Sensation 64), con la tecnica assiale e con lo stesso kernel di ricostruzione (B60). La profondità dei pixel è pari a 12, quindi ogni pixel viene codificato con 12 bit. Le scansioni hanno tutte lo stesso *pixel spacing* (ovvero le dimensioni lungo x e

y), di 0,7 mm, ma non lo stesso *slice thickness* (lo spessore di fetta): alcune possono averlo di 1 mm, altre di 1,5 mm. Per ogni scansione, ci sono sempre almeno due finestre di valori di intensità dei pixel per la visualizzazione delle immagini; i valori dei centri delle finestre non sono costanti ma si aggirano tipicamente a -500 Hounsfield Unit (HU) per la visualizzazione dei polmoni e dei tessuti meno densi, e intorno a 50 HU per vedere il grasso addominale e i tessuti più densi.

In Tabella 3, sono riassunte le caratteristiche che accomunano tutte le immagini:

Parametro	Valore
Formato	DICOM
Scanner di acquisizione	SIEMENS Sensation 64
Kernel di ricostruzione	B60
N. colonne	512 px
N. righe	512 px
Profondità del pixel	12 bit
Pixel spacing	0,7 mm

Tabella 3: caratteristiche comuni a tutte le immagini raccolte.

4.3.1 Ottenimento della ground truth

Insieme all'immagine originale in scala di grigio, sia per il training che per il test, è stata fornita in input alla rete convoluzionale anche un'immagine binaria, con le label assegnate ad ogni pixel; l'immagine binaria rappresenta la segmentazione del tessuto malato all'interno del polmone che, in fase di training, serve alla rete per imparare le caratteristiche discriminanti e in fase di test, invece, dev'essere predetta quanto più correttamente possibile. La maschera binaria è, quindi, una matrice che ha pixel uguali a 1 se appartengono al tessuto malato, uguali a 0 se appartengono al tessuto sano.

Generalmente, la maschera binaria rappresenta il *gold standard*, o *ground truth*, ovvero il risultato corretto da raggiungere. Ci sono diverse opzioni per definire la ground truth:

1. Può essere una maschera disegnata a mano da un operatore esperto, tipicamente un radiologo;
2. Può essere ottenuta con una modalità di imaging invasiva;
3. Può essere ottenuta con una modalità di imaging non invasiva che, allo stato dell'arte, fornisce la migliore qualità diagnostica.

In questo caso, l'opzione definitiva è stata una variante dell'ultima: la ground truth è stata ottenuta dal software Computer-Aided Lung Informatics for Pathology Evaluation and Ratings (CALIPER), ampiamente discusso nel capitolo 2, dal momento che è uno dei pochi validati con successo sui test di funzionalità respiratoria. La scelta di utilizzare CALIPER per estrarre le maschere dei pattern fibrotici nei polmoni è stata costretta anche dall'impossibilità di disporre di segmentazioni manuali fatte dai radiologi.

CALIPER, di ogni paziente, elabora l'intera scansione volumetrica in formato DICOM. Dopo un processing delle slices, più o meno veloce, l'interfaccia si presenta come in Fig 4.5: vi è il glifo, un rendering volumetrico dei polmoni e le visioni triplanari (assiale, coronale e sagittale) delle immagini segmentate che sono esportabili in diversi formati, tra cui il DICOM, che è quello scelto. CALIPER permette solo la visualizzazione dei risultati; l'esportazione viene demandata ad un altro software, Analyze 12.0 (di Analyze Direct), che offre diversi tools per visualizzare, segmentare, registrare oltre che per importare ed esportare immagini. In uscita da Analyze 12.0, le immagini non sono più in scala di grigio, ma a tre canali RGB, quindi a colori (Fig 4.6): verde per il tessuto sano, rosso, arancione e giallo per il tessuto fibrotico (per l'honeycombing, il quadro reticolare e l'opacità a vetro smerigliato, rispettivamente) e blu per l'enfisema. Per gli scopi di questo lavoro, non è stata fatta distinzione tra i vari pattern UIP ma solo tra tessuto sano e tessuto malato; quindi, le zone interessate delle immagini fornite da CALIPER sono l'insieme di zone rosse, arancioni e gialle; è stato ignorato il blu, quindi l'enfisema. Una volta ottenute, le immagini RGB di ogni paziente sono state importate in ambiente Matlab®(R2019b). Qui sono stati valutati i valori dei canali RGB di ogni pixel dell'immagine e una maschera, inizializzata con tutti 0, è stata riempita seguendo i seguenti passaggi:

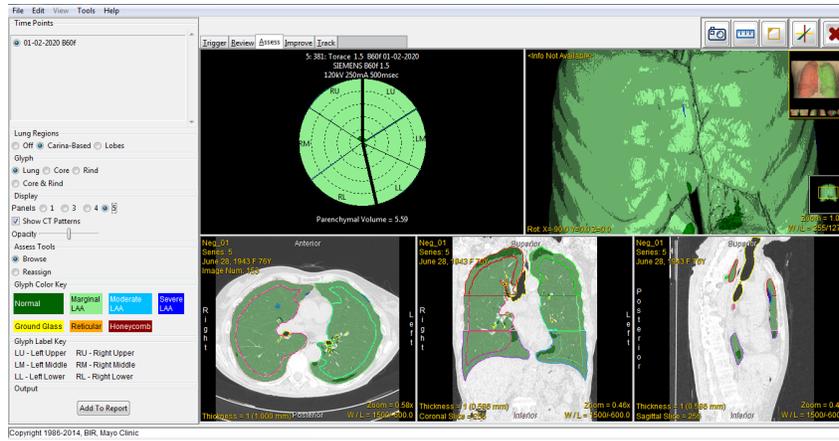


Fig 4.5: interfaccia CALIPER.

- Se il canale rosso del pixel della ground truth era a 1, allora anche il corrispettivo pixel della maschera è stato messo a 1: così sono state ottenute le maschere dell'honeycombing;
- Se il pixel del canale rosso e verde avevano gli stessi valori, ed erano uguali a 1 allora il corrispettivo pixel della maschera è stato messo a 1 perché il pixel della ground truth era giallo o arancione, e rappresentava opacità a vetro smerigliato o quadro reticolare.

Alla fine, per ogni slice, è stata ottenuta una maschera binaria con le stesse dimensioni della ground truth; una volta ottenute le maschere binarie di tutte le slices, sono state impilate a formare un volume di segmentazioni, salvato in formato .mat. Il rendering 3-D del volume finale è in Fig 4.7 ed è stato mostrato a video per controllare la correttezza dei passaggi, a meno dell'interpolazione dei voxel.

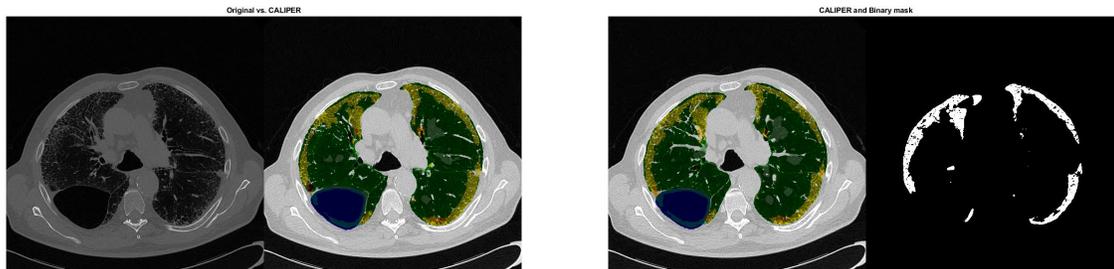


Fig 4.6: a sinistra, confronto fra l'immagine DICOM originale e l'immagine segmentata da CALIPER con le zone gialle, rosse, arancioni, verdi e blu in evidenza; a destra, confronto tra l'immagine segmentata da CALIPER e la maschera binaria ottenuta: i pixel bianchi corrispondono all'insieme di pixel gialli, rossi e arancioni.

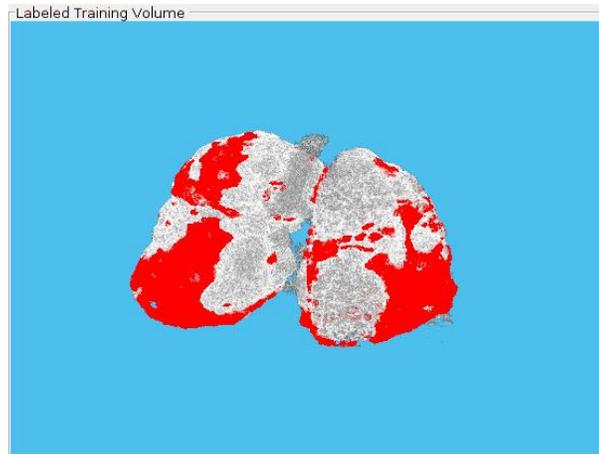


Fig 4.7: rendering volumetrico delle segmentazioni binarie.

4.3.2 Pre-processing delle immagini

Per abbattere il tempo computazionale necessario per il training della rete convoluzionale, è stato eseguito un pre-processing delle immagini. L'obiettivo principale era ridurre non solo le dimensioni, quindi restringere il campo visivo (*Field of View*, FOV), ma anche il numero di pixel da analizzare. In ogni immagine HRCT, infatti, appaiono il fondo, il lettino della TAC e il grasso addominale che non sono utili ai fini dell'apprendimento e che, per questo, potrebbero rallentare la convergenza. Allo stesso tempo, il pre-processing doveva lasciare inalterate le intensità dei pixel appartenenti ai campi polmonari. È stato già discusso nei capitoli precedenti, infatti, come le caratteristiche utili per la discriminazione del tessuto malato da quello sano siano nella texture e, quindi, nella differenza puntuale di intensità tra i livelli di grigio di pixel vicini. Per preservarla, non è stato applicato alcun filtraggio.

In definitiva, per ogni slice, il pre-processing ha:

1. Applicato un algoritmo automatico per la discriminazione sostanziale tra due classi: campi polmonari e fondo, e regione addominale;
2. Isolato i polmoni da tutto il resto.

Per il passaggio descritto al punto 1., propedeutico al passaggio del punto 2., è stato utilizzato un tipo di algoritmo per la segmentazione a soglia automatica delle immagini: *fuzzy C-mean*, o algoritmo di **clustering non esclusivo**.

4.3.2.1 Clustering non esclusivo: algoritmo fuzzy C-means

Nella segmentazione a soglia, ogni pixel viene associato a un tessuto sulla base dell'intensità del segnale; per questo, la segmentazione a soglia consiste nello scegliere una o più soglie nell'istogramma dell'immagine e a classificare i pixel sulla base di queste. In Fig 4.8, è riportato l'istogramma di una delle immagini HRCT a disposizione; sono visibili due picchi: uno predominante, relativo al fondo e ai polmoni, che hanno un segnale molto basso, il cui valore si aggira attorno allo zero e l'altro relativo a strutture a intensità maggiore, il cui valore si aggira attorno a 1000.

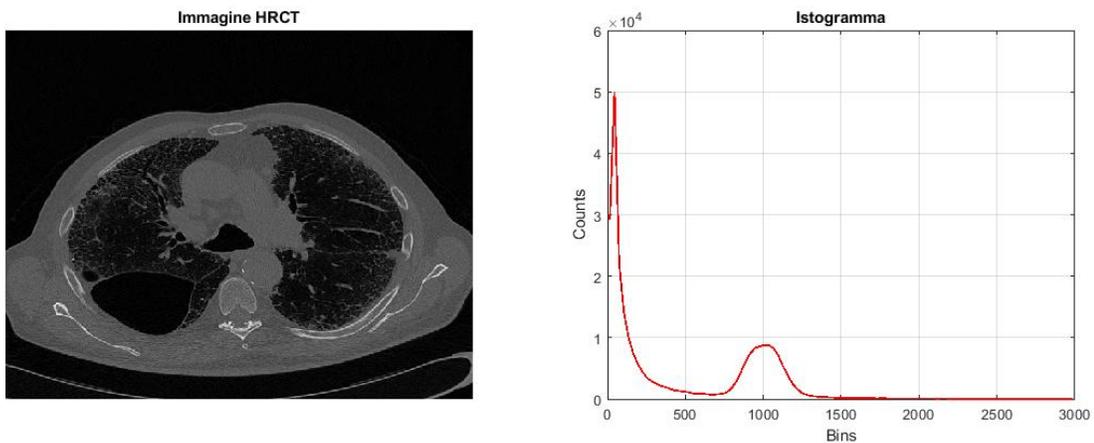


Fig 4.8: esempio di immagine HRCT con il suo istogramma.

L'ispezione visiva dell'istogramma mette in evidenza come i due picchi non siano perfettamente distinti. L'imperfetta separazione dei due picchi può indurre ad una classificazione errata dei pixel. La segmentazione a soglia, infatti, ha dei limiti:

1. Non distingue oggetti topologicamente diversi a cui è assegnato lo stesso segnale. Per questo, sia il fondo che i polmoni vengono attribuiti alla stessa classe;
2. Molti pixel sono spuri perché sono dovuti al rumore. Quanto più rumore c'è, tanto più aumenta la dispersione dei livelli di grigio attorno ai picchi, che si allargano e sono più difficili da separare;
3. La soglia viene definita manualmente, dopo aver visto l'istogramma.

Il problema introdotto dal punto 3., in questo lavoro, è stato risolto attraverso un algoritmo che ricerca autonomamente la soglia facendo ricorso ad una tecnica di *clustering*: l'algoritmo *fuzzy C-means*.

Il clustering è una tecnica di analisi non supervisionata che raggruppa elementi omogenei in un insieme di dati sulla base del numero di classi da trovare, da fornire in input. Oltre al numero di clusters da trovare, l'input dell'algoritmo di clustering è un vettore che ha $N \times M$ elementi dove N è il numero di pixel (o voxel) dell'immagine da analizzare e classificare e M è il numero di acquisizioni disponibili sui pixel (o voxels) stessi. Ci sono tre classi di algoritmi di clustering: clustering *esclusivo*, clustering *non esclusivo* e clustering *probabilistico*.

Nel clustering non esclusivo, un dato può appartenere a più clusters contemporaneamente, con diversi livelli di appartenenza. La somma dei livelli di appartenenza su tutti i possibili clusters è uguale a 1. Infatti, le regole su cui si basano gli algoritmi di clustering non esclusivo fanno ricorso ad una logica cosiddetta *fuzzy*, che rifiuta l'assunto del principio di contraddizione per il quale un elemento non può appartenere contemporaneamente a due insiemi, e che per ogni elemento definisce un valore di appartenenza ad un insieme compreso tra 0 e 1.

L'algoritmo fuzzy C-means (FCM) è basato sulla minimizzazione della seguente funzione obiettivo:

$$J_{FCM} = \sum_{j \in \Omega} \sum_{k=1}^C u_{jk}^m \|y_j - \nu_k\|^2$$

dove m è strettamente maggiore di 1, u_{jk} è il grado di appartenenza di y_j rispetto al cluster k , y_j è il j -esimo di N dati appartenenti all'insieme Ω , ν_k è il centro del cluster k e $\|\cdot\|$ è una distanza. C è il numero di clusters. m è un parametro, detto *fuzzyness*, al crescere del quale il peso dei valori di appartenenza più bassi decresce, fino ad annullarsi per valori di m molto grandi. Per m uguale a infinito, infatti, l'algoritmo non è più di clustering non esclusivo ma di clustering esclusivo.

La funzione di appartenenza u_{jk} viene aggiornata ad ogni iterazione dell'algoritmo, così come i centri dei cluster ν definiti come:

$$u_{jk} = \frac{1}{\sum_{q=1}^C \left(\frac{\|y_j - \nu_k\|}{\|y_j - \nu_q\|} \right)^{\frac{2}{m-1}}}, \nu_k = \frac{\sum_{j=1}^N u_{jk}^m y_j}{\sum_{j=1}^N u_{jk}^m}$$

L'algoritmo iterativo si ferma quando la differenza tra il valore corrente di u e il valore precedente è più piccola di una soglia.

Dal punto di vista implementativo, l'algoritmo fuzzy C-means è stato eseguito in Matlab®(R2019B) attraverso la funzione **fcm**; per ogni immagine, in input è stato fornito il vettore contenente tutti i pixel e il numero di cluster in cui classificarli, fissato a 2. Il numero di cluster è stato scelto sulla base dell'istogramma in Fig 4.8, rappresentativo della distribuzione dei pixel in tutte le immagini. L'algoritmo, una volta giunto a convergenza, ha fornito in output:

1. Un vettore con numero di elementi pari al numero di classi e valori pari ai centri dei cluster;
2. Una matrice con tante righe quanti sono i cluster e tante colonne quanti sono i pixel totali dell'immagine; le colonne contengono il grado di appartenenza di ogni pixel ad ognuna delle due classi.

Poi, sono stati eseguiti diversi passaggi:

1. Dei centri dei due cluster calcolati dall'algoritmo è stato preso il minore, perché associato al fondo e ai polmoni;
2. Tra i pixel raggruppati nel cluster che ha centro nel valore calcolato al passo precedente, sono stati presi quelli che superavano un valore soglia di appartenenza fissato a 0.7. In questo modo, è stata ottenuta una maschera binaria, in cui:
 - Tutti i pixel con un grado di appartenenza al fondo e ai polmoni superiore al 70% hanno valore 1;
 - Tutti i pixel con un grado di appartenenza al fondo e ai polmoni inferiore o uguale al 70% hanno valore 0;

4.3.2.2 Isolamento dei polmoni e ridimensionamento dell'immagine

Una volta ottenuta l'immagine binaria dall'algoritmo fuzzy C-means in cui sono state messe in evidenza le regioni appartenenti al fondo e ai polmoni, sono stati eseguiti i seguenti passaggi affinché, dell'immagine originale, restassero solo i pixel appartenenti ai polmoni:

1. È stato ottenuto il negativo della maschera binaria in uscita dal fuzzy C-mean così che fossero a 1 tutti i pixel del grasso addominale;
2. La maschera ottenuta è stata riempita, con l'istruzione *imfill*, così che apparisse bianco (valore 1) tutto ciò che era del paziente, nero (valore 0) tutto ciò che era al di fuori del paziente;
3. È stato eseguito il prodotto tra la maschera binaria in uscita dal fuzzy C-mean e l'immagine originale;
4. È stato eseguito il prodotto tra la maschera ottenuta al punto 2. e l'immagine a livelli di grigio ottenuta al punto 3.

I passaggi appena descritti e il risultato finale sono riportati in Fig 4.9.

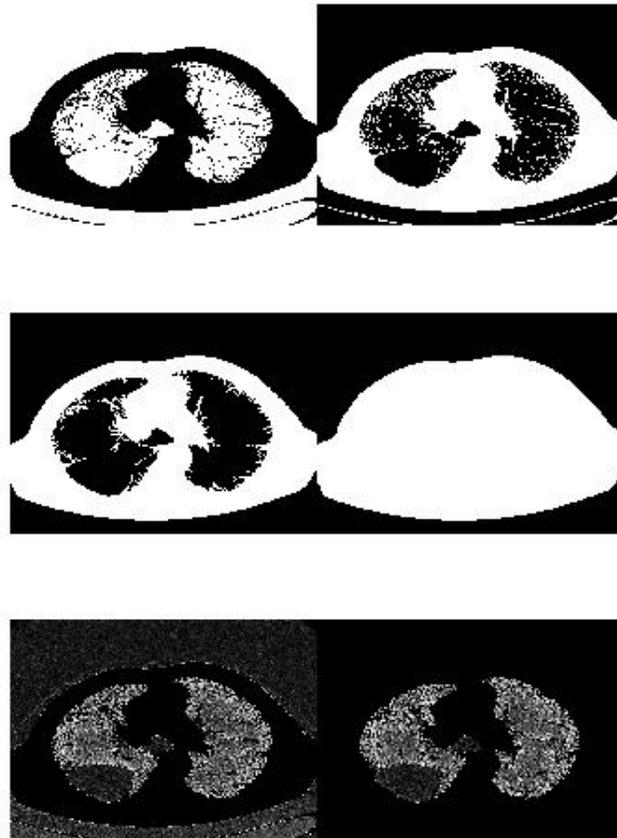


Fig 4.9: pre-processing dell'immagine, prima del ridimensionamento. In alto: maschera binaria in output dall'algoritmo fuzzy C-mean (a sinistra) con il suo negativo (a destra); al centro: isolamento dell'addome; in basso: prodotto tra la maschera binaria del fuzzy C-mean e l'immagine originale (a sinistra) che, moltiplicata per l'immagine binaria che isola l'addome, risulta nell'immagine finale, pronta per training, a meno del ridimensionamento. (a destra)

Una volta ottenuta, l'immagine a livelli di grigio, contenente solo i pixel appartenenti ai polmoni, è stata adattata alle dimensioni dello strato di input della rete. Quindi, da un'immagine 512x512, è stata ottenuta un'immagine 492x492, attraverso un crop centrato nel centro dell'immagine. Il ridimensionamento è stato eseguito anche sulla maschera binaria della ground truth, affinché ci fosse un matching perfetto delle due matrici.

Capitolo 5

Risultati

Nel quinto capitolo, verranno mostrate e commentate le performances della UIP-net sia in fase di training che di test. Verrà, innanzitutto, descritta la strategia di addestramento adottata per prevenire l'overfitting; seguirà la spiegazione di com'è stato eseguito il test e, infine, verranno riportate le analisi quantitative e qualitative dei risultati, in seguito alle quali sono state apportate delle correzioni in post-processing. Infine, vi è una discussione sui risultati ottenuti.

5.1 Training

L'addestramento della rete neurale è stato eseguito in Python 3.7 attraverso Keras, che utilizza come back-end la libreria TensorFlow (versione 1.14), progettata appositamente per l'apprendimento automatico di reti neurali. Per la definizione delle metriche di accuratezza, è stata usata la libreria *segmentation-models* che fornisce strumenti utili per la segmentazione delle immagini.

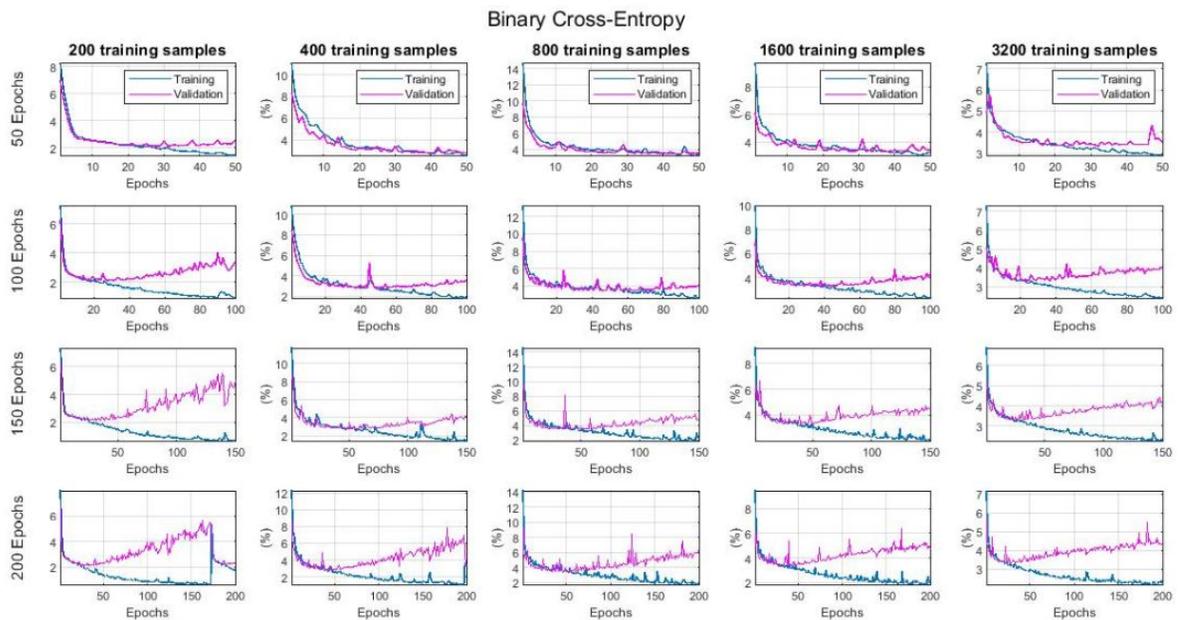
Del dataset a disposizione, sono state utilizzate 20 TC di pazienti diversi: 13 per il training e 7 per il test. Di ogni TC, poi, sono state prese le singole slices pre-processate. La rete è stata progettata per lavorare con le matrici bidimensionali appositamente per aumentare la numerosità dei dati disponibili per l'addestramento. Così facendo, il training set contava un numero totale di immagini di 3200. In input, la rete prende array 4-D in cui la prima dimensione rappresenta il numero di immagini impilate a formare l'insieme, la seconda e la terza rappresentano il nu-

mero di righe e colonne delle immagini, l'ultima il numero dei canali, in questo caso pari a 1. Affinché non ci fosse overfitting, è stato necessario stabilire:

1. Quante immagini fossero sufficienti per il training, in proporzione al numero di parametri della rete da calcolare;
2. Quante epoche fossero sufficienti così da applicare, eventualmente, l'*early stopping*.

Quindi, sono stati fatti 20 addestramenti:

1. Variando la numerosità del dataset e mantenendo fisso il numero di epoche, sono stati forniti alla rete: 200, 400, 800, 1600, 3200 esempi.
2. Variando il numero di epoche e mantenendo fisso il numero di esempi, la rete è stata addestrata per: 50, 100, 150, 200 epoche.



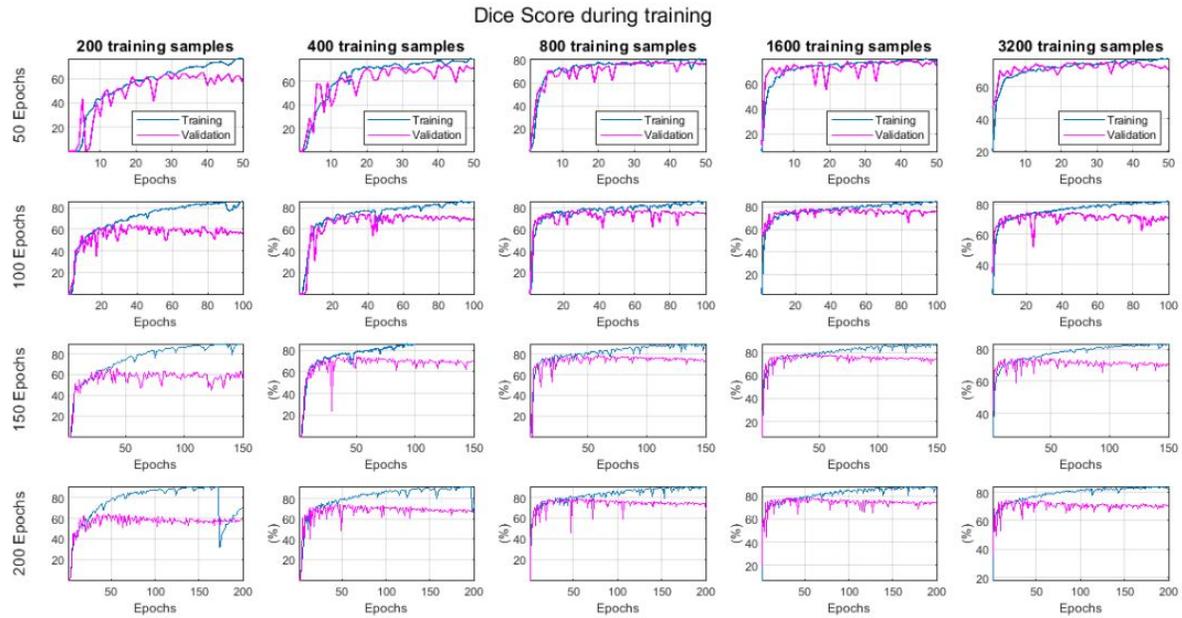


Fig 5.1: in alto, andamento della funzione di perdita durante gli addestramenti; in basso: andamento dell'accuratezza, in termini di indice di Dice.

Tutti gli addestramenti sono stati condotti mantenendo fissa la dimensione del batch a 10. Attraverso la definizione della dimensione del batch, si stabilisce il numero di immagini che vengono mostrate alla rete a ogni iterazione. Un'epoca è conclusa se alla rete sono stati mostrati tutti gli esempi del training set. Quindi, per concludere un'epoca, ci sarà bisogno di un numero di iterazioni pari al rapporto tra la numerosità del training set e quella del batch: quanto più grande sarà la dimensione del batch, tanto maggiore sarà lo spazio di memoria necessario, e fissarla a 10 è un buon compromesso.

In Fig 5.1, sono mostrati gli andamenti della funzione di costo e della metrica di accuratezza, definiti nel capitolo 4, paragrafo 4.2.2.2, durante il training.

Dall'analisi delle curve, il modello migliore in termini di accuratezza, stabilità, e anche a fronte dell'overfitting risulta essere quello addestrato con 800 campioni, per 50 epoche. L'overfitting, infatti, porta l'errore e l'accuratezza sull'insieme di validazione ad essere, rispettivamente, maggiore e minore rispetto all'errore e all'accuratezza misurati sugli insiemi di training, all'aumentare delle epoche. Per quanto riguarda l'insieme di validazione, questo si compone, ad ogni epoca, del 30% dei dati totali di training. Ad ogni epoca, i dati di training vengono

	Dice	Jaccard
Split 1	76.69%	62.36%
Split 2	75.86%	61.21%
Split 3	78.49%	64.7%
Split 4	69.59%	53.61%
Split 5	79.87%	66.58%

Tabella 4: indice di Dice e Jaccard per ogni split della convalida incrociata.

mescolati, così che vengano presi casualmente i campioni del batch, sia per il training che per il validation set.

5.1.1 Convalida incrociata

Una volta scelto di addestrare la rete con 800 campioni per 50 epoche, ne è stata valutata la capacità di generalizzazione, ovvero la capacità di classificare correttamente immagini mai viste durante il training. Questo perché, anche a fronte di una buona performance in fase di addestramento, non è detto che la rete abbia appreso; può succedere, piuttosto, che abbia memorizzato i dati. Per questo, è stato utilizzato il metodo di *cross-validazione*. La cross-validazione, o convalida incrociata, è una tecnica statistica che consiste nella suddivisione del training set in k parti di uguale numerosità. Ad ogni passo, la k -ma parte dell'insieme dei dati viene tenuta per il test mentre con gli altri $(k - 1)$ insiemi avviene l'addestramento. In questo modo, la rete viene allenata per ognuna delle k parti. In definitiva, il training set è stato diviso in gruppi di uguale numerosità, è stato escluso iterativamente un gruppo alla volta e si è cercato di predirlo coi gruppi non esclusi. Nello specifico, il training set della UIP-net è stato diviso in 5 gruppi uguali, per cui per 5 volte la rete è stata addestrata su 640 immagini e testata su 160 mai viste prima. I risultati sulle metriche di valutazione, ovvero gli indici di Dice e di Jaccard per ogni *split* della convalida, sono riassunti in Tabella 4, poi riportati nei box-plot di Fig 5.2.

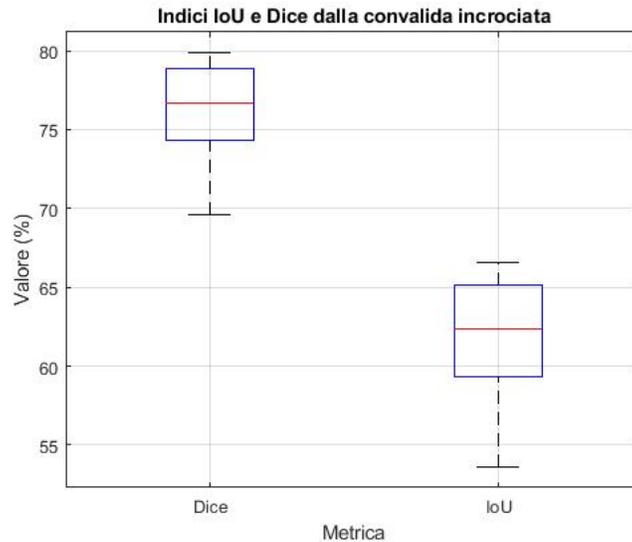


Fig 5.2: boxplot degli indici di Dice e Jaccard valutati durante la convalida incrociata.

Il valore medio e la deviazione standard degli indici in fase di training, quindi, sono:

1. **Indice di Dice: $76,10\% \pm 4,45\%$;**
2. **Indice di Jaccard: $61,69\% \pm 3,54\%$.**

Il valore medio dell'indice di Dice è superiore rispetto a quello dell'indice di Jaccard, perché è meno severo e dà un peso doppio ai veri positivi.

5.2 Test e valutazione della performance

Una volta calcolati e salvati i parametri del modello addestrato con 800 campioni, per 50 epoche e con convalida incrociata, è stato eseguito il test su 7 TC di pazienti diversi. Per fare il test, alla rete sono state fornite, dopo il solito pre-processing, solo le slices originali che sono state segmentate singolarmente; le maschere binarie in uscita dalla rete sono state, poi, esportate per essere confrontate con la ground truth, senza apportarvi nessuna modifica (Fig 5.3).

Dopo averle esportate, la bontà delle segmentazioni è stata valutata attraverso un'analisi quantitativa e una qualitativa. L'analisi quantitativa è stata svolta calcolando cinque indici per il confronto con la ground truth: l'indice di Dice, di Jaccard, di cui è stato già discusso, il BF-score, la sensibilità e la specificità. L'analisi qualitativa, invece, è consistita nel giudizio visivo

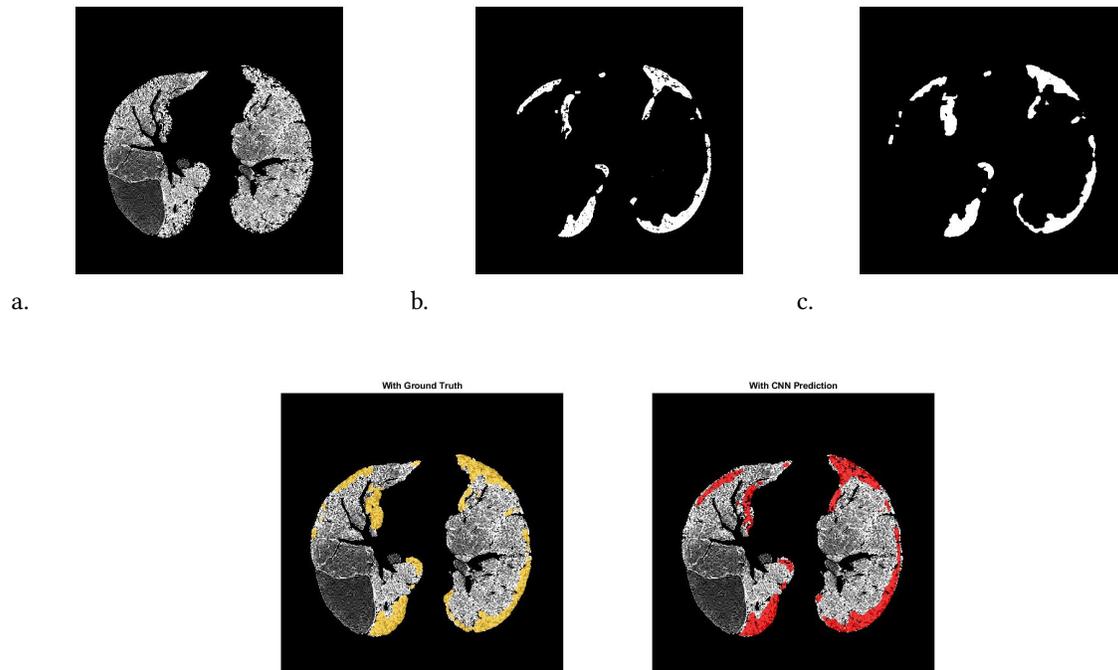


Fig 5.3. In alto: a) immagine originale dopo il pre-processing; b) ground truth; c) segmentazione della rete convoluzionale. In basso: a sinistra, immagine originale con ground truth in giallo sovrapposta; a destra, con segmentazione della rete in rosso sovrapposta

attribuito da un operatore clinico specializzato in radiologia toracica, che opera nell'UO Radiodiagnostica 2, in seguito al quale sono stati applicati delle correzioni in post-processing che hanno, effettivamente, migliorato le metriche sopracitate.

5.2.1 Analisi quantitativa

L'analisi quantitativa dei risultati è stata eseguita in Matlab®, versione R2019b. I risultati sono riportati in Tabella 5.

	P1	P2	P3	P4	P5	P6	P7
Dice	74,94%	73,3%	61,58%	59,4%	58,4%	57%	56,65%
Jaccard	60,67%	59,04%	47,14%	43,87%	44,81%	44%	42,6%
BF-score	78,96%	72,9%	81,16%	78,68%	81,32%	75%	83,73%
Sensitività	83,52%	83,51%	84,57%	74,46%	80,39%	74%	77,03%
Specificità	98,75%	98,33%	97,98%	98,7%	99%	99%	98,17%

Tabella 5: indici di qualità delle segmentazioni sulle immagini del test set; in verde, i valori massimi ottenuti; in rosso, i valori minimi.

Nel capitolo 4, gli indici di Dice e di Jaccard per il confronto tra la segmentazione della rete e la ground truth, sono stati già discussi; in più, è possibile confrontare i contorni. In questo caso, la misura di riferimento è il BF contour matching score, *Boundary F1*, o **BF-score**. Il BF-score viene definito a partire da altre due misure, la precisione (*precision*) e il richiamo (*recall*) che si definiscono come:

$$P = \frac{1}{|C_s|} \sum_{i=1}^{N_s} \|d(C_s(i), C_g) < \theta\|, R = \frac{1}{|C_G|} \sum_{i=1}^{N_G} \|d(C_G(i), C_s) < \theta\|$$

dove d è la distanza euclidea tra un punto del contorno e l'altro contorno, intesa come il minimo delle distanze tra il punto $C_s(i)$ e tutti i punti di C_G ; C_s è il contorno dell'immagine segmentata e C_G è il contorno della ground truth. θ è un valore di soglia pari allo 0.75% della diagonale dell'immagine. Quindi, P è la percentuale di punti del contorno segmentato ad una distanza dalla ground truth minore di θ , e R è la percentuale di punti del contorno della ground truth ad una distanza dal contorno segmentato minore di θ . A partire da queste due misure, il BF-score si ottiene come:

$$F1 = \frac{2PR}{P + R}$$

Il calcolo del BF-score è stato implementato in Matlab® attraverso la funzione *bfscore*.

La **sensitività**, in statistica, rappresenta la capacità di un test di rilevare casi positivi nella popolazione di riferimento; applicata alle immagini e definiti i veri positivi (TP) e i falsi negativi

(FN), anche questi spiegati nel capitolo precedente, la sensibilità si calcola come:

$$Sensitivity = \frac{TP}{(TP + FN)}$$

TP e FN sono stati ottenuti calcolando il numero di indici degli elementi dell'immagine somma e dell'immagine differenza della segmentazione e della ground truth uguali, rispettivamente, a 2 e a 1. Infatti, si ha un vero positivo quando nella stessa posizione l'immagine segmentata e la ground truth hanno pixel a 1; si ha un falso negativo quando, nella stessa posizione, la ground truth ha un pixel a 1 mentre la segmentazione della rete ha un pixel a 0.

La **specificità** si contrappone alla sensibilità e rappresenta la capacità di un test di rilevare veri negativi sul totale dei casi negativi. Dati i veri negativi (TN) e i falsi positivi (FP), la specificità, infatti, si calcola come:

$$Specificità = \frac{TN}{(TN + FP)}$$

In questo caso, TN e FP sono stati ottenuti dal numero degli indici degli elementi dell'immagine somma e dell'immagine differenza della segmentazione e della ground truth uguali, rispettivamente, a 0 e a -1. Si ha un vero negativo quando, nella stessa posizione, l'immagine segmentata e la ground truth hanno pixel a 0, e si ha un falso positivo quando nella stessa posizione la ground truth ha un pixel a 0 mentre la segmentazione della rete ha un pixel a 1.

Il valore medio e la deviazione standard degli indici in fase di test, quindi, sono:

	Dice	Jaccard	BF-score	Sensibilità	Specificità
Valore medio	63,10%	49%	78,8%	79,6%	98,5%
Dev. standard	7%	7%	3%	4,4%	0,4%

Nel box-plot in figura 5.4, è riportata una rappresentazione visiva della dispersione degli indici di confronto.

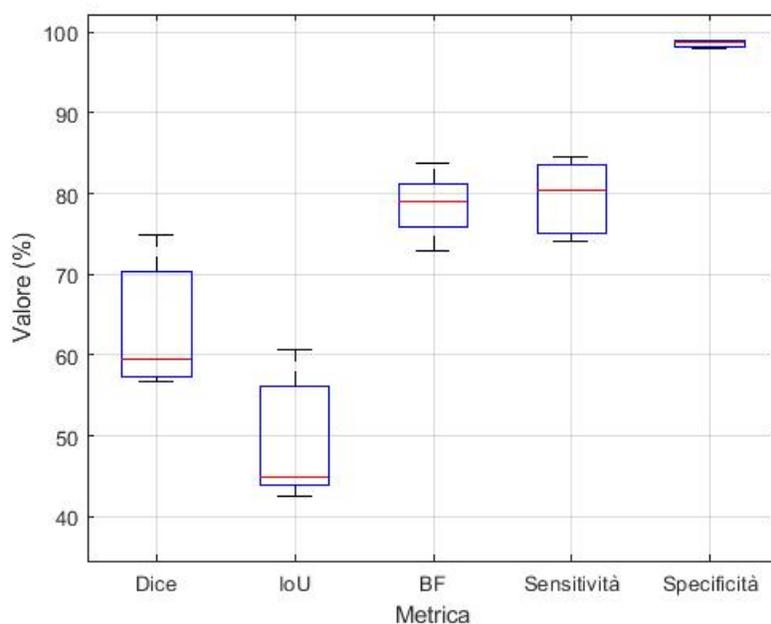


Fig 5.4: indici di confronto tra la performance della rete e la ground truth.

5.2.2 Analisi qualitativa e correzioni in post-processing

Parallelamente all'analisi quantitativa, è stato chiesto ad un operatore clinico specializzato in radiologia toracica, che opera nell'UO Radiodiagnostica 2, un giudizio qualitativo sulla base di un'ispezione visiva dei risultati. In Fig 5.5, è riportato un esempio di campione somministrato che consiste nell'immagine originale, nell'immagine originale con la ground truth, in giallo, e nell'immagine originale con segmentazione della rete, in verde. Il giudizio è stato complessivamente positivo:

1. Rispetto alla ground truth, la rete è più sensitiva; nei riquadri azzurri dell'immagine b. in Fig 5.6, sono evidenziati i pattern che CALIPER non riconosce;
2. Rispetto alla ground truth, la rete compie degli errori; nel riquadro in rosso dell'immagine d. in Fig 5.6, è evidenziata la segmentazione errata della rete. Nello specifico, viene segmentato l'intestino perché contiene aria ma ha composizione diversa, e per questo, la rete rileva fibrosi laddove non c'è. Inoltre, le frecce in rosso, nell'immagine c. di Fig 5.6, indicano i vasi sanguigni: CALIPER segmenta bene le componenti vascolari mentre la rete li agglomera insieme. Questo, se da un lato può essere un errore trascurabile

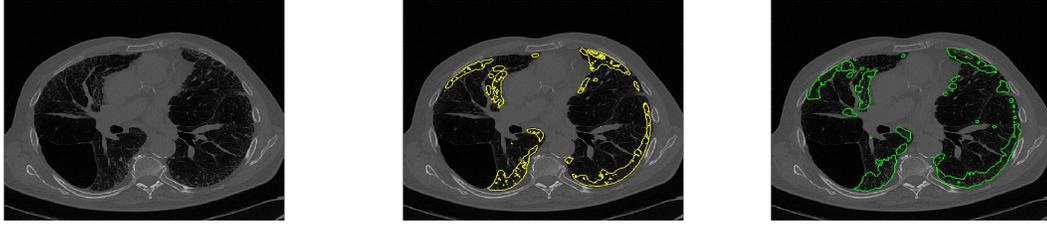


Fig 5.5: esempio di campione per la valutazione visiva. A sinistra: immagine HRCT originale; al centro: immagine originale con segmentazione della ground truth sovrapposta in giallo; a destra: immagine originale con segmentazione della rete sovrapposta in verde.

se la rete ha il compito di indirizzare l'occhio del radiologo per l'assegnamento di uno score visivo, dall'altro può inficiare la misurazione di volume di tessuto malato, e quindi fuorviare la diagnosi, se le segmentazioni servono a estrarre parametri quantitativi.

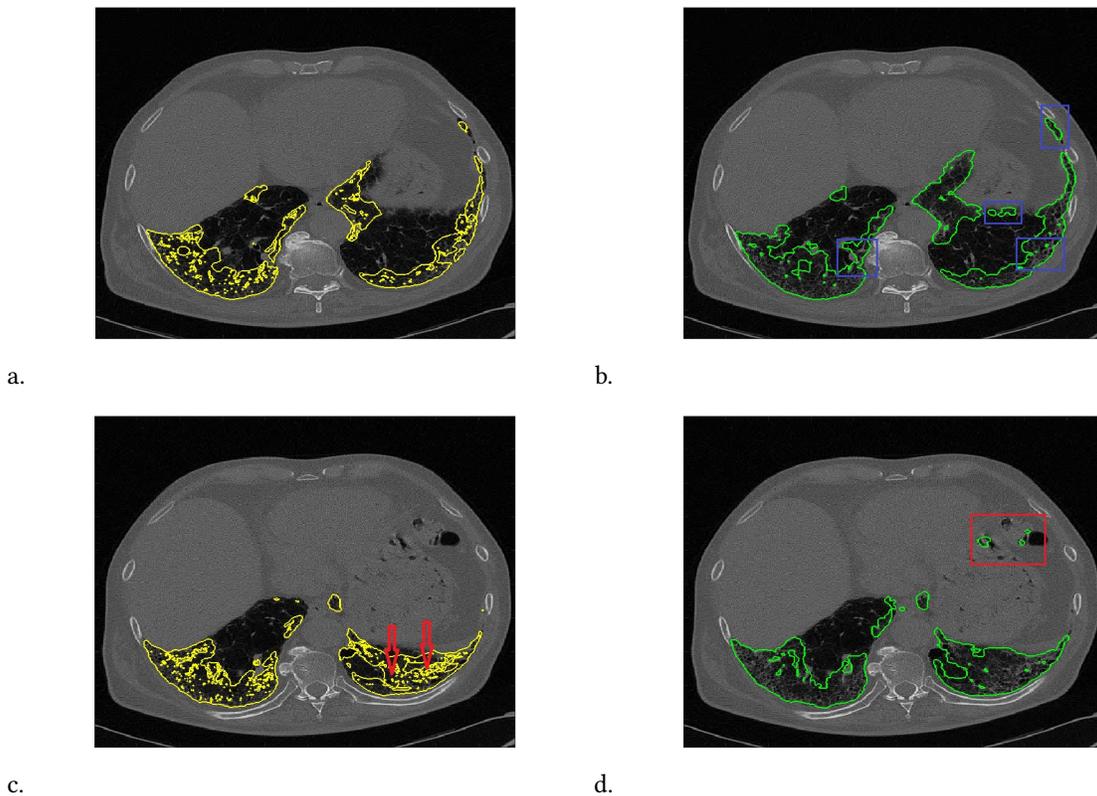


Fig. 5.6: immagini a confronto. A destra: segmentazioni di CALIPER, la ground truth; a sinistra: segmentazioni della rete. In alto, nei riquadri azzurri, vi sono pixel segmentati correttamente dalla rete ma ignorati da CALIPER; in basso: a sinistra, le frecce puntano ai vasi sanguigni correttamente segmentati da CALIPER, erroneamente ignorati dalla rete; a destra, nel riquadro rosso vi è una parte di intestino erroneamente segmentato dalla rete.

5.2.2.1 Eliminazione delle strutture esterne ai polmoni e dei vasi

A seguito dei giudizi visivi, le segmentazioni sono state migliorate per eliminare i vasi sanguigni e le vie aeree nei polmoni, erroneamente inclusi nella segmentazione della fibrosi, e gli organi esterni ai polmoni, quindi trachea, stomaco e intestino.

Per eliminare le segmentazioni fatte erroneamente su organi diversi dai polmoni, è stato utilizzato l'approccio proposto da Guang-Zhong Yang et al.[73]. La singola slice viene, innanzitutto, filtrata con filtro mediano 3×3 ; poi, viene eseguita un'operazione di soglia con il metodo di Otsu, seguita da una morfologica di chiusura che riempie le zone ad elevata attenuazione, non individuate al passaggio precedente, che lasciano parti vuote, che appaiono nere, nella regione ottenuta (tipicamente, queste zone rappresentano i vasi). Per la segmentazione delle vie aeree, invece, viene eseguita la tecnica del region growing delle componenti connesse. La trachea viene ricercata scorrendo dall'ultima slice fino alla prima, rilevando la presenza di un oggetto riconosciuto come *foreground* al centro dell'immagine; la consistenza spaziale assicura che si tratti effettivamente di trachea, così che non vengano presi oggetti spuri. Una volta identificata la posizione, la trachea viene segmentata con un'operazione di soglia a partire dalla quale vengono definiti dei punti (i cosiddetti *seeds*) per l'inizializzazione del region growing, che viene ripetuto iterativamente. Ad ogni iterazione, viene calcolato il volume delle vie aeree e viene calcolata la differenza di volume tra due iterazioni successive: se ricade entro una certa soglia, allora il region growing continua, altrimenti si interrompe perché l'albero bronchiale è stato segmentato tutto. Per distinguere le vie aeree che sono all'interno del parenchima polmonare da quelle che si trovano all'esterno, ogni slice viene analizzata a turno e si applica, anche in questo caso, il region growing delle componenti connesse; se il perimetro di una via aerea risulta circondato per almeno il 75% da un oggetto riconosciuto come polmone, allora anche l'intera componente connessa, disegnata dal region growing, viene etichettata come polmone. Solo alla fine, dalle informazioni sull'orientazione del paziente nel campo dell'immagine contenute nell'header del DICOM, si riesce a stabilire qual è il polmone destro e quale il sinistro e a dividere i polmoni nei rispettivi lobi.

Per la segmentazione dei vasi, da sottrarre alla segmentazione dei campi polmonari, è stato utilizzato l'approccio proposto da Sato et al.[74], mirato all'enfatizzazione della linearità delle

strutture attraverso un filtraggio spaziale in 3-D in grado di:

1. Rilevare strutture lineari di dimensioni diverse, con particolare riguardo a quelle più sottili;
2. Escludere le strutture con andamento non lineare e discontinue;
3. Rimuovere gli effetti dovuti al rumore e agli artefatti.

Per il primo punto, si integrano le risposte di filtri spaziali gaussiani applicati su scale diverse e a profili planari diversi dove la risposta al filtro del singolo profilo è:

$$R(x; \sigma_f) = \left\{ -\frac{d^2}{dx^2} G(X; \sigma_f) \right\} * I(x) \quad (5.1)$$

$G(x; \sigma_f)$ è la funzione gaussiana con deviazione standard σ_f e $*$ indica l'operazione di convoluzione con l'immagine $I(x)$. Per estendere il filtro al caso 3-D, si fa ricorso alla matrice Hessiana, che contiene le derivate seconde parziali dell'immagine, per descrivere la variazione locale delle strutture (l'autovalore massimo della matrice Hessiana indica la direzione di massima variazione). Per ogni elemento della matrice Hessiana, si valuta la risposta al filtro gaussiano; poi, si integrano i risultati per ricostruire le strutture continue che hanno andamento lineare, che identificano i vasi sanguigni o le vie aeree.

Dal punto di vista pratico, entrambi gli approcci sono stati implementati in 3D-slicer, attraverso la libreria *Chest Imaging Platform* (CIP). 3D-slicer è un software gratuito, open-source, scritto in linguaggio Python e Java, che permette la visualizzazione, la manipolazione e l'analisi delle immagini mediche. Attraverso 3D-slicer e CIP, sono state ottenute sia le maschere binarie dei polmoni che quelle dei vasi sanguigni. Le due maschere sono state, poi, sottratte per ottenerne una che comprendesse solo polmoni senza componenti vascolari, vie aeree e altri organi. In seguito, è stato eseguito il prodotto tra la maschera binaria ottenuta al passaggio precedente e la segmentazione fatta dalla UIP-net: così sono stati mantenuti inalterati i pixel appartenenti ai polmoni e annullati, allo stesso tempo, sia quelli dei vasi sanguigni che degli organi esterni ai polmoni.

In Fig 5.7, vi è il confronto tra la segmentazione prima del post-processing e la segmentazione a seguito del post-processing.

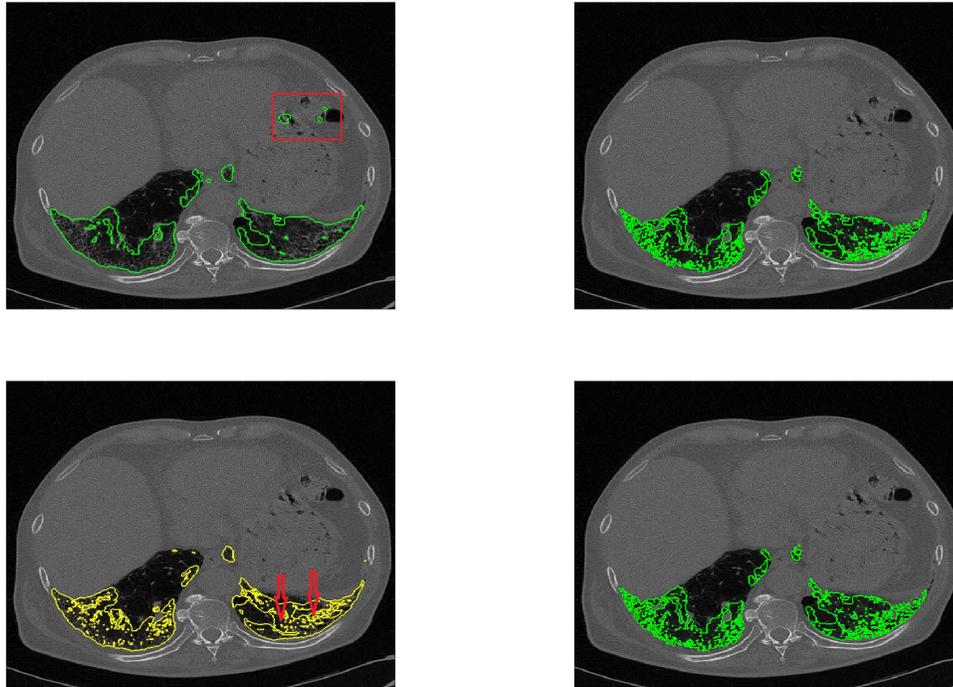


Fig 5.7: in alto, a confronto le segmentazioni della rete senza e con post-processing; in basso, a confronto la segmentazione di CALIPER e la segmentazione della rete dopo il post-processing.

In Tabella 6, sono riportati i valore dei soliti indici ricalcolati a seguito del post-processing.

	P1	P2	P3	P4	P5	P6	P7
Dice	84,43%	77,4%	65,77%	62,28%	75,43%	60,1%	64,58%
Jaccard	75,04%	64%	52,1%	47,44%	62,1%	47%	50,12%
BF-score	96,07%	96,1%	96,71%	91%	93,6%	89,2%	94,14%
Sensitività	75%	84%	85,87%	76,3%	80,7%	75%	78,1%
Specificità	100%	98,33%	98,66%	99,2%	99,3%	99%	98,87%

Tabella 6: indici di qualità della segmentazioni, a seguito del post-processing. In verde: i valori massimi ottenuti; in rosso: i valori minimi ottenuti.

Nel box-plot di figura 5.8, è riportata una rappresentazione visiva della dispersione dei soliti indici di confronto con la ground truth, ricalcolati.

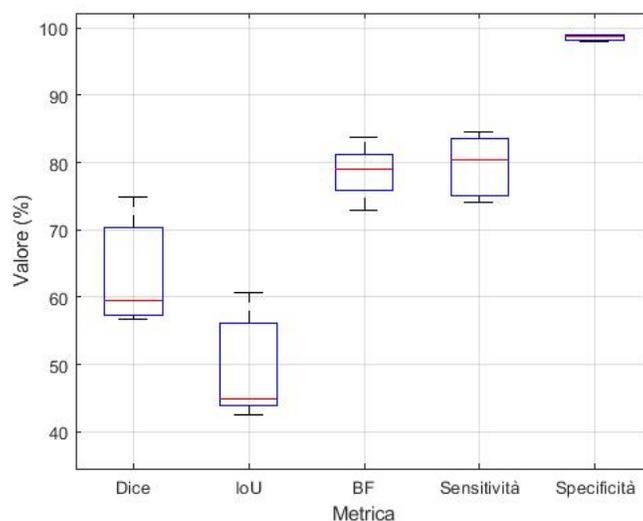


Fig 5.8: indici di qualità delle segmentazioni, a seguito del post-processing.

Le differenze dei valori medi e delle deviazioni standard degli indici di confronto tra le segmentazioni di partenza e quelle ottenute con i passaggi appena descritti, sono:

	Dice	Jaccard	BF-score	Sensitività	Specificità
Senza post-processing	63,10%±7%	49%±7%	78,8%±3%	79%±4,4%	98,5%±0,4%
Con post-processing	70%±9%	56,8%±10%	93,8±2,8%	79,2±4,4%	99±0,5%

5.3 Discussione

I risultati delle analisi quantitative hanno messo in luce diversi aspetti:

1. Confrontando gli indici di Dice e di Jaccard calcolati durante la convalida incrociata della fase di training e quelli calcolati sulle predizioni dell'insieme di test, emerge che il valore dei primi è superiore rispetto al valore dei secondi; questo è coerente con il principio di funzionamento delle reti neurali. Infatti, la rete predice bene i dati del training set perché è attraverso questi che ha imparato a discriminare le features rilevanti delle immagini;

invece, la bontà delle predizioni di dati nuovi, cioè mai visti prima, che sono quelli che compongono il test set, tipicamente diminuisce.

2. Osservando i risultati delle predizioni dell'insieme di test, sui valori dei cinque indici, si può dire che:

- Gli indici di Dice e Jaccard funzionano bene per addestrare la rete, ma non rispecchiano il punteggio visivo assegnato dall'operatore clinico alle segmentazioni delle immagini dell'insieme di test. Gli indici di Dice e Jaccard sono molto indicati per valutare le segmentazioni automatiche di regioni compatte e circoscritte, come i noduli, ma non di malattie sparse, disomogenee e con diffusione random come le interstiziopatie polmonari (Fig 5.9).
- Contrariamente agli indici di Dice e di Jaccard, il BF-score misura il matching dei contorni ed è coerente con il giudizio visivo. Infatti, assume un valore sempre maggiore dei due indici sopracitati.
- Il valore di specificità è sempre molto elevato ma non può essere preso come misura di riferimento per valutare la performance della rete. Com'è stato appena accennato, la malattia ha una distribuzione eterogenea nei polmoni e tende a concentrarsi prevalentemente nei campi medi e inferiori. Dal momento che la specificità conta i veri negativi sui casi negativi totali, i molti pixel a 0 sulle slices apicali dello stesso paziente, pesano sul valore finale di specificità (fig. 5.10).
- La sensibilità, al contrario, può essere presa come misura di riferimento perché conta i veri positivi su tutti i positivi; infatti, tendenzialmente, segue il BF-score.

3. Osservando i risultati sulle predizioni dell'insieme di test a seguito del post-processing, si può dire che:

- I valori di tutti gli indici di qualità aumentano.
- Gli indici di Dice e Jaccard restano ancora penalizzanti, sebbene entrambi aumentino del 7%.

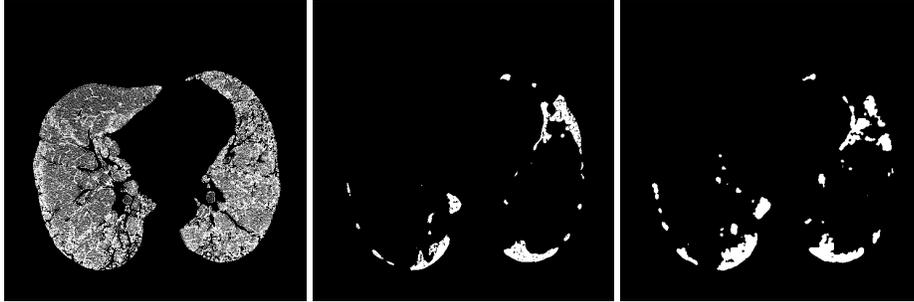


Fig 5.9: esempio di immagine con segmentazione disomogenea e diffusa. A sinistra: immagine originale; al centro: ground truth; a destra: segmentazione della rete.

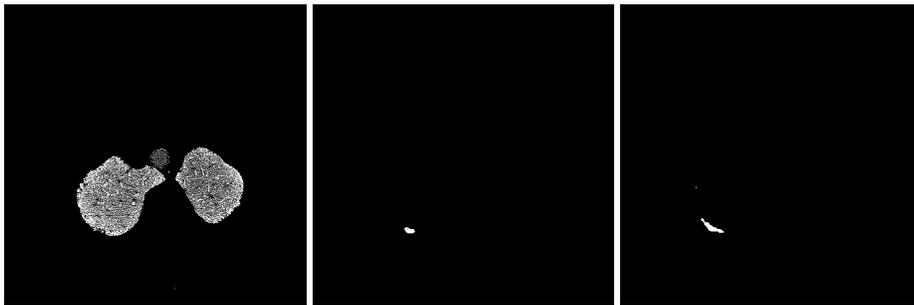


Fig 5.10: esempio di immagine con pochi pixel segmentati con elevata specificità. A sinistra: immagine originale; al centro: ground truth; a destra: segmentazione della rete. L'immagine è delle regioni apicali dei polmoni. La malattia si concentra nei campi medi-inferiori.

- Il BF-score arriva ad un valore massimo del 96,71% e uno medio del 93% e questo è coerente con quanto osservato sulle predizioni dell'insieme di test senza post-processing. L'incremento percentuale dei valori medi tra l'uno e l'altro insieme è del 15%.
- La sensibilità e la specificità restano pressappoco gli stessi, con un incremento dello 0,1% sulla sensibilità e dell'1% sulla specificità, che già prima del post-processing aveva un valore molto elevato. Con il post-processing, la specificità arriva ad un valore del 99%.

Anche le segmentazioni corrette in post-processing sono state valutate dallo stesso operatore clinico: il punteggio visivo dato ha rispecchiato il BF-score ed è stata confermata, ancora una volta, la maggiore sensibilità dei risultati forniti dalla UIP-net rispetto a quelli forniti da CALIPER.

Capitolo 6

Conclusioni

L'obiettivo del lavoro svolto era di indagare la capacità del *deep learning* di segmentare autonomamente le interstiziopatie polmonari in immagini tomografiche ad alta risoluzione del torace. Già in letteratura, si trovano numerosi metodi per la rilevazione automatica dei pattern di *Usual Interstitial Pneumonia* (UIP), che rappresenta un passaggio cruciale per avere una diagnosi; tuttavia, i tools proposti hanno dei limiti dovuti, principalmente, alla dipendenza dal giudizio dell'operatore e dai contesti clinici in cui le immagini radiologiche vengono acquisite. Invece, il deep learning fornisce strumenti indipendenti che imparano, autonomamente e dai dati, a svolgere il compito al quale sono preposti. Tra questi strumenti, l'attenzione è stata focalizzata sulle reti neurali convoluzionali (CNN) perché, come l'occhio umano, colgono le caratteristiche più fini delle immagini e le classificano.

Per riuscire a progettare la UIP-net, una rete convoluzionale per la segmentazione automatica delle interstiziopatie in immagini tomografiche di pazienti affetti da Fibrosi Idiopatica Polmonare, è stato, prima di tutto, condotto un ampio studio della letteratura, volto alla caratterizzazione esatta e precisa del problema e all'identificazione dei limiti degli strumenti allo stato dell'arte. In seguito, sono stati fatti numerosi tentativi: a partire da architetture più complesse e profonde, si è giunti ad un'architettura essenziale che trova, proprio nella sua essenzialità, la chiave per analizzare le caratteristiche salienti delle immagini senza sottrarre informazioni. Una volta definita l'architettura, è stato eseguito un *fine-tuning* degli iper-parametri, al variare dei quali il comportamento della CNN, spesso, diventa imprevedibile. Prima di procedere con il training della UIP-net, sono stati adeguatamente preparati i dati coi quali addestrarla,

attraverso un pre-processing delle immagini volto alla riduzione del contenuto informativo, nell'intento sia di abbattere le risorse computazionali necessarie, sia di trattenere solo le caratteristiche rilevanti; entrambi gli aspetti, infatti, influiscono sulla velocità di convergenza dell'algoritmo. Una volta stabilita l'architettura, definiti gli iper-parametri e pre-processate le immagini, l'indagine è stata indirizzata verso la ricerca della strategia di addestramento migliore per minimizzare il rischio di over-fitting e garantire, di conseguenza, una maggiore riproducibilità. Infine, sono state identificate le metriche più indicate per valutare le performances della rete su di un test set. Gli indici di performance della rete tipicamente misurano il confronto tra la segmentazione di un algoritmo automatico e una cosiddetta *ground truth*, che rappresenta il risultato da raggiungere. Tra questi indici, la sensibilità e il *Boundary F1-score* (o BF-score) si sono rivelati i più coerenti col giudizio visivo. Il BF-score calcola la percentuale di punti del contorno segmentato ad una distanza minima da quello della ground truth: con delle correzioni apportate in post-processing alle segmentazioni fornite dalla UIP-net, esso arriva ad un valore di circa 94%, mentre la sensibilità arriva ad un valore percentuale dell'80%.

Alla luce dei risultati numerici sugli indici di prestazione, si può concludere che il deep learning e, nello specifico, le reti convoluzionali sono strumenti promettenti che possono superare i limiti delle tecniche allo stato dell'arte, primo fra tutto il bias dovuto al giudizio degli operatori clinici di cui necessitano per classificare le immagini.

Ci sono dei limiti, però, che offrono spunti a sviluppi futuri:

1. **La numerosità del dataset.** In futuro, aggiungere campioni al test set può conferire maggior robustezza ai risultati della UIP-net.
2. **Addestrare la UIP-net su un insieme di immagini ricostruite con diversi kernel spaziali e acquisite con macchinari diversi** potrebbe aumentare la riproducibilità dei risultati. In questo lavoro, infatti, la CNN è stata addestrata con immagini ricostruite attraverso un solo kernel (B60) e acquisite da un solo scanner (SIEMENS Sensation 64).
3. **La capacità di differenziare i pattern UIP.** A fronte di una numerosità maggiore anche del training set, l'architettura della UIP-net potrebbe essere arricchita, così da riuscire a rilevare molte più caratteristiche e a ottenere una differenziazione più fine dei

pattern UIP: potrebbero essere distinti, ad esempio, l'honey-combing dall'opacità a vetro smerigliato.

Bibliografia

- [1] Ignazio Fazzari. *Anatomia Sistemica*. UTET, 1978.
- [2] *Medicina Viva, Enciclopedia pratica della medicina moderna*. Fabbri Editore, 1976.
- [3] «American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of The Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 anche by the ERS executive committee». In: *Am Respir Crit Care Med* 165.2 (2001), pp. 277–304.
- [4] Mueller Mang C. Grosse C. Schmid K Stiebellehner L Bankier AA. «What every radiologist should know about idiopathic interstitial pneumonias». In: *Radiographics* 27.3 (2007), pp. 595–615.
- [5] Raghu G. et al. «An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management». In: *Am J Respir Crit Care Med* 183.6 (2011), pp. 788–824.
- [6] Shin KM Lee KS Chung MP Han J Bae YA Kim TS Chung MJ. «Prognostic determinants among clinical, thin-section CT, and histopathologic findings for fibrotic idiopathic interstitial pneumonias: tertiary hospital study». In: *Radiology* 249 (2008), pp. 328–337.
- [7] «American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias, This official statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was approved by the ATS board of directors, June 2013 and by the ERS Steering Committee». In: *Am Respir Crit Care Med* 188.6 (2013), pp. 733–748.

- [8] Raghu G. et al. «Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JR-S/ALAT Clinical Practice Guideline». In: *Am J Respir Crit Care Med* 198.5 (2018), pp. 44–68.
- [9] *Diagnostica per la medicina nucleare*. URL: <http://lem.ch.unito.it/didattica/infochimica/2007Tecnezio/radiodiagnostica.html>.
- [10] *Windowing (CT)*. URL: <https://radiopaedia.org/articles/windowing-ct>.
- [11] Om P Sharma Tatjana Peros-Golubicic. *Clinical Atlas of Interstitial Lung Disease*. Springer, 2006.
- [12] Sverzellati N. «Highlights of HRCT imaging in IPF». In: *Respiratory Research* 13 (2013).
- [13] Ohkubo Hirotsugu e al. «Computer-based quantitative computed tomography image analysis in idiopathic pulmonary fibrosis: A mini review.» In: *Respiratory Investigation* (2017).
- [14] Nicholas D Weatherley et al. «Experimental and quantitative imaging techniques in interstitial lung disease». In: 74.6 (2019), pp. 611–619.
- [15] H Koyama, Y Ohno e Y et al. Yamazaki. «Quantitatively assessed CT imaging measures of pulmonary interstitial pneumonia: effects of reconstruction algorithms on histogram parameters». In: *Eur J Radiol* 74.142 (2010).
- [16] AC Best et al. «Idiopathic pulmonary fibrosis: physiologic tests, quantitative CT indexes, and CT visual scores as predictors of mortality». In: *Radiology* 246.935 (2008).
- [17] SY Ash et al. «Densitometric and local histogram based analysis of computed tomography images in patients with idiopathic pulmonary fibrosis». In: *Respir Res* 18.45 (2017).
- [18] N Sverzellati et al. «Visual score and quantitative CT indices in pulmonary fibrosis: relationship with physiologic impairment.» In: *Radiol Med* 112 (2007), pp. 1160–72.
- [19] Wu Xiaoping et al. «Computed Tomographic Biomarkers in Idiopathic Pulmonary Fibrosis The Future of Quantitative Analysis.» In: *American Journal of Respiratory and Critical Care Medicine* 199.1 (2018), pp. 12–21.

- [20] Y. Ash Samuel e al. «Densitometric and local histogram based analysis of computed tomography images in patients with idiopathic pulmonary fibrosis.» In: *Respiratory Research* 18 (45 2017).
- [21] HJ Kim e al. «Comparison of the quantitative CT imaging biomarkers of idiopathic pulmonary fibrosis at baseline and early change with an interval of 7 months.» In: *Acad Radiol* 22 (2015), pp. 70–80.
- [22] Alan C. Best et al. «Quantitative CT Indexes in Idiopathic Pulmonary Fibrosis: Relationship with Physiologic Impairment». In: *Radiology* 228.2 (2003), pp. 407–414.
- [23] Alan C. Best et al. «Idiopathic Pulmonary Fibrosis: Physiologic Tests, Quantitative CT Indexes, and CT Visual Scores as Predictors of Mortality». In: *Radiology* 246.3 (2008), pp. 935–940.
- [24] Davide Colombi et al. «Visual vs Fully Automatic Histogram-Based Assessment of Idiopathic Pulmonary Fibrosis (IPF) Progression Using Sequential Multidetector Computed Tomography (MDCT)». In: *PLOS ONE* 10.6 (giu. 2015), pp. 1–17.
- [25] SM Humphries e al. «Idiopathic Pulmonary Fibrosis: Data-driven Textural Analysis of Extent of Fibrosis at Baseline and 15-Month Follow-up.» In: *Radiology* 285.1 (ott. 2017), pp. 270–278.
- [26] Mario Anthimopoulos e al. «Semantic Segmentation of Pathological Lung Tissue With Dilated Fully Convolutional Networks.» In: *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS* 23.2 (2019).
- [27] IO Rosas e al. «Automated quantification of high-resolution CT scan findings in individuals at risk for pulmonary fibrosis.» In: *Chest* (2011), pp. 1590–7.
- [28] Ye Xu et al. «Computer-aided Classification of Interstitial Lung Diseases Via MDCT: 3D Adaptive Multiple Feature Method (3D AMFM)». In: *Academic Radiology* 13.8 (2006), pp. 969–978.
- [29] VA Zavaletta e al. «High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis.» In: *Academic Radiology* 14 (2007), pp. 772–87.

- [30] F Maldonado e al. «Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis.» In: *Eur Respir J* 43 (2014), pp. 204–12.
- [31] J Jacob e al. «Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures.» In: *Academic Radiology* 49.8 (2017), pp. 969–978.
- [32] Renuka Uppaluri e al. «Quantification of Pulmonary Emphysema from Lung Computed Tomography Images.» In: *Am J Respir Crit Care Med* 156.1 (1997), pp. 248–54.
- [33] TF Nonnenmacher e al. «Fractals in biology and medicine.» In: (1994).
- [34] Margaret L. Salisbury e al. «Idiopathic Pulmonary Fibrosis: The Association between the Adaptive Multiple Features Method and Fibrosis Outcomes.» In: *Am J Respir Crit Care Med* 195.7 (2017), pp. 921–29.
- [35] S Hu e al. «Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images.» In: *IEEE Trans Med Imaging* 20 (2001), pp. 490–498.
- [36] Hidenori Shikata e al. «Segmentation of Pulmonary Vascular Trees from Thoracic 3D CT Images.» In: *International Journal of Biomedical Imaging* (2009).
- [37] Hidenori Shikata e al. «Serial automated quantitative CT analysis in idiopathic pulmonary fibrosis: functional correlations and comparison with changes in visual CT scores.» In: *Eur Radiol* 28 (2018), pp. 1318–1327.
- [38] Brian J. Bartholomai e al. «Quantitative Computed Tomography Imaging of Interstitial Lung Diseases.» In: *J Thorac Imaging* 28.5 (2013), pp. 298–307.
- [39] J. Jacob e al. «Automated quantitative computed tomography versus visual computed tomography scoring in idiopathic pulmonary fibrosis: validation against pulmonary function.» In: *J Thorac Imaging* 31 (2016), pp. 304–311.
- [40] J Jacob e al. «Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures.» In: *Eur Respir J* 49 (2017).
- [41] TRS Moua e al. «Can progression of fibrosis as assessed by computer-aided lung informatics for pathology evaluation and rating (CALIPER) predict outcomes in patients with idiopathic pulmonary fibrosis.» In: *Chest* 140 (4 2011).

- [42] J. Jacob e al. «Serial automated quantitative CT analysis in idiopathic pulmonary fibrosis: functional correlations and comparison with changes in visual CT scores.» In: *Eur Radiol* 28 (2018), pp. 1318–1327.
- [43] T Watadani e al. «Interobserver variability in the CT assessment of honeycombing in the lungs.» In: *Radiology* 266 (2013), pp. 936–44.
- [44] SL Walsh e al. «Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT.» In: *Thorax* 71 (2016), pp. 45–51.
- [45] SL Walsh e al. «Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study.» In: *Lancet Respir Med* 6 (11 2018), pp. 837–45.
- [46] Mario Anthimopoulos e al. «Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network.» In: *IEEE TRANSACTIONS ON MEDICAL IMAGING* 35.5 (2016).
- [47] Guk Bae Kim e al. «Comparison of Shallow and Deep Learning Methods on Classifying the Regional Pattern of Diffuse Lung Disease.» In: *J Digit Imaging* 31 (2018), pp. 415–424.
- [48] V. Badrinarayanan e al. «SegNet: A deep convolutional encoder-decoder architecture for image segmentation.» In: *Trans. Pattern Anal. Mach. Intell.* 39.12 (2017), pp. 2481–2495.
- [49] O. Ronneberger e al. «Convolutional networks for biomedical image segmentation.» In: *Medical Image Computing and Computer-Assisted Intervention* 9351 (2015).
- [50] S. Agarwala e al. «Deep learning for screening of interstitial lung disease patterns in high-resolution CT images.» In: *Clinical Radiology* 75 (5 2020).
- [51] G. Wang et al. «DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.7 (2019), pp. 1559–1572.
- [52] D. H. Hubel e T. N. Wiesel. «Receptive fields and functional architecture of monkey striate cortex». In: *J Physiol* 195.1 (1968), pp. 215–243.

- [53] CS231n *Convolutional Neural Networks for Visual Recognition*. URL: <https://cs231n.github.io/convolutional-networks/>.
- [54] David Kaeli et al. «Chapter 4 - Examples». In: *Heterogeneous Computing with OpenCL 2.0*. A cura di David Kaeli et al. Boston: Morgan Kaufmann, 2015, pp. 75–109.
- [55] Shepherd A. *Second-Order Methods for Neural Networks*. SpringerVerlag, 1997.
- [56] Anthony M. e P.L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.
- [57] Baum E.B. e D.Haussler. «What size net gives valid generalization?» In: *Neural Computation* 1 (1989), pp. 151–160.
- [58] Bishop C. *Neural Networks for Pattern Recognition*. Clarendon, 1995.
- [59] MF Bobo e al. «Convolutional Neural Networks Improve Abdominal Organ Segmentation.» In: *Proc SPIE Int Soc Opt Eng.* (2018). DOI: 10.1117/12.2293751.
- [60] K. Sharma e al. «Automatic Segmentation of Kidneys using Deep Learning for Total Kidney Volume Quantification in Autosomal Dominant Polycystic Kidney Disease.» In: *Sci Rep.* 7 (2017).
- [61] Wang Xinggang e al. «Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning.» In: *Sci Rep.* 7 (15415 2017).
- [62] M. Abdel Ahmed e al. «Breast cancer classification using deep belief networks.» In: *Expert Systems With Application* 46 (2016), pp. 139–144.
- [63] P Savadjiev e al. «Demystification of AI-driven medical image interpretation: past, present and future.» In: *Eur Radiol.* 29 (2018), pp. 1616–24.
- [64] MP McBee e al. «Deep learning in radiology.» In: *Acad Radiol.* 25 (2018), pp. 1472–80.
- [65] G Zaharchuk e al. «Deep learning in neuroradiology.» In: *Am J Neuroradiol.* 39 (2018), pp. 1776–84.
- [66] A. Tang e al. «Canadian association of radiologists white paper on artificial intelligence in radiology.» In: *Can Assoc Radiol J.* 69 (2018), pp. 120–35.

- [67] D.C. Ciresan e al. «Deep neural networks segment neuronal membranes in electron microscopy images.» In: *NIPS* (2012), pp. 2852–2860.
- [68] Olaf Ronneberger e al. «U-Net: Convolutional Networks for Biomedical Image Segmentation.» In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI* (2015), pp. 34–241.
- [69] Fang Liu e al. «Deep Convolutional Neural Network and 3D Deformable Approach for Tissue Segmentation in Musculoskeletal Magnetic Resonance Imaging.» In: *Magnetic Resonance in Medicine* 79 (2018), pp. 2379–2391.
- [70] Y. Bengio. *Learning Deep Architectures for AI: Foundations and Trends in Machine Learning*. Boston, 2009, p. 127.
- [71] Y LeCun, Y Bengio e Hinton G. «Deep learning». In: *Nature* 521.7 (2019), pp. 436–44.
- [72] Kingma Diederik e Jimmy Ba. «Adam: A method for stochastic optimization.» In: (2014).
- [73] JC Ross e al. «Lung Extraction, Lobe Segmentation and Hierarchical Region Assessment for Quantitative Analysis on High Resolution Computed Tomography Images. Medical image computing and computer-assisted intervention: MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention.» In: *Medical Image Analysis*. 12 (2009), pp. 690–8.
- [74] Y. Sato e al. «Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images.» In: *Medical Image Analysis*. 2 (2 1998), pp. 143–68.