



UNIVERSITÀ
POLITECNICA
DELLE MARCHE

FACOLTÀ DI INGEGNERIA

CORSO DI LAUREA IN INGEGNERIA INFORMATICA E DELL'AUTOMAZIONE

Un approccio Multi-task Deep Learning per la valutazione del Covid-19 nelle ecografie polmonari

Multi-task Deep Learning for lung ultrasound Covid-19 detection

Candidato:

Lorenzo Federici

Relatore:

Prof. Primo Zingaretti

Correlatori:

Maria Chiara Fiorentino, PhD

Riccardo Rosati, PhD

Anno Accademico 2022-2023



UNIVERSITÀ
POLITECNICA
DELLE MARCHE

FACOLTÀ DI INGEGNERIA

CORSO DI LAUREA IN INGEGNERIA INFORMATICA E DELL'AUTOMAZIONE

Un approccio Multi-task Deep Learning per la valutazione del Covid-19 nelle ecografie polmonari

Multi-task Deep Learning for lung ultrasound Covid-19 detection

Candidato:
Lorenzo Federici

Relatore:
Prof. Primo Zingaretti

Correlatori:
Maria Chiara Fiorentino, PhD
Riccardo Rosati, PhD

Anno Accademico 2022-2023

UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA
CORSO DI LAUREA IN INGEGNERIA INFORMATICA E DELL'AUTOMAZIONE
Via Brecce Bianche – 60131 Ancona (AN), Italy

A mio Fratello, ai miei genitori e a Kisela

Ringraziamenti

Vorrei esprimere la mia profonda gratitudine a mio fratello Gianluca e ai miei genitori, Federico e Cinzia. Il loro amore incondizionato, il costante sostegno e la fiducia in me sono stati la forza motivante dietro ogni mio sforzo. Grazie per avermi insegnato a affrontare le sfide e i problemi sempre con un sorriso. Non posso che essere grato per avere una famiglia così straordinaria al mio fianco, pronta a sostenermi in ogni fase della mia vita. L'impatto che avete avuto sulla mia vita va ben oltre le parole e rimarrà per sempre nel mio cuore.

Un ringraziamento speciale va al Prof. Primo Zingaretti, per avermi concesso l'opportunità di sviluppare la mia tesi in linea con le mie passioni accademiche. Le sue preziose osservazioni, hanno notevolmente contribuito ad arricchire la qualità del mio lavoro di ricerca. Un profondo ringraziamento va anche a Riccardo e Maria Chiara, che hanno svolto un ruolo fondamentale come correlatori di questa tesi. Il loro prezioso sostegno non solo ha costantemente elevato la qualità di questo lavoro, ma ha anche contribuito in modo significativo a migliorare le mie capacità e competenze. La loro guida attenta e i loro feedback costruttivi hanno rappresentato una fonte di ispirazione, permettendomi di crescere e migliorare.

Carissimi Edoardo e Andrian, è impossibile non riconoscere il vostro ruolo fondamentale nel mio successo accademico, e vorrei dedicare questo momento per esprimervi la mia gratitudine dal profondo del cuore. Grazie per aver condiviso con me gioie e dolori di questo percorso, per essere stati sempre presenti nei miei e nei NOSTRI momenti critici e per aver condiviso le NOSTRE vittorie con entusiasmo e gioia. La vostra amicizia è stata una luce costante nel mio percorso universitario, un faro che mi ha guidato attraverso le NOSTRE sfide e le NOSTRE vittorie di questi anni. La convivenza con voi è stata un regalo prezioso, i ricordi indelebili che abbiamo creato insieme saranno per sempre impressi nel mio cuore. Siete i miei sparring partner, i compagni di viaggio che hanno contribuito a plasmare la mia determinazione e a farmi crescere come individuo. Grazie per aver creduto in me e per avermi ispirato a dare sempre il massimo. Spero tanto di continuare a lavorare insieme.

Vorrei inoltre dedicare qualche parola a Lorenzo. Sin dalle elementari, sei stato al mio fianco, un amico prezioso che ha condiviso con me gioie e avventure. Grazie per avermi spronato ad uscire quando i pensieri pesavano troppo e grazie per avermi regalato momenti di leggerezza e spensieratezza.

Infine, ma non meno importante, vorrei ringraziare Kisela, la mia fidanzata. Nonostante la distanza e le innumerevoli sfide che affronti ogni giorno sei sempre stata al mio fianco con il tuo sorriso, anche quando le mie ansie rendevano difficile per me fare altrettanto. Grazie per aver abbattuto i miei muri e per spingermi costantemente a essere la mia versione migliore. Udhëtimi ynë së bashku vazhdon dhe mezi pres të fillojmë aventura të reja përkrah njeri tjetrit dhe të të shikoj që vazhdon të shkëlqesh. Faleminderit zemra ime.

Ancona, Febbraio 2024

Lorenzo Federici

Abstract

Deep Learning (DL) has established itself as an extremely effective method in the field of medical imaging. This interest has grown significantly during the recent COVID-19 pandemic, pushing for further research and advanced solutions focused on improving early diagnosis. While most research has concentrated on computed tomography (CT) scans, this study is dedicated to the application of DL techniques in lung ultrasound (LUS), a safer and more accessible imaging methodology compared to CT. The paper introduces an innovative multitask learning architecture, specifically designed for the classification and segmentation of lung damage caused by COVID-19. This architecture constitutes a novel and unexplored approach in the current state of the art. Its effectiveness has been validated through the analysis of images from the ICLUS dataset, provided by the ULTRa Laboratory of Ultrasound at the University of Trento, yielding promising results. In classification, the model achieved an accuracy of 66%, exceeding the 63% obtained with single-task approaches. For segmentation, the model achieved a Dice similarity coefficient of 49%, a slight improvement over the 47% achieved with traditional techniques. These results underscore the complexity in analyzing lung ultrasound images, characterized by non-distinct geometric patterns, which present significant challenges in classification but especially in segmentation. It was observed that a multitask approach, with careful parameter tuning and an appropriate choice of loss function, can enhance the performance of individual tasks, even in contexts characterized by highly variable and complex images like those analyzed. It turned out that sharing features between classification and segmentation can enhance predictions and strengthen the overall performance of the model. This, in turn, allows for a more accurate and clearly visible diagnosis of patients' clinical conditions.

Sommario

Il Deep Learning (DL) si è affermato come un metodo estremamente efficace nel campo dell'imaging medico. Questo interesse è cresciuto notevolmente durante la recente pandemia di COVID-19, motivando una crescente attività di ricerca e sviluppo di soluzioni avanzate focalizzate sul miglioramento della diagnosi precoce. Mentre la maggior parte degli studi si sono concentrati sulle tomografie computerizzate (TC), questo elaborato si dedica all'applicazione delle tecniche di DL nelle ecografie polmonari (LUS); una metodologia di imaging più sicura e accessibile rispetto alla TC. L'elaborato introduce un'innovativa architettura di apprendimento multitask, progettata per la classificazione e la segmentazione dei danni polmonari dovuti al COVID-19. L'architettura proposta costituisce un approccio innovativo e inesplorato nel panorama attuale dello stato dell'arte. La sua efficacia è stata validata attraverso l'analisi di immagini provenienti dal dataset ICLUS, fornito dal Laboratorio di Ecografia dell'Università di Trento (ULTRa), ottenendo risultati promettenti. Nella classificazione, il modello ha raggiunto un'accuratezza del 66%, superiore al 63% ottenuto con approcci basati su singolo task. Per quanto riguarda la segmentazione, il modello ha conseguito un indice di somiglianza (Dice) del 49%, un leggero miglioramento rispetto al 47% ottenuto con tecniche tradizionali. Questi risultati sottolineano la complessità nell'analisi delle immagini ecografiche polmonari, caratterizzate da pattern geometrici non distinti, che presentano sfide significative nella classificazione ma soprattutto nella segmentazione. Si è osservato che un approccio multitask, con un'attenta regolazione dei parametri e una scelta appropriata della funzione di perdita, può migliorare le prestazioni dei singoli task, anche in contesti caratterizzati da immagini altamente variabili e complesse come quelle analizzate. È emerso che la condivisione delle features tra classificazione e segmentazione può potenziare le predizioni e rafforzare le prestazioni complessive del modello. Questo, a sua volta, permette di fornire una diagnosi più accurata e chiaramente visibile delle condizioni cliniche dei pazienti.

Indice

1	Introduzione	1
1.1	Medical Imaging	2
1.1.1	Ecografia polmonare (LUS)	3
1.2	Deep Learning	5
1.2.1	Multitask	6
1.3	Obiettivo Tesi	7
1.4	Struttura della Tesi	7
2	Related Work	9
2.1	Segmentazione	10
2.2	Classificazione	10
2.3	Multitask	12
3	Materiali e Metodi	15
3.1	Dataset	15
3.1.1	Dataset per la segmentazione	18
3.1.2	Preprocessing del Dataset	19
3.1.3	Sfide del dataset	22
3.2	Reti neurali convoluzionali - CNN	24
3.2.1	UNet	26
3.2.2	Unet++	28
3.2.3	ResNet	29
3.3	Multitask Learning	30
3.4	Procedura Sperimentale	33
3.4.1	Funzioni di loss	33
3.4.2	Metriche	35
3.4.3	Framework sperimentale	36
3.4.4	Grid Search	37
3.4.5	Data Augmentation	38
4	Risultati e Discussioni	41
4.1	Classificazione	42
4.1.1	Resnet18	42
4.1.2	Resnet50	43
4.2	Segmentazione	45

Indice

4.3	Multitask learning	49
4.3.1	ResUnet18 e ResUnet18++	49
4.3.2	ResUnet50 e ResUnet50++	60
4.3.3	Studio della Loss	61
4.3.4	Confronti	63
5	Conclusioni	67
5.1	Sviluppi futuri	68

Elenco delle figure

1.1	Differenze tra tipologie di sonde ecografiche	3
1.2	Confronto schemi sigle-task e multitask	6
2.1	Architettura proposta da Roy et al.	11
2.2	Integrazione di conoscenza del dominio di Frank et al.	11
2.3	Architettura multitask proposta da Y. Zhou et al.	12
2.4	Architettura multitask proposta da J. Chowdary et al.	13
3.1	Struttura del dataset ICLUS	16
3.2	Impatto della variazione degli score sulle immagini ecografiche	17
3.3	Organizzazione del dataset	18
3.4	Contenuto file xls	18
3.5	Struttura del dataset per la segmentazione	19
3.6	Preprocessing dataset	20
3.7	Distribuzioni delle classi nei tre set	21
3.8	Distribuzioni dei centri medici nei tre set	22
3.9	Esempio incosistenza maschere di verità tra frame dello stesso video	23
3.10	Esempio incosistenza maschere in due pazienti	23
3.11	Esempio differenza di forma delle maschere nei frame dello stesso video	24
3.12	Architettura classica di una CNN	24
3.13	Esempio di convoluzione: filtro 3x3 che passa sopra ad un'immagine producendo una nuova matrice	25
3.14	Differenza tra max pooling e average pooling	26
3.15	Architettura UNet	27
3.16	Architettura UNet++	28
3.17	Blocco residuo	29
3.18	Architettura ResNet18	29
3.19	Rappresentazione grafica blocchi residui	30
3.20	Architettura ResUnet per Multitask	31
3.21	Architettura ResUnet++ per Multitask	33
3.22	Rappresentazione grafica metrica Dice	36
3.23	Processo grafico del Framework sperimentale	37
3.24	Data Augmentation	39
4.1	Matrice di confusione ottenuta con una Resent18 per la classificazione	43
4.2	Matrice di confusione ottenuta con una Resent50 per la classificazione	45

Elenco delle figure

4.3	Esempio segmentazione sbagliata con accuratezza alta	46
4.4	Esempio predizioni segmentazione	48
4.5	Matrice di confusione ResUnet18 con focus alla classificazione	51
4.6	Predizioni ResUnet18 con focus alla classificazione	52
4.7	Matrice di confusione ResUnet18++	54
4.8	Predizioni ResUnet18++ su immagini complesse	55
4.9	Andamento predizioni nell'addestramento di ResUnet18++ epoca 0 e 5	57
4.10	Andamento predizioni nell'addestramento di ResUnet18++ epoca 10 e 15	58
4.11	Andamento predizioni nell'addestramento di ResUnet18++ epoca 20 e 25	59
4.12	Confronto matrici di confusione tra approcci a singolo task e multitask	65

Elenco delle tabelle

3.1	Iperparametri	38
3.2	Parametri Data Augmentation On the fly	38
4.1	Risultati classificazione: Resnet18	42
4.2	Risultati classificazione: Resnet50	44
4.3	Risultati segmentazione: Unet con backbone Resnet18	45
4.4	Risultati segmentazione: Unet++ con backbone Resnet18	46
4.5	Risultati segmentazione: Unet con backbone Resnet50	49
4.6	Risultati segmentazione: Unet++ con backbone Resnet50	49
4.7	Risultati multitask learning - focus classificazione: Unet con backbone Resnet18	50
4.8	Risultati multitask learning - focus segmentazione: Unet con backbone Resnet18	50
4.9	Risultati multitask learning - focus classificazione: Unet++ con backbone Resnet18	53
4.10	Risultati multitask learning - focus segmentazione: Unet++ con backbone Resnet18	53
4.11	Risultati multitask learning: Unet con backbone Resnet50	60
4.12	Risultati multitask learning: Unet con backbone Resnet50++	60
4.13	Iperparametri migliori per l'addestramento di ResUnet18 e ResUnet18++	61
4.14	Studio del peso lambda nella loss del modello ResUnet18	62
4.15	Studio del peso lambda nella loss del modello ResUnet18++	62
4.16	Iperparametri migliori per l'addestramento di ResUnet50 e ResUnet50++	62
4.17	Studio del peso lambda nella loss del modello ResUnet50	63
4.18	Studio del peso lambda nella loss del modello ResUnet50++	63
4.19	Confronto risultati ResUnet18, Resnet18 e Unet	64
4.20	Confronto risultati ResUnet18++, Resnet18 e Unet++	64

Capitolo 1

Introduzione

Nel contesto di un panorama globale costantemente minacciato dalla nascita di nuove patologie infettive, la garanzia di un'ampia disponibilità di test diagnostici assume un ruolo cruciale nella gestione efficace della diffusione dei virus. Un esempio paradigmatico è stato la diffusione su scala planetaria del virus SARS-CoV-2, il quale ha notevolmente compromesso la capacità di effettuare test diagnostici, a causa della diffusa carenza di apparecchiature mediche a livello internazionale. La pandemia di COVID-19 ha rappresentato dunque, una sfida senza precedenti per il sistema sanitario globale, mettendo in luce e aggravando diverse carenze e problemi già esistenti. La pandemia ha anche esercitato una pressione immensa sui sistemi sanitari [1] aumentando la necessità di personale sanitario adeguato, ventilatori, test diagnostici, protezione individuale (PPE), materiali per dialisi e farmaci [2], influenzando negativamente la conduzione delle sperimentazioni cliniche. Questa situazione ha sottolineato la vulnerabilità dei sistemi sanitari e la necessità di una migliore preparazione e risposta in caso di future emergenze sanitarie ma, al contempo, ha stimolato la comunità scientifica a sviluppare soluzioni sempre più innovative. Un esempio significativo è la creazione di maschere filtranti riutilizzabili stampate in 3D, fondamentali per superare le carenze di dispositivi di protezione individuale (PPE) [3]. Contemporaneamente, si è verificata una svolta significativa nell'uso dell'intelligenza artificiale nelle applicazioni diagnostiche. In particolare, la letteratura scientifica evidenzia l'utilità dell'analisi automatica nell'imaging medico, attraverso l'impiego di metodi di Machine Learning (ML) e Deep Learning (DL). L'adozione di queste tecnologie rappresenta un sostegno fondamentale e prezioso per gli operatori sanitari, facilitandogli l'identificazione di pattern e biomarcatori associati alle specifiche condizioni mediche in esame. Questo contribuisce ad importanti progressi, quali la riduzione dei tempi di analisi, l'incremento della precisione diagnostica, l'integrazione di dati multimodali e la diminuzione della necessità di esami ripetitivi e non essenziali. Inizialmente, l'uso di queste applicazioni si è concentrato principalmente sulla tomografia computerizzata del torace (TC), un metodo di imaging medico ampiamente utilizzato per esaminare le condizioni polmonari. Tuttavia, la TC presenta alcuni notevoli svantaggi. In primo luogo, l'esposizione a radiazioni rappresenta un fattore di rischio significativo, specialmente in caso di uso frequente. In secondo luogo, la disponibilità limitata di questo strumento in

determinate aree geografiche ne compromette l'efficacia e l'accessibilità come metodo diagnostico a livello globale. Questi limiti hanno stimolato la ricerca di alternative più sicure e accessibili nell'ambito diagnostico. Tra queste, l'ecografia polmonare (LUS) è emersa come una tecnica particolarmente promettente grazie alla completa assenza di radiazioni e alla sua portabilità che consente di eseguire esami direttamente a letto del paziente[4][5]. In questo capitolo introduttivo, verrà esplorato il contesto in cui si inserisce il lavoro di tesi, delineando il panorama emergente della diagnostica per immagini in un'epoca segnata dalla pandemia di COVID-19 e sottolineando il potenziale rivoluzionario delle tecnologie IA nell'ambito della LUS. L'obiettivo della tesi è dunque quello di studiare se un approccio multitask possa essere più efficace di un approccio a singolo task. Per fare ciò, la tesi presenta l'implementazione sia di task singoli, quali la classificazione e la segmentazione, sia del multitask, permettendo così un confronto diretto tra queste due metodologie. Attraverso questa analisi comparativa, si mira a valutare le potenzialità e le limitazioni dell'approccio multiplo, con l'obiettivo di fornire un contributo significativo al campo della diagnostica per immagini ecografiche del polmone.

1.1 Medical Imaging

Il medical imaging è una disciplina fondamentale nell'ambito della diagnostica e del monitoraggio delle condizioni di salute umana. Essa consiste nell'uso di tecniche avanzate per acquisire, analizzare e interpretare immagini del corpo umano al fine di diagnosticare patologie, guidare trattamenti e monitorare l'efficacia delle terapie.

Nel contesto specifico dei polmoni, questa disciplina gioca un ruolo cruciale nella diagnosi precoce e nella gestione delle malattie polmonari, tra cui infezioni, tumori e patologie croniche. Diverse tecniche di imaging possono essere applicate per ottenere informazioni dettagliate sulla struttura e sulla funzione dei polmoni come:

1. Radiografia del torace: tecnica più comune, fornisce una visione bidimensionale della struttura polmonare. È utilizzata per identificare anomalie strutturali, come masse o infiltrati. È particolarmente utile per la diagnosi di lesioni più piccole, la valutazione delle vie aeree e la caratterizzazione delle lesioni sospette.
2. Tomografia computerizzata (TC): fornisce immagini tridimensionali ad alta risoluzione dei polmoni.
3. Risonanza magnetica (RM): meno comune rispetto alla TC ma può offrire dettagli senza l'uso di radiazioni ionizzanti. È spesso utilizzata per la valutazione di specifiche condizioni, come le malattie interstiziali polmonari.
4. Imaging nucleare: include tecniche come la scintigrafia polmonare e la tomografia ad emissione di positroni (PET). La scintigrafia è utile per la valutazione della ventilazione e della perfusione polmonare, mentre la PET può essere utilizzata per identificare l'attività metabolica in tessuti polmonari.

5. Ecografia polmonare (LUS): tecnica che utilizza ultrasuoni per generare immagini dei polmoni. È una tecnica in tempo reale caratterizzata da facile accessibilità, costi contenuti e sicurezza. Questa metodologia è comunemente adottata in contesti di point-of-care per la valutazione dei disturbi respiratori.

1.1.1 Ecografia polmonare (LUS)

L'ecografia polmonare è una tecnica di imaging non invasiva che utilizza gli ultrasuoni per visualizzare e valutare le strutture polmonari e la pleura. Come anticipato precedentemente, rispetto ad altre tecniche di imaging come la radiografia toracica o la tomografia computerizzata (TC), ha il vantaggio di essere priva di radiazioni, relativamente economica, e può essere eseguita a letto del paziente[4]. Attualmente, grazie all'innovazione tecnologica, la sonda ecografica può essere connessa a computer portatili, smartphone o tablet[5]. Questa versatilità la rende estremamente preziosa in contesti urgenti o nelle unità di terapia intensiva.

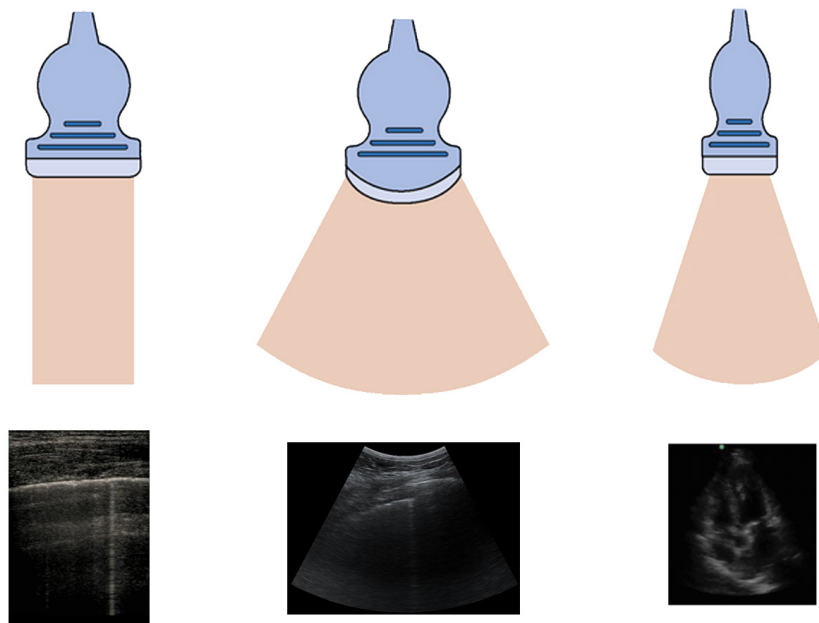


Figura 1.1: Differenze tra tipologie di sonde ecografiche

Nella tecnologia ecografica, si assiste all'impiego di onde sonore ad alta frequenza, tipicamente variabili tra 1 e 18 MHz. Queste onde sono generate da trasduttori o sonde, che possono essere configurati in diverse forme, come lineari, convessi o settoriali (Figura 1.1). La scelta della forma del trasduttore è fondamentale poiché influenza la risoluzione dell'immagine e la profondità di penetrazione delle onde sonore. Una volta emesse dalla sonda, le onde ultrasoniche attraversano il corpo del paziente e interagiscono con vari tessuti. Un aspetto cruciale è la velocità di propagazione delle onde, che varia in base alla densità e all'elasticità dei tessuti. Per esempio, in

tessuti densi e compatti come l'osso, si propagano a una velocità maggiore rispetto a quanto non facciano in tessuti più morbidi. Quando le onde ultrasoniche incontrano un'interfaccia tra tessuti di diversa densità, si verifica un fenomeno di riflessione: parte dell'onda sonora viene riflessa indietro verso il trasduttore, generando un eco. Questi echi sono fondamentali per l'ecografia, in quanto vengono catturati dal trasduttore e trasformati poi in segnali elettrici. La sonda, quindi, oltre ad essere un generatore di onde sonore è anche un ricevitore. I segnali elettrici ricevuti sono poi inviati a un computer che attraverso complessi algoritmi di elaborazione, vengono convertiti in immagini che rappresentano una mappatura spaziale dei tessuti esaminati, fornendo all'esperto di dominio una visione dettagliata della struttura intera del polmone. L'intensità e il tempo di ritorno degli echi vengono analizzati per determinare la posizione e le caratteristiche dei tessuti, consentendo di distinguere strutture diverse come fluidi, tessuti morbidi e ossa. È importante sottolineare che l'ecografia polmonare si distingue per la sua capacità di fornire immagini in tempo reale in grado di offrire informazioni grazie a pattern fondamentali per la diagnosi di varie condizioni polmonari. Tra questi, alcuni dei più rilevanti sono:

1. Linee A: linee iperecogene orizzontali che rappresentano un artefatto da riverbero, tipico del contenuto aereo del polmone. La loro presenza è normale e indica che l'onda ultrasonica sta viaggiando attraverso l'aria.
2. Linee B: linee iperecogene perpendicolari alla linea pleurica, mobili con le escursioni respiratorie, sono indicative di varie condizioni patologiche.
3. Lung Pulse: movimento ritmico della pleura sincronizzato con il ritmo cardiaco. Questo fenomeno è più evidente nelle aree adiacenti all'area cardiaca e può essere indicativo di specifiche condizioni polmonari.

L'identificazione accurata di questi pattern permette una valutazione accurata delle condizioni polmonari come:

1. Versamento Pleurico: L'accumulo di liquido tra i polmoni e la parete toracica appare come un'area scura (ipoecogena) nelle immagini ecografiche, in quanto i fluidi trasmettono gli ultrasuoni meglio dell'aria.
2. Consolidamenti: Le aree di polmone infettato o solidificato riflettono più ultrasuoni rispetto al polmone normale, apparendo come aree più luminose (iperecogene) sullo schermo.
3. Edemi: Anche questi possono alterare il modo in cui il tessuto polmonare riflette gli ultrasuoni.
4. Pneumotorace: La presenza di aria extra alveolare nel cavo pleurico può essere identificata dalla mancanza di movimento tipico del polmone o dalla presenza di linee artefattuali specifiche.

La sua crescente popolarità è dovuta alla sua capacità di fornire informazioni immediate e dettagliate sullo stato del polmone, specialmente in situazioni in cui altre modalità di imaging potrebbero non essere disponibili o praticabili. Tuttavia, l'efficacia dell'ecografia polmonare dipende in larga misura dall'esperienza e dalla formazione dell'operatore, dato che l'interpretazione delle immagini ecografiche può essere difficile a causa della natura complessa del tessuto polmonare e della presenza di aria. In questo contesto, l'impiego di tecniche di Deep Learning diventa cruciale, offrendo supporto nell'analisi e interpretazione delle immagini, aiutando così a superare le limitazioni legate alla variabilità delle competenze umane.

1.2 Deep Learning

L'era moderna dell'intelligenza artificiale (AI) è stata profondamente segnata dall'emergere del Deep Learning (DL), un sottoinsieme del Machine Learning (ML) che ha introdotto metodi e applicazioni all'avanguardia che mirano a potenziare l'automazione riducendo al minimo la necessità di intervento umano. Il Deep Learning rappresenta un avanzamento specifico all'interno del ML, caratterizzato dall'uso di reti neurali profonde, ovvero reti composte da tre o più strati. Queste, sono progettate per emulare le funzioni del cervello umano, consentendo ai sistemi di apprendere ed evolvere attraverso l'analisi di grandi volumi di dati. Mentre una rete neurale con un solo strato può eseguire previsioni di base, l'aggiunta di strati nascosti supplementari incrementa notevolmente la capacità del sistema di perfezionare e affinare la sua precisione. In questo studio si pone una particolare attenzione alle applicazioni del DL nel settore medico, esplorando in modo dettagliato i due task principali del DL (classificazione e segmentazione) e la loro collaborazione (multitask).

- **Classificazione:** nell'elaborazione delle immagini e nell'apprendimento automatico è il processo di identificazione e categorizzazione di un'immagine o di una parte di essa in una classe specifica. In generale, la classificazione si riferisce all'assegnazione di una label a un'intera immagine o a un oggetto all'interno dell'immagine. Il DL ha rivoluzionato il modo in cui le macchine interpretano e categorizzano i dati. Utilizzando complesse reti neurali è possibile identificare e classificare pattern in vasti set di dati con un'accuratezza e velocità, aprendo così nuove strade in numerosi domini, dalla diagnosi medica alla visione artificiale.
- **Segmentazione:** nell'elaborazione delle immagini è il processo di suddivisione di un'immagine in più segmenti o regioni, spesso per isolare o identificare oggetti o confini specifici all'interno dell'immagine. Questo task è fondamentale in numerose applicazioni, come la segmentazione di immagini mediche per l'identificazione di patologie o l'analisi di immagini satellitari per il monitoraggio ambientale.

1.2.1 Multitask

Prima di esplorare il concetto di Multitask Learning (MTL), è essenziale definire cos'è un singolo task. Secondo Kim-Han Thung, un task è generalmente descritto come "l'apprendimento di un output utilizzando una singola sorgente di input" [6] (trad. mia). Se la sorgente di input consiste in una singola feature, l'analisi verrà classificata come univariata, se consiste in più features, in analisi multivariata.

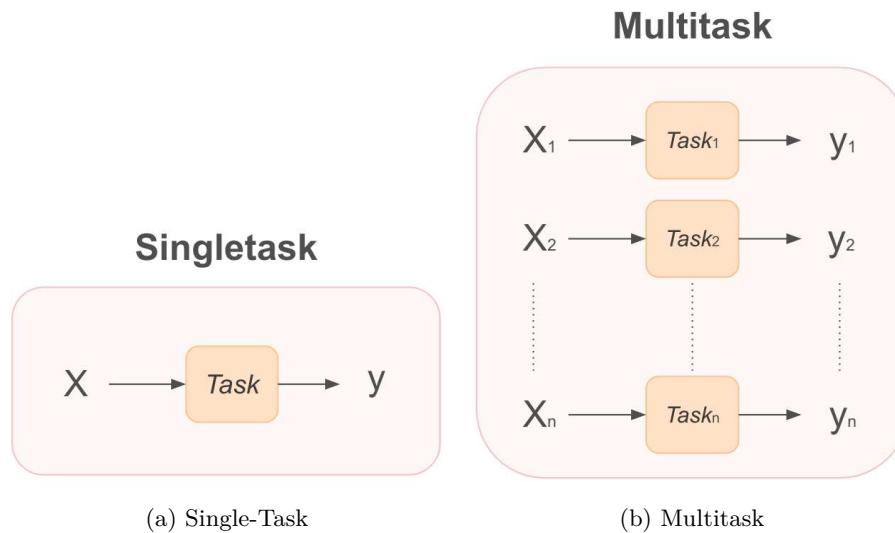


Figura 1.2: Confronto schemi sigle-task e multitask

Passando al Multitask Learning, una descrizione chiave viene fornita da Rich Caruana, il quale lo definisce in questi termini:

«Il Multitask-learning è un approccio di trasferimento induttivo che migliora la generalizzazione tramite l'impiego delle informazioni di dominio presenti nei segnali di addestramento di compiti correlati, usati come bias induttivo. Questo metodo apprende compiti in parallelo utilizzando una rappresentazione condivisa a bassa dimensionalità, permettendo che ciò che si impara in un compito possa facilitare l'apprendimento negli altri» [7]

In sintesi, l'MTL è una tecnica di apprendimento automatico che mira a risolvere più task simultaneamente, ottimizzando l'uso delle informazioni disponibili. Tale approccio è particolarmente utile in scenari caratterizzati da alta dimensionalità dei dati e scarsità di esempi di addestramento come ad esempio, in ambiti come l'analisi di immagini mediche, dove l'etichettatura manuale è onerosa e richiede tempo e competenze specifiche. Infatti, senza una conoscenza pregressa, gli algoritmi di apprendimento automatico possono riscontrare difficoltà nell'ottimizzare con precisione i parametri del modello, portando a prestazioni sub-ottimali. L'MTL quindi affronta queste sfide, facilitando una migliore stima dei parametri attraverso la condivisione di informazioni tra i vari task correlati.

1.3 Obiettivo Tesi

Questa tesi mira ad esplorare un approccio multitask fondato sul Deep Learning nel contesto dell'Ecografia Polmonare, una tecnica di imaging con numerosi vantaggi ma complessa da interpretare. L'obiettivo principale è indagare se condividere le features tra una rete di classificazione e una rete di segmentazione possa non solo migliorare le prestazioni individuali di ciascun task, ma anche facilitare una comprensione più dettagliata e completa delle diagnosi, affrontando efficacemente la complessità interpretativa delle immagini ecografiche polmonari. Questo approccio potrebbe apportare notevoli miglioramenti nella gestione e nella successiva cura dei pazienti, specialmente in situazioni critiche come la pandemia da COVID-19. Tale metodologia ambisce ad apportare un contributo alla letteratura scientifica, introducendo una nuova prospettiva nell'analisi delle immagini ottenute tramite ecografia polmonare (LUS).

1.4 Struttura della Tesi

La tesi si articola nei seguenti capitoli, ciascuno dedicato a un aspetto specifico del lavoro di ricerca:

- **Related Work:** panoramica esaustiva dei lavori rilevanti nel dominio di interesse, si concentra su tutti i contributi significativi realizzati nei settori della segmentazione, della classificazione e del multitask nell'ambito dell'imaging medico. Attraverso questa disamina, vengono identificate le fondamenta teoriche e pratiche che hanno ispirato e sostenuto lo sviluppo dell'intero lavoro di tesi.
- **Materiali e Metodi:** esplorazione e discussione delle strategie selezionate per affrontare l'obiettivo posto. Viene data particolare attenzione alle architetture dei modelli neurali impiegati, dettagliando le loro specifiche e il modo in cui sono stati adattati o sviluppati per il progetto di ricerca. Inoltre viene presentato il dataset su cui questa tesi ha fatto riferimento.
- **Esperimenti e Risultati:** vengono presentati i risultati ottenuti dagli esperimenti condotti. Oltre alla semplice esposizione dei dati, vi è un'approfondita analisi di questi risultati, includendo riflessioni critiche e interpretazioni. L'obiettivo è di valutare l'efficacia delle metodologie applicate e di discutere le implicazioni dei risultati nel contesto più ampio del dominio di ricerca.
- **Conclusioni:** Capitolo finale riassume i punti chiave emersi dalla ricerca, riflettendo sui successi raggiunti e sulle limitazioni incontrate. Vengono inoltre proposte direzioni future per la ricerca in questo dominio, suggerendo come il lavoro svolto possa essere esteso o migliorato.

Capitolo 2

Related Work

In questo capitolo si approfondisce la letteratura scientifica relativa all'applicazione di tecniche di apprendimento automatico nelle immagini ecografiche polmonari (LUS), ponendo particolare enfasi sul contesto della pandemia COVID-19. Si osserva che, in questo specifico dominio (immagini LUS), prevale un orientamento verso solo approcci di single task, concentrati specificamente sulla segmentazione o sulla classificazione delle immagini. La segmentazione in LUS è cruciale per l'identificazione accurata delle aree di interesse, facilitando così diagnosi più precise. Allo stesso tempo, la classificazione assume un ruolo essenziale nella distinzione tra diverse condizioni patologiche, inclusi i casi di COVID-19. Nonostante l'importanza di entrambi questi aspetti, si rileva una mancanza nell'attuale panorama di ricerche per quanto riguarda l'integrazione sinergica dei due task sopracitati in un unico framework di multitask-learning. Tale lacuna rappresenta un ambito di ricerca promettente e poco esplorato, nel quale questa tesi intende inserirsi.

Di conseguenza, questo capitolo esplora lo stato dell'arte riguardante i due task cruciali nell'elaborazione delle immagini LUS: la segmentazione e la classificazione. Attraverso un esame approfondito dei lavori esistenti, si mira a stabilire una comprensione solida dei metodi e dei risultati ottenuti finora nei due singoli ambiti. Questo passaggio è fondamentale per costruire una base di conoscenza robusta e aggiornata, essenziale per l'indirizzo di ricerca proposto in questa tesi. Successivamente, l'attenzione si sposta verso l'analisi di ricerche significative nel multitask learning, estendendosi oltre il dominio specifico delle immagini LUS, vista la mancanza di quest'ultima. L'obiettivo di questa sezione è di esplorare come approcci multitask siano stati implementati e quali benefici abbiano portato in altri ambiti dell'apprendimento automatico su immagini ecografiche. Questa indagine aiuterà a delineare le strategie adottate e i vantaggi derivanti dall'utilizzo di tali approcci, fornendo così un contesto più ampio che motiva e supporta l'idea di applicare metodologie simili nel trattamento delle immagini ecografiche polmonari, specialmente in un periodo critico come quello della pandemia di COVID-19.

L'insieme di queste analisi contribuirà a definire un quadro chiaro dello stato attuale della ricerca e a identificare potenziali percorsi innovativi per il trattamento delle immagini LUS, sottolineando l'importanza e la rilevanza di esplorare nuove frontiere attraverso l'applicazione di approcci multitask in questo campo.

2.1 Segmentazione

Il Deep Learning (DL) ha dimostrato un notevole successo in vari ambiti della visione artificiale, spaziando dal riconoscimento e rilevamento di oggetti fino alla segmentazione semantica. Questi traguardi hanno generato un interesse sempre più marcato nell'utilizzo della DL soprattutto in ambito medico. Il DL infatti offre strumenti avanzati per il supporto decisionale e diagnostico agli specialisti del settore, migliorando significativamente l'accuratezza e l'efficienza nelle pratiche mediche.

Lo sviluppo in questo ambito ha avuto inizio con i lavori pionieristici di Ronneberger et al. nel 2015, i quali hanno utilizzato la rete neurale UNet per la segmentazione di immagini biomediche[8]. In seguito, nel 2017, Rajpurkar et al. hanno contribuito notevolmente con il loro lavoro su CheXNet[9], una rete neurale progettata per la rilevazione della polmonite tramite radiografie del torace. Questi studi hanno sottolineato come, grazie all'accesso a dati sufficienti, il DL possa svolgere un ruolo fondamentale nell'assistenza medica e nell'automazione delle diagnosi preliminari, evidenziando il potenziale di superare le capacità umane in specifici compiti diagnostici.

Tuttavia, l'applicazione del DL nelle immagini ecografiche comporta sfide uniche, dovute alla loro bassa qualità e alla loro alta variabilità, donandogli una difficile interpretazione, come evidenziato nella sezione Ecografia polmonare (LUS) e da Liu et al. nel 2019[10]. Questa complessità è ulteriormente amplificata quando si considera la natura intricata del polmone come oggetto di studio, rendendo l'applicazione della segmentazione in questo contesto molto complessa e poco esplorata nello stato dell'arte. Di conseguenza, ci sono pochi studi che affrontano con successo questa sfida specifica. Recentemente, nel 2020, Roy et al.[11] hanno condotto una ricerca sull'uso del DL nell'analisi delle immagini LUS per la diagnosi assistita di COVID-19. Utilizzando il dataset ICLUS, protagonista anche di questa tesi, hanno sviluppato un metodo per prevedere maschere di segmentazione sfruttando una gamma di architetture di rete ampiamente note in letteratura, tra cui UNet, UNet++ e DeepLab v3, per effettuare la segmentazione delle immagini ottenendo risultati promettenti.

2.2 Classificazione

I metodi tradizionali per la classificazione delle immagini ecografiche si affidano all'estrazione manuale di features acustiche, modelli di eco, confini delle lesioni, margini, orientamento, forma e caratteristiche della texture. Un approccio notevole è stato proposto da Okolo, Katsigiannis e Ramzan (2022)[12], dove hanno sviluppato un metodo multi-modale che integra le caratteristiche estratte dalle immagini ultrasound con le informazioni sulla sonda utilizzata. Zhang et al.[13] nel 2020, invece, ha esplorato l'uso di diverse reti neurali, come VGG, ResNet ed EfficientNet, per la classificazione delle immagini ultrasound di polmonite. I loro risultati indicano che il modello EfficientNet supera gli altri in termini di accuratezza, fornendo un

metodo promettente per la classificazione delle immagini in diverse fasi cliniche della polmonite. Nel 2020, il lavoro presentato da Roy et al.[11] ha introdotto l'uso delle Spatial Transformer Networks (STN) in un modello supervised che permette di predire utilizzando etichette basate su frame. A differenza dello studio di G. Soldati et al.[14], che si concentra su un problema di classificazione, Roy et al.[11] ha ampliato l'approccio includendo anche la regressione ordinale. Ciò ha permesso non solo di rilevare la presenza di artefatti legati al COVID-19, ma anche di attribuire un punteggio che indica la gravità della malattia. Inoltre, ha fatto un ulteriore passo avanti rispetto agli studi precedenti proponendo un modello di previsione a livello di video, basandosi su una metodologia frame-by-frame, piuttosto che limitarsi alla classificazione di singoli frame isolati (architettura mostrata in Fig 2.1).

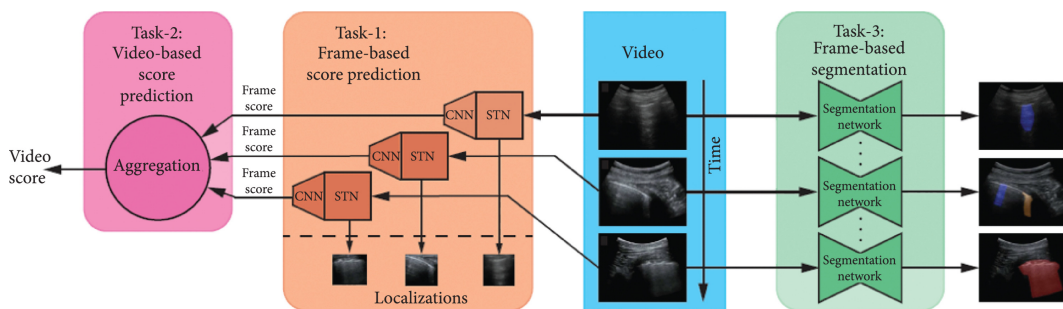


Figura 2.1: Architettura proposta da Roy et al.

Per affrontare la necessità di personale qualificato nell'interpretazione delle ecografie, che limita il loro utilizzo, lo studio pubblicato da Oz Frank et al.[15] ha proposto un approccio innovativo. Questo consiste nello sviluppo di reti neurali per l'interpretazione delle immagini ottenute tramite LUS, allo scopo di ampliarne l'applicazione e l'efficacia. Questo approccio prevede l'integrazione di conoscenze di dominio nelle reti neurali, inclusi aspetti anatomici e artefatti, attraverso l'uso di canali aggiuntivi (come mostrato nella Figura 2.2).

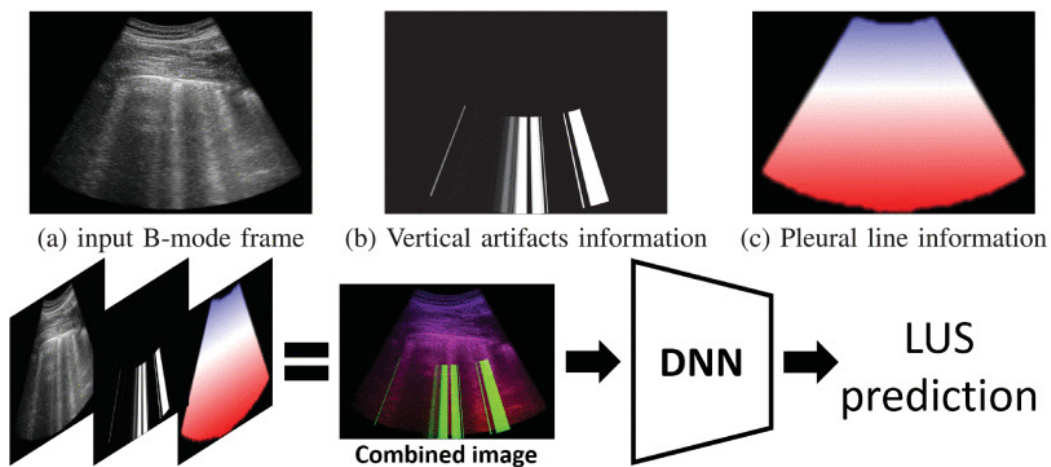


Figura 2.2: Integrazione di conoscenza del dominio di Frank et al.

2.3 Multitask

Come evidenziato nell'introduzione di questo capitolo, il multitask learning non ha ancora trovato applicazione nel campo delle ecografie polmonari. Tuttavia, ha mostrato risultati promettenti in altri ambiti legati all'ecografia, come nello studio delle immagini ecografiche per il cancro al seno. In queste applicazioni si è dimostrato di essere un metodo efficace per migliorare la precisione nella diagnosi. La sinergia tra i vari task di un approccio multitask permette alle diverse reti di collaborare per compensare le sfide poste dall'interpretazione complessa delle immagini. Un esempio significativo è lo studio condotto da Y. Zhou et al.[16], che si concentra sull'applicazione del multitask learning per migliorare sia la segmentazione che la classificazione dei tumori nelle ecografie tridimensionali del seno, note come Automated Breast Ultrasound Systems (ABUS). Lo studio affronta le problematiche legate alle notevoli variazioni morfologiche dei tumori al seno e alle caratteristiche tecniche sfidanti delle immagini ecografiche, come basso contrasto e rapporto segnale/rumore limitato.

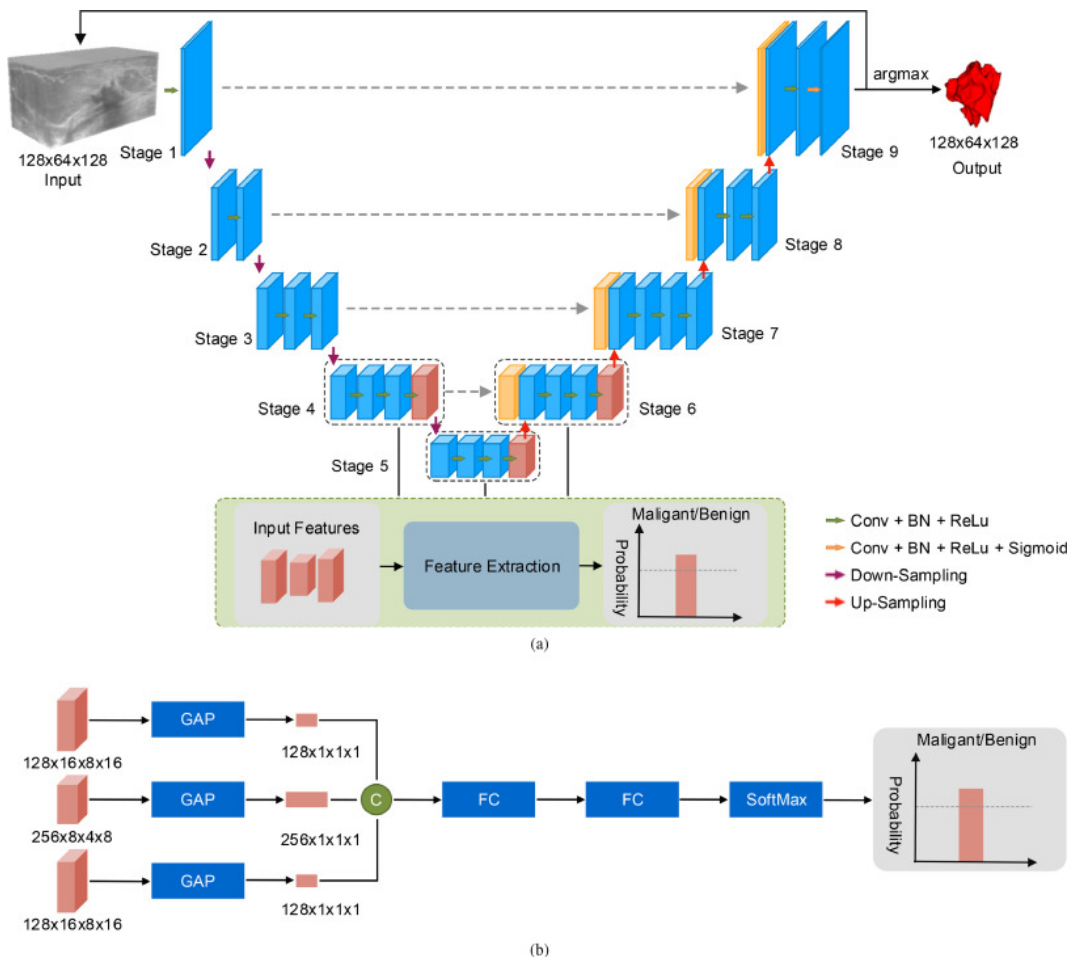


Figura 2.3: Architettura multitask proposta da Y. Zhou et al.

Il modello proposto integra due sottoreti (come si può notare dalla Fig. 2.3):

- UNet classica per la parte della segmentazione
- Rete multi-scala leggera per la classificazione.

Inoltre viene implementato una strategia di addestramento iterativo per affinare le mappe delle features con l'aiuto di mappe di probabilità ottenute dalle iterazioni precedenti. I risultati sperimentali su un dataset clinico di 170 volumi ABUS 3D raccolti da 107 pazienti hanno dimostrato che l'approccio multitasking proposto migliora la segmentazione e la classificazione dei tumori rispetto alle metodologie basate su apprendimento a singolo task. Un altro esempio è la ricerca di J. Chowdary et al.[17], pubblicata nel 2022, che punta a migliorare la diagnosi precoce del cancro al seno e, di conseguenza, a ridurre il tasso di mortalità tramite un approccio innovativo di multitask learning. Questo studio si focalizza sulla segmentazione e classificazione dei tumori al seno attraverso immagini ecografiche in 2D, a differenza del lavoro precedente. Come su Y. Zhou et al. il framework è composto da una Unet per la segmentazione, e un percorso di classificazione densa per l'identificazione dei tumori (vedere Fig. 2.4). Anche in questo caso i risultati hanno dimostrato un miglioramento sia nella precisione che nella sensibilità della segmentazione e classificazione.

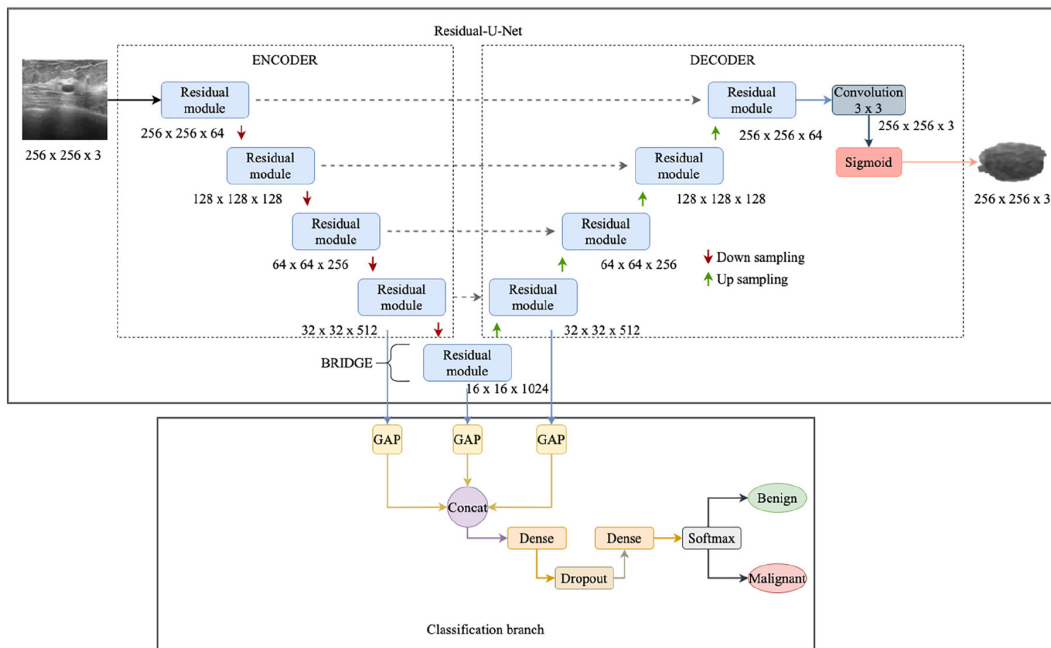


Figura 2.4: Architettura multitask proposta da J. Chowdary et al.

Capitolo 3

Materiali e Metodi

Il presente capitolo mira a delineare in modo dettagliato il processo sperimentale e metodologico adottato.

- Si inizia con una descrizione dettagliata del dataset impiegato, evidenziando le sue caratteristiche principali, la provenienza e la sua rilevanza nel contesto della segmentazione e classificazione. Segue poi un'analisi delle tecniche di preprocessing utilizzate e altre procedure essenziali per la preparazione dei dati al task di segmentazione.
- In seguito, vengono esplorate le specifiche architetture delle reti neurali convoluzionali (CNN) selezionate per questo studio. Questa sezione mette in luce la rilevanza delle CNN nel contesto dell'elaborazione di immagini mediche, descrivendo come queste reti possano catturare le caratteristiche spaziali delle immagini per la segmentazione e classificazione.
- Dopo aver introdotto le CNN, ci si concentra sul modello multitask proposto. Qui, viene dettagliata la struttura del modello, evidenziando come esso sia stato progettato per affrontare simultaneamente le sfide della segmentazione e della classificazione.
- Infine, il capitolo si conclude con una descrizione dettagliata della configurazione del framework utilizzato per implementare il modello, insieme alle varie metriche e funzioni di loss utilizzate nel training della rete multitask proposta.

L'obiettivo di questo capitolo è fornire una base solida e dettagliata per la comprensione degli aspetti tecnici e metodologici che stanno alla base del modello proposto, facilitando così la comprensione dei risultati ottenuti nei capitoli successivi.

3.1 Dataset

L'ecografia polmonare rappresenta uno strumento diagnostico fondamentale per l'esame delle patologie polmonari. Il suo ruolo è particolarmente rilevante nell'analisi della linea pleurica, una sottile struttura rilevabile nell'ecografia che costituisce l'interfaccia tra la pleura viscerale e quella parietale del polmone. Variazioni nella

sua continuità e struttura sono indicatori chiave di diverse condizioni polmonari. L'accurata valutazione di queste variazioni è fondamentale per la diagnosi e il monitoraggio di patologie polmonari, specialmente in contesti critici come la pandemia di COVID-19.

In questo contesto, il dataset ICUS, rilasciato dal Laboratorio di Ecografia dell'Università di Trento (ULTRa), emerge come una risorsa preziosa. Comprende 277 video di ecografia polmonare, con un totale di 58.924 frame, che includono immagini ottenute sia da sonde lineari (13.364 frame) sia convesse (45.560 frame). Questi dati sono stati raccolti in vari ospedali italiani durante la pandemia e comprendono un campione di 35 pazienti, di cui 17 positivi al COVID-19 (49%), 4 sospetti (11%) e 14 sani e asintomatici (40%)(vedere Fig. 3.1).

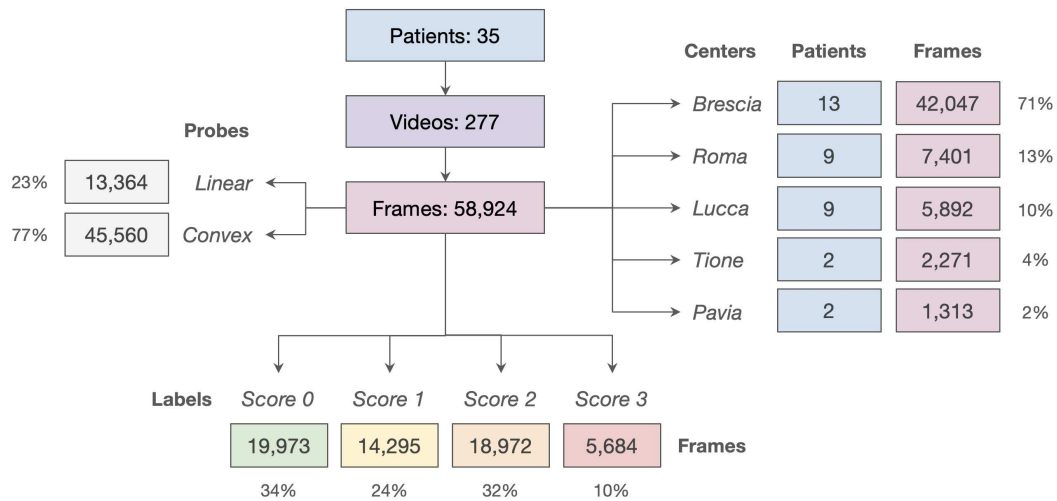


Figura 3.1: Struttura del dataset ICLUS

Come mostrato dalla Fig.3.3 il dataset è strutturato in base ai pazienti, dove per ciascun soggetto e per ogni sessione di diagnosi polmonare, sono presenti tre tipologie di file:

- video originale in formato .avi dove ogni frame può essere circondato da un rettangolo colorato in base all'etichettatura associata.
- matrice di etichette salvata come file .mat. Si presenta come una matrice 2D con dimensioni (3 x F). In questa rappresentazione, il numero 3 corrisponde ai possibili punteggi, che vanno da 0 a 3, mentre "F" rappresenta il numero totale di frame nel video. In pratica, ogni singolo frame è associato a un vettore tridimensionale (3 x 1), il quale indica il punteggio assegnato a quel frame. Quindi, un vettore (0,0,0) suggerisce che il frame è stato classificato con un punteggio di 0, un vettore (1,0,0) indica un punteggio di 1, un vettore (1,1,0) corrisponde a un punteggio di 2, e infine, un frame con un vettore (1,1,1) è classificato con il punteggio massimo di 3.

- versione del video originale in formato .mat. I dati sono organizzati come una matrice 4D ($R \times C \times 3 \times F$), dove "R" sta per il numero di righe e "C" per il numero di colonne presenti in ciascun frame del video. Il numero 3 rappresenta i canali di colore, e "F" indica il totale dei frame contenuti nel video. Il video fornito è stato modificato per eliminare qualsiasi informazione personale del paziente.

I frame sono classificati secondo un sistema di valutazione a quattro livelli che descrive accuratamente lo stato della linea pleurica:

- Score 0: Linea pleurica continua e regolare con artefatti orizzontali.
- Score 1: Linea pleurica indentata con aree bianche verticali sottostanti.
- Score 2: Linea pleurica interrotta con aree consolidate di varie dimensioni e aree bianche associate (polmone bianco).
- Score 3: Polmone bianco esteso con o senza grandi consolidazioni, coprendo almeno il 50% della linea pleurica.

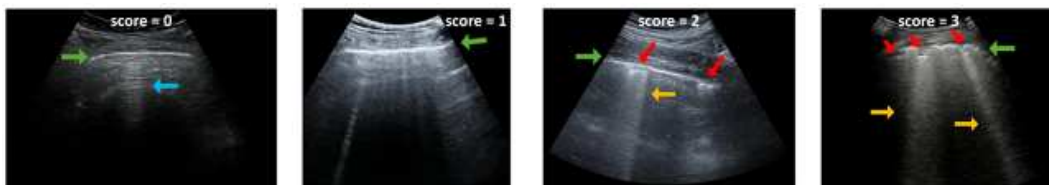


Figura 3.2: Impatto della variazione degli score sulle immagini ecografiche

Per assicurare la massima precisione e oggettività nelle annotazioni, il processo di valutazione dei frame è stato rigorosamente strutturato in quattro fasi distinte:

1. Assegnazione del punteggio a ogni frame da parte di quattro studenti esperti in ecografia.
2. Convalida dei punteggi da un dottorando con competenze in ecografia polmonare (LUS).
3. Seconda convalida da un ingegnere biomedico con oltre 10 anni di esperienza in LUS.
4. Terza convalida e accordo tra medici con oltre 10 anni di esperienza in LUS.

L'approccio meticoloso adottato nel dataset ICUS assicura un elevato standard di qualità e precisione nell'interpretazione delle immagini ecografiche polmonari, rendendolo uno strumento di valore per la ricerca e la pratica clinica.

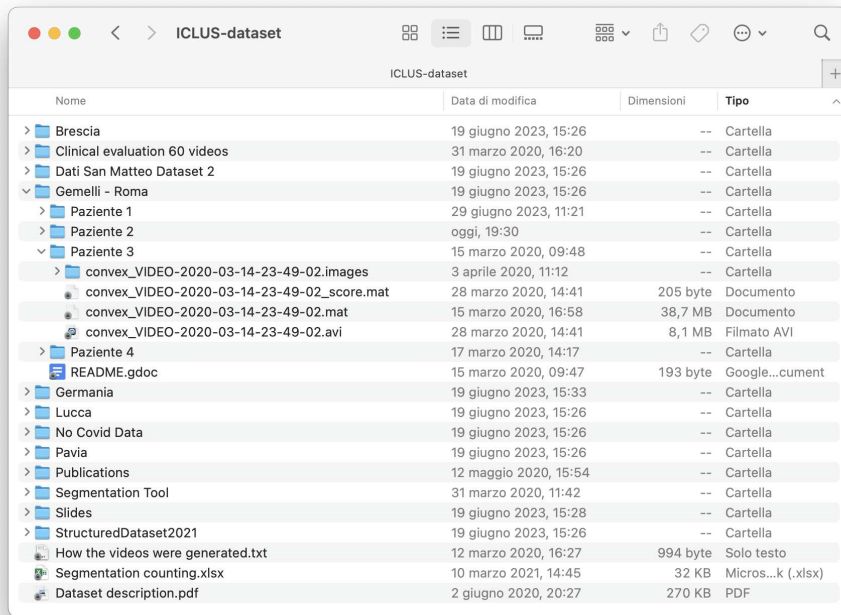


Figura 3.3: Organizzazione del dataset

3.1.1 Dataset per la segmentazione

Il dataset ICLUS, oltre a fornire informazioni necessarie per la classificazione, è predisposto anche per applicazioni di segmentazione. Questo è possibile grazie all'annotazione dei frame con delle maschere di verità realizzate con il supporto di esperti del dominio. È importante notare che non tutti i frame di un video sono segmentati, pertanto la loro identificazione selettiva è un passo essenziale. Il processo di filtraggio è possibile grazie ad un file Excel fornito dall'Università di Trento che contiene tutti i path relativi ai video con annotazioni (vedi Fig. 3.4). Ogni video nel dataset è associato ad una cartella contenente tanti file JSON quanti frame annotati.

Hospital/Folder	Paziente	File Name	Score 0	Score 1	Score 2	Score 3	Number of segmented frames
Brescia	Paziente 9	convex_202003161646210215ABD	0	13	0	0	13
Brescia	Paziente 9	linear_202003161642300208SMP	0	0	5	6	11
Brescia	Paziente 9	linear_202003161643020209SMP	0	0	16	14	30
Gemelli - Roma	Paziente 1	convex_3.1	0	0	13	3	16
Gemelli - Roma	Paziente 1	convex_3.2	0	0	11	0	11
Gemelli - Roma	Paziente 1	convex_3.3	0	0	12	0	12
Gemelli - Roma	Paziente 1	convex_3.4	0	0	11	0	11
Gemelli - Roma	Paziente 1	convex_3.5	0	0	11	0	11
Gemelli - Roma	Paziente 1	linear_2.5	0	0	0	7	7
Gemelli - Roma	Paziente 1	linear_2.6	0	0	11	0	11
Gemelli - Roma	Paziente 1	linear_2.7	7	0	0	0	7
Gemelli - Roma	Paziente 2	convex_movie1	0	0	3	3	6
Gemelli - Roma	Paziente 2	convex_movie2	10	0	0	0	10
Gemelli - Roma	Paziente 2	convex_movie4	0	0	6	0	6
Gemelli - Roma	Paziente 2	convex_movie5	3	1	3	0	7
Gemelli - Roma	Paziente 2	convex_movie6	5	5	0	0	10
Gemelli - Roma	Paziente 3	convex_VIDEO-2020-03-14-23-49-02	0	0	4	0	4
Lucca	Paziente 1	convex_posveryearly	1	0	6	0	7
Lucca	Paziente 1	convex_posveryearly2	8	8	2	0	18
Lucca	Paziente 1	convex_posveryearly3	11	0	2	0	13

Figura 3.4: Contenuto file xls

L'estrazione delle maschere di segmentazione avviene grazie a questi passaggi:

1. **Interpretazione dei Dati JSON:** I file JSON descrivono le forme geometriche

della maschera e le loro posizioni nelle immagini. Quindi vengono analizzati ed interpretati per estrarre informazioni pertinenti come punti, tipi di forma e etichette.

2. **Generazione delle Maschere:** Per ogni forma geometrica descritta nel JSON, viene creata una maschera binaria che evidenzia la posizione all'interno dell'immagine. Le forme possono essere: cerchi, rettangoli, linee e punti.
3. **Segmentazione Semantica e di Istanza:** Le forme vengono classificate in classi semantiche utilizzando etichette specifiche.
4. **Aggregazione delle Maschere:** Infine, tutte le maschere generate per un video vengono aggregate in una struttura complessiva che rappresenta tutte le annotazioni per quell'immagine.

Il dataset risultante sarà un sottocampione dell'originale e conterrà circa 2000 frame quindi il 4% dei dati contenuti all'interno di quello originale. La struttura del dataset per la segmentazione è mostrata in Fig. 3.5.

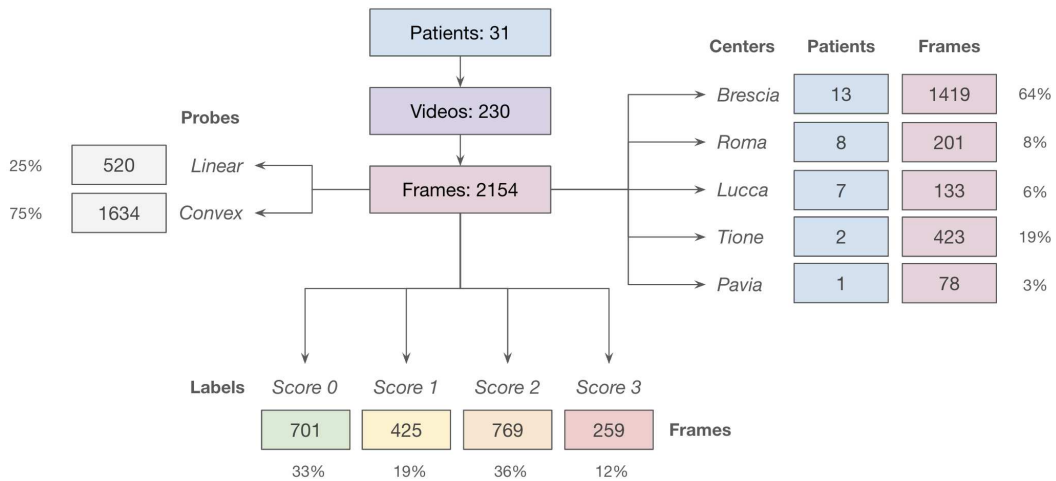


Figura 3.5: Struttura del dataset per la segmentazione

3.1.2 Preprocessing del Dataset

Il preprocessing è una fase fondamentale nella preparazione del dataset per l'addestramento di una rete neurale. L'obiettivo principale è trasformare i dati grezzi in un formato adatto per alimentare un modello di deep learning. Questo processo comporta una serie di operazioni mirate a rendere i dati più strutturati e facilmente accessibili, ottimizzando così il processo di apprendimento della rete. Data la natura strutturale del dataset utilizzato, si è optato per la conversione in formato *HDF5*. Hierarchical Data Format version 5 o *HDF5* è un formato file estremamente versatile utilizzato ampiamente in ambito scientifico, in particolare per la gestione di dataset multidimensionali complessi, come matrici e tabelle, al quale è possibile associare

metadati che forniscono informazioni aggiuntive sul contenuto dei dati. *HDF5* è noto per la sua capacità di gestione di grandi quantità di informazioni organizzate gerarchicamente. Questa capacità di archiviazione è fondamentale per mantenere la tracciabilità e la comprensione dei dati durante il processo di addestramento di una rete neurale, rendendo il dataset facilmente accessibile e manipolabile. Il file *HDF5* risultante include l'organizzazione dei video e delle etichette. Tuttavia, sebbene *HDF5* offra vantaggi in termini di organizzazione dei dati, si è rivelato poco pratico per l'addestramento della rete neurale a causa della sua pesantezza e della lentezza nella lettura dei dati. Per superare queste limitazioni, il dataset è stato successivamente convertito in *TfRecord*, che è specificamente progettato per essere utilizzato in TensorFlow, popolare framework di deep learning. *TfRecord* è un formato di file binario che offre vantaggi considerevoli in termini di efficienza di lettura sequenziale dei dati. Questa caratteristica è particolarmente preziosa quando si tratta di lavorare con grandi quantità di dati, come video, durante l'addestramento di reti neurali. La capacità di caricare dati in modo efficiente e gestire la memoria in modo ottimale è cruciale per garantire un processo di addestramento più veloce e stabile. La conversione dei dati in formato *TfRecord* ha comportato l'encoding dei frame video, delle relative etichette e delle maschere di segmentazione in una forma più compatta e ottimizzata. Ogni record *TfRecord* contiene quindi un frame video, la corrispondente etichetta e la maschera di segmentazione associata. Questa trasformazione dei dati ha portato a significativi miglioramenti nelle prestazioni di addestramento della rete. La riduzione del carico di dati e la maggiore efficienza nella lettura di essi hanno accelerato notevolmente il processo di addestramento, rendendolo più veloce ed efficiente.

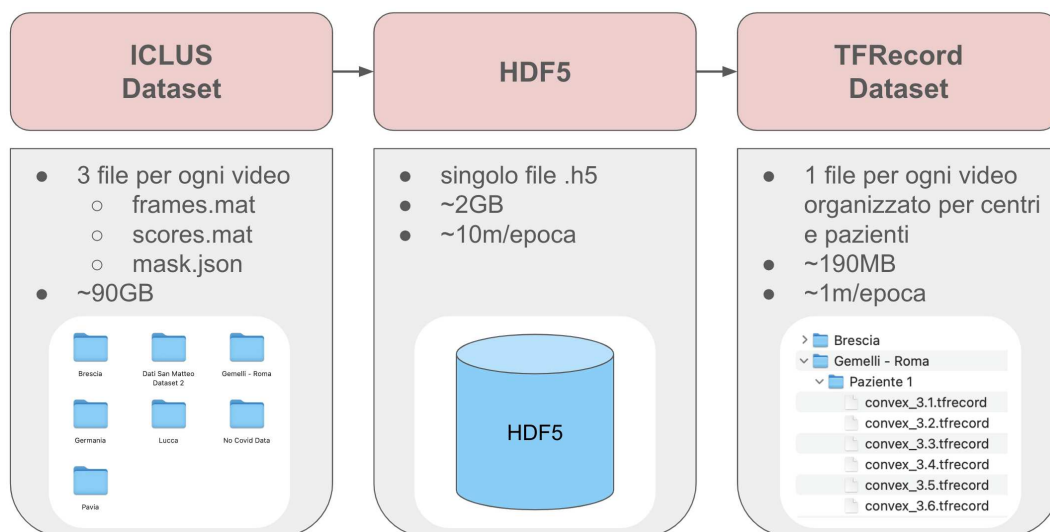


Figura 3.6: Preprocessing dataset

Acquisito il dataset in formato *TfRecord*, si è resa indispensabile la standardizzazione di tutte le immagini presenti. Questa necessità è emersa a causa di una

notevole eterogeneità nelle risoluzioni dei frame e di conseguenza, nelle maschere associate. Tale variabilità è stata attribuita sia all'utilizzo di strumentazioni diverse (vedi Sez.1.1.1) sia alla diversità dei centri medici coinvolti nello studio. In risposta a tale problematica, sono stati ridimensionati tutti i frame a una risoluzione uniforme di 224x224x3. Tale scelta è stata guidata dalla volontà di garantire uniformità nelle dimensioni spaziali e nella profondità di colore dei frame, consentendo analisi e comparazioni più consistenti e affidabili all'interno del dataset. In aggiunta è stato implementato un ulteriore passaggio fondamentale: la normalizzazione dei frame. Questo processo ha comportato l'adeguamento dei valori dei pixel per ottenere distribuzioni uniformi in tutto il dataset. Ciò si è rivelato essenziale per minimizzare potenziali bias dovuti a differenze nell'illuminazione, nel contrasto o in altre variabili visive che potrebbero influenzare l'analisi.

Un ulteriore aspetto cruciale è lo splitting effettuato a **livello di paziente (patient-level)**. Questo approccio è stato adottato per prevenire la creazione di bias dovuti alla possibile correlazione tra campioni provenienti dallo stesso individuo. In altre parole, assicurando che tutti i campioni provenienti da uno stesso paziente fossero allocati in un unico set (sia esso di training, validation, o test), si minimizza il rischio che i modelli apprendessero semplicemente a riconoscere caratteristiche specifiche dei pazienti piuttosto che generalizzare a partire dalle caratteristiche patologiche o cliniche di interesse. Questa scelta metodologica è essenziale per garantire che i risultati ottenuti siano non solo accurati, ma anche clinicamente validi e applicabili in un contesto reale.

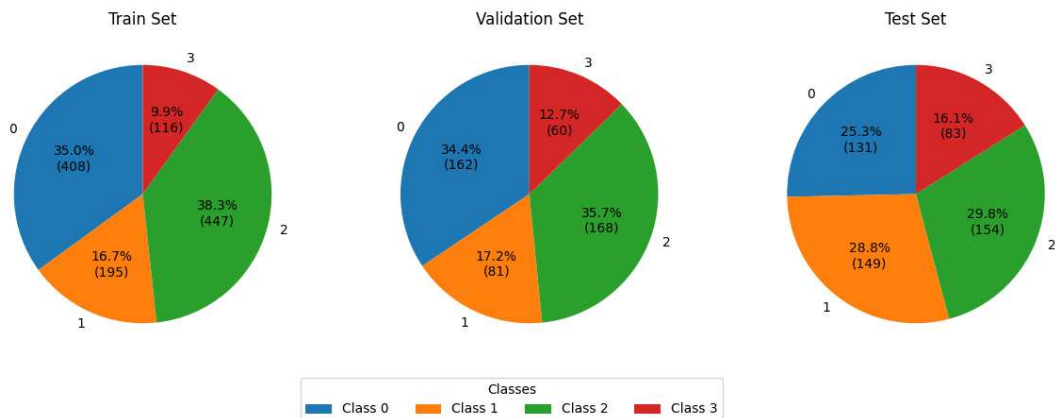


Figura 3.7: Distribuzioni delle classi nei tre set

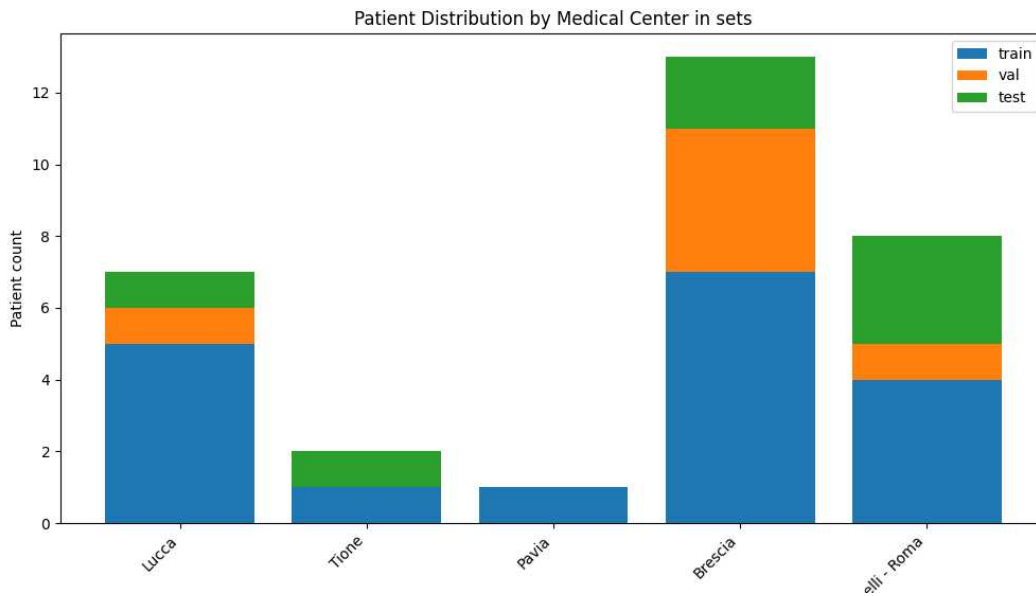


Figura 3.8: Distribuzioni dei centri medici nei tre set

3.1.3 Sfide del dataset

L'utilizzo delle reti neurali convoluzionali (CNN) per analizzare e valutare le ecografie polmonari è un campo di ricerca in forte crescita. Questa tecnologia promette di fornire strumenti diagnostici sofisticati e un supporto decisionale in tempo reale (vedi Sez. 2). Tuttavia, l'applicazione delle CNN in questa tipologia di immagini incontra delle sfide significative a causa della loro natura. Queste sfide sono ulteriormente amplificate nel caso specifico del polmone, a causa delle sue peculiarità strutturali, come la forma della linea pleurica, la caratterizzazione del tessuto polmonare e la presenza di aria all'interno dei polmoni. In particolare, l'interazione tra il tessuto polmonare e l'aria, produce riflessi e artefatti visivi che possono distorcere l'immagine, rendendo più difficile per gli esperti del dominio creare annotazioni precise e affidabili. Tutto ciò contribuisce alla mancanza di pattern geometrici chiari e consistenti nelle ecografie polmonari, rendendo più difficile l'addestramento delle CNN, che tradizionalmente si affidano a tali pattern per l'apprendimento.

Questa problematica è possibile notarla nel dataset ICLUS, dove le maschere di segmentazione non sono costanti tra i vari frame di un singolo video e la sottile distinzione tra i vari stati patologici della linea pleurica porta una differenza tra gli score vicini ad essere quasi impercettibile. Nel primo caso, come mostrato in Fig. 3.9, il primo frame presenta due maschere di segmentazione, una a sinistra e l'altra a destra, nel frame successivo, la maschera di verità sulla sinistra scompare. Questa inconsistenza nelle annotazioni può causare una diminuzione delle prestazioni nella valutazione di un modello, poiché le annotazioni non corrispondono alle maschere predette dal modello. Di conseguenza, ciò può portare a livelli più bassi nelle metriche

di valutazione. Altri esempi di errori di annotazioni sono mostrati in Fig 3.10 e 3.11.

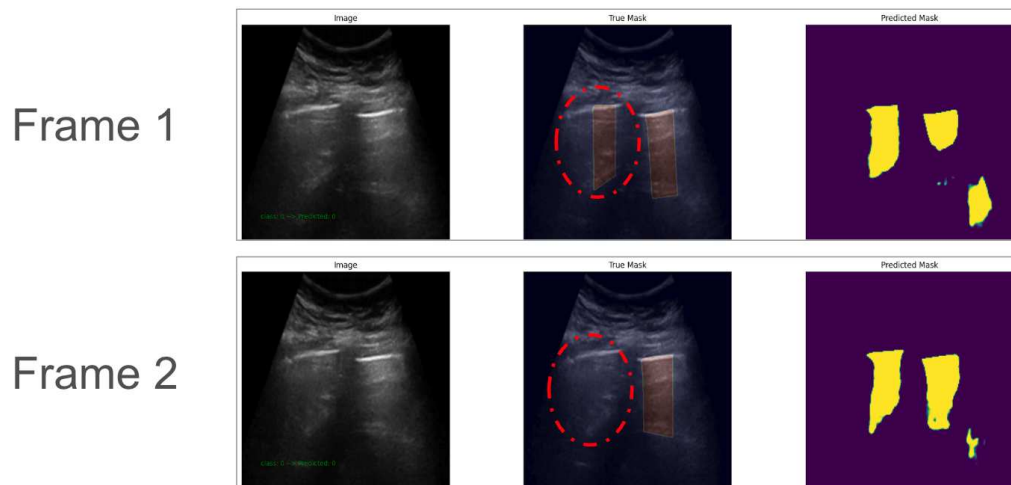


Figura 3.9: Esempio incosistenza maschere di verità tra frame dello stesso video

Il problema della mancanza di coerenza nelle maschere di segmentazione tra frame consecutivi può presentarsi in altri pazienti. Come illustrato nella Fig. 3.10, in cui:

- Nel primo frame del paziente A manca la maschera sulla parte sinistra dell'immagine, che invece è presente nel secondo frame
- Nell'ultimo frame del paziente B manca la maschera a sinistra, che invece è presente sia nel primo che nel secondo frame.

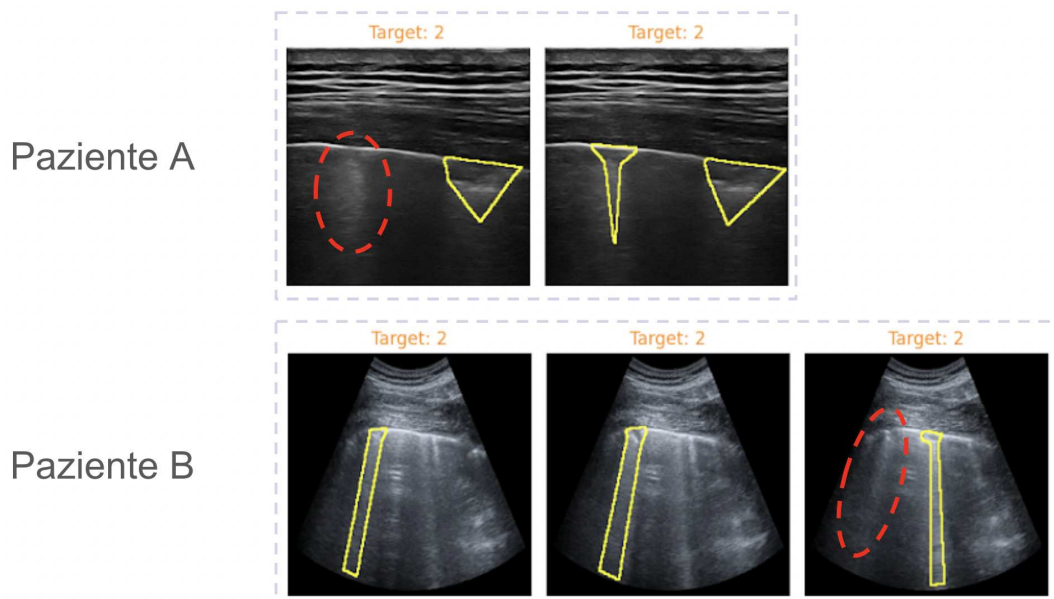


Figura 3.10: Esempio incosistenza maschere in due pazienti

Infine, un altro problema relativo alle maschere di segmentazione è la loro inconsistenze nella forma, come mostrato nella Fig. 3.11, in cui nello stesso paziente in frame consecutivi l'annotazione varia sia di posizione che di forma.

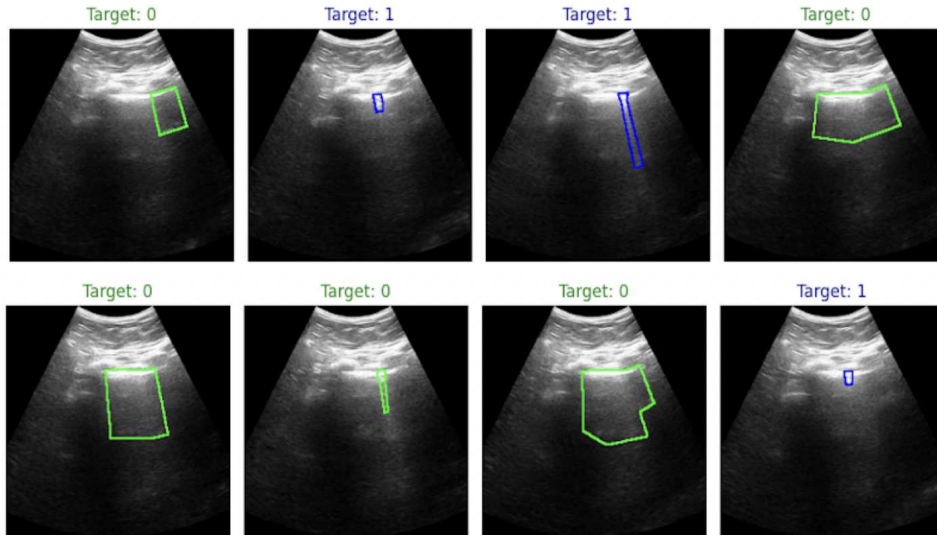


Figura 3.11: Esempio differenza di forma delle maschere nei frame dello stesso video

3.2 Reti neurali convoluzionali - CNN

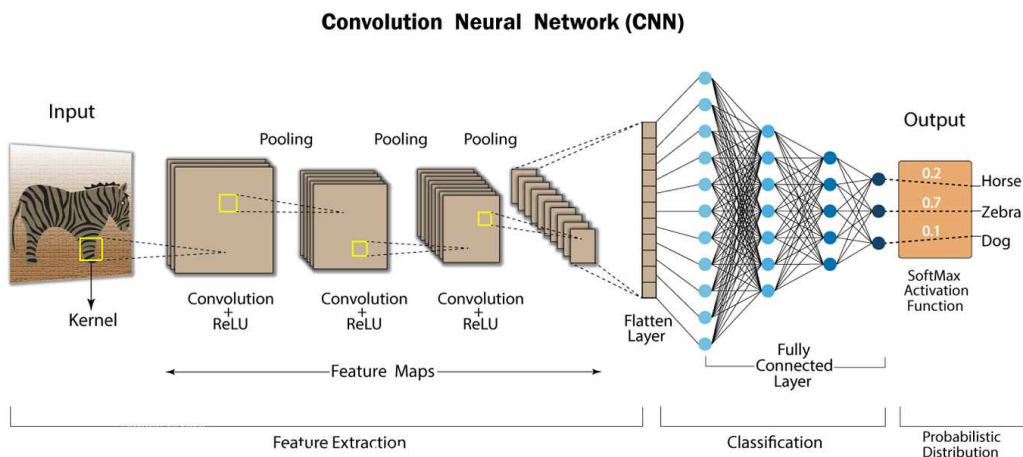


Figura 3.12: Architettura classica di una CNN

Le Reti Neurali Convoluzionali rappresentano una pietra miliare nel campo del Deep Learning. Queste, si distinguono per la loro capacità di emulare la complessa struttura neurale del cervello umano, permettendo un apprendimento automatico e adattivo di features spaziali dai dati. La forza di queste reti risiede nella loro flessibilità e potenza di elaborazione, rendendole ideali per una vasta gamma di

applicazioni; dal riconoscimento di immagini e oggetti, alla diagnosi medica avanzata. Per gestire in modo efficiente ed efficace le immagini e i dati visivi, l'architettura di una CNN è divisa in due componenti (Fig. 3.12): una dedicata all'estrazione delle caratteristiche e un'altra focalizzata sulla classificazione.

Estrazione delle features

Nella fase di estrazione delle features le immagini di input, che possono essere in formati diversi come Scala di grigi, RGB, HSV, o CMYK, vengono analizzate per identificare ed isolare un insieme di feature significative dell'immagine. Questa operazione è cruciale quando si lavora con dati di grandi dimensioni, poiché aiuta a ridurre la complessità computazionale e a semplificare la gestione delle risorse. La parte di rete adibita a questa fase è nota come backbone ed è generalmente composta da:

- Strati convoluzionali: la convoluzione (Fig. 3.13) è l'operazione fondamentale, inizia con una piccola matrice chiamata kernel o filtro, di solito 3x3 o 5x5, applicata all'immagine di input in modo sistematico. Il filtro si sposta su tutta l'immagine attraverso una serie di passi noti come 'strides', e in ogni posizione, viene effettuata una moltiplicazione punto per punto tra i valori dell'immagine e i valori del filtro. I risultati poi sono sommati per creare un valore singolo, che viene collocato in una nuova matrice chiamata feature map. Questa operazione cattura informazioni locali, come bordi, angoli e texture.

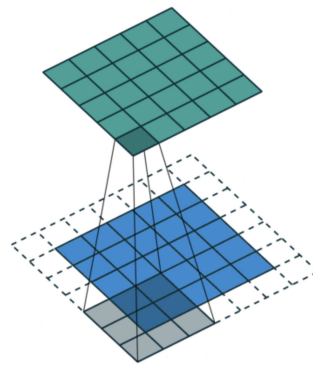


Figura 3.13: Esempio di convoluzione: filtro 3x3 che passa sopra ad un'immagine producendo una nuova matrice

- Strati di attivazione: dopo la convoluzione, viene applicata una funzione di attivazione, (come la Sigmoid, la Tangente Iperbolica e la Rectified Linear Unit (ReLU)), che introduce non linearità nella rete e contribuisce a far apprendere alla rete neurale le relazioni complesse e le dipendenze nei dati.
- Pooling: Dopo uno o più strati di convoluzione e attivazione, è comune applicare uno strato di pooling che va a ridurre la dimensione della mappa delle caratteristiche, semplificando così la rete e riducendo il numero di parametri. Il

pooling può essere effettuato tramite operazioni come il max pooling o l'average pooling, che selezionano il valore massimo o medio all'interno di piccole regioni della mappa delle caratteristiche (come mostrato in Fig. 3.14). Questo riduce la sensibilità della rete alle piccole variazioni nell'immagine e favorisce l'invarianza alle traslazioni.

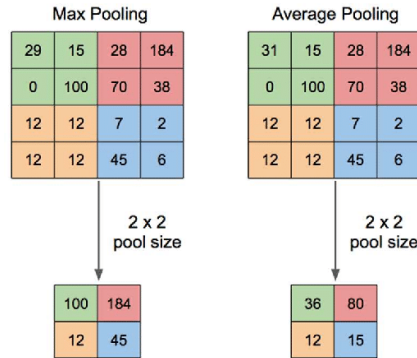


Figura 3.14: Differenza tra max pooling e average pooling

Classificazione

Dopo l'estrazione delle caratteristiche, si passa alla fase di classificazione. In questa fase, solitamente vengono aggiunti uno o più strati completamente connessi, noti anche come strati densi, simili a quelli di una rete neurale tradizionale. Ogni input di questi strati è connesso a ciascun output attraverso dei pesi che vengono appresi durante la fase di addestramento. Questi strati densi vengono alimentati da uno strato chiamato 'flattening' il quale trasforma le mappe delle feature in uscita dalla parte principale della rete in un array monodimensionale. Gli strati densi utilizzano poi le caratteristiche estratte per effettuare previsioni sulla classe dell'immagine in ingresso. In genere, lo strato denso finale è configurato con un numero di neuroni di uscita uguale al numero di classi considerate nel problema di classificazione. Successivamente, si applica una funzione di attivazione in base alla natura del problema di classificazione. Nel caso di una classificazione multi-classe, una delle funzioni di attivazione comunemente utilizzate è la 'Softmax.' Questa funzione normalizza i valori reali prodotti all'uscita dell'ultimo strato per ciascuna classe, convertendoli in valori compresi tra 0 e 1. Un aspetto importante della Softmax è che la somma di tutti i valori ottenuti è sempre uguale a 1, fornendo così una distribuzione di probabilità su cui basare le previsioni della rete per le diverse classi.

3.2.1 UNet

La UNet è una CNN progettata per la segmentazione delle immagini. Questa architettura si è affermata come particolarmente efficace nel campo della biomedicina[8],

grazie alla sua capacità di dettagliare con precisione le strutture complesse presenti nelle immagini mediche. Tuttavia, la sua versatilità le ha permesso di essere impiegata con successo anche in altri ambiti dell'elaborazione delle immagini, dimostrando la sua ampia applicabilità al di là del solo dominio biomedico. L'architettura della UNet, il cui design ricorda la forma della lettera "U" (Fig. 3.15), si distingue per la sua struttura composta da due componenti principali:

- Encoder (Downsampling): Situato sul lato sinistro del modello, è una rete neurale convoluzionale convenzionale che esegue operazioni di downsampling sull'immagine di input. Questa sezione è cruciale per la cattura del contesto all'interno dell'immagine. Attraverso una serie di strati convoluzionali, si verifica una riduzione progressiva delle dimensioni spaziali dell'immagine, mentre aumenta il numero di canali delle features. Ogni strato nell'encoder è tipicamente seguito da una funzione di attivazione e da operazioni di pooling, che contribuiscono a condensare le informazioni e ridurre l'overfitting.
- Decoder (Upsampling): Posizionato simmetricamente sul lato destro del modello, si occupa dell'upsampling delle features apprese, con l'obiettivo di ricostruire l'immagine di output con dettagli precisi a livello di pixel (localizzazione precisa). Questo processo è fondamentale per ottenere una segmentazione di alta risoluzione. Nel decoder, le mappe delle features subiscono un processo inverso rispetto all'encoder, con l'aumento delle dimensioni spaziali e la riduzione del numero di canali. Ciò viene realizzato attraverso strati di convoluzioni trasposte o operazioni di upsampling.

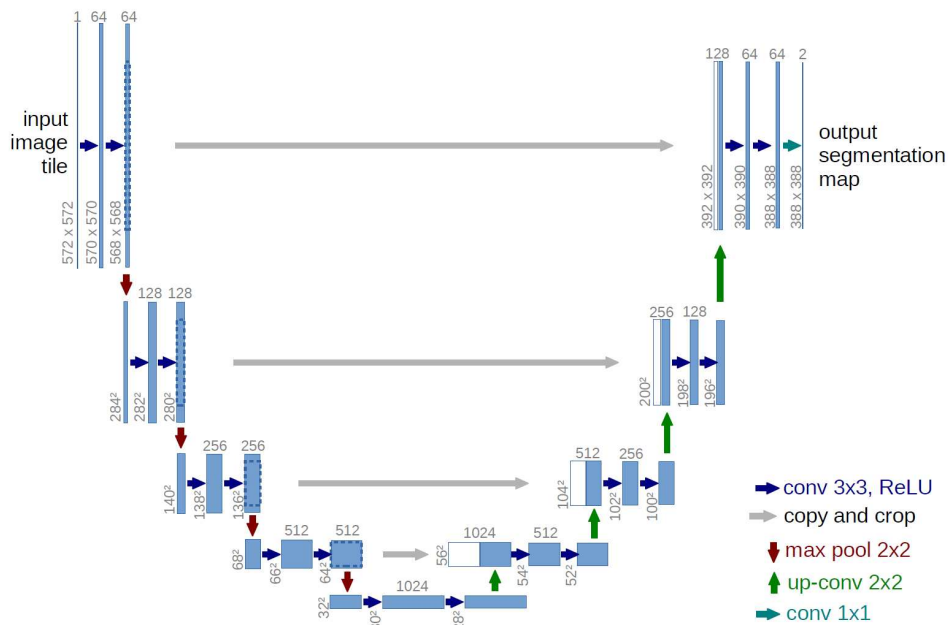


Figura 3.15: Architettura UNet

Altro punto chiave di questa rete convoluzionale sono gli skip connection. Questi, collegano direttamente i livelli del lato di encoder con i corrispondenti livelli del lato di decoder, aiutando a recuperare le informazioni spaziali perse durante il downsampling, consentendo alla rete di fare previsioni più precise.

3.2.2 Unet++

La UNet++, illustrata nella Figura 3.16, rappresenta un'importante evoluzione della rete UNet. Questa architettura è stata progettata e proposta da Zhou et al.[18] (2019) per affrontare alcune delle limitazioni della UNet standard, migliorando la precisione nella segmentazione delle immagini andando a modificare la struttura del modello originale introducendo il concetto di connessioni dense tra i livelli della rete. Rispetto alla UNet, che presenta un'architettura encoder-decoder con collegamenti diretti, quest'ultima introduce una serie di percorsi di collegamento aggiuntivi che creano una rete più intricata. Questi percorsi aiutano a migliorare sia il flusso di informazioni sia i gradienti attraverso la rete, facilitando l'apprendimento. Un altro aspetto chiave della UNet++ è l'implementazione di moduli di nidificazione. Questi moduli permettono di catturare e combinare le features a vari livelli di risoluzione, migliorando la capacità della rete di gestire variazioni nelle dimensioni e nelle forme degli oggetti da segmentare. Questa caratteristica è particolarmente utile nel contesto della segmentazione di immagini mediche, dove le strutture da identificare possono variare notevolmente in dimensioni e forma.

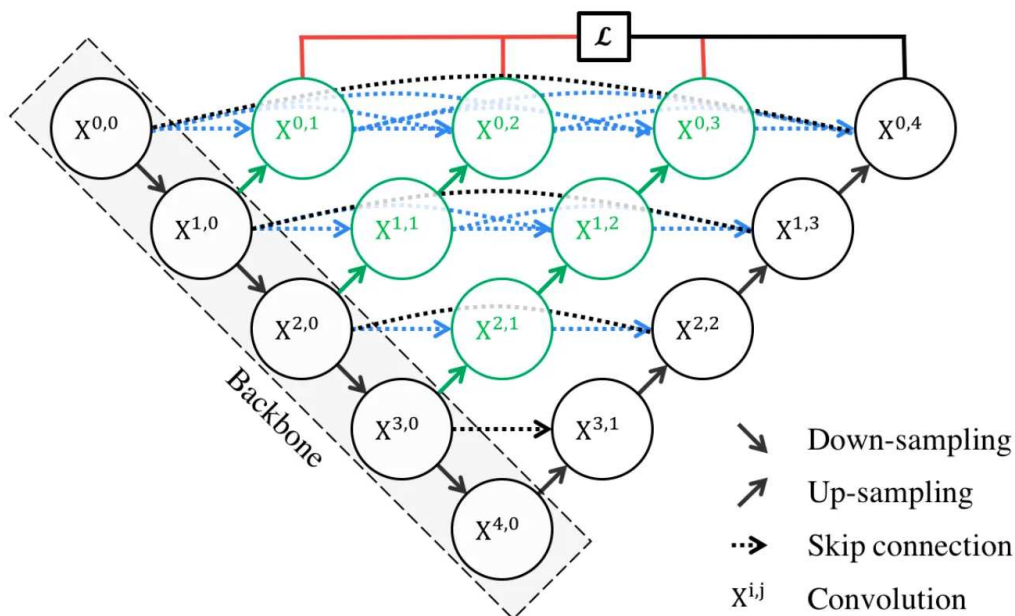


Figura 3.16: Architettura UNet++

3.2.3 ResNet

Il problema del *Vanishing/Exploding gradient* si verifica quando i gradienti delle funzioni di errore diventano estremamente piccoli (prossimi a zero) o eccessivamente grandi mentre si propagano attraverso i numerosi strati di una rete profonda durante il processo di addestramento. Questo fenomeno ha conseguenze negative sulle prestazioni della rete, poiché rende difficile l'aggiornamento dei pesi delle connessioni e può portare a un aumento del tasso di errore sia durante l'addestramento che sul set di test. In pratica, questo significa che mentre l'aggiunta di più strati inizialmente può portare a una migliore rappresentazione di essi e a una riduzione dell'errore, i gradienti possono diventare troppo piccoli per effettuare aggiornamenti significativi dei pesi, rallentando o addirittura ostacolando l'apprendimento.

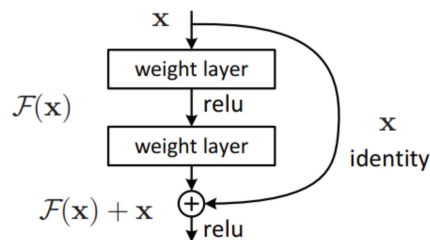


Figura 3.17: Blocco residuo

Per superare questo problema, sono state introdotte nuove architetture come la ResNet, abbreviazione di Residual Network (Rete Residua) ed introdotta da Kaiming He et al. nel 2015 [19]. Queste reti utilizzano le *skip connection* e *blocchi residui* per consentire un addestramento efficace di reti neurali molto profonde senza incorrere nei problemi del *Vanishing/Exploding gradient*. La *skip connection* connette le attivazioni di uno strato a ulteriori strati, bypassandone uno o più intermedi, andando a formare un *blocco residuo* (Fig. 3.17).

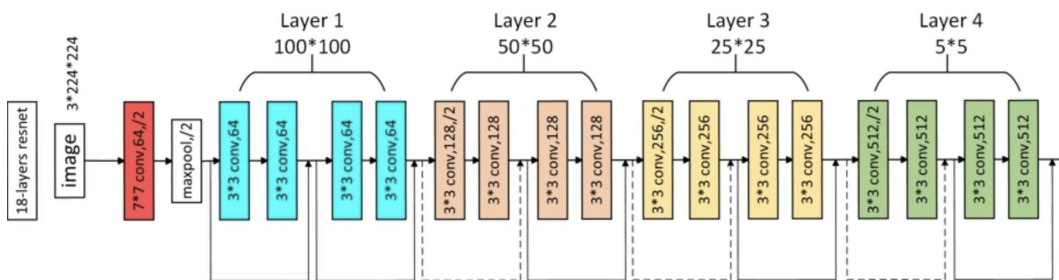


Figura 3.18: Architettura ResNet18

Le ResNet dunque, sono create impilando insieme questi *blocchi residui*. Il vantaggio di aggiungere questo tipo di skip connection è che se uno strato danneggia le prestazioni dell'architettura, verrà saltato tramite la regolarizzazione. La loro progettazione ha permesso di costruire reti molto più profonde rispetto a quanto fosse possibile prima, con architetture che vanno da ResNet-18 (con 18 strati) a ResNet-152

(con 152 strati) e oltre. In questa tesi, per una fair comparison con lo stato dell'arte, vengono prese in considerazione solamente la versione a 18 e 50 strati della ResNet. L'architettura dei due modelli fondamentalmente inizia con un layer di convoluzione iniziale seguito da Batch Normalization e attivazione ReLU. Successivamente, c'è uno strato di Max Pooling per ridurre le dimensioni dell'immagine. Gli strati successivi costituiscono blocchi residuali che utilizzano convoluzioni e Batch Normalization. La rete termina con uno strato di Global Average Pooling (GAP) che riduce le dimensioni delle feature a un vettore di output.

3.3 Multitask Learning

Nel processo di sviluppo del modello multitask, diverse combinazioni di architetture sono state analizzate. Inizialmente, seguendo i modelli proposti nella letteratura scientifica, citate nella Sez.2.3, è stata progettata una rete che integra elementi delle architetture Residuali e U-Net: **ResUnet**. Questo modello offre due benefici chiave:

- L'uso di Unità Residuali per migliorare l'efficienza dell'apprendimento.
- Le Skip Connections per una migliore propagazione delle informazioni.

Inoltre, consente di utilizzare una *ResNet* come *encoder* all'interno dell'architettura UNet. Questo aspetto riveste una grande importanza in quanto ci permette di effettuare un confronto diretto con i risultati ottenuti nell'ambito della classificazione, presenti nello stato dell'arte, utilizzando anch'essi una rete *ResNet*.

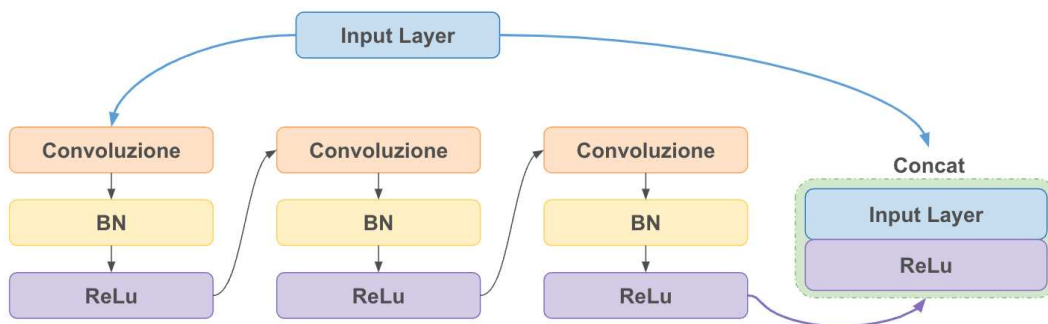


Figura 3.19: Rappresentazione grafica blocchi residui

L'architettura **ResUnet** (Fig.3.20) ha un layer di input concepito per essere alimentato da immagini 2D a colori (RGB) con una dimensione, in questo caso, di 224x224 pixel. Successivamente, il modello si sviluppa con nove livelli organizzati secondo il concetto fondamentale di Unet, ovvero la suddivisione in: *encoder*, *bridge* e *decoder*.

- L'*encoder* ha il compito di trasformare l'immagine di input in una rappresentazione compatta

- Il *decoder* si occupa di ricostruire la rappresentazione fornita dall'*encoder* in una classificazione pixel per pixel
- Il *bridge* è il collegamento tra l'*encoder* e il *decoder*.

I tre componenti sono costruiti utilizzando unità residue mostrate nella Fig. 3.19, in cui ciascun blocco consiste in 3 livelli convoluzionali con filtri 3×3 , 3 livelli di normalizzazione batch, 3 livelli di attivazione *ReLU* e un mappaggio di identità. Come menzionato in precedenza, il blocco di *encoding*, utilizza una *ResNet* 18 o 50 il quale si distingue per l'impiego di blocchi convoluzionali che includono connessioni residue (vedi Sez. 3.2.3). Queste connessioni permettono di aggiungere direttamente l'input di un blocco al suo output, facilitando così la propagazione del gradiente attraverso la rete; caratteristica particolarmente utile per apprendere funzioni identità, cruciali per prevenire il problema della scomparsa del gradiente nelle reti neurali profonde. Dunque, l'*encoder* sarà composto da 4 unità residue, che hanno l'obiettivo di ridurre la dimensionalità dell'immagine per estrarre informazioni semantiche di alto livello. Simmetricamente il percorso del *decoder* è composto anch'esso da quattro unità residue, ma tra ciascuna di queste unità avviene una concatenazione tra la mappa delle features proveniente dal percorso di *encoding* del layer corrispondente e la mappa delle features che è stata upscalata dall'unità precedente del *decoder*. Infine, l'ultima parte della rete, dopo l'ultima unità del *decoder*, consiste in un layer di *upsampling* e uno strato convoluzionale 3×3 , che riportano la mappa delle features alla dimensione spaziale originale. L'output finale è una singola mappa di features, prodotta da un livello di attivazione *sigmoidale*, che rappresenta la maschera di segmentazione predetta dalla rete per l'immagine in ingresso.

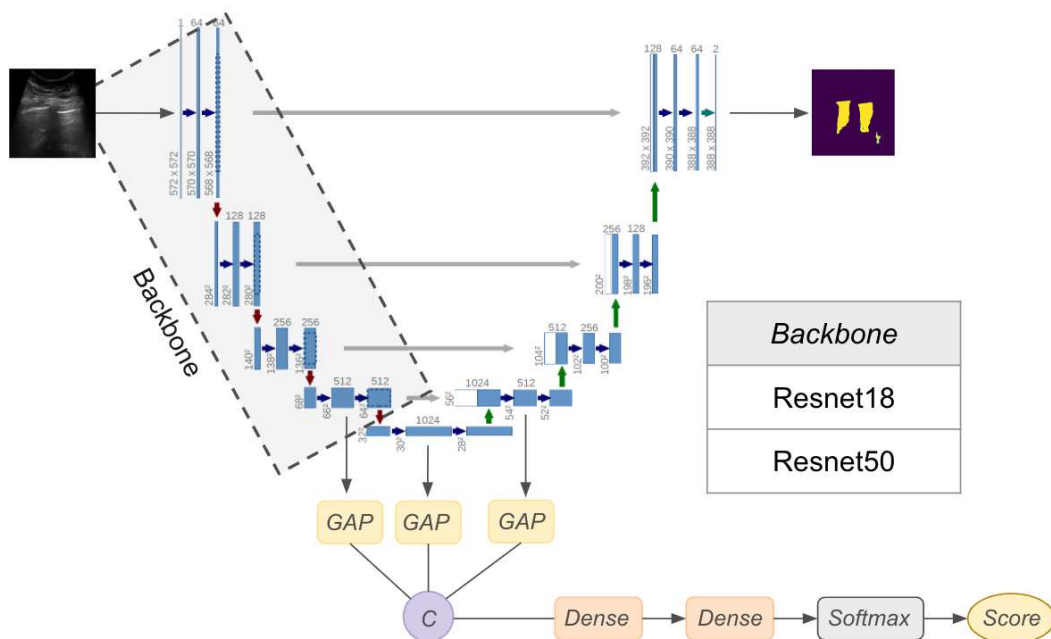


Figura 3.20: Architettura ResUnet per Multitask

Nella classificazione delle immagini, è comune utilizzare mappe di features di alto livello da modelli di reti neurali convoluzionali (CNN) come VGG (Simonyan e Zisserman, 2014[20]) e ResNet (He et al., 2015[19]). Un'osservazione interessante in questo contesto è stata fatta da Zhou et al.[16] nel loro lavoro. Essi hanno condiviso le mappe di features di alto livello del loro modello di segmentazione, consentendo così di estrarre features comuni per scopi di classificazione e segmentazione all'interno del loro approccio multitask. Questa condivisione ha dimostrato di essere efficace nel migliorare le prestazioni di entrambi i compiti senza duplicare i calcoli di estrazione delle features. Seguendo questa idea, si è scelto di estendere l'architettura **ResUnet** con un ramo dedicato alla classificazione; come illustrato nella Fig. 3.20. Il ramo riceve in input le mappe delle features estratte da diverse parti della rete: l'ultimo blocco dell'encoder, il bridge e il primo blocco del decoder. Sono state scelte le fasi 4, 5 e 6 della **ResUnet** per alimentare il classificatore in quanto rappresentano le parti più profonde del modello, estraggono feature rappresentative e astratte, e catturano informazioni complesse. Questo approccio combina informazioni contestuali profonde con dettagli più fini, migliorando la capacità di classificazione di immagini complesse come le ecografie. La selezione delle tre fasi anziché singole fasi o combinazioni di esse è stata fatta per catturare una vasta gamma di feature rappresentative. Per garantire che queste mappe abbiano le stesse dimensioni e possano essere concatenate in modo diretto, viene applicata l'operazione di *Global Average Pooling* (GAP) alla fine di ciascun blocco. Infine, le features fuse vengono inserite nel ramo di classificazione composto da due strati completamente connessi (Dense). Il primo strato denso ha 256 unità ed è attivato tramite la funzione *ReLU*. L'ultimo, invece, ha 4 unità ed è attivato tramite la funzione *softmax*. Quest'ultimo, viene utilizzato per effettuare la previsione finale della classe dell'immagine LUS in input, assegnando uno dei quattro punteggi indicati nella Sez.3.1. Complessivamente, il modello esibisce un numero di parametri di circa 1.94 milioni.

Per affrontare le complessità che caratterizzano le immagini LUS, come la loro difficile interpretazione e l'assenza di pattern geometrici chiari, questa tesi introduce un'evoluzione dell'architettura **ResUNet**, nota come **ResUNet++**. Basandosi sulla struttura base di **ResUnet**, **ResUnet++** è progettata specificatamente per migliorare la precisione nella segmentazione di immagini dettagliate come le ecografie polmonari. **ResUnet++** conserva l'approccio multitask di **ResUnet** ma integra nella sua componente di segmentazione la *Residual UNet++*. Quest'ultima versione introduce connessioni residuali più profonde, migliorando notevolmente la precisione, l'accuratezza e il dettaglio della segmentazione. Inoltre, **ResUnet++** può impiegare come backbone sia *ResNet* 18 che *ResNet* 50, similmente all'architettura precedente. Il modello sviluppato presenta un totale di circa 5.37 milioni di parametri, un numero significativamente maggiore rispetto alla versione precedente, garantendo così un maggior dettaglio e precisione.

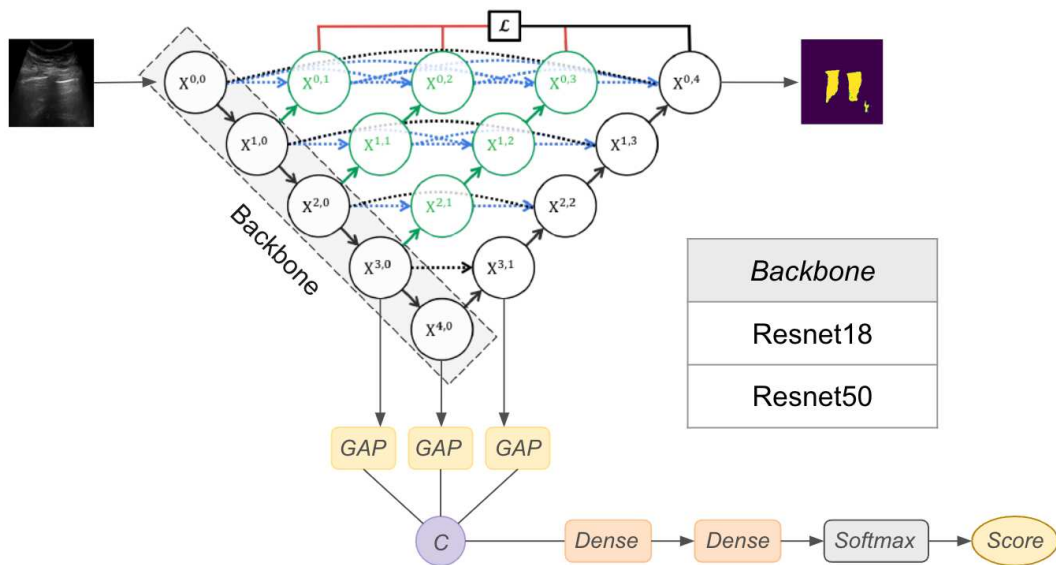


Figura 3.21: Architettura ResUnet++ per Multitask

3.4 Procedura Sperimentale

L'addestramento dei modelli sul dataset ICLUS è stato condotto attraverso l'uso di diverse configurazioni. Grazie al framework descritto nella Sez. 3.4.4, è stata adottata una strategia di modularità per le diverse architetture, al fine di facilitarne lo studio e l'analisi. Questo approccio ha permesso una maggiore flessibilità e personalizzazione nella configurazione dei modelli, rendendo il processo di addestramento e valutazione più efficace ed efficiente. L'addestramento è stato effettuato con un numero variabile di epoche: da 8, 16 a 32 epoche. È stato inoltre studiato il Fine Tuning per implementare il trasferimento di apprendimento, utilizzando i pesi pre-addestrati da ImageNet. Infine, sono stati testati due tipi di ottimizzatori: SGD e ADAM

3.4.1 Funzioni di loss

Segmentazione

Nel contesto della segmentazione delle immagini, uno dei problemi più comuni è il bias causato dalla differenza tra lo sfondo (background) e l'oggetto di interesse (foreground). Questo bias può portare a risultati inaccurati, dove parti significative dell'immagine possono essere erroneamente segmentate o trascurate. In particolare, nelle situazioni in cui il background e il foreground hanno caratteristiche di contrasto o intensità simili, gli algoritmi di segmentazione tradizionali possono avere difficoltà a discriminare correttamente queste due regioni. Per affrontare questo problema, si è scelto di utilizzare una funzione di loss che mira a massimizzare il coefficiente di **Dice**. È particolarmente adatta per i dati con una distribuzione disuguale delle classi, come spesso si verifica nelle immagini mediche. La **Dice Loss** è formulata come in

Eq. 3.1, e fornisce una misura diretta di quanta variazione c'è tra la segmentazione del modello e la verità di base. Essendo una misura di dissimilarità, un valore più basso di **Dice Loss** indica una migliore performance di segmentazione.

$$L_{seg} = DSC_{loss} = 1 - DSC \quad (3.1)$$

Classificazione

Il *class imbalance* rappresenta una sfida significativa nell'analisi dell'imaging medico, specialmente nei modelli di apprendimento automatico utilizzati per scopi diagnostici. Questo problema è particolarmente evidente anche nel contesto del dataset utilizzato. Come si può notare nella figura 3.5 si hanno pochi valori per lo score 3 mentre se ne hanno tanti per gli score 0 e 2. Questo sbilanciamento potrebbe portare ad un bias durante l'addestramento della rete che potrebbe poi scaturire in falsi positivi o peggio ancora in contesti clinici in falsi negativi. Per contrastare questo squilibrio, si propone l'utilizzo della loss basata sulla **Weighted Categorical Cross-Entropy** per la classificazione nel dataset ICLUS. Questa funzione modifica la **Categorical Cross-Entropy Loss** utilizzata nei problemi di classificazione multi-classe, introducendo dei pesi diversi per le classi nel calcolo della loss. Nella *Categorical Cross-Entropy* standard, la loss per un singolo esempio è data da:

$$CCE_{loss} = - \sum_{c=1}^M y_{o,c} \log(p_{o,c}) \quad (3.2)$$

dove CCE_{loss} è la loss totale per un'osservazione, M è il numero di classi, $y_{o,c}$ è un indicatore che vale 1 se la classe reale dell'osservazione o è c , altrimenti 0, e $p_{o,c}$ è la probabilità predetta che l'osservazione o appartenga alla classe c . In quella pesata invece, viene introdotto un peso w_c per ciascuna classe, ottenendo la formula:

$$L_{cls} = WCCE_{loss} = - \sum_{c=1}^M w_c y_{o,c} \log(p_{o,c}) \quad (3.3)$$

I pesi w_c sono tipicamente proporzionali al reciproco della frequenza delle classi nel dataset di training. Questo approccio assicura che ogni classe contribuisca in modo più equilibrato alla loss complessiva, aiutando il modello ad imparare in modo più bilanciato da tutte le classi. Infatti, assegnando pesi maggiori alle classi meno frequenti, la funzione di loss "punisce" di più il modello per errori su queste classi, incoraggiandolo a prestare maggiore attenzione alle classi meno rappresentate.

Multitask

La combinazione di queste due loss in una funzione loss totale fornisce un equilibrio efficace tra la necessità di segmentazione precisa e quella di classificazione accurata, essenziale per applicazioni di diagnosi medica basate su immagini. La combinazione

è lineare ed è ottenuta tramite un iperparametro λ . Dunque la loss utilizzata nel modello multitask è definita in base al peso della task di classificazione e alla perdita di segmentazione, con λ nell'intervallo $[0, 1]$.

$$L_{multitask} = \lambda L_{cls} - (1 - \lambda) L_{seg} \quad (3.4)$$

3.4.2 Metriche

Come per le funzioni di loss, sono state utilizzate sia metriche di valutazione per la classificazione sia per la segmentazione per poi utilizzarle insieme nel modello Multitask proposto. Nella classificazione di immagini, l'**accuratezza** è una metrica fondamentale, ed è definita come il rapporto tra il numero di predizioni corrette e il numero totale di predizioni effettuate. Formalmente, può essere espressa come:

$$ACC = \frac{\text{N. predizioni corrette}}{\text{N. Tot predizioni}} \quad (3.5)$$

Per un modello di classificazione, una "predizione corretta" significa che la categoria assegnata dall'algoritmo corrisponde alla categoria reale (verità di base) dell'oggetto o dell'immagine. Questa metrica è particolarmente utile per fornire una visione generale dell'efficacia del modello. Per una fair comparison con la letteratura, si è scelto di adottare l'**accuratezza** anche nella valutazione delle reti di segmentazione. Tuttavia, è essenziale sottolineare alcune considerazioni riguardo al suo utilizzo. In particolare, nell'ambito della segmentazione, l'**accuratezza** viene calcolata tenendo conto anche della predizione del background. Questo approccio tende a generare valori elevati durante l'addestramento e la validazione, dovuti alla predominanza del background nelle immagini di questo tipo. Questa peculiarità richiede un'attenta interpretazione dei risultati per assicurare una valutazione realistica dell'efficacia della rete.

Un'altra metrica utilizzata per la segmentazione è il coefficiente di **Dice**. Noto anche come indice di **Sørensen-Dice**, rappresenta una metrica statistica fondamentale utilizzata per valutare la performance nelle applicazioni di segmentazione delle immagini. Questo indice quantifica la similarità e la corrispondenza tra due insiemi di dati, in questo caso, l'output dell'algoritmo di segmentazione e il ground truth. Il coefficiente valuta quanto efficacemente l'algoritmo di segmentazione riesce a discriminare il foreground dal background. È calcolato basandosi sul numero di pixel correttamente identificati nel foreground (veri positivi), rispetto alla somma totale dei pixel classificati come foreground sia dall'algoritmo che nella realtà. Matematicamente, è definito come:

$$DSC = \frac{2 \times |X \cap Y|}{|X| + |Y|} \quad (3.6)$$

Dove X rappresenta l'insieme dei pixel identificati come foreground dall'algoritmo, e

Y l'insieme dei pixel del foreground effettivo. La formula calcola il doppio dell'intersezione dei due insiemi ($X \cap Y$), ovvero i veri positivi, e lo divide per la somma totale dei pixel presenti in entrambi gli insiemi ($|X| + |Y|$). Questo indice è particolarmente utile per quantificare la precisione di un algoritmo di segmentazione nell'identificare accuratamente le aree di interesse in un'immagine.

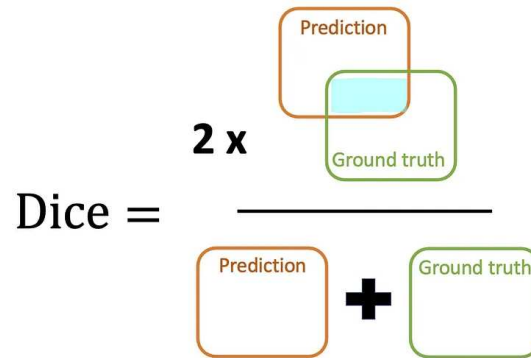


Figura 3.22: Rappresentazione grafica metrica Dice

3.4.3 Framework sperimentale

Le architetture discusse sono progettate per essere valutate con un framework automatizzato. Questo strumento facilita la valutazione di modelli neurali sul dataset ICLUS, organizzando il processo in "esperimenti". Ogni esperimento corrisponde a una configurazione specifica di un modello, definita tramite un file JSON. Questo file stabilisce le impostazioni per l'addestramento e la valutazione dei modelli, e include i seguenti campi:

- Architettura da utilizzare:
 - Resnet18/50: Resnet classica solo per compiti di classificazione.
 - unet: unet standard, utilizzabile sia per la segmentazione che per compiti multitask.
 - unetpp: Simile a unet, ma con l'architettura "Nasted".
 - ResUnet18/50: Primo modello progettato per compiti multitask.
 - ResUnetpp18/50: Secondo modello sviluppato per il multitask.
- Tipo di compito: Segmentazione, Classificazione o Multitask.
- Numero di epoche.
- Splitting Ratio: Determina la proporzione di dati usati per l'addestramento, la validazione e il test.
- Ottimizzatore: Adam o SGD.

- Learning Rate: Imposta il tasso di apprendimento.
- Data Augmentation: Flag per decidere se applicare o meno la data augmentation.
- Transfer Learning su Imagenet: Opzione per applicare il trasferimento di apprendimento utilizzando il dataset Imagenet.
- Layer da "frizzare": Specifica fino a quali layer bisogna bloccare l'apprendimento nel caso si opti per il Transfer Learning.
- Dropout

Questo framework consente quindi di personalizzare e ottimizzare i modelli neurali per specifiche esigenze e tipologie di compiti.

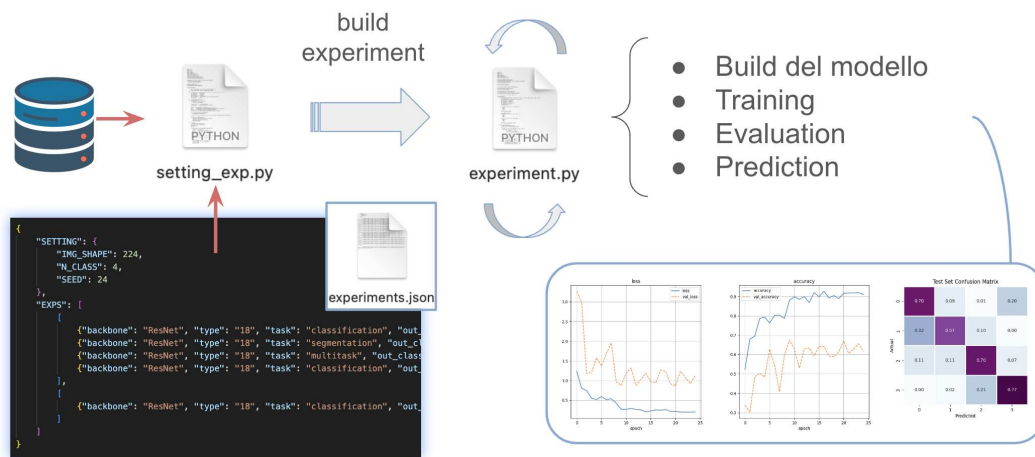


Figura 3.23: Processo grafico del Framework sperimentale

3.4.4 Grid Search

I risultati ottenuti e mostrati nel capitolo successivo, sono il frutto di un'accurata analisi di vari iperparametri, illustrati nella Tab. 3.1, punto di riferimento cruciale per la replicabilità e la validità degli esperimenti. In particolare, nella Tab. 3.1 sono inclusi tutti gli iperparametri utilizzati per lo studio condotto mediante l'approccio del *grid search bayesiano* implementato in Keras. A differenza del tradizionale *grid search*, che valuta sistematicamente ogni combinazione possibile di iperparametri, quello *bayesiano* adotta un approccio probabilistico. Questo metodo utilizza i risultati delle valutazioni precedenti per formulare una distribuzione di probabilità, indicando quali combinazioni di parametri potrebbero portare a prestazioni migliori. In sostanza, 'impara' dalle iterazioni precedenti, ottimizzando la ricerca verso le aree dello spazio degli iperparametri che promettono risultati migliori. Questo approccio non solo è spesso più efficiente in termini di tempo rispetto al *grid search standard* ma può

anche portare a una scoperta più accurata del set di iperparametri ottimale per un dato modello.

Split Ratio	60-20-20
Batch size	4-8-16-32
Ottimizzatore	SGD, Adam
LR	[0.01, 0.0001]
Epoche	50
Dropout	[.0, .3]

Tabella 3.1: Iperparametri

3.4.5 Data Augmentation

Nella presente tesi si è scelto di adottare la tecnica di On the fly Data Augmentation, anche detta Online. In questo approccio, a differenza della metodologia tradizionale dove le trasformazioni vengono applicate al dataset prima dell'inizio dell'addestramento, le modifiche sono effettuate in tempo reale durante il processo di addestramento stesso. Le trasformazioni implementate sono state scelte seguendo le raccomandazioni dello stato dell'arte[21], inoltre sono state tutte applicate con una probabilità dello 0.5%, garantendo così di non influenzare eccessivamente il contenuto informativo dei dati originali. Le trasformazioni utilizzate sono:

- Rotazione Casuale: Le immagini sono ruotate casualmente entro un intervallo per simulare variazioni di orientamento e migliorare la generalizzazione del modello.
- Luminosità Casuale: La luminosità delle immagini è variata casualmente all'interno di un intervallo per aumentare la robustezza del modello a differenti condizioni di illuminazione.
- Ritaglio Centrale Casuale
- Flip orizzontale

Vedi Tab. 3.2 per i paramtri utilizzati.

Trasformazioni	Parametri
Random rotation	[-23, +23]
Random crop	[0.5,0.8]
Random brightness	[0, 0.10]
Random horizontal flip	-

Tabella 3.2: Parametri Data Augmentation On the fly

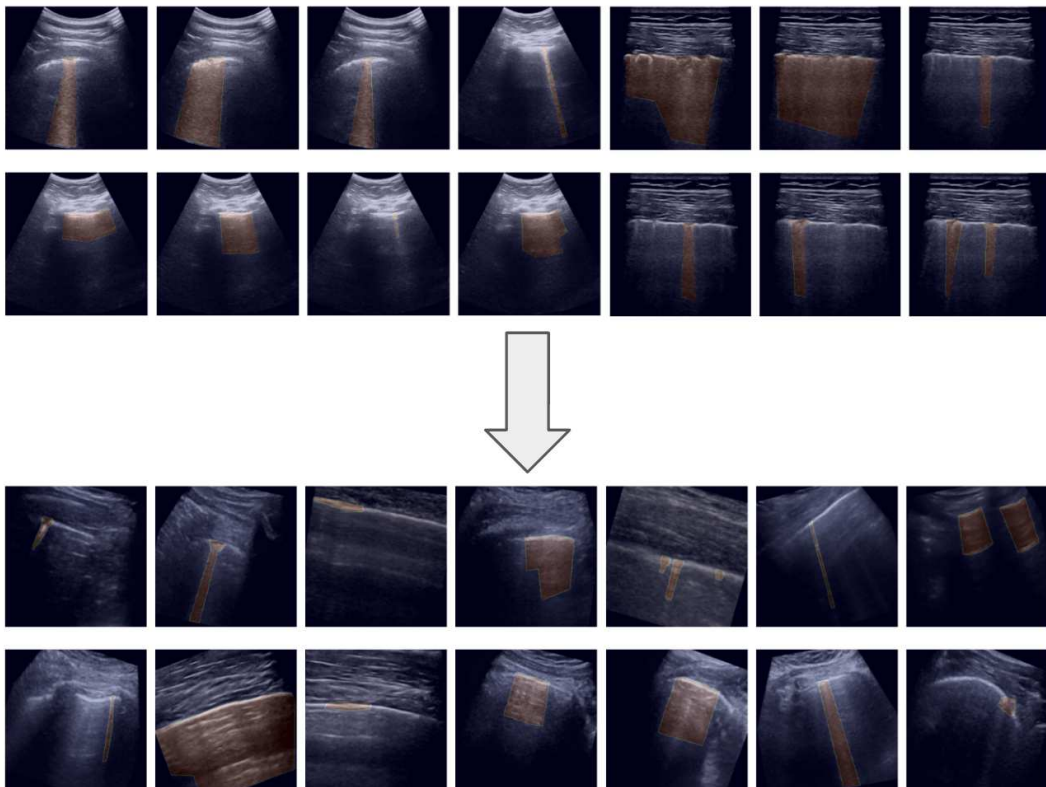


Figura 3.24: Data Augmentation

Capitolo 4

Risultati e Discussioni

In questo capitolo, si esamina con attenzione i risultati emersi dallo studio svolto. L'analisi si articola in una struttura tripartita, che specchia le principali aree di indagine considerate: **classificazione**, **segmentazione** e **multitask**. La strutturazione di questo capitolo in tre sezioni è motivata:

- Dall'assenza, nella letteratura, di un approccio **multitasking** nell'analisi di immagini ecografiche del polmone.
- Dal non utilizzo del dataset per la *segmentazione* (Sez. 3.1.1) nei modelli di riferimento dello stato dell'arte[11] per la **classificazione**.
- Da una mancata chiarezza riguardo lo splitting dei dati utilizzato nei lavori proposti dallo stato dell'arte per la **segmentazione**[11]. Questa ambiguità porta a risultati non completamente riproducibili ed impedisce una comparazione equa, poiché le differenze nella suddivisione nei vari set di dati può influenzare significativamente i risultati ottenuti.

Di conseguenza, si è optato per un approccio comparativo indiretto; sono stati implementati i modelli single-task di riferimento del SOTA, calcolando i risultati sul set di dati utilizzato. Questi risultati sono stati poi confrontati con quelli ottenuti nei due compiti dal modello *multitask*. Tale metodologia ha permesso di stabilire basi di confronto più eque e coerenti, fornendo una visione chiara dell'efficacia dell'approccio *multitask* rispetto alle soluzioni single-task, nel contesto specifico delle immagini ecografiche polmonari. Attraverso questa struttura analitica, il capitolo mira a fornire una comprensione dettagliata delle prestazioni dei diversi modelli in relazione ai compiti di *classificazione* e *segmentazione*, concludendo con un'analisi dell'approccio *multitask*, che rappresenta l'elemento innovativo della ricerca nel dominio LUS.

È fondamentale sottolineare che tutti gli addestramenti sono stati conseguiti utilizzando la stessa suddivisione del dataset. Tale splitting è stato selezionato in modo specifico per garantire che le distribuzioni delle classi nei tre set (training, validation, e test) fossero le più uniformi possibili (vedi Fig. 3.7). Questa scelta metodologica è stata guidata dall'obiettivo di evitare qualsiasi forma di bias derivante da una distribuzione disomogenea delle classi, che potrebbe altrimenti influenzare negativamente l'accuratezza e la generalizzabilità dei modelli.

4.1 Classificazione

Nella presente sezione si discutono le performance ottenute attraverso l'impiego di una rete *ResNet* a 18 e 50 strati, nel singolo task di classificazione.

4.1.1 Resnet18

Nella Tab. 4.1 sono riportati i quattro set di parametri che hanno mostrato le migliori prestazioni in termini di *accuratezza* (ACC) utilizzando una *ResNet18*. Dalla tabella emerge che la configurazione con un batch size pari a 4, ottimizzatore Adam con learning rate di 0.001 e nessun dropout ha conseguito l'*accuratezza* più alta, pari a **0.636**.

Batch size	opt	lr	Dropout	ACC
4	Adam	0.001	0.0	0.636
4	Adam	0.001	0.3	0.605
8	Adam	0.001	0.3	0.624

Tabella 4.1: Risultati classificazione: Resnet18

Interessante notare come l'introduzione del dropout abbia influenzato le prestazioni: per esempio, aumentando il dropout a 0.3 con lo stesso batch size, l'*accuratezza* si è leggermente ridotta a **0.605**. Allo stesso modo, anche l'aumento della dimensione del batch a 8 ha portato a risultati leggermente inferiori, con un picco di *accuratezza* del *0.624* per un dropout del 0.3. Da sottolineare è il fatto che configurazioni con batch size più grandi, come 16 e 32, non hanno prodotto risultati ottimali. Questa mancanza di efficacia può essere parzialmente attribuita alla possibilità che batch più ampi conducano a una stima meno precisa del gradiente, influenzando negativamente l'efficacia dell'addestramento. Infatti, considerando la natura complessa e altamente variabile delle immagini ecografiche, un batch size maggiore potrebbe non essere il più adatto per gestire le sfide specifiche presentate da questi dati. In un batch di dimensioni ridotte, ogni immagine contribuisce in modo più significativo al calcolo del gradiente durante l'addestramento, consentendo alla rete di adattarsi e imparare in modo più efficace da ciascun esempio. Ciò favorisce una migliore focalizzazione sui dettagli fini e specifici di ciascuna immagine, riducendo il rischio di generalizzazioni eccessive e "confusione" che potrebbero emergere con batch più grandi. Va inoltre sottolineato che si è tentato anche di utilizzare una versione della rete *ResNet18* pre-addestrata sul dataset ImageNet. Tuttavia, questa strategia non ha portato a miglioramenti significativi nei risultati. Questo può essere attribuito alla notevole differenza tra le caratteristiche delle immagini ecografiche analizzate e quelle presenti in ImageNet. Infatti, quelle ecografiche presentano sfide uniche in termini di qualità dell'immagine e di caratteristiche visive, che sono sostanzialmente diverse e più complesse rispetto a quelle generalmente rappresentate nel dataset ImageNet. Pertanto, l'applicazione nella classificazione di modelli pre-addestrati su un set di

dati così difficile ha mostrato una limitata efficacia, evidenziando la necessità di strategie di addestramento e ottimizzazione più specifiche per questa tipologia di immagini. In sintesi, le configurazioni con batch size minori e un livello di dropout bilanciato hanno dimostrato prestazioni superiori, riflettendo l'importanza di una scelta accurata di questi parametri nel contesto della classificazione di immagini ecografiche con *ResNet18*.

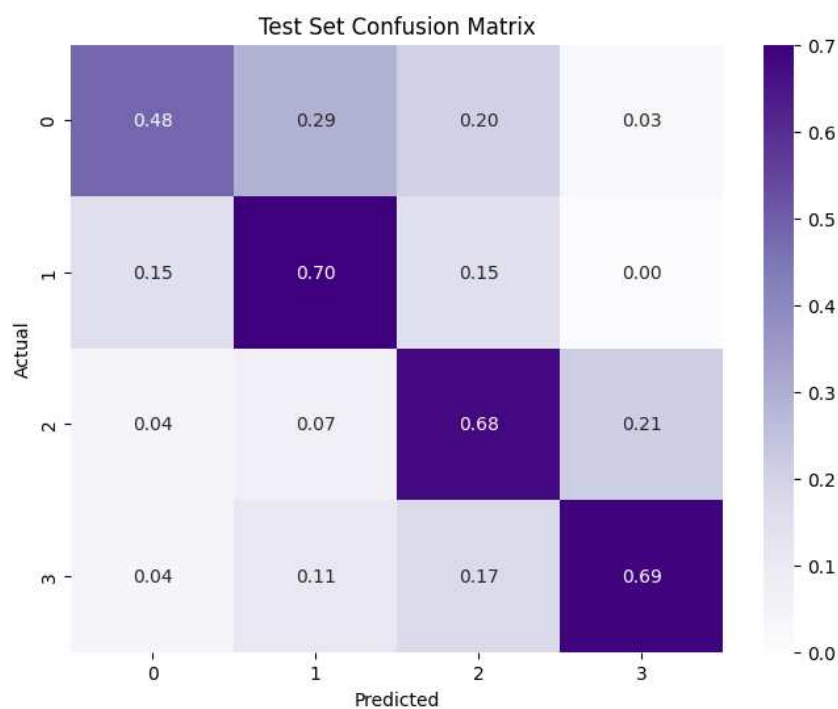


Figura 4.1: Matrice di confusione ottenuta con una Resnet18 per la classificazione

La matrice di confusione mostrata nella Fig. 4.1 riflette l'efficacia del miglior modello di classificazione sul set di test, evidenziando la capacità di preservare l'ordine naturale delle classi. La diagonale principale mostra valori elevati di precisione per le classi individuali: 70% per la classe 1, 68% per la classe 2 e 69% per la classe 3. Questi dati indicano che la maggior parte delle predizioni del modello sono corrette. Errori di classificazione sono presenti ma tendono a verificarsi tra classi adiacenti, il che è meno problematico in un contesto biomedico a cui questa tesi fa riferimento. Significativamente, le classi distanti non sono state confuse tra loro, come dimostrato dall'assenza di valori elevati nelle celle distanti dalla diagonale. Questo conferma che il modello non solo classifica con precisione ma è anche sensibile alla gerarchia di gravità intrinseca delle classi.

4.1.2 Resnet50

La *ResNet50*, è un'architettura di rete neurale convoluzionale significativamente più profonda rispetto alla sua controparte a 18 strati. Questa maggiore profondità

introduce una complessità aggiuntiva che, sebbene possa essere vantaggiosa per dataset ampi e variegati, può risultare controproducente quando applicata a dataset di dimensioni ridotte o con una limitata diversità di esempi. In questi casi, la rete più profonda può tendere all'overfitting, adattandosi eccessivamente ai dati di addestramento e perdendo la capacità di generalizzare a nuovi dati. I risultati ottenuti con l'uso della *ResNet50* sembrano rispecchiare questa problematica. Come si può notare nella Tab. 4.2, le accuratezze raggiunte sono in media inferiori a quelle osservate con la *ResNet18*, suggerendo che la maggiore profondità della rete non si traduce in una migliore prestazione di classificazione per il dataset in questione.

Per rafforzare ulteriormente quest'ipotesi, si è osservato che l'adozione di un livello di dropout più elevato si è rivelata una tattica efficace per migliorare le prestazioni della rete. L'incremento di questo parametro si è dimostrato un passaggio critico nel moderare la complessità di una struttura profonda come quella della *ResNet50*. Questo approccio di regolarizzazione agisce come un fattore di semplificazione essenziale, disincentivando la rete dal dipendere troppo da singole connessioni neuronali e favorendo invece una rappresentazione dei dati più generalizzata e robusta. Tale dinamica è evidente nell'analisi dei dati presentati nella tabella, dove le configurazioni con un livello di dropout più alto hanno conseguito risultati migliori, sottolineando il delicato equilibrio tra la profondità del modello e la sua capacità di generalizzare, specialmente in scenari dove i dati sono limitati o particolarmente specifici.

Batch size	opt	lr	Dropout	ACC
8	Adam	0.001	0.2	0.622
8	Adam	0.001	0.3	0.589
16	Adam	0.001	0.3	0.593

Tabella 4.2: Risultati classificazione: Resnet50

L'analisi della matrice di confusione ottenuta tramite l'uso di una rete *ResNet50* per la classificazione rivela una performance complessivamente inferiore rispetto al modello precedente. Questo è evidente dalla diagonale principale, dove i valori, pur rappresentando la maggioranza delle previsioni corrette, sono meno dominanti rispetto alla matrice precedente, indicando una precisione ridotta nella classificazione. Nonostante la "diagonale sporca", che indica una maggior presenza di errori di classificazione, si osserva un aspetto positivo: la matrice mostra errori trascurabili tra classi con una distanza significativa (ad esempio, tra la classe 0 e la classe 3). Questo suggerisce che, sebbene la *ResNet50* non abbia raggiunto un'alta precisione complessiva, il modello è comunque riuscito a evitare grossi errori ordinando le classi in modo appropriato. In altre parole, anche in presenza di una performance non ottimale, la rete è stata capace di mantenere un certo grado di sensibilità rispetto alla distanza tra le classi. fattore importante in molteplici contesti applicativi in cui gli errori tra classi vicine sono meno gravi di quelli tra classi lontane.

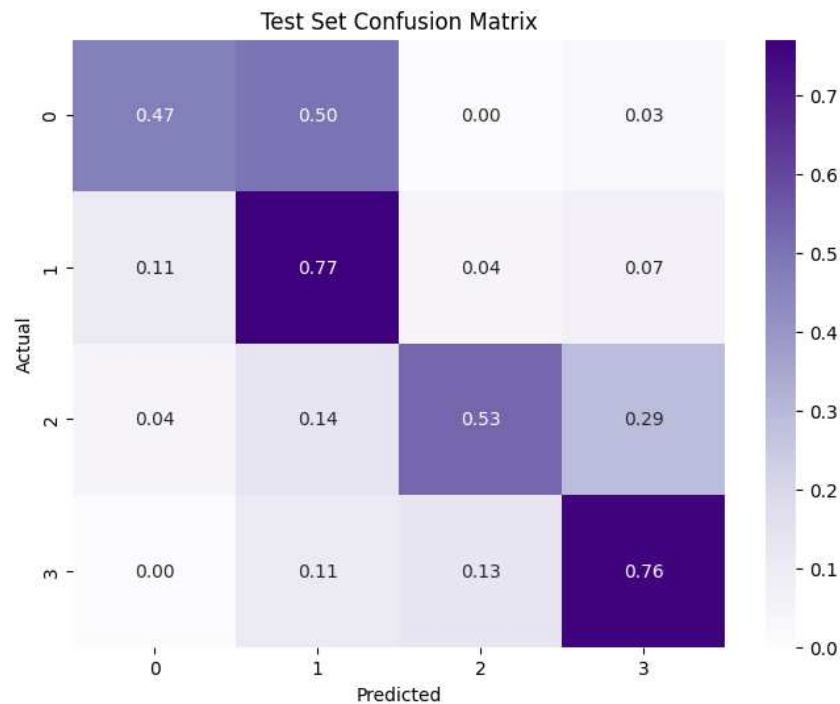


Figura 4.2: Matrice di confusione ottenuta con una Resnet50 per la classificazione

4.2 Segmentazione

Nella sezione dedicata all'analisi dei risultati ottenuti nel compito di segmentazione utilizzando due diverse backbone per le reti **Unet** e **Unet++**. Si pone particolare enfasi sull'esame della backbone **Resnet18**, utilizzata come rete convolutiva principale dalle precedenti ricerche nella letteratura. Parallelamente, si utilizza la backbone **Resnet50** come termine di confronto. Quest'ultimo serve principalmente a validare le ipotesi e i risultati ottenuti con il suo omologo a 18 strati, fornendo così una prospettiva comparativa sulle prestazioni e sulle differenze che emergono dall'uso di ciascun backbone nei due modelli menzionati.

Resnet18

Batch size	TL	opt	lr	Dropout	ACC	Dice
4	from Scratch	Adam	0.0001	0.0	0.923	0.468
4	from Scratch	Adam	0.0001	0.2	0.914	0.469
4	Imegenet	Adam	0.0001	0.2	0.918	0.467

Tabella 4.3: Risultati segmentazione: Unet con backbone Resnet18

In generale, per quanto riguarda la rete Unet i risultati (Tab. 4.3) mostrano che le prestazioni del modello non variano drammaticamente tra le diverse configurazioni, indicando una certa robustezza. Rivolgendo l'attenzione alla Tab. 4.4, si osserva

Batch size	TL	opt	lr	Dropout	ACC	Dice
4	from Scratch	Adam	0.0001	0.2	0.902	0.479
4	Imegenet	Adam	0.0001	0.2	0.924	0.474
4	Imegenet	Adam	0.0001	0.3	0.914	0.473

Tabella 4.4: Risultati segmentazione: Unet++ con backbone Resnet18

che la Unet++, una versione più complessa e nidificata della Unet standard, in generale mostra prestazioni superiori, superando la versione standard di alcuni punti percentuali. I risultati suggeriscono che la Unet++ è capace di catturare dettagli più fini rispetto alla standard, il che la rende una scelta migliore per applicazioni che richiedono una segmentazione molto precisa. Inoltre, si osserva che aumentando il valore del parametro Dropout, i risultati migliorano. Questo suggerisce che la UNet++ necessita di una regolarizzazione più forte per prevenire l'overfitting, probabilmente a causa della sua architettura più complessa che tende a imparare dettagli molto specifici del dataset di addestramento. L'impiego di un batch size ridotto, in particolare di grandezza 4, risulta essere significativamente più efficace, consentendo al modello di concentrarsi meglio sulle aree di interesse, apprendendo dettagli più minuziosi durante la fase di segmentazione. Questo approccio è particolarmente vantaggioso per modelli di maggiore complessità e per dataset difficili e dettagliati, come quello dell'ICLUS.



Figura 4.3: Esempio segmentazione sbagliata con accuratezza alta

Nel contesto dell'analisi delle immagini mediche, è importante considerare attentamente le metriche utilizzate per valutare la segmentazione. Come discusso nelle Sez. 3.4.2, l'accuratezza potrebbe essere un indice fuorviante a causa della sua natura intrinseca. Questa metrica misura quanti pixel sono stati correttamente identificati come background e quanti come parte della maschera di interesse. Tuttavia, nelle immagini mediche, il background occupa generalmente una porzione molto più ampia rispetto alla regione di interesse. Di conseguenza, anche una segmentazione imprecisa può risultare in un'alta accuratezza, semplicemente perché la maggior parte dei pixel appartiene al background. La Figura 4.3 mostra che la maschera predetta copre

un'area significativamente più ampia rispetto alla verità di base portando ad una predizione imprecisa. In questo scenario, erratamente, l'accuratezza della predizione risulterà notevolmente più elevata rispetto alla metrica Dice. Questo è dovuto al fatto che il background, essendo di dimensioni maggiori, è stato predetto con maggiore accuratezza. D'altra parte, il coefficiente Dice, metrica di valutazione alternativa, non è riuscito a superare il 50% (come si può notare nella Tab. 4.4). Questo è principalmente dovuto ai problemi associati al dataset, che sono stati dettagliati nelle sezioni precedenti 3.1.3. Le immagini del dataset ICLUS presentano una notevole difficoltà interpretativa, il che le rende particolarmente sensibili alla conoscenza e all'esperienza di chi ha creato le maschere di verità. Queste maschere hanno mostrato inconsistenze significative, anche tra diversi frame dello stesso paziente, contribuendo a una drastica riduzione delle prestazioni del modello. Questo sottolinea l'importanza di un'accurata preparazione e revisione delle maschere di verità in studi di questo tipo, dove la qualità del dato di partenza è fondamentale per il successo dell'analisi.

La Fig. 4.4 mostra varie predizioni fatte dal miglior modello Unet++. L'immagine mostra che il modello è generalmente in grado di individuare correttamente la regione di interesse. Tuttavia, si sono osservate alcune difficoltà specifiche. In particolare, il modello presenta difficoltà nella segmentazione di maschere di dimensioni ridotte, come evidenziato dai risultati sulle *Test Image 2* e *Test Image 3*. Questo suggerisce una possibile limitazione del modello nell'elaborazione di dettagli più fini o elementi minori all'interno dell'immagine che potrebbero variare da frame a frame, portando rumore nel dataset. Ulteriori problematiche emergono nella segmentazione di immagini acquisite tramite sonda lineare, come dimostrato dai risultati su *Test Image 4*. Le immagini prodotte da questo tipo di sonda sono risultate particolarmente complesse e confusionarie per il modello. Queste difficoltà potrebbero derivare dalla natura del segnale acquisito, che presenta un'elevata variabilità e una bassa qualità del segnale, influenzando negativamente la capacità del modello di segmentare accuratamente la regione d'interesse.

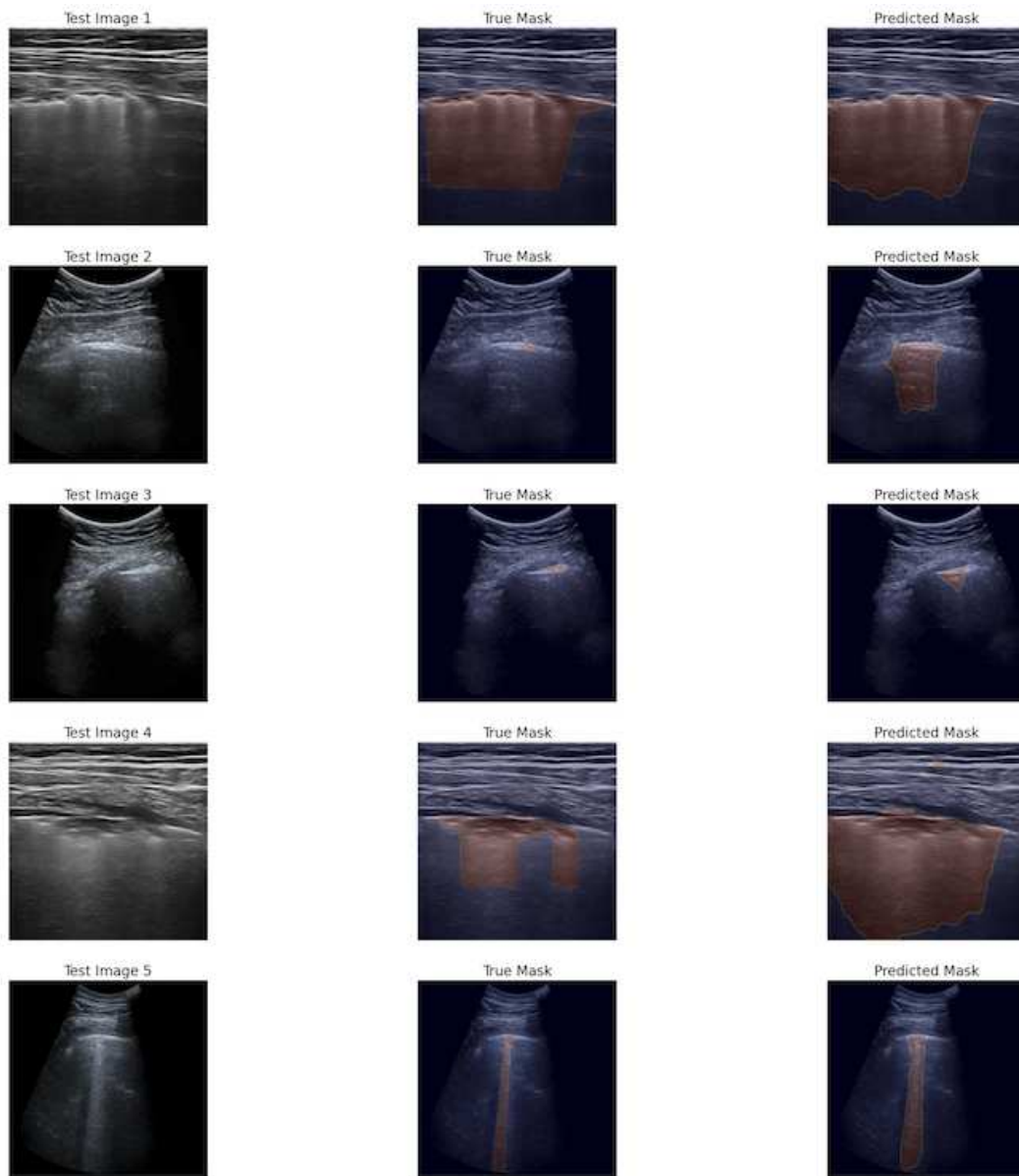


Figura 4.4: Esempio predizioni segmentazione

Resnet50

Nella fase successiva della ricerca, sia il modello U-Net sia il modello U-Net++ sono stati testati utilizzando una struttura ResNet50 come backbone, per verificare se le conclusioni tratte dall'analisi effettuata precedentemente con la ResNet18 fossero applicabili anche in questo nuovo scenario. I risultati ottenuti hanno confermato le ipotesi iniziali: è stato constatato che il modello U-Net++ ha mostrato prestazioni superiori rispetto al modello U-Net tradizionale; tuttavia, per raggiungere tale livello di efficacia, quest'ultimo necessita di un'intervento di regolarizzazione attraverso l'uso del dropout. Questo dettaglio sottolinea una differenza cruciale nell'approccio di ottimizzazione dei due modelli, evidenziando come il U-Net++, pur superior in termini di performance, richieda considerazioni aggiuntive nell'ambito della regolarizzazione per massimizzare il suo potenziale.

Batch size	TL	opt	lr	Dropout	ACC	Dice
16	Imegenet	Adam	0.001	0.1	0.944	0.513
8	Imegenet	Adam	0.001	0.1	0.930	0.503
16	from Scratch	Adam	0.001	0.2	0.922	0.501

Tabella 4.5: Risultati segmentazione: Unet con backbone Resnet50

Batch size	TL	opt	lr	Dropout	ACC	Dice
16	Imegenet	Adam	0.001	0.1	0.930	0.515
16	Imegenet	Adam	0.001	0.2	0.921	0.510
16	Imegenet	Adam	0.001	0.1	0.921	0.508

Tabella 4.6: Risultati segmentazione: Unet++ con backbone Resnet50

4.3 Multitask learning

4.3.1 ResUnet18 e ResUnet18++

ResUnet18

Nell'analisi dei risultati ottenuti dall'adozione di un approccio multitask per la classificazione e la segmentazione, mediante l'utilizzo di un modello Unet con backbone Resnet18, è stato osservato che le prestazioni del metodo multitask superano di alcuni punti percentuali quelle ottenute impiegando i task in modo indipendente.

Dall'analisi delle tabelle sopra riportate è emerso che l'ottimizzazione degli iperparametri tramite grid search ha un impatto, non indifferente, sui risultati ottenuti, con variazioni osservabili in termini di punti percentuali tra la segmentazione e la classificazione del modello multitask. Tale sensibilità può essere attribuita alla complessità e alla varietà delle immagini presenti nel dataset ICLUS, viste in precedenza (Sez. 3.1.3), che rappresentano una sfida significativa per il modello nella

BS	TL	opt	lr	DP	ACC CLS	ACC SEG	Dice
8	from Scratch	Adam	0.001	0.0	0.639	0.928	0.456
8	Imegenet	Adam	0.001	0.1	0.642	0.927	0.469
16	from Scratch	Adam	0.001	0.3	0.620	0.923	0.428

Tabella 4.7: Risultati multitask learning - focus classificazione: Unet con backbone Resnet18

BS	TL	opt	lr	DP	ACC CLS	ACC SEG	Dice
8	from Scratch	Adam	0.001	0.1	0.582	0.911	0.474
16	Imegenet	Adam	0.001	0.0	0.564	0.913	0.482
16	Imegenet	Adam	0.001	0.3	0.504	0.930	0.490

Tabella 4.8: Risultati multitask learning - focus segmentazione: Unet con backbone Resnet18

sua capacità di identificare caratteristiche comuni utili sia per la classificazione sia per la segmentazione. In questo contesto, sono state presentate due tabelle distinte: la prima evidenzia i risultati migliori ottenuti nel multitask per la classificazione (Tab .4.7), mentre la seconda mostra i migliori risultati per la segmentazione (Tab .4.8). Si osserva che, quando gli iperparametri sono ottimizzati per favorire un particolare task, tendono a penalizzare l'altro, determinando variazioni di alcuni punti percentuali nelle performance. Questo fenomeno sottolinea un trade-off intrinseco nell'ottimizzazione degli iperparametri per task multipli, dove il miglioramento in un'area può comportare un leggero decremento nelle prestazioni in un'altra. Tale osservazione è cruciale nell'ambito dell'apprendimento multiplo, poiché indica la necessità di un equilibrio delicato nella scelta e nell'aggiustamento degli iperparametri per massimizzare l'efficacia complessiva del modello su compiti diversificati ma soprattutto complessi (come quelli esposti dal dataset ICLUS). L'impiego del transfer learning da ImageNet ha dimostrato di essere benefico, particolarmente per il miglioramento dei risultati nei task di segmentazione. L'inizializzazione dei pesi con valori preaddestrati anziché casuali sembra conferire al modello una base di conoscenza più ampia e pertinente, facilitando l'apprendimento e l'adattamento ai dati specifici del problema in esame. Per quanto riguarda la dimensione del batch (BS) e il tasso di dropout, è emerso che valori medi di questi parametri contribuiscono a ottimizzare le prestazioni del modello. Un BS di 8 e un tasso di dropout di 0.1 sembrano bilanciare efficacemente l'apprendimento, evitando sia l'overfitting che l'underfitting, come evidenziato dalle prestazioni nelle due tabelle.

La Figura 4.5 presenta la matrice di confusione generata utilizzando il modello ResUnet18, che ha conseguito i migliori risultati in termini di classificazione. Analogamente a quanto osservato nella matrice di confusione relativa alla classificazione (vedi Figura 4.1), la diagonale principale appare pronunciata, fatta eccezione per le coppie di score 0 - 1 e 2 - 3, le quali rappresentano etichette particolarmente sfidanti

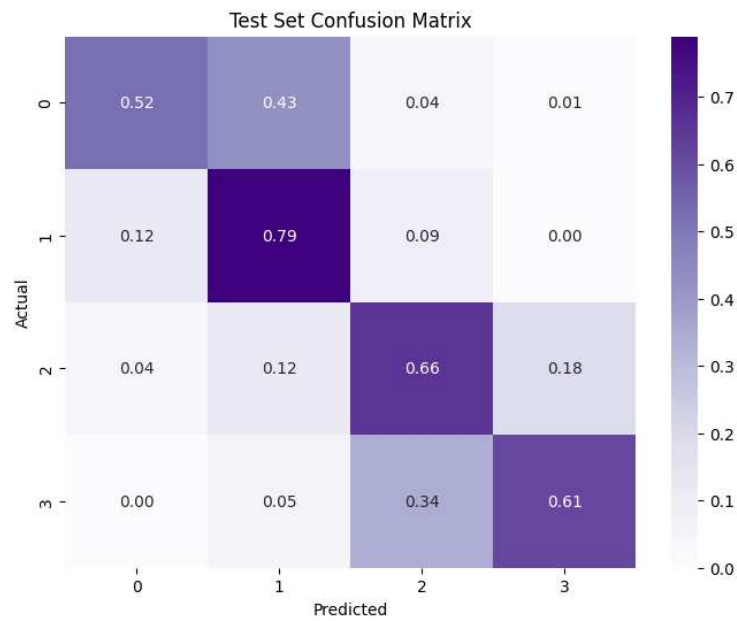


Figura 4.5: Matrice di confusione ResUnet18 con focus alla classificazione

da distinguere, anche visivamente da un occhio esperto. Questa difficoltà si riflette anche nella Figura 4.6, che illustra le predizioni effettuate dal medesimo modello su immagini di particolare complessità.

Dall'analisi delle predizioni sopra citate si evince che, quando il modello commette errori nella classificazione, emergono parallelamente difficoltà anche nella segmentazione. Nonostante ciò, il modello riesce nella maggior parte dei casi a identificare la regione di interesse. Inoltre, gli errori si concentrano principalmente sulle coppie di score che sono state precedentemente identificate come più difficili da distinguere.

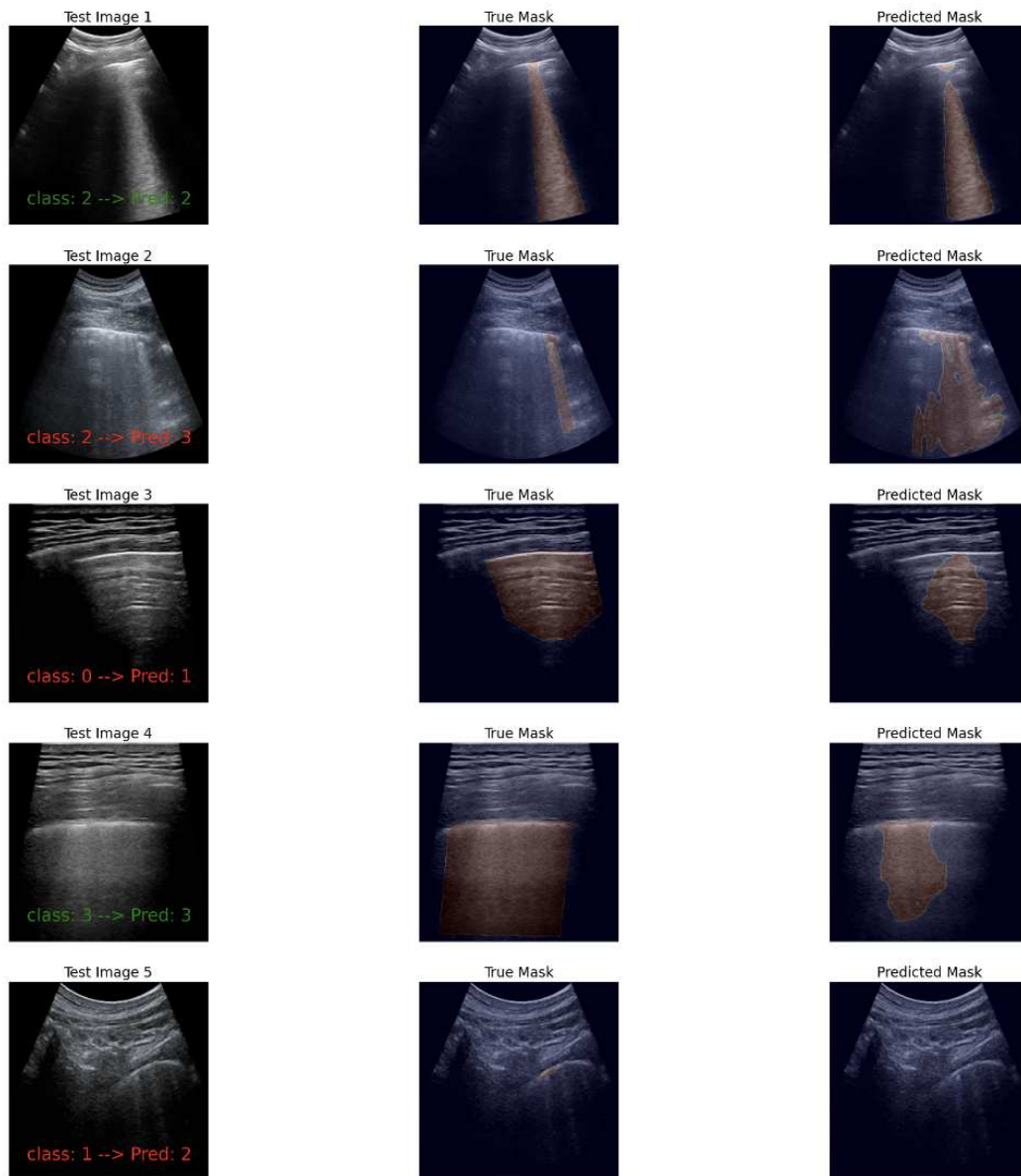


Figura 4.6: Predizioni ResUnet18 con focus alla classificazione

ResUnet18++

BS	TL	opt	lr	DR	ACC CLS	ACC SEG	Dice
8	from Scratch	Adam	0.001	0.2	0.600	0.928	0.459
8	Imegenet	Adam	0.001	0.3	0.660	0.913	0.492
16	Imegenet	Adam	0.001	0.0	0.601	0.929	0.432

Tabella 4.9: Risultati multitask learning - focus classificazione: Unet++ con backbone Resnet18

BS	TL	opt	lr	DR	ACC CLS	ACC SEG	Dice
8	from Scratch	Adam	0.001	0.2	0.600	0.928	0.459
8	Imegenet	Adam	0.001	0.3	0.660	0.913	0.492
16	from Scratch	Adam	0.001	0.2	0.570	0.919	0.463

Tabella 4.10: Risultati multitask learning - focus segmentazione: Unet++ con backbone Resnet18

L'esame dei risultati ottenuti dal modello ResUnet18++ (Unet++ con backbone Resnet8) rivela un trend di miglioramento coerente con i progressi osservati nel singolo task di segmentazione. La configurazione di iperparametri che include un batch size di 8, un tasso di dropout di 0.3 e l'uso di pesi preaddestrati da ImageNet si distingue per la sua efficacia, evidenziando miglioramenti nelle prestazioni sia nella classificazione sia nella segmentazione rispetto alle configurazioni valutate nei modelli precedenti. Interessante notare è la stabilità dei risultati tra le due tabelle. Indipendentemente dal focus posto sulla classificazione (Tab. 4.9) o sulla segmentazione (Tab. 4.10), emergono due configurazioni di iperparametri che si dimostrano vincenti in entrambe le situazioni. Questo suggerisce una solidità del modello nei confronti dei due task, un fattore cruciale per un approccio multitask. L'architettura Unet++, rivelatasi più precisa rispetto alla Unet tradizionale, quando combinata con una backbone Resnet18 e un appropriato tasso di dropout, mostra un'efficacia notevole. Questa combinazione non solo supera le prestazioni dei modelli dedicati ai singoli task ma, nel caso della classificazione, supera anche i risultati ottenuti dal modello multitask precedente. In termini di generalizzazione, il tasso di dropout di 0.3 appare ottimale nel contrastare l'overfitting, permettendo ad un modello complesso come la Unet++ di mantenere una buona capacità di apprendimento senza perdere la capacità di generalizzare da dati non visti.

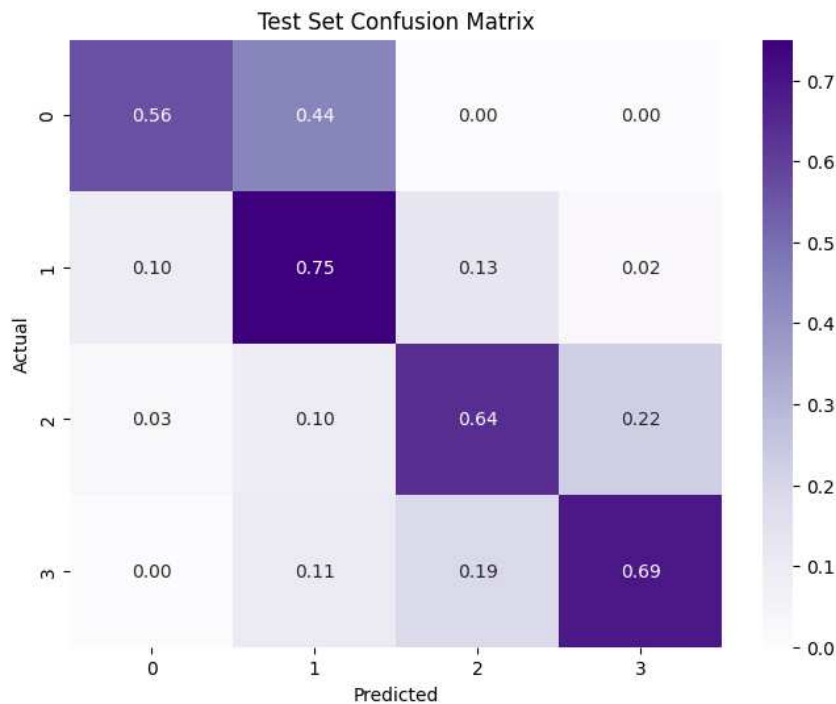


Figura 4.7: Matrice di confusione ResUnet18++

La matrice di confusione del modello più performante mostra risultati incoraggianti nella classificazione (Fig. 4.7). Analogamente agli esempi precedenti, la diagonale principale della matrice rivela una notevole concentrazione, segnalando l'efficacia del modello nel prevedere accuratamente le varie categorie. Tuttavia, si nota una problematica ricorrente: gli errori di classificazione tra le categorie 0 e 1, un fenomeno già riscontrato in analisi precedenti. Questo fenomeno potrebbe essere attribuito alla limitata quantità di dati etichettati come categoria 1, rispetto a una presenza più ampia di dati etichettati come categoria 0. Questo sbilanciamento nella distribuzione dei dati, insieme alla complessità nell'interpretazione delle ecografie del polmone, potrebbe creare una sfida nell'effettuare una distinzione accurata tra le due categorie.

Nella Fig. 4.8, si osserva la capacità del modello di elaborare immagini complesse, in particolare quelle caratterizzate da maschere di segmentazione di piccole dimensioni e prodotte da sonde lineari. Queste caratteristiche, come detto in precedenza, rappresentano tipicamente punti critici per la rete, a causa della difficoltà nel riconoscere pattern geometrici chiari. Risulta evidente che il modello non solo mostra limitazioni nella classificazione accurata di tali immagini, ma anche nella loro corretta segmentazione. Più specificatamente, il modello tende a sovra-segmentare le regioni di interesse, piuttosto che a identificarle con precisione.

4.3 Multitask learning

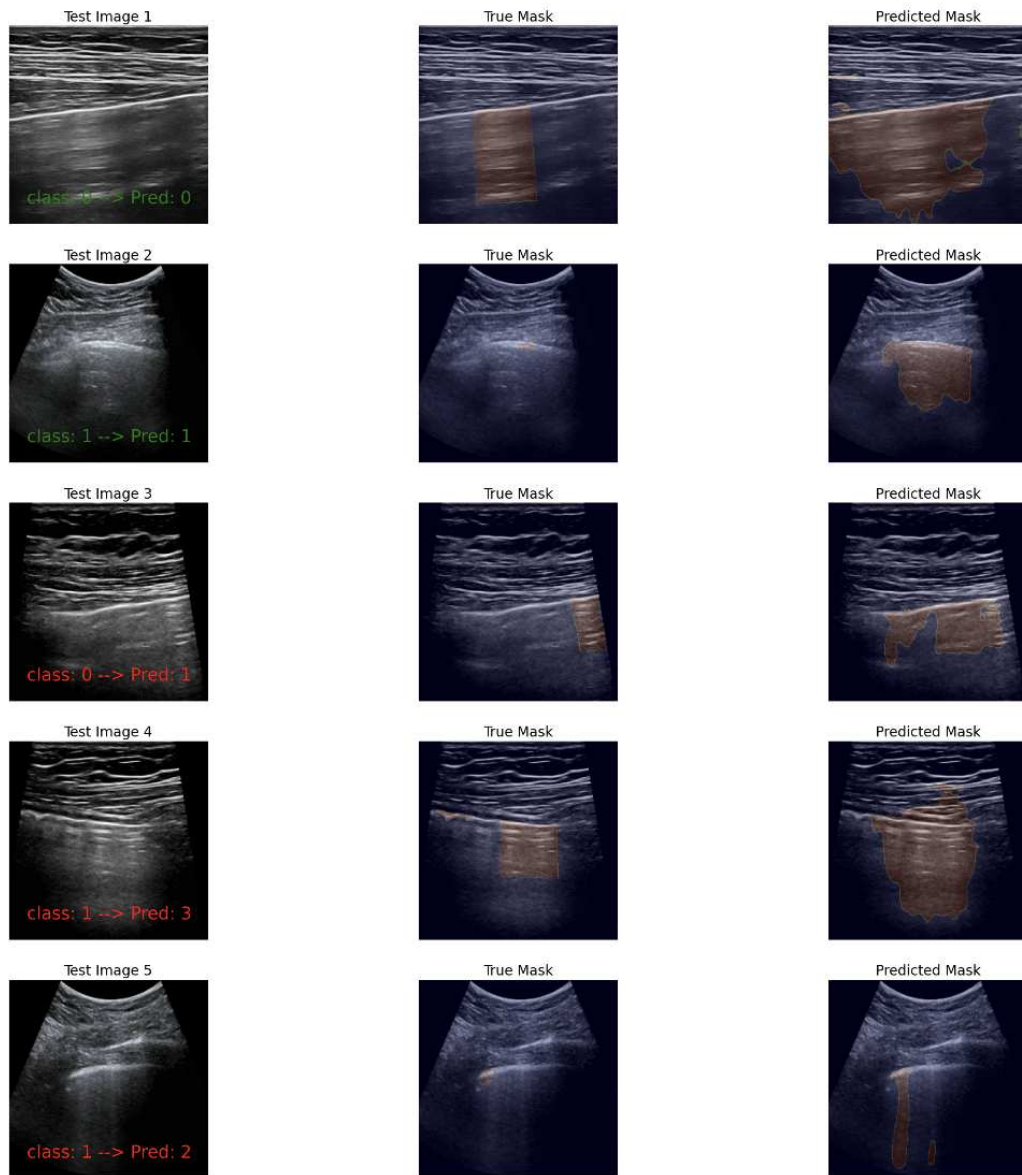


Figura 4.8: Predizioni ResUnet18++ su immagini complesse

Nelle seguenti pagine, verranno presentate alcune figure che illustrano l'andamento delle predizioni per epoca durante il processo di addestramento del modello preso in esame. Queste figure consentono di osservare l'evoluzione dell'addestramento e forniscono non solo le predizioni per la classificazione e la segmentazione, ma anche delle mappe di calore (heatmap) che risultano fondamentali per comprendere le aree in cui la rete neurale focalizza la sua attenzione al fine di catturare le caratteristiche più significative delle immagini, che saranno poi tradotte in maschere di segmentazione. Dall'epoca 0 (Fig. 4.9), è evidente che il modello iniziale riscontra difficoltà nell'individuare la regione d'interesse, concentrandosi inizialmente in modo casuale sulle caratteristiche salienti dell'immagine. Tuttavia, man mano che l'addestramento prosegue, il modello mostra un notevole miglioramento nelle sue capacità predittive per entrambi i compiti. Già dall'epoca 10 (Fig. 4.10), il modello riesce a discriminare con precisione le classi delle tre immagini di test e, allo stesso modo, già dall'epoca 2 è in grado di delineare con successo le regioni di interesse nelle immagini. Prendendo in esame la segmentazione, con il progredire delle epoche, è possibile osservare come il modello raffina progressivamente la maschera predetta fino a raggiungere una versione quasi identica al ground truth. Questo processo dimostra l'abilità del modello nell'apprendere e catturare le features delle immagini nel corso dell'addestramento, permettendo una segmentazione sempre più accurata delle regioni d'interesse. Le predizioni effettuate durante il processo di addestramento sono acquisite ogni 5 epoche anziché alla migliore epoca, questo significa che il monitoraggio delle prestazioni del modello avviene a intervalli regolari nel corso dell'addestramento. Tuttavia, poiché l'addestramento di reti neurali può comportare oscillazioni nelle prestazioni, ci sono momenti in cui le predizioni possono essere meno accurate rispetto ad altre. Un esempio di questo fenomeno può essere osservato all'epoca 15 (Fig. 4.10), in cui la segmentazione dell'immagine di test 3 ha prodotto un risultato inaspettato, segmentando completamente un'altra regione anziché quella d'interesse desiderata. Questo rappresenta un caso in cui il modello ha mostrato una significativa fluttuazione nelle sue prestazioni, tipico di modelli addestrati su un dataset altamente complesso e variabile. Inoltre, all'epoca 25 (Fig. 4.11), nel test image 1, il modello ha espanso l'area prevista ricoprendo l'intera regione più luminosa dell'immagine, erroneamente interpretata come zona d'interesse. Questo episodio mette in evidenza come l'addestramento possa portare a interpretazioni errate in alcune circostanze dettate dalla complessa natura delle immagini di input.

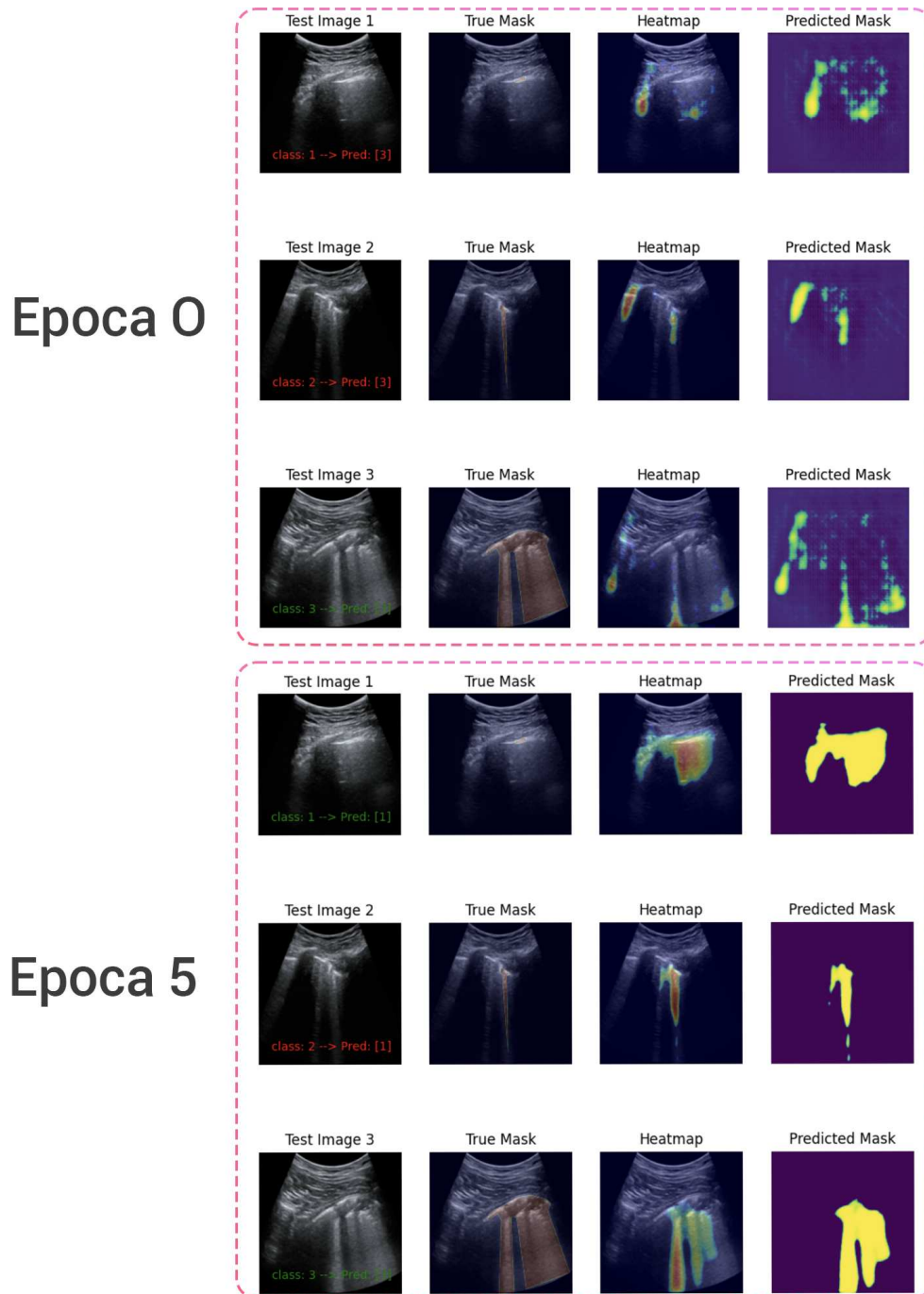


Figura 4.9: Andamento predizioni nell'addestramento di ResUnet18++ epoca 0 e 5

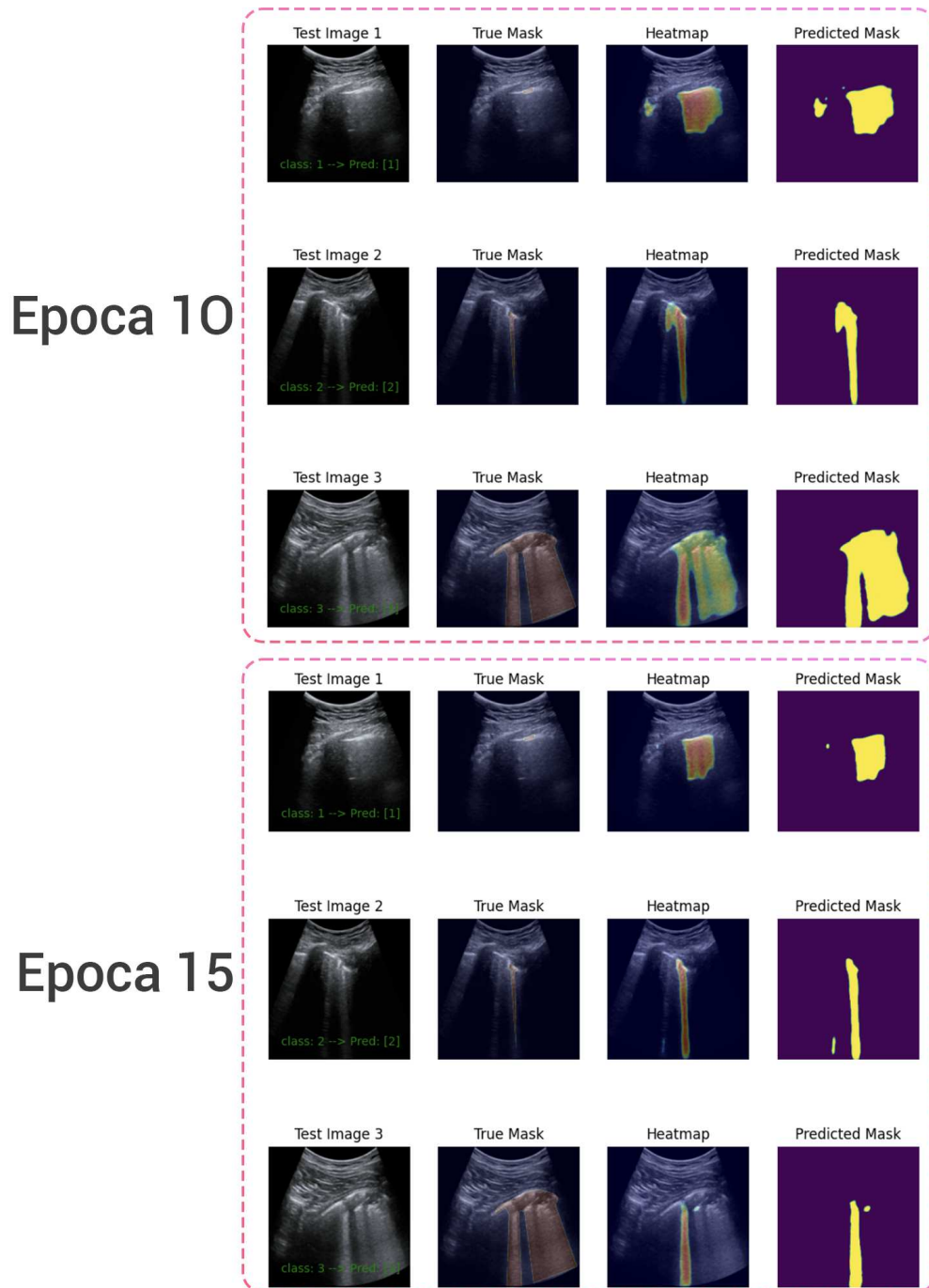


Figura 4.10: Andamento predizioni nell'addestramento di ResUnet18++ epoca 10 e 15



Figura 4.11: Andamento predizioni nell'addestramento di ResUnet18++ epoca 20 e 25

4.3.2 ResUnet50 e ResUnet50++

Successivamente, si procede con una breve analisi dei risultati ottenuti tramite l'impiego dei modelli ResUnet50 e ResUnet50++.

ResUnet50

BS	TL	opt	lr	DP	ACC CLS	ACC SEG	Dice
16	from Scratch	Adam	0.001	0.2	0.631	0.904	0.491
4	Imegenet	Adam	0.001	0.2	0.615	0.930	0.484
16	Imegenet	Adam	0.001	0.2	0.603	0.932	0.491

Tabella 4.11: Risultati multitask learning: Unet con backbone Resnet50

Nella valutazione delle performance del modello multitask ResUnet50, è emerso che, sebbene i risultati in termini assoluti non superano quelli ottenuti con una backbone Resnet18, il modello dimostra una maggiore stabilità. Questo si evince dall'analisi degli iperparametri ottimali che conducono a risultati migliorativi sia nella segmentazione che nella classificazione. Inoltre, l'approccio multitask supera di pochi punti percentuali il modello basato su singolo task, evidenziando anche in questo caso un leggero vantaggio dell'approccio multiplo. Tale constatazione è supportata dai risultati ottenuti, che mostrano come i migliori iperparametri stabiliscano un equilibrio efficace tra i due task, contribuendo a una performance complessiva più robusta e affidabile. Analizzando la Tab. 4.11 si deduce che, data la maggiore complessità architettonica della rete Resnet50 rispetto al suo omologo meno complesso Resnet18, si rende necessario un incremento del parametro di dropout per garantire risultati performanti. Infatti, un dropout impostato a 0.2 sembra contribuire positivamente ai risultati del modello, permettendo di raggiungere un equilibrio tra la capacità di generalizzazione e la capacità di apprendimento del dettaglio fornita dalla maggiore profondità della rete.

ResUnet50++

BS	TL	opt	lr	DP	ACC CLS	ACC SEG	Dice
16	from Scratch	Adam	0.001	0.2	0.630	0.923	0.501
16	from Scratch	Adam	0.001	0.1	0.613	0.916	0.484
16	Imegenet	Adam	0.001	0.1	0.623	0.918	0.481

Tabella 4.12: Risultati multitask learning: Unet con backbone Resnet50++

I risultati presentati nella Tab. 4.12 riflettono l'efficacia di un approccio multitask che combina le architetture Unet++ e Resnet50, qui chiamato come ResUnet50++. Questo modello sembra aver ereditato la stabilità e le prestazioni del modello precedente, ResUnet50, offrendo miglioramenti significativi. Anche in questo caso il

parametro di dropout, usato come metodo di regolarizzazione, aiuta a prevenire il sovra-adattamento del modello, consentendo una miglior generalizzazione sui nuovi dati. Il batch size di 16, che è relativamente piccolo per i moderni standard di addestramento di reti neurali, suggerisce che il modello ResUnet50++ è capace di generalizzare efficacemente anche senza la necessità di grandi batch size. Ciò potrebbe indicare che la complessità intrinseca di ResUnet50++ consente di apprendere rappresentazioni robuste dei dati, anche quando gli esempi forniti in ogni iterazione dell'addestramento sono limitati.

4.3.3 Studio della Loss

I risultati precedentemente presentati sono stati ottenuti impiegando una loss multitask non ponderata. In tale approccio, le loss relative ai due task di classificazione e segmentazione sono state sommate direttamente, senza applicare un peso specifico a ciascuna, vedi formula sotto riportata.

$$L_{multitask} = L_{cls} + L_{seg} \quad (4.1)$$

Nella presente sezione, l'attenzione è rivolta sul bilanciamento delle loss relative ai due compiti e sull'utilità di questo approccio. Questo viene realizzato variando il peso lambda, il quale può assumere valori compresi tra 0 e 1.

$$L_{multitask} = \lambda L_{cls} + (1 - \lambda) L_{seg} \quad (4.2)$$

Resnet18

Per l'analisi dei modelli multitask con backbone Resnet18 i risultati sono stati conseguiti utilizzando i migliori iperparametri identificati nelle tabelle precedenti (riportati nella Tab 4.13). Tale analisi mira a comprendere in che misura la variazione del peso lambda possa influire sull'efficacia complessiva del modello multitask nelle operazioni di classificazione e segmentazione.

BS	TL	opt	lr	Dropout
8	Imagenet	Adam	0.001	0.1

Tabella 4.13: Iperparametri migliori per l'addestramento di ResUnet18 e ResUnet18++

L'indagine condotta si è concentrata sull'esame dei modelli di rete neurale ResUnet18 (Tab. 4.14) e ResUnet18++ (Tab. 4.15). Dall'analisi effettuata, è emerso che un bilanciamento equo tra le funzioni di classificazione e segmentazione è fondamentale per migliorare le prestazioni complessive dei modelli multitask. Si è osservato che in entrambi i modelli, una diminuzione del valore di lambda (λ) tende a favorire la segmentazione, mentre un incremento dello stesso valore privilegia la classificazione. È tuttavia essenziale sottolineare che un eccessivo sbilanciamento verso uno dei due

compiti, come nel caso di un valore di lambda pari a 1 o 0, porta a una riduzione significativa delle prestazioni complessive in entrambi i modelli, andando ad azzerare una delle due componenti del calcolo della loss. La ragione di ciò può essere attribuita alla complessità intrinseca delle immagini ecografiche che rappresenta una sfida per il modello nell'identificazione di feature utili per entrambi i compiti. Di conseguenza, un'alterazione del peso della loss favorisce lo sviluppo di features specifiche per un singolo compito, a discapito dell'altro, limitando così la capacità del modello di apprendere rappresentazioni utili per entrambi i task. Concludendo, un approccio bilanciato sembra essere più efficace nel massimizzare le prestazioni dei modelli multitask, specialmente in studi che coinvolgono immagini mediche complesse.

λ	ACC CLS	ACC SEG	Dice
0	0.290	0.915	0.456
0.25	0.563	0.922	0.440
0.5	0.594	0.924	0.503
0.75	0.572	0.923	0.423
1	0.576	0.320	0.150

Tabella 4.14: Studio del peso lambda nella loss del modello ResUnet18

λ	ACC CLS	ACC SEG	Dice
0	0.226	0.905	0.466
0.25	0.501	0.921	0.471
0.5	0.601	0.922	0.491
0.75	0.532	0.920	0.484
1	0.535	0.301	0.124

Tabella 4.15: Studio del peso lambda nella loss del modello ResUnet18++

Resnet50

Successivamente, l'attenzione si focalizza sull'analisi dei modelli multitask con backbone ResNet50. I parametri impiegati sono riportati nella Tab. 4.16. L'obiettivo principale è indagare le variazioni nel peso della funzione di loss, al fine di determinare la coerenza di questi risultati con quelli precedentemente ottenuti utilizzando la backbone Resnet18.

BS	TL	opt	lr	Dropout
16	From Scratch	Adam	0.001	0.2

Tabella 4.16: Iperparametri migliori per l'addestramento di ResUnet50 e ResUnet50++

Le evidenze presentate nelle due tabelle, sotto riportate, supportano le ipotesi formulate precedentemente riguardo al modello in esame. È chiaro che, anche in

queste situazioni, l'impostazione della funzione di loss con un bias a favore di un task rispetto all'altro non risulta vantaggiosa per l'efficacia complessiva del modello multitasking. Questo fenomeno è evidente dalle tabelle analizzate, dove si osserva che un approccio equilibrato migliora le prestazioni sia nella classificazione che nella segmentazione.

λ	ACC CLS	ACC SEG	Dice
0	0.160	0.919	0.501
0.25	0.541	0.932	0.454
0.5	0.593	0.938	0.525
0.75	0.475	0.902	0.450
1	0.580	0.800	0.136

Tabella 4.17: Studio del peso lambda nella loss del modello ResUnet50

λ	ACC CLS	ACC SEG	Dice
0	0.222	0.927	0.510
0.25	0.540	0.931	0.448
0.5	0.609	0.906	0.511
0.75	0.313	0.542	0.233
1	0.550	0.748	0.116

Tabella 4.18: Studio del peso lambda nella loss del modello ResUnet50++

4.3.4 Confronti

Nella presente, si procede con un'analisi dei confronti tra i risultati conseguiti dal modello multitask e quelli ottenuti attraverso l'utilizzo di task singoli. Entrambi gli approcci si avvalgono di Resnet18 come backbone principale. Questo dettaglio è cruciale, in quanto permette di stabilire un punto di confronto omogeneo tra i due metodi, focalizzandosi sulle differenze prestazionali attribuibili alla natura multitask rispetto all'approccio basato su singoli task. Questa sotto-sezione è articolata in due parti distinte: la prima è dedicata al modello multitask ResUnet, mentre la seconda esplora la sua variante implementata con Unet++. Per tutti e due i modelli, inizialmente, si procede con la presentazione dei valori delle metriche, fondamentali per valutare l'efficacia dei modelli in esame. Successivamente, si procede al confronto delle matrici di confusione, strumenti essenziali nell'analisi di modelli di classificazione, in quanto permettono di visualizzare le performance del modello in termini di corrette classificazioni e di tipi di errori commessi. Attraverso questo confronto, è possibile individuare punti di forza e debolezza dei modelli multitask ResUnet e ResUnet++ rispetto all'approccio a singolo task, fornendo una visione più approfondita e completa dell'efficacia del modello proposto nell'ambito di questo elaborato.

Esaminando i dati presentati nelle Tabelle 4.19 e 4.20, si osserva come l'impiego dei modelli multitask influenzi i risultati in termini di classificazione (ACC CLS) e

Modello	ACC CLS	ACC SEG	Dice
ResUnet18	0.642	0.927	0.469
Resnet18	0.636	-	-
Unet	-	0.914	0.469

Tabella 4.19: Confronto risultati ResUnet18, Resnet18 e Unet

Modello	ACC CLS	ACC SEG	Dice
ResUnet18++	0.660	0.913	0.492
Resnet18	0.636	-	-
Unet++	-	0.902	0.479

Tabella 4.20: Confronto risultati ResUnet18++, Resnet18 e Unet++

segmentazione (ACC SEG e Dice) delle ecografie polmonari. Nella prima (Tab. 4.19), il focus è rivolto verso l'utilizzo di una Unet classica per la parte di segmentazione, mentre la seconda (Tab. 4.20) evidenzia l'utilizzo del modello Unet++ in tale contesto. Analizzando i risultati ottenuti, emerge che l'impiego dei modelli multitask ha portato a un miglioramento complessivo delle prestazioni. Nello specifico, nel campo della segmentazione, il beneficio apportato è relativamente modesto. Tuttavia, nel contesto della classificazione, si osserva un aumento significativo dell'accuratezza: ResUnet18 ha contribuito a un incremento dell'1%, mentre ResUnet18++ ha registrato un miglioramento del 3%. La maggiore efficacia dei risultati ottenuti con la Unet++ può essere ricondotta alla sua architettura sofisticata. Questa struttura fornisce al modello una precisione superiore e la capacità di identificare caratteristiche più dettagliate e descrittive, particolarmente preziose in ambiti dove le immagini non presentano pattern geometrici definiti, una sfida notevole per le CNN tradizionali. Nonostante l'adozione di tali modelli avanzati, la difficoltà di segmentare le immagini nel dominio specifico è evidente. In entrambi i casi, sia nella soluzione multitask che a singolo task, il valore del coefficiente Dice non supera il 50%. Nel caso dell'approccio multiplo, potrebbe indicare che le features estratte in questo particolare dominio non hanno una forte correlazione diretta tra classificazione e segmentazione, suggerendo che, sebbene ci sia un miglioramento in termini generali di prestazione, questo non è particolarmente marcato. Pertanto, si può concludere che, sebbene il modello multitask e le innovazioni nella struttura della Unet++ portino a dei miglioramenti, queste tecniche da sole non riescono a superare determinate limitazioni intrinseche nell'analisi di immagini complesse nel dominio in esame.

L'analisi delle matrici di confusione presentate in Fig. 4.12 rivela differenze significative tra gli approcci di classificazione singolo task e multitask. È possibile osservare che le tre matrici mostrano una distribuzione uniforme e una diagonale principale ben definita, indicando un'alta percentuale di corretta classificazione. Ciò che distingue gli approcci multitask dal singolo task è la difficoltà riscontrata nel distinguere tra le

immagini etichettate con 0 e 1. Tuttavia, le predizioni nei modelli multitask sono più concentrate lungo la diagonale, risultando in una completa assenza di errori nella predizione di valori estremi come 0 e 3. Aspetto particolarmente rilevante in ambito medico, dove la precisione di tale distinzione può essere cruciale. Al contrario, l'approccio singolo task mostra piccole percentuali di errore in tali previsioni. Questa differenza potrebbe essere attribuita al fatto che l'approccio multitask, condividendo le caratteristiche con una rete di segmentazione, possa discriminare più efficacemente le immagini con livelli di gravità distanti. Probabilmente, le features estratte dalla segmentazione enfatizzano meglio i punti critici dell'immagine, focalizzando l'apprendimento sulle aree problematiche. Tale sinergia tra classificazione e segmentazione può portare a una maggiore precisione nelle predizioni, specialmente in quelle situazioni in cui è essenziale distinguere correttamente tra diverse gravità di condizioni mediche.

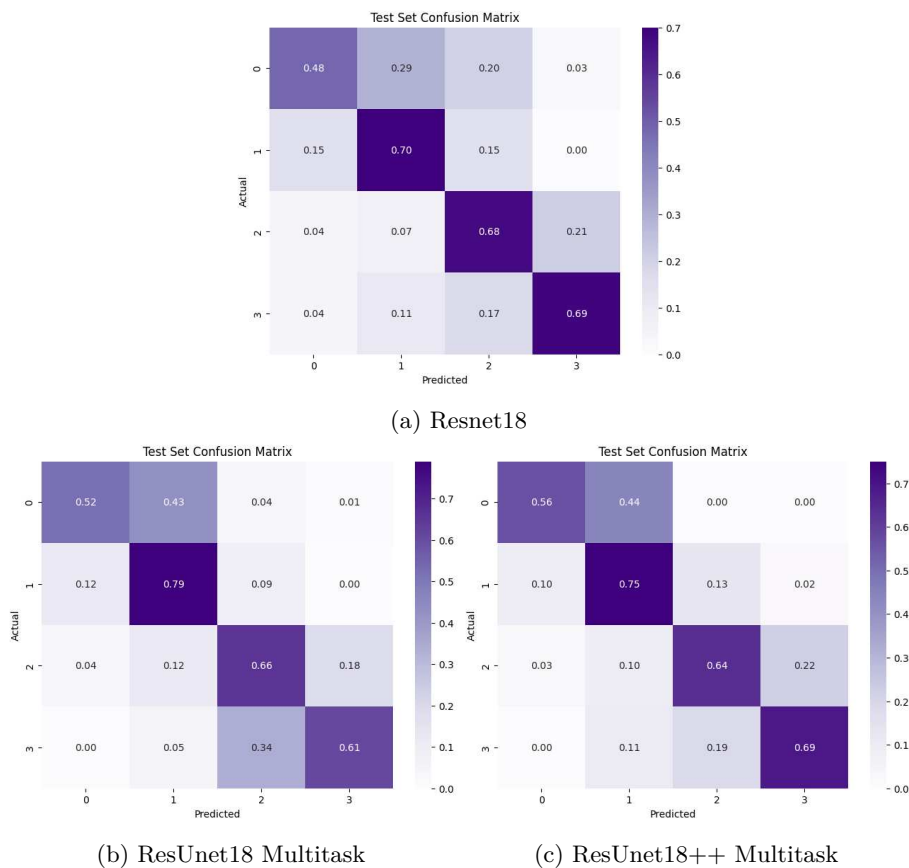


Figura 4.12: Confronto matrici di confusione tra approcci a singolo task e multitask

Capitolo 5

Conclusioni

La complessità e la varietà dei frames ecografici polmonari costituiscono una sfida reale, la quale rende necessario esplorare metodologie avanzate per migliorare l'analisi e l'elaborazione degli stessi. Nell'ambito di questa ricerca, l'obiettivo è valutare l'efficacia di un approccio multitasking per la classificazione e la segmentazione delle ecografie polmonari (LUS). Utilizzando il dataset ICLUS, si è voluto investigare se l'implementazione di tale metodologia potesse migliorare le prestazioni rispetto a strategie basate su singoli task. In particolare, l'attenzione si è focalizzata sulla possibilità che la condivisione di features tra un classificatore e una rete di segmentazione potesse apportare benefici in contesti caratterizzati da immagini di difficile interpretazione.

L'approccio multitasking nella classificazione e segmentazione di ecografie polmonari rappresenta un terreno di ricerca innovativo e fertile, come evidenziato dalla mancanza di precedenti studi nello stato dell'arte. Pertanto, per validare l'efficacia di tale metodologia, si è reso necessario adottare un approccio comparativo indiretto. A tale scopo, sono stati implementati i modelli single-task di riferimento dello stato dell'arte, successivamente sono stati testati sul medesimo set di dati impiegato anche per l'approccio multitasking. I risultati ottenuti dai modelli single-task hanno fornito un punto di riferimento fondamentale per valutare le prestazioni del modello proposto. Questa analisi comparativa ha costituito il fulcro della validazione dei risultati ottenuti, offrendo una panoramica sull'efficienza, sull'efficacia e sui limiti del multitask nel dominio specifico dello studio.

I risultati ottenuti nel corso di questo elaborato, sono frutto di un'approfondita ottimizzazione degli iperparametri attraverso l'impiego del Grid Search Bayesiano. Questo approccio ha consentito di individuare gli insiemi ottimali di parametri per ciascun modello, sia nella configurazione single task che multitask, contribuendo così a massimizzare le performance. La fase di valutazione ha coinvolto diverse metriche chiave per l'analisi delle prestazioni. La metrica *Dice* è stata impiegata per valutare la bontà della segmentazione, mentre l'*accuratezza* è stata utilizzata per valutare la classificazione. Questa scelta ha permesso un'indagine completa, considerando sia la precisione nella delineazione delle regioni d'interesse sia l'efficacia nella corretta classificazione delle immagini. Inoltre, al fine di ottenere una visione più dettagliata delle performance nei singoli contesti di classificazione, sono state

impiegate le *matrici di confusione*. Queste ultime hanno permesso di analizzare l'efficacia delle predizioni nei quattro score di gravità del dataset ICLUS, offrendo un'analisi più mirata e dettagliata. Per quanto riguarda la segmentazione, vengono esibite le maschere predette da frame complessi per evidenziare le sfide incontrate dal modello, aiutando a comprendere i frame in cui le performance potrebbero risultare inferiori.

Dal momento che l'approccio adottato per la validazione dei risultati si basa su un confronto indiretto, è essenziale considerare attentamente le limitazioni e gli aspetti metodologici che possono influenzare l'interpretazione dei dati. L'assenza di un punto di riferimento nello stato dell'arte, può limitare la bontà delle conclusioni e compromettere la capacità di valutare appieno l'efficacia dell'approccio multitasking proposto. Inoltre, l'utilizzo di un singolo split del dataset, sebbene svolto con la massima cura per garantire rappresentatività, presenta intrinsecamente delle limitazioni; è importante riconoscere che tale scelta potrebbe comportare alcune restrizioni in termini di robustezza dell'analisi. La scelta di uno specifico split potrebbe influenzare i risultati in modo non rappresentativo dell'intera variabilità del dataset. Pertanto, è cruciale considerare la possibilità di applicare approcci più mirati, come l'utilizzo di tecniche di cross-validation, al fine di garantire una valutazione più robusta e generalizzabile delle prestazioni del modello.

Viene dimostrato dunque l'efficacia delle architetture **ResUnet** e **ResUnet++** nell'analisi di immagini ecografiche polmonari per la diagnosi precoce del COVID-19. I modelli proposti hanno portato a miglioramenti sia nella classificazione che nella segmentazione dei danni polmonari, superando i metodi basati su singoli compiti. Questi risultati evidenziano la possibilità di adottare strategie multiple in contesti complessi come l'imaging medico. La condivisione delle features tra diversi task ha dimostrato di potenziare le capacità predittive del modello, promuovendo una diagnosi più accurata e dettagliata delle condizioni patologiche analizzate.

Nonostante le sfide legate alla variabilità e alla complessità delle immagini LUS, i modelli hanno dimostrato prestazioni promettenti sottolineando le difficoltà ma anche il potenziale di queste tecniche in questo ambito. Questo lavoro contribuisce alla letteratura fornendo nuove intuizioni sull'uso di reti neurali multitask nell'analisi delle immagini ecografiche polmonari, aprendo la strada a future ricerche. La capacità di tali sistemi di affrontare le sfide poste dall'analisi di immagini medicali complesse rappresenta un passo avanti significativo verso il miglioramento della diagnosi precoce, con potenziali implicazioni positive sulla salute pubblica e sui sistemi sanitari a livello globale.

5.1 Sviluppi futuri

Ci sono molteplici prospettive per migliorare i risultati ottenuti dai modelli multitask proposti. In primo luogo, risulta imperativo incentivare una più accurata annotazione dei dati da parte degli esperti di dominio, in quanto tale elemento è

stato individuato come una delle principali ragioni alla base dei bassi punteggi di segmentazione.

Si potrebbe valutare anche l'adozione di tecniche avanzate di integrazione del dominio[15] per ampliare la capacità del modello nella ricerca di features significative e consentire una segmentazione più precisa delle regioni di interesse. Inoltre, l'impiego di modelli Transformer, che hanno dimostrato risultati eccellenti nella letteratura sul deep learning, potrebbero portare ad un passo avanti il task della segmentazione, migliorando ulteriormente la precisione della localizzazione delle regioni di interesse in immagini complesse come quelle del dataset ICLUS.

Per quanto riguarda la classificazione, è possibile introdurre modifiche mirate alla gestione degli score, valutando l'ordinalità delle etichette[22][23]. Questo approccio considera la naturale successione delle etichette, stabilendo che un errore tra etichette distanti, come ad esempio tra 3 e 0, abbia un impatto maggiore rispetto a errori tra etichette vicine. Integrare questa componente potrebbe migliorare la coerenza della classificazione e rendere il modello più sensibile alla struttura intrinseca delle etichette.

Infine, un ulteriore sviluppo interessante potrebbe essere l'implementazione di una cross-validation, al fine di ottenere risultati più solidi e indipendenti dalla suddivisione specifica dei dati. Questo approccio potrebbe contribuire a garantire la robustezza dei risultati ottenuti e a rendere il modello più affidabile.

Bibliografia

- [1] M. Asaad, N. K. Habibullah, and C. Butler. The impact of covid-19 on clinical trials. *Annals of Surgery*, 272:e222 – e223, 2020.
- [2] F. Miller, S. Young, M. Dobrow, and K. Shojania. Vulnerability of the medical product supply chain: the wake-up call of covid-19. *BMJ Quality Safety*, 30:331 – 335, 2020.
- [3] Brandon K.K. Fields, Natalie L Demirjian, C. DeBoer, Daniel Stemen, and D. Hwang. 3d printing novel ppe for response to covid-19 related shortages. 11583:1158300 – 1158300–9, 2020.
- [4] Maroš Hliboký, Marek Bundzel, and Stanislav Husár. State-of-the-art in lung ultrasound processing - brief review. In *2023 IEEE 21st World Symposium on Applied Machine Intelligence and Informatics (SAMII)*, pages 000091–000096, 2023.
- [5] Rachel Liu, Christopher Moore, and Vivek Tayal. *Ultrasound Equipment and Purchase*, pages 145–176. 01 2018.
- [6] Kim-Han Thung and Chong-Yaw Wee. A brief review on multi-task learning. *Multimedia Tools and Applications*, 77(22):29705–29725, 2018.
- [7] Rich Caruana. Multitask learning. In Sebastian Thrun and Lorien Pratt, editors, *Learning to Learn*, pages 95–133. Springer US, Boston, MA, 1998.
- [8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
- [9] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, and Andrew Y. Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning, 2017.
- [10] Shengfeng Liu, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, and Tianfu Wang. Deep learning in medical ultrasound analysis: A review. *Engineering*, 5(2):261–275, 2019.
- [11] Subhankar Roy, Willi Menapace, Sebastiaan Oei, Ben Luijten, Enrico Fini, Cristiano Saltori, Iris Huijben, Nishith Chennakeshava, Federico Mento, Alessandro

Bibliografia

- Sentelli, Emanuele Peschiera, Riccardo Trevisan, Giovanni Maschietto, Elena Torri, Riccardo Inchingolo, Andrea Smargiassi, Gino Soldati, Paolo Rota, Andrea Passerini, and Libertario Demi. Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound. *IEEE Transactions on Medical Imaging*, PP, 05 2020.
- [12] Gabriel Iluebe Okolo, S. Katsigiannis, and N. Ramzan. Multi-modal lung ultrasound image classification by fusing image-based features and probe information. *2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 45–50, 2022.
- [13] Jiaqi Zhang, Chin-Boon Chng, Xuan Chen, Chunshuang Wu, Mao Zhang, Yuqi Xue, Jianxin Jiang, and C. Chui. Detection and classification of pneumonia from lung ultrasound images. *2020 5th International Conference on Communication, Image and Signal Processing (CCISP)*, pages 294–298, 2020.
- [14] G. Soldati et al. Towards computer aided lung ultrasound imaging for the management of patients affected by covid-19, under submission. *ZAMM*, 8:161–185, 1928.
- [15] Oz Frank, Nir Schipper, Mordehay Vaturi, Gino Soldati, Andrea Smargiassi, Riccardo Inchingolo, Elena Torri, Tiziano Perrone, Federico Mento, Libertario Demi, Meirav Galun, Yonina C. Eldar, and Shai Bagon. Integrating domain knowledge into deep networks for lung ultrasound with applications to covid-19. *IEEE Transactions on Medical Imaging*, 41(3):571–581, 2022.
- [16] Yue Zhou, Houjin Chen, Yanfeng Li, Qin Liu, Xuanang Xu, Shu Wang, D. Shen, and P. Yap. Multi-task learning for segmentation and classification of tumors in 3d automated breast ultrasound images. *Medical image analysis*, 70:101918, 2020.
- [17] Jignesh Chowdary, P. Yogarajah, Priyanka Chaurasia, and Velmathi Guruviah. A multi-task learning framework for automated segmentation and classification of breast tumors from ultrasound images. *Ultrasonic Imaging*, 44:3 – 12, 2022.
- [18] Zongwei Zhou, M. R. Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39:1856–1867, 2019.
- [19] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2015.
- [20] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

- [21] Subhankar Roy, Willi Menapace, Sebastiaan Oei, Ben Luijten, Enrico Fini, Cristiano Saltori, Iris Huijben, Nishith Chennakeshava, Federico Mento, Alessandro Sentelli, Emanuele Peschiera, Riccardo Trevisan, Giovanni Maschietto, Elena Torri, Riccardo Inchingolo, Andrea Smargiassi, Gino Soldati, Paolo Rota, Andrea Passerini, Ruud J. G. van Sloun, Elisa Ricci, and Libertario Demi. Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound. *IEEE Transactions on Medical Imaging*, 39(8):2676–2687, 2020.
- [22] Javier Barbero-Gómez, Pedro-Antonio Gutiérrez, Víctor-Manuel Vargas, Juan-Antonio Vallejo-Casas, and César Hervás-Martínez. An ordinal cnn approach for the assessment of neurological damage in parkinson’s disease patients. *Expert Systems with Applications*, 182:115271, November 2021.
- [23] Víctor-Manuel Vargas, Pedro-Antonio Gutiérrez, and César Hervás-Martínez. Cumulative link models for deep ordinal classification, 2019.