



**UNIVERSITÀ DEGLI STUDI DI CAGLIARI**  
**FACOLTÀ DI SCIENZE**

Corso di Laurea Triennale in Informatica

**Un approccio basato sul Deep Learning per  
la diagnosi del COVID-19**

**Relatore**

Prof.ssa Cecilia Di Ruberto

Dott. Andrea Loddo

**Studente**

Fabio Pili

ANNO ACCADEMICO 2019/2020



To date, COVID-19 has more than one hundred and sixty million confirmed cases and a quick diagnosis is critical to controlling the spread of the disease. RT-PCR is the standard technique for virus detection, however many countries are unable to provide enough tests as the disease spreads quickly, therefore, faster and cheaper screening techniques are needed. Computed tomography represents a valid alternative, in fact, since the beginning of the pandemic, it has become increasingly important for the diagnosis and management of pneumonia caused by COVID-19. Deep Learning methods have been widely used in medical imaging, in particular, CNNs have been used for classification and segmentation problems regarding CT images. The aim of this work is to propose a classifier capable of detecting cases of COVID-19, through Deep Learning techniques and the use of CT image datasets. Three different classifications were performed: ternary, patient-oriented ternary and binary; the first achieved excellent results, especially from the VGG and Residual networks, with Accuracies exceeding 98%. The patient-oriented classification has reached 100% accuracy in some classes. The cross-dataset binary classification has led to below average results, which can be improved through future developments.



Ad oggi, il COVID-19 conta più di centosessanta milioni di casi confermati e una diagnosi rapida è fondamentale per controllare la diffusione della malattia. L'RT-PCR è la tecnica standard per il rilevamento del virus, tuttavia molti paesi non sono in grado di fornire un numero sufficiente di test in quanto la malattia si diffonde velocemente, pertanto, sono necessarie tecniche di screening più veloci ed economiche. La tomografia computerizzata rappresenta una valida alternativa, infatti, dall'inizio della pandemia, essa è diventata sempre più importante per la diagnosi e la gestione della polmonite causata da COVID-19. I metodi del Deep Learning sono stati ampiamente utilizzati nell'imaging del campo medico, in particolare, le CNN sono state utilizzate per problemi di classificazione e segmentazione riguardo immagini CT. L'obiettivo di questo lavoro è proporre un classificatore capace di rilevare casi di COVID-19, attraverso tecniche di Deep Learning e l'utilizzo di dataset di immagini CT. Sono state eseguite tre diverse classificazioni: ternaria, ternaria orientata ai pazienti e binaria; la prima ha ottenuto ottimi risultati, soprattutto da parte delle reti VGG e Residual, con delle Accuracy che superano il 98%. La classificazione orientata ai pazienti, ha raggiunto l'accuratezza del 100% in alcune classi. La classificazione binaria cross-dataset ha portato a dei risultati sotto la media, migliorabili attraverso sviluppi futuri.



# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Stato dell'Arte</b>	<b>5</b>
2.1	Tecniche di classificazione . . . . .	5
2.2	Reti Neurali Convoluzionali . . . . .	7
2.3	Transfer Learning . . . . .	8
2.4	Fondamenti . . . . .	9
<b>3</b>	<b>Materiali e metodi</b>	<b>13</b>
3.1	Dataset . . . . .	13
3.1.1	COVIDx CT-2A . . . . .	13
3.1.2	COVID-CT-dataset . . . . .	14
3.2	Reti utilizzate . . . . .	14
3.2.1	AlexNet . . . . .	14
3.2.2	GoogLeNet . . . . .	14
3.2.3	Inception V3 . . . . .	15
3.2.4	VGG . . . . .	16
3.2.5	ShuffleNet . . . . .	17
3.2.6	MobileNet V2 . . . . .	18
3.2.7	ResNet . . . . .	18
<b>4</b>	<b>Analisi sperimentale</b>	<b>21</b>
4.1	Training delle reti . . . . .	21
4.2	Test svolti . . . . .	23
4.2.1	Metriche . . . . .	23
4.2.2	Classificazione ternaria COVIDx CT-2A . . . . .	25
4.2.3	Classificazione ternaria orientata ai pazienti di COVIDx CT- 2A . . . . .	27
4.2.4	Classificazione binaria COVID-CT . . . . .	30

<b>5 Conclusioni</b>	<b>33</b>
<b>6 Sviluppi futuri</b>	<b>35</b>
<b>References</b>	<b>39</b>



# Capitolo 1

## Introduzione

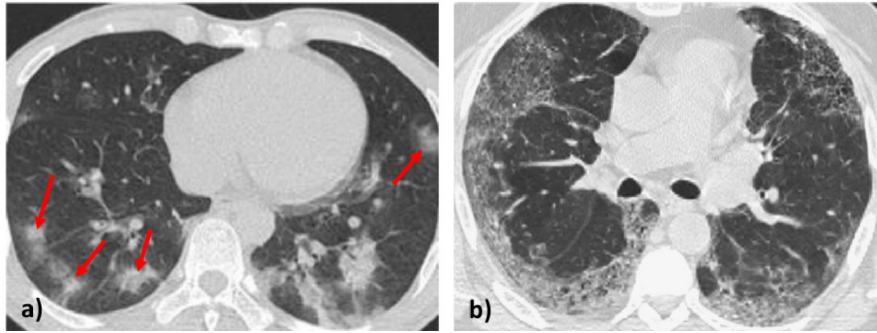
Il Coronavirus (COVID-19) è una malattia dovuta al virus SARS-CoV-2 diffusa a livello mondiale, dichiarata pandemia dall'Organizzazione Mondiale della Sanità l'11 Marzo 2020. Ad oggi, il COVID-19 conta più di centosessanta milioni di casi confermati, di cui più di tre milioni decessi (quindi con un tasso di mortalità dell'1.9%) e novantasette milioni di guariti in tutto il mondo[5].

Una diagnosi rapida è fondamentale per controllare la diffusione della malattia, aumenta l'efficacia delle cure mediche e, di conseguenza, la possibilità di sopravvivenza senza la necessità di terapie intensive, un punto cruciale, in quanto gli ospedali posseggono una disponibilità limitata delle apposite attrezzature. L'RT-PCR (Reverse transcriptase-polymerase chain reaction) è la tecnica standard per il rilevamento del virus, tuttavia molti paesi non sono in grado di fornire un numero sufficiente di test in quanto la malattia si diffonde velocemente. Vengono perciò testate solamente le persone con sintomi evidenti e occorrono diverse ore per fornire il risultato. Pertanto sono necessarie tecniche di screening più veloci e affidabili che potrebbero essere ulteriormente confermate dal test PCR o sostituirlo completamente.

La tomografia computerizzata (CT) rappresenta una valida alternativa per rilevare i casi di COVID-19 con una sensibilità maggiore, fino al 98% rispetto al 71% della RT-PCR. Dall'inizio della pandemia, la CT è diventata sempre più importante per la diagnosi e la gestione della polmonite da COVID-19. Le prime ricerche mostrano risultati esatti in particolar modo se il paziente viene sottoposto a una CT due o più giorni dopo lo sviluppo di sintomi. Tuttavia, il principale ostacolo che i radiologi sperimentano nell'analisi delle immagini radiografiche è la scansione visiva di piccoli dettagli. Inoltre, un gran numero di immagini CT deve essere valutato in un tempo molto breve, aumentando così la probabilità di classificazioni errate. Ciò

giustifica l'utilizzo di approcci intelligenti in grado di classificare automaticamente le immagini CT del torace.

I metodi del Deep Learning sono stati ampiamente utilizzati nell'imaging del campo medico. In particolar modo, le reti neurali convoluzionali (CNN) sono state utilizzate sia per problemi di classificazione che di segmentazione, anche per quanto riguarda immagini CT. Tuttavia, le immagini CT di polmoni, riferite a COVID-19 e non, possono essere facilmente classificate erroneamente soprattutto quando sono presenti contemporaneamente danni dovuti a polmoniti di cause diverse. I risultati principali di CT toraciche di pazienti positivi al COVID-19 indicano tracce di opacità a vetro smerigliato, noto come GGO (ground-glass opacity). Ad esempio, due scansioni CT di COVID-19 e non COVID-19 sono riportate nelle figure 1.1a e 1.1b [25].



**Figura 1.1:** (a) CT dei polmoni di un paziente affetto da COVID-19 dove sono presenti evidenti tracce di GGO indicate con frecce rosse. (b) CT dei polmoni di un paziente non affetto da COVID-19 con opacità diffusa nelle parti esterne dei polmoni. Immagini estratte da uno dei dataset utilizzati [29].

Fino ad ora, la maggior parte dei dataset esistenti per lo studio del COVID-19 contengono un numero ridotto di immagini CT. Per questo motivo, durante la fase di training è necessario limitare, o al meglio evitare, l'overfitting, fenomeno che indica che la CNN non sta imparando le caratteristiche discriminanti delle scansioni CT per COVID-19 ma le sta solo memorizzando.

Un altro punto critico è che l'inferenza delle CNN richiede molta potenza di calcolo: solitamente, infatti, le CNN vengono eseguite su GPU particolarmente costose, dotate di specifici sistemi di accelerazione hardware. Ad ogni modo, questo tipo di GPU sono ancora l'eccezione e non la norma, infatti le unità di elaborazione più comuni sono basate sulle CPU. Inoltre, questo tipo di macchine potrebbe non essere disponibile negli ospedali, specialmente in situazioni di emergenza [28].

L'obiettivo di questa tesi è quello di studiare e proporre un sistema automatico per la diagnosi di COVID-19, attraverso tecniche di Deep Learning e l'utilizzo di immagini CT relative a casi di COVID-19, casi di polmoniti non causate da COVID-19 e condizioni polmonari normali.



# Capitolo 2

## Stato dell'Arte

### 2.1 Tecniche di classificazione

Il pattern recognition è un sottoinsieme dell'apprendimento automatico. Nel pattern matching applicato alle immagini, si considera un pattern come se fosse una disposizione spaziale di caratteristiche, e una classe di pattern come un insieme di pattern che condividono alcune caratteristiche comuni. Il pattern matching comprende tecniche per l'assegnazione automatica dei pattern alle rispettive classi. Ciò significa che dati uno o più pattern la cui classe è sconosciuta, il compito di un sistema di pattern matching è quello di assegnare un'etichetta di classe a ciascuno dei suoi pattern in ingresso.

Sono quattro le fasi principali del riconoscimento: acquisizione, preelaborazione, estrazione delle caratteristiche e classificazione. Il rilevamento riguarda la generazione di segnali in un formato spaziale 2D o superiore. La preelaborazione utilizza tecniche per operazioni come riduzione del rumore, miglioramento, restauro e segmentazione. La fase di estrazione delle caratteristiche è quella che permette al sistema di riconoscere e apprendere i caratteri distintivi dei pattern. La classificazione si occupa dell'utilizzo di un insieme di funzionalità come base per l'assegnazione di etichette di classe a immagini di input sconosciute.

I tre approcci di base utilizzati per la classificazione delle immagini sono: classificazione basata sulla corrispondenza di pattern sconosciuti con prototipi specifici, classificatori statistici ottimali e reti neurali. La differenza tra questi approcci risiede nel livello di "ingegnerizzazione" richiesto per trasformare i dati grezzi in formati adatti all'elaborazione al computer.

- Nella classificazione basata su prototipi, l'obiettivo è quello di rendere le caratteristiche facilmente rilevabili, in modo da semplificare la classificazione.
- Nella seconda categoria, la classificazione viene effettuata in termini decisionali, teorici e statistici e la classificazione si basa sulla scelta di parametri che consentano di ottenere prestazioni ottimali in senso statistico.
- Nella terza categoria, la classificazione viene effettuata utilizzando reti neurali. Questi sistemi hanno la capacità unica di generare, a partire da dati grezzi, rappresentazioni adatte al riconoscimento.

I tre approcci descritti sono basati su parametri da specificare o che possono essere appresi da pattern che rappresentano il problema di riconoscimento che si intende risolvere. I pattern possono essere etichettati, ossia attribuire a ognuno di essi una classe di appartenenza, oppure non etichettati, il che significa che i dati sono noti per essere pattern ma non è nota la loro classe di appartenenza. Quando si lavora con dati etichettati, il dataset viene generalmente suddiviso in tre sottoinsiemi: training set, validation set e test set. Il processo mediante il quale un training set viene utilizzato per generare i parametri del classificatore è chiamato, appunto, training. In questo processo, ad un classificatore viene data l'etichetta di classe di ciascun pattern con l'obiettivo di apportare modifiche ai parametri se il classificatore commette un errore nell'identificare la classe del pattern dato.

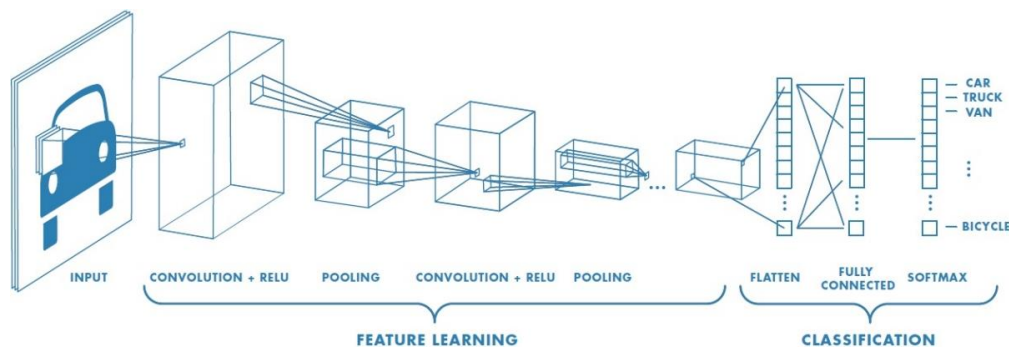
Al termine del training, viene utilizzato il validation set per confrontare i risultati con il pattern dato. In generale, sono necessarie diverse iterazioni di training/validation per ottenere delle prestazioni soddisfacenti. Se i risultati di training/validation sono accettabili, ma i risultati dei test non lo sono, allora il modello accusa il fenomeno di "overfitting", il che significa che esso si adatta troppo ai parametri del sistema rispetto ai dati disponibili, quindi solamente memorizzando e non apprendendo le caratteristiche discriminanti dei pattern. In caso di overfit, sono necessari ulteriori lavori sull'architettura del sistema [27].

## 2.2 Reti Neurali Convoluzionali

Le reti neurali convoluzionali, o CNN, sono un tipo specializzato di rete neurale adatto all'elaborazione di array di dati strutturati, come ad esempio le immagini. L'utilizzo delle CNN per il Deep Learning è diventato sempre più comune grazie a tre importanti fattori:

- Eliminano la necessità dell'estrazione manuale delle caratteristiche, poiché queste ultime vengono apprese direttamente dalla CNN.
- Producono risultati di riconoscimento all'avanguardia.
- Possono essere ri-addestrate per operare nuove attività di riconoscimento su set di dati diversi, consentendo così di sfruttare reti preesistenti [20].

Le reti neurali convoluzionali hanno svolto un ruolo importante nella storia del Deep Learning. Furono infatti uno dei primi Deep Models a produrre dei risultati veritieri, prima ancora che tali modelli venissero considerati attendibili. Sono state tra le prime reti neurali a risolvere importanti problemi applicativi e rimangono ancora oggi in prima linea nelle applicazioni commerciali del Deep Learning.



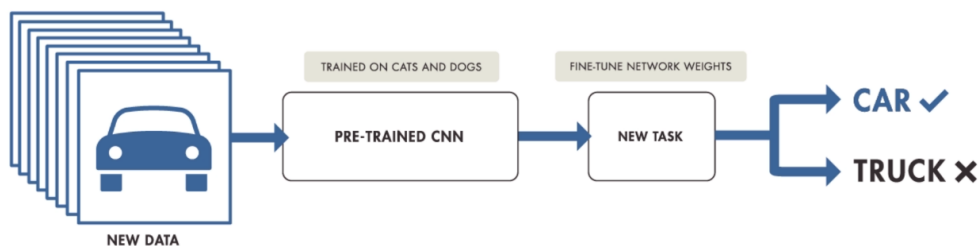
**Figura 2.1:** *Struttura di una rete neurale convoluzionale*

Il nome "Rete Neurale Convoluzionale" indica che la rete impiega un tipo specializzato di operazione lineare, chiamata appunto convoluzione. Le CNN sono semplicemente reti neurali che utilizzano la convoluzione al posto della moltiplicazione della matrice generale in almeno uno dei loro livelli [13]. Una CNN può avere decine o centinaia di livelli, ognuno dei quali impara a rilevare diverse caratteristiche da un'immagine. A ciascuna immagine vengono applicati diversi filtri e l'output di ogni immagine su cui è stata effettuata la convoluzione viene utilizzato come input per il livello successivo. I filtri possono essere inizialmente caratteristiche molto semplici, per poi assumere forme via via più complesse che definiscono in modo univoco l'oggetto [20].

Un tipico livello di una rete convoluzionale è costituito da tre fasi. Nella prima fase, il livello esegue diverse convoluzioni in parallelo per produrre un insieme di attivazioni lineari. Nella seconda fase, ogni attivazione lineare viene eseguita tramite una funzione di attivazione, come ad esempio l'Unità Lineare Rettificata (ReLU), che consente un addestramento più rapido ed efficace, mappando i valori negativi a zero e mantenendo i valori positivi. Nella terza fase, viene utilizzata una funzione di pooling che semplifica l'output, eseguendo il sottocampionamento non lineare e riducendo così il numero di parametri che la rete deve apprendere. Queste operazioni vengono ripetute su decine o centinaia di livelli, ognuno dei quali impara a identificare diverse caratteristiche. Dopo aver appreso le caratteristiche nei vari livelli, la rete passa alla classificazione. Il penultimo livello è un livello completamente connesso che genera un vettore di dimensione  $K$ , in cui  $K$  è il numero di classi che la rete sarà in grado di prevedere. Questo vettore contiene le probabilità per ogni classe di qualsiasi immagine classificata. Il livello finale dell'architettura utilizza un livello di classificazione, come softmax, per fornire l'output di classificazione [13].

## 2.3 Transfer Learning

Il Transfer Learning è una tecnica di apprendimento supervisionato, che riutilizza parti di un modello precedentemente addestrato su una nuova rete, che deve risolvere un problema diverso ma comunque simile. L'utilizzo di un modello pre-addestrato riduce in modo significativo il tempo necessario per la progettazione e il training delle funzionalità. Il primo passo è selezionare un modello di origine, idealmente uno con un set di dati di grandi dimensioni, con cui effettuare il training [6].



**Figura 2.2:** *Schema di funzionamento del Transfer Learning*

Il passo successivo consiste nel decidere quali livelli riutilizzare nella propria rete. Tipicamente i livelli più profondi vengono riutilizzati poiché questi tendono ad essere più generali, mentre i livelli superiori tendono ad adattarsi più finemente su un problema particolare. Infine, il nuovo modello viene addestrato sul nuovo set



di dati. Il vantaggio è che il modello tende a convergere molto più velocemente e quindi sono necessari meno dati e tempo di calcolo [6].

## 2.4 Fondamenti

In questa sezione verranno descritti, in ordine cronologico alcuni lavori che hanno influenzato direttamente lo sviluppo di questa tesi, evidenziandone i fatti principali come l'estrazione delle caratteristiche, in maniera artigianale o automatica, modelli di classificazione, dataset utilizzati, tipi di immagini utilizzate negli esperimenti e tipologie di polmoniti studiate.

Il primo lavoro è presentato da Nanni et al. [17] nel 2010. Qui gli autori hanno confrontato una serie di descrittori di pattern artigianali derivati dal Local Binary Pattern (LBP), considerando il loro uso specifico in applicazioni mediche. Le varianti di LBP valutate furono il Local Ternary Pattern (LTP), Elliptical Binary Pattern (EBP) ed Elongated Quinary Pattern (EQP). Questi descrittori sono stati valutati su tre diverse applicazioni mediche: rilevamento dell'espressione del dolore, classificazione delle immagini del fenotipo cellulare e Pap Test (Papanikolaou Test) finalizzato alla diagnosi del cancro cervicale. Dopo una serie completa di esperimenti utilizzando il classificatore SVM (Support Vector Machine), gli autori hanno constatato che il descrittore EQP, o gli insiemi creati utilizzando variazioni di EQP, funzionassero meglio per tutte le attività affrontate.

Sempre nel 2010, Parveen e Sthik [26] hanno affrontato il rilevamento della polmonite nelle immagini CXR: l'estrazione delle caratteristiche era ancora fortemente accoppiata alla prospettiva artigianale, tuttavia, vale la pena menzionare gli sforzi compiuti dagli autori per trovare descrittori utili per acquisire adeguatamente le informazioni sui diversi tipi di infezioni polmonari.

Sebbene l'epidemia di COVID-19 sia un evento abbastanza recente, ha attirato molta attenzione dalla società e in particolare anche dalla comunità di ricerca sull'analisi di immagini, vista l'urgenza della questione. Zhou et al.[30] hanno pubblicato uno studio che descrive un modello di Deep Learning per differenziare la nuova polmonite da coronavirus (NCP) e la polmonite influenzale tramite immagini di tomografie computerizzate (CT) del torace. Questo è uno dei lavori pionieristici che ha portato alla luce alcune prove scientifiche riguardanti la particolare pandemia che ha drammaticamente colpito il mondo. È importante evidenziare che Zhou et al. abbiano adottato delle immagini CT in quanto abbiamo una migliore capacità nel mostrare i dettagli dell'infezione polmonare.

Nel protocollo sperimentale, Zhou et al. hanno creato il training set utilizzando immagini CT relative a 35 pazienti positivi al COVID-19 e 1138 pazienti sospetti. Tra queste immagini sono state incluse quelle di 361 pazienti con polmonite virale, di cui 156 polmonite influenzale. In sintesi, lo studio dimostrò che il 96,6% delle lesioni NCP sono più grandi di 1cm, e per il 76,6% delle lesioni l'intensità era inferiore a -500Hu, deducendo che queste lesioni hanno un minore consolidamento di quelle provocate dall'influenza, la cui dimensione dei nodi varia da 5 a 10 mm. Per quanto riguarda i risultati della classificazione, il deep model creato ha ottenuto un tasso superiore a 0,93 relativo alla distinzione di NCP e influenza, considerando la metrica AUC.

Un lavoro simile è stato condotto da Li et al. [19], con lo scopo di identificare casi di COVID-19 su immagini CT del torace mediante l'utilizzo di tecniche di intelligenza artificiale. A tal fine, gli autori hanno utilizzato un dataset composto da immagini CT di pazienti positivi al COVID-19, pazienti con polmoniti virali e pazienti a cui non è stata diagnosticata alcuna polmonite. Le immagini sono state fornite da sei ospedali cinesi e il dataset creato è composto da 2969 immagini per il training set, di cui 400 da pazienti positivi, 1396 da pazienti con polmonite virale e 1173 da pazienti senza alcun tipo di polmonite. È stato inoltre creato un test set indipendente con immagini di 68 pazienti positivi, 155 pazienti con polmonite virale e 130 pazienti senza alcun tipo di polmonite, per un totale di 353 immagini CT.

Tramite l'utilizzo di ResNet 50 come base, è stato creato il modello 3D COVNet, capace di classificare immagini CT secondo le tre classi COVID-19, polmonite virale, non polmonite. Dopo la sperimentazione gli autori hanno ottenuto, secondo la metrica AUC, il valore di 0,96 per COVID-19 e 0,95 per le polmoniti virali [11].

Lo studio di Gozes et al. [7] si è occupato del rilevamento di COVID-19 e del monitoraggio dei pazienti utilizzando modelli di Deep Learning su immagini CT. Per monitoraggio del paziente, gli autori intendono l'evoluzione della malattia sul paziente nel tempo, utilizzando un volume 3D, generando quello che loro definiscono "Corona score". Gli autori affermano che il loro lavoro è il primo sviluppato per rilevare, caratterizzare e tracciare la progressione del COVID-19. Lo studio è stato sviluppato utilizzando immagini CT di 157 pazienti, dalla Cina e dagli Stati Uniti. Gli autori hanno fatto uso di robusti modelli di Deep Learning 2D e 3D, e hanno inoltre modificato e adattato i modelli di IA esistenti, combinando i risultati con la comprensione clinica. I risultati della classificazione, con l'obiettivo di differenziare le immagini del coronavirus rispetto alle immagini non coronavirus, hanno mostrato un valore di 0,996 sulla metrica AUC.

Wang e Wong [18] hanno svolto un lavoro simile, creando COVID-Net, una Deep Neural Network open source creata appositamente per rilevare il COVID-19 sulle immagini CT. A tal fine gli autori hanno creato COVIDx, un dataset creato esclusivamente per supportare lo sviluppo di COVID-Net. Il dataset è composto da 16.756 immagini CT del torace di 13.645 pazienti diversi, prelevate da due archivi distinti. Il prototipo di progettazione iniziale della rete è stato creato sulla base dei principi e delle migliori pratiche di progettazione guidata dall'uomo, combinati con l'esplorazione della progettazione guidata dalla macchina per produrre l'architettura di rete. Gli autori confermano che il modello sviluppato ha ottenuto un ottimo compromesso tra accuratezza e complessità computazionale. In termini di prestazioni di riconoscimento, nel complesso ha ottenuto un'accuratezza del 92,4%, mentre nello specifico il tasso di sensibilità ottenuto per ogni tipo di immagine è di 95% per i pazienti "normali", 91% per polmoniti virali e 80% per polmoniti da COVID-19.

Poiché l'identificazione di COVID-19 nelle immagini CT e CXR è un argomento tutt'ora esplorato, nel momento in cui si scrive, è bene precisare che nuovi lavori allo stato dell'arte emergono ogni giorno e che lo stato dell'arte è in continua evoluzione.



# Capitolo 3

## Materiali e metodi

### 3.1 Dataset

Uno degli elementi fondamentali per l'applicazione del Deep Learning è sicuramente il dataset. Esso è un insieme di dati, nel nostro caso immagini, che viene utilizzato per le fasi di training, validation e testing. Maggiore è la dimensione del dataset, più è possibile cercare di ottenere modelli accurati. Di seguito verranno descritti i dataset utilizzati in questa tesi.

#### 3.1.1 COVIDx CT-2A

COVIDx CT-2A è un dataset open access, creato inizialmente come COVIDx, il dataset che venne utilizzato per lo sviluppo di COVID-Net da Wang e Wong [18]. Inizialmente il dataset contava 16.756 immagini CT ma nel corso dei mesi la sua dimensione è aumentata notevolmente fino a raggiungere la dimensione attuale di 194.930 immagini CT di 4501 pazienti provenienti da 15 nazioni differenti. Al momento è utilizzato come base per lo sviluppo di diversi modelli di Deep Learning, in particolare per lo sviluppo di COVID-Net CT-2[10].

Il dataset si presenta in due versioni: COVIDx CT-2A, appena descritto e che verrà utilizzato come base per il lavoro di questa tesi, e COVIDx CT-2B. Mentre COVIDx CT-2A è formato interamente da immagini la cui classe di appartenenza è confermata, COVIDx CT-2B è formato dalle stesse immagini di COVIDx CT-2A con l'aggiunta di altre immagini la cui classe di appartenenza non è veramente affidabile. Inoltre COVIDx CT-2B possiede una licenza più restrittiva [9].

### **3.1.2 COVID-CT-dataset**

COVID-CT-dataset è stato creato da Zhao et al. [29] con lo scopo di essere utilizzato per ricerche e lavori futuri relativi alla diagnosi di COVID-19 tramite immagini CT. Le immagini sono state raccolte tra il 19 Gennaio 2020 e il 25 Marzo 2020 per un totale di 470 immagini CT di pazienti differenti. Di queste immagini, 275 sono CT toraciche di pazienti positivi al COVID-19, mentre le restanti 195 provengono da pazienti non positivi. I loro lavori si basarono sul Transfer Learning applicato alla rete DenseNet e portarono ad un Accuracy finale di 84.7%.

## **3.2 Reti utilizzate**

In questa sezione verranno brevemente descritti i diversi modelli di reti neurali convoluzionali che sono attualmente allo stato dell'arte e che sono stati utilizzati nella fase sperimentale di questa tesi.

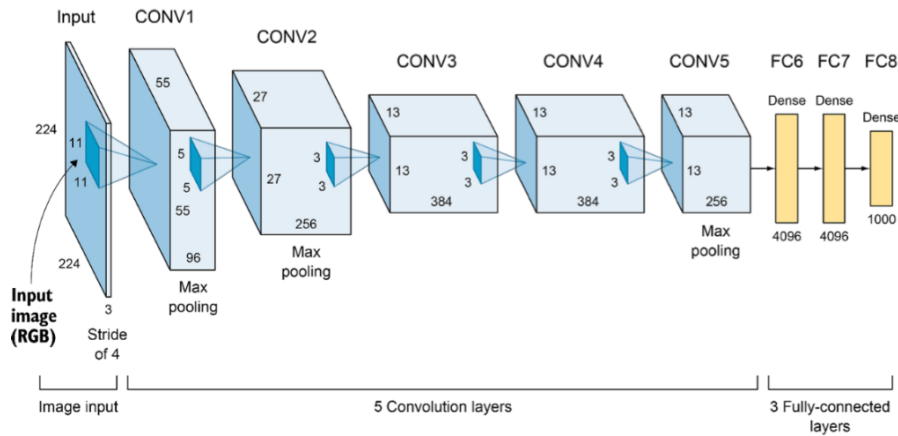
### **3.2.1 AlexNet**

AlexNet è una delle architetture di rete neurale più popolari al giorno d'oggi. È stata proposta da Alex Krizhevsky per l'ImageNet Large Scale Visual Recognition Challenge (ILSVRC), raggiungendo un tasso di errore del 15.3%. L'architettura è composta in totale da otto strati, di cui i primi 5 sono strati convoluzionali e gli ultimi 3 sono completamente connessi. I primi due strati sono collegati a strati di max-pooling sovrapposti, in modo così da estrarre un numero massimo di caratteristiche. Il terzo, quarto e quinto strato convoluzionale sono direttamente collegati agli strati completamente connessi. Tutti gli output degli strati convoluzionali e di quelli completamente connessi, sono collegati alla funzione di attivazione non lineare ReLU. Lo strato finale di output è collegato ad uno strato di attivazione softmax, che produce una distribuzione di 1.000 etichette di classe [15].

Le dimensioni di input per la rete sono di  $227 \times 227 \times 3$ , il che significa che l'immagine di input per AlexNet è un'immagine a colori RGB (3 canali) di  $227 \times 227$  pixel. Inoltre, per ridurre l'overfitting durante il processo di training, la rete utilizza strati di dropout [15].

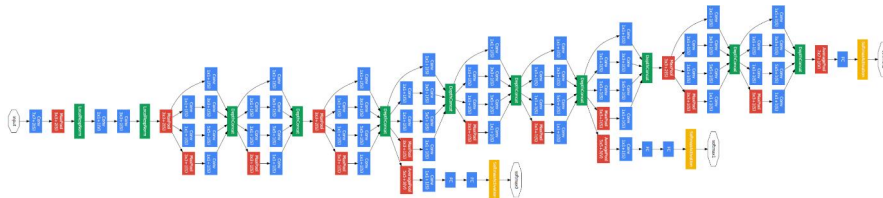
### **3.2.2 GoogLeNet**

La rete Inception è stata una delle maggiori scoperte nel campo delle reti neurali, in particolare per le CNN. La prima versione, denominata GoogLeNet, ha visto il suo esordio nel 2014 e, come suggerisce il nome, è stata sviluppata da un



**Figura 3.1:** *Architettura AlexNet*

team di Google. Se una rete è costituita da molti strati profondi, può incorrere nel problema dell'overfitting. Per risolvere questo problema è stata proposta l'architettura GoogLeNet, con l'idea di avere filtri con più dimensioni che possono operare sullo stesso livello. Con questa idea, la rete diventa in realtà più ampia piuttosto che più profonda. L'operazione di convoluzione viene eseguita su input con tre dimensioni di filtro: (1x1), (3x3) e (5x5). Insieme alle convoluzioni viene eseguita anche un'operazione di max-pooling che viene poi inviata al successivo modulo di inception.



**Figura 3.2:** *Architettura GoogLeNet*

L'architettura GoogLeNet è profonda 22 strati, con 27 strati di pooling inclusi, e riceve in input immagini RGB con dimensioni di 224x224 pixel. Sono presenti 9 moduli di inception linearmente impilati in totale. Le estremità dei moduli di inception sono poi collegate allo strato di pooling medio globale [15].

### 3.2.3 Inception V3

Inception v3 si concentra principalmente sulla riduzione della potenza di calcolo modificando le precedenti versioni di Inception. Questa idea è stata proposta nel documento "Rethinking the Inception Architecture for Computer Vision", pubblicato nel 2015. È stato scritto in collaborazione con Christian Szegedy, Vincent

Vanhoucke, Sergey Ioffe e Jonathon Shlens. Rispetto alle reti VGG, le Inception (GoogLeNet / Inception v1) si sono dimostrate più efficienti dal punto di vista computazionale, sia in termini di numero di parametri generati dalla rete che in costo di memoria e risorse. Nel modello Inception V3, sono state suggerite diverse tecniche per l'ottimizzazione della rete per allentare i vincoli e facilitare l'adattamento del modello. Le tecniche includono convoluzioni fattorizzate, regolarizzazione, riduzione delle dimensioni e calcoli parallelizzati.

L'architettura di una rete Inception v3 viene costruita progressivamente, passo dopo passo tramite l'utilizzo di Convoluzioni fattorizzate, più piccole e asimmetriche, che aiutano a ridurre l'efficienza computazionale, mantengono la rete efficiente ed essendo più piccole portano a un allenamento più veloce. Viene utilizzato anche un classificatore ausiliario, ossia una piccola CNN inserita tra i livelli durante la fase di training, che funge da classificatore [16].

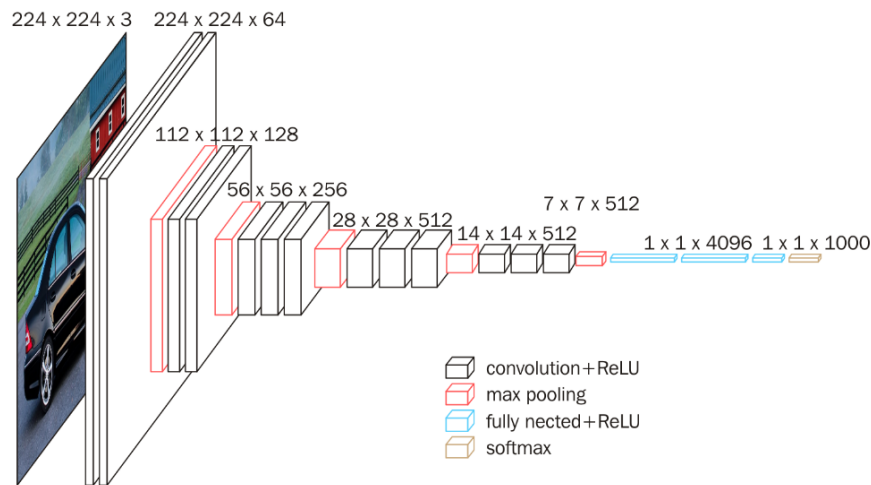
### **3.2.4 VGG**

VGG è una popolare architettura di rete neurale, proposta da Karen Simonyan ed Andrew Zisserman dell'Università di Oxford. Si basa anch'essa sulle CNN ed è stata applicata all'ImageNet Challenge nel 2014, ottenendo un'accuracy del 92.7%. I principali miglioramenti di VGG, rispetto ad AlexNet, includono l'uso di filtri kernel-sized di grandi dimensioni (dimensioni di 11 e 5, rispettivamente nel primo e nel secondo strato convoluzionale), combinati con filtri kernel-sized multipli (3x3) posti uno dopo l'altro.

#### **VGG 16**

L'architettura riceve in input immagini con dimensioni di 224x224 pixel. In una fase di preelaborazione, il valore RGB medio viene sottratto da ogni pixel di ciascuna immagine. Al termine della preelaborazione le immagini vengono passate ad una pila di strati convoluzionali con piccoli filtri di tipo receptive-field, di dimensioni 3x3. Il pooling spaziale è effettuato da cinque strati di max-pooling, che seguono diversi strati convoluzionali. Il max-pooling viene eseguito su una finestra di dimensione 2x2, con la dimensione del passo impostata a 2.





**Figura 3.3:** *Architettura VGG 16*

Per quanto riguarda la configurazione degli strati completamente connessi, i primi due hanno 4096 canali ciascuno, il terzo esegue la classificazione ILSVRC a 1000 vie, e il livello finale è lo strato softmax. Tutti i livelli nascosti per la rete VGG sono seguiti dalla funzione di attivazione ReLU [15].

## VGG 19

VGG19 è una CNN con più di un milione di immagini provenienti dal database ImageNet. A differenza di VGG16 che ha 16 livelli di profondità, VGG 19 ne ha 19 e quindi possiede un'architettura più complessa. Può classificare immagini in 1000 classi diverse di oggetti e animali[24].

### 3.2.5 ShuffleNet

Megvii Inc ha introdotto ShuffleNet, che afferma di essere un'architettura estremamente efficiente in termini di calcolo, progettata per dispositivi mobili con potenza di calcolo di 10 - 150 MFLOPs. ShuffleNet utilizza la convoluzione a gruppi puntuali e lo shuffle dei canali per ridurre i costi di calcolo, mantenendo comunque una buona precisione. Riesce ad ottenere un errore top-1 inferiore rispetto al sistema MobileNet sulla classificazione ImageNet, e raggiunge una velocità effettiva circa 13 volte superiore ad AlexNet, mantenendo una precisione comparabile. La rete è composta da 172 livelli complessivi e riceve in input immagini RGB con dimensioni di 224x224 pixel [3].

### 3.2.6 MobileNet V2

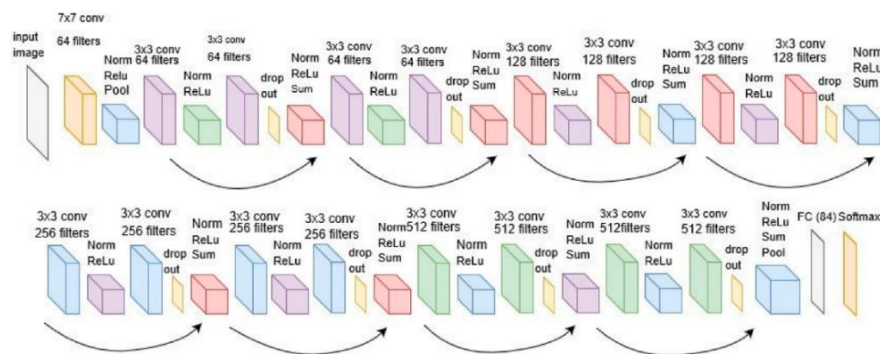
MobileNet V2 è una rete sviluppata da Google e si tratta di una versione più raffinata della precedente, MobileNet V1, che la rende ancora più efficiente e potente. MobileNet V2 utilizza convoluzioni separabili in profondità. Ci sono tre strati convoluzionali, di cui gli ultimi due effettuano una convoluzione in profondità che filtra gli input, seguita da uno strato convoluzionale puntuale  $1 \times 1$ . Quest'ultimo strato ha lo scopo di ridurre il numero di canali e viene chiamato anche strato di proiezione, poiché proietta i dati con un numero elevato di dimensioni in un tensor con un numero di dimensioni molto inferiore. L'architettura completa di MobileNet V2 è quindi costituita da 17 di questi blocchi, in fila. Seguono una convoluzione regolare  $1 \times 1$ , uno strato di pooling medio globale e uno strato di classificazione [12].

### 3.2.7 ResNet

Poiché le reti neurali profonde sono sia lunghe da addestrare che inclini all'overfitting, un team di Microsoft ha introdotto un framework di apprendimento residuo per migliorare il training delle reti che sono sostanzialmente più profonde di quelle utilizzate in precedenza. Quando si addestrano reti profonde arriva un punto in cui l'aumento della profondità provoca una saturazione dell'accuratezza, seguita da un rapido degrado. Questo è chiamato "problema del degrado" ed evidenzia come non tutte le architetture di reti neurali siano facili da ottimizzare. Resnet utilizza una tecnica chiamata "residual mapping" per combattere questo problema. In generale, nelle architetture di reti neurali convenzionali, diversi strati vengono addestrati per svolgere il proprio compito. Così facendo, la rete apprende diverse caratteristiche di varia complessità alla fine dei suoi livelli. Resnet, attraverso il "residual mapping", invece di provare ad apprendere alcune caratteristiche, prova da apprenderne alcuni residui. Il residuo può essere semplicemente inteso come sottrazione della caratteristica appresa dall'input di quel livello. ResNet lo fa collegando direttamente l'input dell' $n$ -esimo strato ad un qualche strato  $n+x$ . Molti problemi possono essere risolti utilizzando ResNet. Sono facili da ottimizzare e raggiungono una maggiore precisione quando la profondità della rete aumenta, producendo risultati migliori rispetto alle reti precedenti. Esistono molte varianti dell'architettura ResNet, in cui il concetto di funzionamento rimane lo stesso, ma quello che varia è il numero di strati. Abbiamo ResNet 18, ResNet 34, ResNet 50, ResNet 101 ... e così via [16].

## ResNet 18

ResNet 18 è una rete neurale convoluzionale appartenente alle Residual Network, profonda 18 strati. La rete, pre-addestrata sul database di immagini ImageNet, può classificare le immagini in 1000 categorie di oggetti, come tastiera, mouse, matita e varie specie animali. Come risultato, la rete ha appreso rappresentazioni ricche di caratteristiche per una vasta gamma di immagini. La rete riceve in input immagini RGB con dimensioni di 224x224 pixel [22].



**Figura 3.4:** *Architettura ResNet 18*

## ResNet 50

Resnet 50 è una rete neurale convoluzionale appartenente alle Residual Network, profonda 50 strati e addestrata su un milione di immagini di 1000 categorie del database ImageNet. Inoltre, il modello ha oltre 23 milioni di parametri addestrabili, il che lo rende migliore per il riconoscimento delle immagini.

In particolare, il modello ResNet 50 è costituito da 5 stadi, ciascuno con un blocco residuo. Ogni blocco residuo ha 3 strati con entrambe le convoluzioni 1x1 e 3x3 [4]. La rete riceve in input immagini RGB con dimensioni di 224x224 pixel [23].

## ResNet 101

Anche ResNet 101 è una rete neurale appartenente alle Residual Network, con una profondità di 101 strati. È stata addestrata sul database di immagini ImageNet e può classificare le immagini in 1000 categorie di oggetti. Riceve in input immagini RGB con dimensioni di 224x224 pixel [21].



# Capitolo 4

## Analisi sperimentale

In questo capitolo verranno descritti i metodi utilizzati e i risultati ottenuti. Come ambiente di sviluppo (IDE) è stato utilizzato Matlab R2020b, comunemente utilizzato per applicazioni scientifiche, analisi numerica e per la simulazione di sistemi dinamici. Per le tecniche di classificazione, è stato utilizzato sia l'approccio from scratch, in quanto nonostante siano state utilizzate delle reti pre-addestrate, nessuna di loro era precedentemente capace di classificare immagini CT relative a questo lavoro, sia il Transfer Learning, in quanto una volta addestrate, le reti sono state sottoposte a un fine tuning. Per ciascuna rete utilizzata è stato installato il relativo plug-in su Matlab che ne consente l'utilizzo al suo interno.

### 4.1 Training delle reti

Tramite l'utilizzo di fogli di calcolo forniti dai creatori dei dataset, il primo passo è stato suddividere i dataset in sottocartelle in base alla loro classe di appartenenza:

- nel caso del dataset COVIDx CT-2A [18] le classi sono "covid-19" che indica immagini CT di pazienti positivi al COVID-19, "pneumonia" che indica immagini CT di pazienti affetti da polmonite non causata da COVID-19, "normal" che indica immagini CT di pazienti in condizioni normali, non affetti da alcun tipo di polmonite;
- nel caso del dataset COVID-CT di Zhao et al. [29] le classi sono "covid-19", che come nel precedente dataset indica immagini CT di pazienti positivi al COVID-19, e "nonCovid-19" che indica immagini CT di pazienti non affetti da COVID-19, raggruppando immagini CT di pazienti affetti da polmonite non causata da COVID-19 e di pazienti in condizioni normali.

È importante evidenziare come entrambi i dataset siano sbilanciati verso la classe "covid-19": la classe, infatti, possiede un alto numero di immagini rispetto a le due rimanenti e ciò influisce sui risultati dei test.

Successivamente, i dataset sono stati importati singolarmente all'interno dell'IDE e attraverso le opzioni "IncludeSubfolders" e "LabelSource: foldername", le immagini vengono automaticamente etichettate col nome della cartella in cui sono contenute. Essendo COVIDx CT-2A il dataset più ricco, le sue immagini sono state suddivise in base alle seguenti percentuali, indicate anche nello studio effettuato da Wang e Wong [18]: 70% training, 20% validation e 10% testing.

Per quanto riguarda il dataset COVID-CT, è stato utilizzato in due modi: la prima volta è stato preso per intero come testing set, mentre per la seconda è stato suddiviso allo stesso modo del COVIDx CT-2A in modo da essere utilizzato per le operazioni di fine tuning.

Una volta creati i set di immagini, è il momento di preparare le reti alla fase di training. I livelli convoluzionali delle reti si occupano di estrarre le caratteristiche di ogni immagine che l'ultimo Fully Connected Layer (livello di apprendimento) e il Classification Output Layer (livello di classificazione finale) utilizzano per classificare l'immagine data in input. Questi due livelli contengono informazioni su come combinare le funzionalità estratte dalla rete per ottenere una probabilità di classe, un valore di perdita e le etichette predette. Per addestrare una rete alla classificazione di nuove immagini, bisogna sostituire questi due livelli con dei nuovi, adattati appositamente al nuovo set di dati.

Ogni rete richiede in input immagini con una dimensione precisa, ma le immagini contenute dei datastore hanno dimensioni diverse. Tramite la funzione "Augmented Image Datastore", le immagini vengono ridimensionate in base alle richieste delle reti e viene inoltre specificata l'opzione "Color Preprocessing: gray2rgb": essendo tutte le immagini in scala di grigi, tramite questa opzione si garantisce che tutte le immagini in output abbiano il numero di canali richiesti dall'Image Input Layer.

Vengono poi definite le opzioni per il training. Innanzitutto, è importante fornire alcune definizioni:

- Risolutore: è l'algoritmo di apprendimento che aggiorna i parametri di rete (pesi e bias) per minimizzare la loss function, facendo piccoli passi ad ogni iterazione;
- Epoche: sono i passaggi in cui l'algoritmo di apprendimento viene esposto al training set;

- Mini Batch: indica il sottoinsieme del training set che il risolutore utilizza per l'apprendimento.

In particolare, il training delle reti è stato effettuato utilizzando il risolutore Adaptive Moment Estimation (ADAM), utilizzando una dimensione fissa del Mini Batch e variando i valori di Max Epochs. Il training effettivo avviene tramite la funzione "Train Network" a cui vengono passati come parametri l'Augmented Datastore, la rete adattata e le opzioni, che includono anche il Validation Dataset.

A training terminato, quando i valori di Training e Validation Accuracy sono elevati e quelli della loss function tendono allo zero, si prosegue alla classificazione delle immagini del Testing Dataset.

## 4.2 Test svolti

Sono state progettate diverse tipologie di esperimenti: sul dataset COVIDx CT-2A si è eseguita prima una classificazione ternaria di immagini, successivamente una orientata ai pazienti. Sul dataset COVID-CT, dopo le operazioni di fine tuning, si è eseguita una classificazione binaria.

### 4.2.1 Metriche

Per i risultati, è stato tenuto conto delle seguenti metriche:

- Accuracy: si tratta del rapporto tra il numero di predizioni corrette e il numero totale di predizioni [8];

$$Accuracy = \frac{TP + TF}{TP + TF + FP + FN},$$

- Precision: è il rapporto tra il numero dei "veri positivi" e il numero totale di positivi predetti, formato da "veri positivi" e "falsi positivi" [8];

$$Precision = \frac{TP}{TP + FP}$$

- Recall: simile a Precision, è il rapporto tra il numero di "veri positivi" e il numero totale di positivi, formato da "veri positivi" e "falsi negativi". Un alto valore di Recall indica che sono presenti un alto numero di veri positivi e un basso numero di falsi negativi. In ambito medico ed economico i modelli

con alto valore di Recall sono i più richiesti [14];

$$Recall = \frac{TP}{TP + FN}.$$

- Specificity: analogamente a Recall, indica il rapporto tra il numero di "veri negativi" e il numero totale di negativi, formato da "veri negativi" e "falsi positivi" [14];

$$Specificity = \frac{TN}{FP + TN},$$

- F-measure: si tratta di una media ponderata tra i valori di Precision e Recall. Benché indichino la stessa cosa, l'F-measure di una classificazione binaria viene calcolata in maniera differente rispetto a quella di una classificazione multi-classe [8];

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

- MAvG: è la Macro-average geometric, definita come la media geometrica dell'Accuracy parziale di ciascuna classe.

$$MAvG = \left( \prod_{i=1}^J Acc_i \right)^{\frac{1}{J}},$$

- MAvA: è la Macro-average arithmetic, definita come la media aritmetica dell'Accuracy parziale di ciascuna classe.

$$MAvA = \frac{\sum_{i=1}^J Acc_i}{J}$$

Tutti questi valori sono stati calcolati sia con il metodo Micro-average che con il Macro-average [1], ad eccezione di MAvG e MAvA, che da come fa intuire il nome, sono calcolati solo tramite il secondo metodo. È importante specificare che questi due valori vanno presi in considerazione solo nelle classificazioni multi-classe, specialmente quando i dataset sono sbilanciati e propendono ad avere molte più immagini di una o più classi rispetto alle altre, come nel caso del dataset COVIDx CT-2A, che possiede un elevato numero di immagini CT appartenenti alla classe "covid-19" rispetto alle altre due [2].



### 4.2.2 Classificazione ternaria COVIDx CT-2A

Cominciamo con il considerare i risultati ottenuti in fase di training:

Rete	Risolutore	Max Epochs	Mini Batch size	Validation Frequency	Training Accuracy	Validation Accuracy
AlexNet	adam	50	32	8010	89,73%	90,40%
GoogLeNet	adam	20	32	8010	87,50%	87,51%
Inception V3	adam	20	32	8010	96,00%	91,30%
VGG16	adam	50	32	8010	94,67%	94,30%
VGG19	adam	20	32	8010	100,00%	97,01%
ShuffleNet	adam	20	32	8010	96,88%	93,61%
MobileNet V2	adam	20	32	8010	100,00%	90,61%
ResNet 18	adam	20	32	8010	100,00%	92,82%
ResNet 50	adam	20	32	8010	100,00%	86,53%
ResNet 101	adam	20	32	8010	100,00%	92,33%

**Tabella 4.1:** *Training e Validation con COVIDx CT-2A*

Nonostante i modelli abbiano pressoché avuto tutti le stesse opzioni in fase di training, reti come GoogLeNet, VGG19 e le varie ResNet hanno ottenuto risultati diversi: si può vedere, infatti, come la prima abbia raggiunto soltanto l'87,50% di Training Accuracy, mentre le altre, con le stesse opzioni, siano riuscite ad ottenere il 100,00%. È importante notare però che per quanto riguarda i valori di Validation Accuracy, GoogLeNet non si discosti affatto dai risultati precedenti, mentre ResNet 50 mostra un grande dislivello sulle percentuali. Tutte le restanti reti si comportano discretamente, con risultati che oscillano, sia in training che in validation, sempre attorno al 90%. I risultati migliori sono stati ottenuti invece da VGG19, che anche raggiunge il 97,01% di Validation Accuracy.

Nella tabella successiva verranno rappresentati i risultati ottenuti da ogni rete secondo il metodo Micro-Average: questo metodo aggrega i contributi di tutte le classi per calcolare il valore medio per ogni metrica utilizzata. In una configurazione di classificazione multi-classe, la Micro-Average è preferibile se si sospetta che ci possa essere uno squilibrio di classe.

Rete	Precision	Recall	Specificity	Accuracy	F-measure
AlexNet	93,07%	93,07%	96,54%	93,07%	93,07%
GoogLeNet	90,39%	90,39%	95,19%	90,39%	90,39%
Inception V3	95,46%	95,46%	97,73%	95,46%	95,46%
VGG16	96,90%	96,90%	98,45%	96,90%	96,90%
VGG19	98,15%	98,15%	99,08%	98,15%	98,15%
ShuffleNet	95,92%	95,92%	97,96%	95,92%	95,92%
MobileNet V2	94,56%	94,56%	97,28%	94,56%	94,56%
ResNet 18	96,57%	96,57%	98,28%	96,57%	96,57%
ResNet 50	91,93%	91,93%	95,97%	91,93%	91,93%
ResNet 101	94,79%	94,79%	97,39%	94,79%	94,79%

**Tabella 4.2:** *Performance in Micro-average per il testing su COVIDx CT-2A*

Come si può vedere dai valori in tabella, tutte le reti hanno ottenuto ottimi risultati, con alcune di esse che spiccano più di altre. Reti come AlexNet, Inception V3, ShuffleNet, MobileNet V2 e ResNet 101 ottengono tutte valori simili che si aggirano tra il 93% e il 95%. I valori più bassi sono stati ottenuti da ResNet 50 e GoogLeNet, le cui Validation Accuracy nella fase di training erano le più basse. Mentre le migliori restano ResNet 18 e le reti VGG, in particolare VGG19 che ottiene i risultati più alti e soddisfacenti. È importante notare che tutti i valori di Specificity sono più alti rispetto a quelli delle altre metriche.

Rete	Precision	Recall	Specificity	Accuracy	F-measure	MAvA	MAvG
AlexNet	91,88%	92,50%	96,64%	95,38%	92,17%	91,75%	91,88%
GoogLeNet	89,46%	88,96%	95,13%	93,59%	89,08%	89,46%	89,39%
Inception V3	95,48%	94,34%	97,54%	96,97%	94,84%	95,48%	95,48%
VGG16	96,65%	96,57%	98,44%	97,93%	96,58%	96,65%	96,63%
VGG19	97,85%	97,87%	99,08%	98,87%	97,86%	97,85%	97,84%
ShuffleNet	95,36%	94,92%	97,94%	97,28%	95,13%	95,36%	95,34%
MobileNet V2	94,24%	93,05%	97,31%	96,38%	93,38%	94,24%	94,14%
ResNet 18	96,41%	96,67%	98,22%	97,71%	95,98%	96,41%	96,40%
ResNet 50	92,19%	89,53%	95,61%	94,62%	90,45%	92,19%	92,16%
ResNet 101	94,99%	93,06%	97,16%	96,53%	93,08%	94,99%	94,97%

**Tabella 4.3:** *Performance in Macro-average per il testing su COVIDx CT-2A*

I valori riportati in questa tabella sono stati ottenuti, invece, tramite il metodo Macro-Average, quindi verranno riportati anche i valori di MAvA e MAvG. Questo

metodo permette di calcolare i valori medi delle metriche in modo indipendente per ogni classe.

Nel dettaglio, è possibile notare come GoogLeNet, ancora la rete con i risultati peggiori, abbia ottenuto dei valori sempre al di sotto del 90%, eccezion fatta per Specificity e Accuracy, la prima infatti è molto più alta dei valori medi ottenuti da questa rete, mentre la seconda risulta essere più alta a causa della dipendenza dagli altri valori, e quindi anche la Specificity. Anche ResNet 50, che riporta valori abbastanza diversi tra loro, scende fino all'89,53% in Recall, indicando che la rete ha avuto qualche difficoltà a distinguere i veri positivi con accuratezza.

Per quanto riguarda il resto dei modelli, presentano tutti risultati omogenei tra loro, come era deducibile anche dai valori ottenuti in fase di training. Anche in questo caso VGG19 presenta risultati molto alti, arrivando al 99,08% in Specificity, mentre gli altri valori superano sempre il 97%. Risultati simili sono stati ottenuti da VGG16 e ResNet 18; quest'ultima riportava inizialmente valori simili a MobileNet V2, ma in fase di testing ha dimostrato di essere più accurata.

Senza entrare nello specifico, in quanto ne verrà discusso nella sezione opportuna, si può evincere che la rete che ha portato a risultati migliori sia VGG19, avendo valori sempre molto alti sotto ogni punto di vista. Seguono VGG16 e ResNet 18, con risultati ottimi poco al di sotto di quelli raggiunti da VGG19; il fine tuning, che verrà eseguito per preparare i modelli alla classificazione del dataset COVID-CT, riguarderà infatti proprio questi tre modelli, più MobileNet V2 allo scopo di poter desumere un'eventuale adattabilità di reti specifiche in ambito mobile per questo task.

### **4.2.3 Classificazione ternaria orientata ai pazienti di COVIDx CT-2A**

Per questo lavoro sono stati utilizzati i modelli allenati nella classificazione ternaria e si è deciso di procedere in questo modo: il testing dataset, formato da 25.658 immagini, è stato suddiviso, secondo i fogli di calcolo forniti dai creatori del dataset, in 426 pazienti. Durante la fase di preparazione, è stato accertato che ogni paziente avesse solo ed esclusivamente immagini appartenenti a una sola classe, poiché in caso contrario, ciò avrebbe portato alla vanificazione del test. Per quanto riguarda il testing, ogni paziente è stato caricato come singolo dataset e si è proceduto al caricamento una classe per volta, analizzando prima i pazienti "covid-19", successivamente i pazienti "normal" e infine i pazienti "pneumonia". Una volta esaminate le immagini, il modello avrebbe prodotto dei risultati come quelli della classificazione ternaria; si è quindi deciso di utilizzare l'Accuracy relativa alla clas-

se di appartenenza come chiave di lettura degli esiti: se essa fosse stata maggiore del 50%, il paziente sarebbe risultato appartenente alla sua classe di appartenenza, in caso contrario sarebbe stato classificato come errato.

In questo modo si è potuto vedere come ogni modello si comportasse con ogni singola classe ed infine ne è stata tratta una Accuracy media che descrive il livello di accuratezza della rete.

Rete	COVID-19	Normal	Pneumonia	AVG
AlexNet	88,89%	96,83%	94,40%	93,97%
GoogLeNet	73,10%	92,06%	94,90%	86,52%
Inception V3	83,63%	98,41%	98,40%	93,48%
VGG16	90,64%	97,62%	100,00%	96,09%
VGG19	95,91%	98,41%	97,60%	97,31%
ShuffleNet	85,96%	98,41%	96,00%	93,46%
MobileNet V2	77,78%	98,41%	100,00%	92,06%
ResNet 18	82,46%	98,41%	98,40%	93,09%
ResNet 50	66,08%	99,21%	96,00%	87,10%
ResNet 101	71,93%	99,21%	99,20%	90,11%

**Tabella 4.4:** *Testing orientato ai pazienti con COVIDx CT-2A*

Si può osservare come AlexNet abbia qualche difficoltà nel classificare le immagini CT dei pazienti appartenenti alla classe "covid-19". Da un'analisi più approfondita dei risultati è emerso che molti dei pazienti classificati in maniera errata venissero classificati come appartenenti alla classe "pneumonia", questo è dovuto alla similarità delle immagini, in quanto entrambe le classi rappresentano delle polmoniti, anche se di origini differenti. Questa particolarità si ripresenterà anche nei test effettuati con le altre reti, in alcuni casi in forma più accentuata e in altri meno. Per quanto riguarda la classe "normal" invece, è stato raggiunto un risultato abbastanza elevato, segno che pochi pazienti siano stati classificati erroneamente; discorso simile per la classe "pneumonia" che dove alcuni pazienti sono stati classificati sia come appartenenti alla classe "covid-19" sia alla classe "normal". In generale, il modello ha ottenuto una Accuracy media discreta.

Anche nel caso di GoogLeNet, il risultato ottenuto per i pazienti della classe "covid-19" è basso, molto più basso rispetto ad AlexNet e, nonostante le Accuracy delle altre classi siano nella media, possiamo vedere come il suo peso si ripercuota sull'Accuracy media della rete.

Per quanto riguarda Inception V3, l'Accuracy della classe "covid-19" è più alta rispetto ad GoogLeNet, ma rimane comunque in forte dislivello se comparata ai risultati delle altre classi. In compenso l'Accuracy media della rete ottiene un risultato positivo.

VGG16 ottiene il 100% nella classificazione dei pazienti della classe "pneumonia", il risultato più alto finora nonché quello ottimale. Anche la classe "normal" ottiene un risultato molto alto, mentre la classe "covid-19", nonostante abbia ottenuto ottimi risultati, ha qualche leggera difficoltà nella classificazione. In compenso la rete si comporta egregiamente. Per quanto riguarda VGG19, benché rispetto a VGG16 non abbia ottenuto il 100% nella classe "pneumonia", questo modello raggiunge risultati elevati e uniformi: il 95,91% della classe "covid-19" infatti è il più alto finora, nonché difficile da raggiungere. Le altre classi ottengono anch'esse ottimi risultati e l'Accuracy media della rete lo dimostra col suo 97,31%.

Anche ShuffleNet si comporta come le reti nella media, portando a un risultato relativamente basso per la classe "covid-19" rispetto alle altre due classi che invece danno ottimi risultati. In ogni caso l'Accuracy della rete rimane comunque nella media vista finora.

Come si può vedere dai risultati in tabella, MobileNet V2 si comporta ottimamente nelle classi "pneumonia" e "normal", in compenso lo scarso risultato ottenuto nella classe "covid-19" si ripercuote nell'Accuracy media della rete.

Per quanto riguarda le reti ResNet: ResNet 18 ottiene alti risultati nelle classi "normal" e "pneumonia" e rimane nella media per quanto riguarda la classe "covid-19". Nella media anche l'Accuracy finale della rete; nonostante entrambe le reti ResNet 50 e ResNet 101 abbiano ottenuto altissimi risultati nelle due classi "pneumonia" e "normal", così come ha fatto ResNet 18, quelli ottenuti nella classe "covid-19" sono drasticamente bassi, i più bassi finora. Ciò rende ResNet 50 la rete con una Accuracy media più bassa, causata soprattutto dai risultati ottenuti dalla classificazione dei pazienti relativi alla classe "covid-19"; ResNet 101 è di poco superiore, ma comunque con risultati poco soddisfacenti.

Considerando i risultati ottenuti con VGG19, che è la rete con il minor numero di pazienti di classe "covid-19" classificati errati, si è andati a verificare nello specifico quali fossero per tentare di trarne delle conclusioni. VGG19 ha classificato in maniera errata sette pazienti e, di questi, non tutti sono stati classificati in maniera errata dalle altre reti; ciò indica che possa servire presumibilmente una rete ad hoc per questo tipo di lavoro.

#### 4.2.4 Classificazione binaria COVID-CT

Per questo lavoro si è proceduto in due tempi: inizialmente si è utilizzato l'intero dataset COVID-CT come testing dataset dei modelli ottenuti dal lavoro precedente con il dataset COVIDx CT-2A ma i risultati ottenuti non erano affatto soddisfacenti e ben lontani da quelli ottenuti durante il primo test. Successivamente si è deciso di eseguire un fine tuning solamente sui modelli più promettenti, quali VGG19, dati i suoi risultati in entrambi gli esperimenti, MobileNet V2, in quanto una delle reti più leggere e per aver portato ottimi risultati nella classificazione dei pazienti delle classi "normal" e "pneumonia", e infine, VGG16 e ResNet 18, in quanto essere le due reti con i risultati migliori dopo VGG19. Il dataset è stato quindi diviso in training, validation e testing, secondo le percentuali fornite dai fogli di calcolo dei creatori.

Per semplicità, verranno riportati solo i risultati pre e post fine tuning delle reti sopra citate.

Rete	Precision	Recall	Specificity	Accuracy	F-measure
VGG19	52,82%	52,82%	52,82%	52,82%	52,82%
VGG16	47,18%	47,18%	47,18%	47,18%	47,18%
ResNet 18	51,74%	51,74%	51,74%	51,74%	51,74%
MobileNet V2	50,94%	50,94%	50,94%	50,94%	50,94%

**Tabella 4.5:** *Performance in Micro-average per il testing su COVID-CT pre fine tuning*

Rete	Precision	Recall	Specificity	Accuracy	F-measure
VGG19	52,23%	52,13%	52,13%	52,82%	51,88%
VGG16	45,28%	45,94%	45,94%	47,18%	44,43%
ResNet 18	49,70%	49,81%	49,81%	51,74%	45,75%
MobileNet V2	49,11%	49,33%	49,33%	50,94%	46,72%

**Tabella 4.6:** *Performance in Macro-average per il testing su COVID-CT pre fine tuning*

Come si può vedere dalle tabelle, nonostante VGG19 fosse la migliore nei test precedenti, non ha prodotto gli stessi risultati con questo dataset. Per quanto riguarda VGG16, essendo che nei precedenti test ha ottenuto punteggi inferiori a VGG19, anche nella classificazione binaria di COVID-CT ha ottenuto risultati più bassi e poco soddisfacenti. Benché nei test precedenti VGG16 e ResNet 18 si comportassero in modo simile, in questo caso, anche prima del fine tuning ResNet 18 raggiunge

risultati superiori che si avvicinano a quelli ottenuti da VGG19. Discorso analogo per MobileNet V2, che pur non avendo raggiunto altissimi risultati precedentemente, in questo caso arriva a risultati poco al di sotto di ResNet 18. In sintesi, in questa classificazione cross-dataset, la peggiore è stata VGG16.

Procediamo ad analizzare i dati post fine tuning:

Rete	Risolutore	Max Epochs	Mini Batch size	Validation Frequency	Training Accuracy	Validation Accuracy
VGG19	adam	100	8	250	75,00%	73,28%
VGG16	adam	100	8	250	100,00%	69,83%
ResNet 18	adam	100	8	250	100,00%	70,69%
MobileNet V2	adam	100	8	250	62,50%	73,28%

**Tabella 4.7:** *Fine tuning delle reti con COVID-CT*

Il fine tuning ha permesso di raggiungere una Training Accuracy del 100,00% per le reti VGG16 e ResNet 18, ma, se si tiene conto della Validation Accuracy, si evince che nonostante VGG19 non abbia ottenuto risultati altissimi, rimanendo attorno al 75%, è l'unica che ottiene delle percentuali simili per entrambe le Accuracy; le prime due reti nominate, infatti, a prescindere dalla Training Accuracy ottenuta, hanno raggiunto risultati molto più bassi nella Validation Accuracy. MobileNet V2, invece, nonostante una Training Accuracy estremamente bassa, ha raggiunto un valore più alto nella Validation. Concludiamo con l'analisi dei risultati ottenuti in fase di testing:

Rete	Precision	Recall	Specificity	Accuracy	F-measure
VGG19	69,15%	69,15%	69,15%	69,15%	69,15%
VGG16	61,19%	61,19%	61,19%	61,19%	61,19%
ResNet 18	70,15%	70,15%	70,15%	70,15%	70,15%
MobileNet V2	67,16%	67,16%	67,16%	67,16%	67,16%

**Tabella 4.8:** *Performance in Micro-average per il testing su COVID-CT post fine tuning*

Grazie al fine tuning i risultati sono migliorati notevolmente, sempre al di sotto dei risultati raggiunti con il precedente dataset, ma migliorati sotto ogni aspetto. Questi risultati sotto la media possono essere dovuti al fatto che il dataset COVID-CT originale, risalente a inizio 2020, è stato lentamente modificato nel corso dei mesi con l'aggiunta di nuove immagini CT di scarsa qualità o compromesse da elementi in sovrapposizione. Ciò spiega perché le reti non riescano a classificare correttamente le immagini, in quanto essendo state allenate con le immagini di alta qualità

Rete	Precision	Recall	Specificity	Accuracy	F-measure
VGG19	70,19%	68,70%	68,70%	69,15%	68,40%
VGG16	61,19%	60,94%	60,94%	61,19%	60,84%
ResNet 18	70,16%	70,01%	70,01%	70,15%	70,02%
MobileNet V2	67,13%	67,05%	67,05%	67,16%	67,07%

**Tabella 4.9:** *Performance in Macro-average per il testing su COVID-CT post fine tuning*

presenti nel dataset COVIDx CT-2A, esse si trovino in difficoltà nel classificare con accuratezza questi nuovi elementi. Oltre al fine tuning, quindi, sarebbe necessario un lavoro di pre-processing che si presume possa aiutare notevolmente le reti. Nel dettaglio, i risultati di entrambe le tabelle non si discostano molto e si può affermare che in questo caso ResNet 18 sia stata capace di superare VGG19, con dei valori che si aggirano sempre al 70%. Quest'ultima ottiene dei valori accettabili assieme a MobileNet V2, mentre VGG16 continua ad avere difficoltà riguardo a ogni metrica di valutazione.



## Capitolo 5

### Conclusioni

L'obiettivo di questa tesi era quello di proporre una metodologia di classificazione per la diagnosi del COVID-19 attraverso tecniche di Deep Learning applicate su immagini CT. Per raggiungere tale fine, è stato realizzato un estensivo studio comparativo delle principali architetture CNN esistenti.

I test eseguiti sui due dataset hanno mostrato risultati molto distinti tra loro. I risultati ottenuti con il dataset COVIDx CT-2A sono ottimi per tutti i modelli utilizzati, in particolare spicca VGG19 per gli alti valori ottenuti non solo nella metrica Specificity, ma soprattutto in Precision e Recall, risultati che nessun'altra rete ha raggiunto. È importante dire, però, che anche reti come VGG16 e ResNet 18 abbiano raggiunto risultati più che soddisfacenti. Riguardo le altre reti, le meno adatte sono state GoogLeNet e ResNet 50, che si sono sempre discostate parecchio dai valori medi ottenuti. Volendo fare un confronto con le reti utilizzate nello studio di riferimento [18], possiamo affermare che i risultati ottenuti raggiungono quelli dello stato dell'arte.

Anche la classificazione orientata ai pazienti ha portato ottimi risultati, con dei valori di accuratezza della classe "covid-19" elevati, e, in alcuni casi, vi sono state delle accuratezze del 100% riguardo la classe "pneumonia". La rete migliore rimane in ogni caso VGG19, essendo quella con una Accuracy media più alta e, quindi, per aver classificato in maniera errata pochi pazienti rispetto alle altre reti. Tramite l'analisi dei pazienti classificati errati, si è dedotto che per poter ottenere un risultato ancora migliore, è necessario creare una rete ad hoc sfruttando le architetture neurali esistenti.

Per quanto riguarda il dataset COVID-CT invece, i risultati non raggiungono i precedenti e, anzi, si è verificato un calo di quasi il 50% delle prestazioni. Solamente

con il fine tuning si è riusciti a rimediare, incrementando i valori ottenuti del 20%, una cifra che non compensa comunque il dislivello venutosi a creare. Il problema può essere dovuto principalmente alla qualità delle immagini del secondo dataset, spesso compromesse o di pessima qualità.

Considerando l'obiettivo iniziale di questa tesi, in conclusione è stato raggiunto grazie all'analisi effettuata sulla classificazione ternaria, la variante orientata ai pazienti e su quella binaria, sebbene vi siano larghi margini di miglioramento, come dettagliato nella sezione successiva.

# Capitolo 6

## Sviluppi futuri

Considerati i promettenti risultati ottenuti grazie a questo lavoro, si possono delineare molteplici sviluppi futuri, orientati al miglioramento di quanto ottenuto finora, ad esempio, un lavoro di pre-processing potrebbe riportare i risultati ottenuti con il dataset COVID-CT ai valori inizialmente raggiunti con COVIDx CT-2A. Condurre nuovi test con dataset diversi, aventi immagini di qualità, potrebbe portare a ulteriori conferme se i risultati dovessero essere soddisfacenti.

Per migliorare ulteriormente i risultati, considerata l'assenza di dataset più bilanciati, sarebbe fondamentale bilanciare le classi in esame attraverso tecniche di undersampling o di oversampling, in quanto come già visto, in questo caso entrambi i dataset possedevano più immagini appartenenti alla classe "covid-19".

Infine la soluzione migliore sarebbe adottare un approccio ibrido, in cui unire i vantaggi delle reti VGG a quelli delle Residual Networks, che hanno dato fondamentalmente i risultati migliori.



# Bibliografia

- [1] Abracadabra. Micro- and macro-average of precision, recall and f-score. <https://tomaxent.com/2018/04/27/Micro-and-Macro-average-of-Precision-Recall-and-F-Score/>, 2018.
- [2] Roberto Alejo, J. A. Antonio, Rosa Maria Valdovinos, and J. H. Pacheco-Sanchez. Assessments metrics for multi-class imbalance learning: A preliminary study. In *Pattern Recognition - 5th Mexican Conference, MCPR 2013*, volume 7914, pages 335–343, 2013.
- [3] H. Yang B. Liu. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. <https://medium.com/syncedreview/shufflenet-an-extremely-efficient-convolutional-neural-network-for-mobile-devices-72c6f5b01651>, Medium.
- [4] N. Danielsen. Simple Image Classification with Resnet-50. <https://medium.com/@nina95dan/simple-image-classification-with-resnet-50-334366e7311a>, Medium.
- [5] Our World Data. Coronavirus Pandemic (COVID-19) - the data. ["https://ourworldindata.org/coronavirus-data"](https://ourworldindata.org/coronavirus-data).
- [6] DeepAI. Transfer Learning. <https://deepai.org/machine-learning-glossary-and-terms/transfer-learning>.
- [7] O. Gozes and M. Frid-Adar et. al. Rapid ai development cycle for the coronavirus (COVID-19) pandemic: initial results for automated detection & patient monitoring using deep learning CT image analysis. *arXiv:2003.05037*, 2020.
- [8] Margherita Grandini, Enrico Bagli, and Giorgio Visani. Metrics for multi-class classification: an overview. <https://arxiv.org/pdf/2008.05756.pdf>, arXiv 2008.05756, 2020.
- [9] H. Gunraj. COVID-Net Open Source Initiative - COVIDx CT-2 Dataset. <https://www.kaggle.com/hgunraj/covidxct>, 2020.

- [10] H. Gunraj and A. Sabri et. al. COVID-Net CT-2: Enhanced Deep Neural Networks for Detection of COVID-19 from Chest CT Images Through Bigger, More Diverse Learning. *arXiv:2101.07433*, 2021.
- [11] K. He and X. Zhang et. al. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [12] M. Hollemans. MobileNet version 2. <https://machinethink.net/blog/mobilenet-v2>, MachineThink.
- [13] Aaron Courville Ian Goodfellow, Yoshua Bengio. Deep Learning. *MIT Press*, 2016.
- [14] Ajitesh Kumar. ML metrics: Sensitivity vs. specificity. <https://dzone.com/articles/ml-metrics-sensitivity-vs-specificity-difference>, DZone.com, 2018.
- [15] V. Kurama. A Review of Popular Deep Learning Architectures: Alex-Net, VGG16, and GoogleNet. <https://blog.paperspace.com/popular-deep-learning-architectures-alexnet-vgg-googlenet/>, PaperSpace.
- [16] V. Kurama. A Review of Popular Deep Learning Architectures: ResNet, InceptionV3 and SqueezeNet. *PaperSpace*.
- [17] S. Brahnem L. Nanni, A. Lumini. Local binary patterns variants as texture descriptors for medical image analysis. *Artif. Intell. Med.*, 49 (2):117–125, 2010.
- [18] A. Wong L. Wang. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. *arXiv:2003.09871*, 2020.
- [19] L. Li and L. Qin et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest ct. *Radiology*, page 200905, 2020.
- [20] MathWorks. Rete Neurale Convolutionale, 3 cose da sapere. <https://it.mathworks.com/solutions/deep-learning/convolutional-neural-network.html>.
- [21] Help Center MathWorks. Resnet101. <https://it.mathworks.com/help/deeplearning/ref/resnet101.html>.
- [22] Help Center MathWorks. Resnet18. <https://it.mathworks.com/help/deeplearning/ref/resnet18.html>.
- [23] Help Center MathWorks. Resnet50. <https://it.mathworks.com/help/deeplearning/ref/resnet50.html>.

- [24] Help Center MathWorks. VGG 19. <https://it.mathworks.com/help/deeplearning/ref/vgg19.html>.
- [25] Giuseppe Placidi Matteo Polsinelli, Luigi Cinque. A light CNN for detecting COVID-19 from CT scans of the chest. *Pattern Recognition Letters*, 140:95–100, 2020.
- [26] M.M. Sathik N. Parveen. Detection of pneumonia in chest x-ray images. *J. Xray Sci. Technol.*, 19 (4):423–428, 2011.
- [27] Richard E. Woods. Rafael C. Gonzalez. Digital Image Processing. *Pearson*, 2007.
- [28] M.Z. Mao V. Vanhoucke, A. Senior. Improving the speed of neural networks on CPUs . <https://covid-19.conacyt.mx/jspui/handle/1000/4157>, 2011.
- [29] J. Zhao and Y. Zhang et. al. COVID-CT-dataset: a CT scan dataset about COVID-19. <https://arxiv.org/pdf/2003.13865v1.pdf>, 2020.
- [30] M. Zhou and Y. Chen et. al. Improved deep learning model for differentiating novel coronavirus pneumonia and influenza pneumonia. *medRxiv*, 2020.





# Elenco delle figure

1.1	<i>(a) CT dei polmoni di un paziente affetto da COVID-19 dove sono presenti evidenti tracce di GGO indicate con frecce rosse. (b) CT dei polmoni di un paziente non affetto da COVID-19 con opacità diffusa nelle parti esterne dei polmoni. Immagini estratte da uno dei dataset utilizzati [29]. . . . .</i>	2
2.1	<i>Struttura di una rete neurale convoluzionale . . . . .</i>	7
2.2	<i>Schema di funzionamento del Transfer Learning . . . . .</i>	8
3.1	<i>Architettura AlexNet . . . . .</i>	15
3.2	<i>Architettura GoogLeNet . . . . .</i>	15
3.3	<i>Architettura VGG 16 . . . . .</i>	17
3.4	<i>Architettura ResNet 18 . . . . .</i>	19



## Elenco delle tabelle

4.1	Training e Validation con COVIDx CT-2A . . . . .	25
4.2	Performance in Micro-average per il testing su COVIDx CT-2A . . .	26
4.3	Performance in Macro-average per il testing su COVIDx CT-2A . .	26
4.4	Testing orientato ai pazienti con COVIDx CT-2A . . . . .	28
4.5	Performance in Micro-average per il testing su COVID-CT pre fine tuning . . . . .	30
4.6	Performance in Macro-average per il testing su COVID-CT pre fine tuning . . . . .	30
4.7	Fine tuning delle reti con COVID-CT . . . . .	31
4.8	Performance in Micro-average per il testing su COVID-CT post fine tuning . . . . .	31
4.9	Performance in Macro-average per il testing su COVID-CT post fine tuning . . . . .	32



# Ringraziamenti

Giunto al termine, desidero innanzitutto ringraziare i miei relatori, la professoressa Cecilia Di Ruberto, per la professionalità, la cortesia e l'avermi dato la possibilità di svolgere questo lavoro, e il Dott. Andrea Loddo, per l'infinita disponibilità, il costante supporto e per l'enorme aiuto fornitomi, senza il quale non avrei potuto scrivere questa tesi.

Il ringraziamento più grande va alla mia famiglia, in particolare a mia madre, la donna più forte che conosca, che mi è sempre stata accanto, avendomi supportato in ogni scelta, senza mai pretendere nulla, per avermi confortato nei momenti più difficili e per essere stata la roccia su cui aggrapparmi. Ringrazio Gaia, con cui condivido gioie e disgrazie, per avermi fatto evadere dalla realtà nei momenti di stress. Ringrazio i miei nonni, per avermi sempre spronato a dare il meglio e per aver sempre gioito di ogni mio traguardo.

Come non ringraziare le mie due ancore: Melissa, anzi Meli, la mia compagna di vita, con cui ho condiviso pomeriggi di disperazione e notti insonni, ma soprattutto le gioie e i dolori di questa esperienza e di tutto quello che c'è al di fuori; e la mia Gommai Giulia, amica prima ancora di nascere e la mia supporter numero uno, una super donna fonte di gioie e ammirazione, distante nel mondo ma vicina a me in ogni situazione. E naturalmente ringrazio Davide, per aver portato un'altalena di emozioni che hanno colorato la mia vita e un disordine del quale non potrei vivere senza.

Ringrazio i miei amici, che non avrei mai pensato di averne così tanti da non poterli ringraziare uno ad uno: ringrazio gli amici storici, gli amici sparsi per il mondo, quelli con me dai tempi delle superiori e quelli venuti dopo, e ultimi ma non per importanza, i miei colleghi, senza i quali non avrei potuto vivere un'esperienza così bella in questi anni.

A tutti loro auguro il meglio nella vita e di sentirsi realizzati in tutto quello che per loro è importante, così come mi sento io oggi.

