

POLITECNICO DI TORINO

---

Corso di Laurea in Ingegneria Biomedica

Tesi di Laurea Magistrale

**Sviluppo di un classificatore  
basato su reti neurali  
convoluzionali per il  
riconoscimento del melanoma  
cutaneo**



**Relatore**

Prof. Kristen Meiburger

**Candidato**

Francesco BRANCIFORTI

**Correlatori:**

Prof. Filippo Molinari

Ing. Massimo Salvi

---

MARZO 2020





# Indice

<b>Elenco delle figure</b>	<b>5</b>
<b>Elenco delle tabelle</b>	<b>9</b>
<b>1 Introduzione</b>	<b>1</b>
1.1 Anatomia e fisiologia della cute . . . . .	1
1.1.1 Composizione della cute . . . . .	2
1.1.2 Funzioni della cute . . . . .	2
1.1.3 La melanina . . . . .	4
1.2 Lesioni melanocitiche . . . . .	5
1.2.1 Nevi melanocitici . . . . .	5
1.2.2 Melanoma maligno . . . . .	6
1.3 Lesioni non melanocitiche . . . . .	10
1.4 Dermatoscopia . . . . .	11
1.4.1 Criteri dermoscopicci . . . . .	15
1.4.2 Modelli diagnostici . . . . .	15
1.4.3 Esame dermoscopicco . . . . .	18
1.4.4 Follow up dermoscopicco . . . . .	19
<b>2 Materiali e Metodi</b>	<b>20</b>
2.1 Archivio ISIC 2019 . . . . .	20
2.2 Dataset Novara . . . . .	21
2.2.1 Strumento utilizzato . . . . .	21
2.3 Nurugo Derma . . . . .	23
2.3.1 Protocollo di acquisizione . . . . .	26
2.3.2 Algoritmo di segmentazione . . . . .	26
2.4 Cropping delle immagini . . . . .	28
2.4.1 MultiCrop . . . . .	30
2.5 Classificazione di lesioni cutanee . . . . .	31

2.5.1	Deep Learning e CNN . . . . .	31
2.5.2	Architettura della rete e allenamento . . . . .	37
2.5.3	Test Sets . . . . .	39
2.6	Metriche di validazione . . . . .	40
2.6.1	Confusion Matrix . . . . .	40
2.6.2	Curva ROC . . . . .	42
2.6.3	t-SNE . . . . .	44
2.6.4	Occlusion Sensitivity Map . . . . .	44
<b>3</b>	<b>Risultati</b>	<b>46</b>
3.1	Individuazione del crop ottimale . . . . .	46
3.1.1	Metodo M1 . . . . .	46
3.2	Architetture a cascata . . . . .	47
3.2.1	Metodo M2 . . . . .	49
3.2.2	Metodo M3 . . . . .	53
3.3	Confronto tra le tre metodologie . . . . .	57
3.3.1	Analisi delle soglie . . . . .	62
3.4	Prestazioni sul test-set Dermoscopico . . . . .	63
3.5	Prestazioni sul test-set Nurugo Derma . . . . .	65
<b>4</b>	<b>Conclusioni e sviluppi futuri</b>	<b>71</b>
4.1	Conclusioni . . . . .	71
4.2	Sviluppi futuri . . . . .	72
4.2.1	Miglioramento del classificatore . . . . .	72
4.2.2	Progettazione di un dispositivo "Nurugo-like" . . . . .	73
	<b>Bibliografia</b>	<b>77</b>

# Elenco delle figure

1.1	Struttura della cute [3]. . . . .	3
1.2	A sinistra immagine al microscopio ottico di melanociti. E' possibile osservare i pigmenti di melanina rilasciati dai melanociti. A destra raffigurazione schematica di un melanocita [2]. . . . .	4
1.3	Esempio di lesioni melanocitiche osservate in epiluminescenza. <b>a</b> , Immagini di nevo benigno. <b>b</b> , Immagini di melanoma maligno. [4]. . . . .	5
1.4	Numero stimato dei casi di incidenza del melanoma considerando entrambi i sessi e tutte le età [9]. . . . .	8
1.5	Tassi di mortalità normalizzati stimati a livello globale nel 2018. [9]. . . . .	8
1.6	Esempio di lesioni non melanocitiche osservate in epiluminescenza. <b>a</b> , Immagini di cheratosi seborroica. <b>b</b> , Immagini di carcinoma basocellulare. [4]. . . . .	11
1.7	A sinistra immagine digitale acquisita al dermatoscopio di nevo benigno. A destra esame dermoscopico [17]. . . . .	12
1.8	Cross polarizzazione[18]. . . . .	13
1.9	Nevo melanocitico osservato tramite <i>fotografia clinica</i> (A), <i>NPD</i> (B), <i>PCD</i> (C) e <i>PNCD</i> (D). E' possibile osservare come le strutture di regressione, tra cui il velo blu-bianco e le cisti, siano maggiormente distinguibili se osservate in NPD e quasi inosservabili se viste in PNCD [18]. . . . .	13
1.10	Melanoma maligno osservato tramite <i>fotografia clinica</i> (A), <i>NPD</i> (B), <i>PCD</i> (C) e <i>PNCD</i> (D). Le zone striate bianco-lucide sono osservabili unicamente tramite tecniche che sfruttano la polarizzazione della luce [18]. . . . .	14

1.11	Cheratosi seborroica osservata tramite fotografia clinica (A), <i>NPD</i> (B), <i>PCD</i> (C) e <i>PNCD</i> (D). L' acquisizione in <i>NPD</i> , al contrario delle restanti, permette di apprezzare maggiormente le aperture "comedolike" e le cisti. [18]. . . . .	14
1.12	Schematizzazione del metodo ABCD [19]. . . . .	16
1.13	Evoluzione nel tempo di un melanoma a crescita superficiale [21] . . . . .	19
2.1	Panoramica del dermoscopio HEINE Delta 20T [23]. . . . .	22
2.2	Esempio di crop su immagine dermoscopica. . . . .	23
2.3	Nurugo Derma montato su iPhone 6s. . . . .	24
2.4	A sinistra immagine di nevo benigno acquisita con Nurugo: in arancione l'effetto ombra, in rosso il glare effect. A destra stessa lesione cutanea acquisita con Nurugo con vetrino in condizione di epilinescenza. E' possibile notare l'artefatto di riflessione che ricopre gran parte dell'immagine. . . . .	25
2.5	Rappresentazione dei primi due step della trasforma di Hough: a) costruzione del vettore di accumulo b) stima del centro [24].	27
2.6	Diagramma di flusso dell'algorithmo di segmentazione. . . . .	29
2.7	Applicazione dell'algorithmo ad immagine di nevo benigno acquisita con Nurugo. . . . .	29
2.8	Esempio di crop su immagine di melanoma maligno con raggio pari a 1.14 mm . . . . .	30
2.9	Esempio di MultiCrop da 3 mm su immagine di cheratosi seborroica. In blu il crop centrale, in verde i crop considerati idonei, in rosso i crop scartati. . . . .	31
2.10	Struttura tipica di una rete neurale [25]. . . . .	32
2.11	Struttura tipica di una CNN [25]. . . . .	33
2.12	Diverse strategie di allenamento [26]. . . . .	34
2.13	Architettura di AlexNet [28]. . . . .	35
2.14	Panoramica comportamenti di una rete durante la fase di training. . . . .	36
2.15	Schematizzazione dei tre metodi proposti. . . . .	38
2.16	Esempio di confusion matrix per classificazione binaria. . . . .	41
2.17	Esempio di curva ROC. . . . .	43
2.18	Esempio di grafico t-SNE. . . . .	44
2.19	Esempio di visualizzazione tramite OSM: in questo caso sono analizzati i pixel che contribuiscono al riconoscimento del corno musicale [31]. . . . .	45

3.1	Distribuzioni t-SNE del training del metodo M1 a diversi valori di crop. . . . .	48
3.2	Distribuzione t-SNE della rete <b>M1b</b> sul testset1. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata. . . . .	49
3.3	Distribuzione t-SNE relativa all'allenamento della CNN 2.1 . . . . .	50
3.4	Confusion Matrix relativa alla CNN 2.1 . . . . .	50
3.5	Distribuzione t-SNE della rete CNN 2.1 sul testset-1. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata. . . . .	51
3.6	Distribuzione t-SNE relativa all'allenamento della CNN 2.2 . . . . .	52
3.7	Confusion Matrix relativa alla CNN 2.2 . . . . .	52
3.8	Distribuzione t-SNE della rete CNN 2.2 sul testset 1. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata. . . . .	53
3.9	Distribuzione t-SNE relativa all'allenamento della CNN 3.1 . . . . .	54
3.10	Confusion Matrix relativa alla CNN 3.1 . . . . .	54
3.11	Distribuzione t-SNE della rete CNN 3.1 sul testset 1. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata. . . . .	55
3.12	Distribuzione t-SNE relativa all'allenamento della CNN 3.2 . . . . .	56
3.13	Confusion Matrix relativa alla CNN 3.2 . . . . .	56
3.14	Distribuzione t-SNE della rete CNN 3.2 sul testset 1. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata. . . . .	57
3.15	Confusion Matrix relativa all'architettura M1, valutata sul test-set 1. . . . .	58
3.16	Confusion Matrix relativa all'architettura M2, valutata sul test-set 1. . . . .	58
3.17	Confusion Matrix relativa all'architettura M3, valutata sul test-set 1. . . . .	59
3.18	Curva ROC relativa all'architettura M1. Viene valutata la capacità della rete di discriminare il melanoma maligno. . . . .	60
3.19	Curve ROC relative all'architettura M2. La curva ROC a sinistra valuta la capacità della rete di discriminare le lesioni melanocitiche dalle lesioni non melanocitiche. La curva a destra invece analizza la capacità di differenziare il melanoma maligno dal nevo benigno. . . . .	60

3.20	Curve ROC relative all'architettura M3. La curva ROC a sinistra valuta la capacità della rete di discriminare le lesioni maligne dalle lesioni benigne. La curva a destra invece analizza la capacità di differenziare il nevo benigno dalla cheratosi seborroica. . . . .	61
3.21	Esempio di Occlusion Sensitivity Maps. . . . .	62
3.22	Confusion Matrix relativa all'architettura M1, valutata sul test-set 2 . . . . .	64
3.23	Confusion Matrix relativa all'architettura M2, valutata sul test-set 2 . . . . .	64
3.24	Confusion Matrix relativa all'architettura M3, valutata sul test-set 2 . . . . .	65
3.25	Confusion Matrix relativa all'architettura M1, valutata sul test-set 3. . . . .	66
3.26	Confusion Matrix relativa all'architettura M2, valutata sul test-set 3. . . . .	67
3.27	Confusion Matrix relativa all'architettura M3, valutata sul test-set 3. . . . .	67
3.28	Distribuzione t-SNE dell'architettura M1 sul testset 3. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata. E' possibile osservare come la maggior parte delle misclassificazioni siano collocate all'interno del cluster delle cheratosi seborroiche, in giallo. . . . .	69
3.29	Esempi di immagini di nevo benigno e melanoma maligno classificate erroneamente come cheratosi seborroica. E' possibile osservare i diversi artefatti, bolle d'aria, mancanza di messa a fuoco e striature, che inficiano sulla qualità dell'immagine. . .	70
4.1	Prototipo "Nurugo-like". . . . .	74

# Elenco delle tabelle

1.1	Principali criteri dermoscopic [16] . . . . .	15
2.1	Caratteristiche NIKON E4500. . . . .	22
2.2	Caratteristiche immagini acquisite con Nurugo. . . . .	26
2.3	Composizione del dataset per il training delle diverse strutture proposte . . . . .	39
2.4	Parametri di Data Augmentation utilizzati . . . . .	39
2.5	Composizione dei diversi dataset di test. . . . .	40
3.1	Performance delle tre reti sul testset 1. . . . .	47
3.2	Performance delle tre metodologie sul testset 1. . . . .	57
3.3	Performance dei tre dermatologi sul testset 1. . . . .	59
3.4	AUC per le diverse CNN. . . . .	61
3.5	Confronto tra le prestazioni delle architetture M2 e M3 sul test-set1, utilizzando criteri di calcolo della predizione diversi. . . . .	63
3.6	Performance delle tre architetture sul test-set 2 . . . . .	63
3.7	Performance delle tre architetture sul test-set 3 . . . . .	66

## Sommario

Il melanoma è una pericolosa ed aggressiva forma di tumore maligno cutaneo responsabile di migliaia di morti ogni anno in tutto il mondo. La diagnosi precoce, condotta attraverso l'esame visivo dermoscopico, risulta fondamentale per evitare la progressione della malattia e ridurre conseguentemente l'impatto fisico e psicologico della terapia sul paziente. Questa tipologia d'esame presenta tuttavia diverse criticità essendo una tecnica diagnostica che richiede grande esperienza dell'operatore. L'interpretazione della lesione e l'applicazione dei diversi criteri diagnostici non è infatti immediata, e può spesso condurre a misclassificazioni. Da qui la necessità di affiancare al dermatologo strumenti di diagnosi intelligenti in grado di aumentare l'accuratezza e rendere la valutazione più robusta, diminuendo così da un lato i casi di melanomi cutanei non riconosciuti e dall'altro lato il numero di escissioni non necessarie.

Questo lavoro di tesi si focalizza sulla realizzazione di un classificatore basato su Reti Neurali Convoluzionali (CNN) in grado di distinguere tre differenti tipologie di lesioni cutanee: nevo cutaneo, melanoma e cheratosi seborroica. Il problema di classificazione viene affrontato proponendo tre strutture aventi architetture differenti. Il primo metodo, M1, esegue una classificazione ternaria one-shot, distinguendo le tre lesioni in un unico passaggio. Il secondo metodo, M2, propone invece un'architettura a cascata: la prima rete distingue le lesioni in melanocitiche e non melanocitiche, la seconda rete invece in nevo benigno e melanoma maligno; questo approccio ricalca lo standard della visita dermatologica. Infine il terzo metodo, M3, anch'esso con architettura a cascata, differenzia inizialmente le lesioni in benigne e maligne per poi effettuare una distinzione tra nevo benigno e cheratosi seborroica.

Le CNN dei tre metodi sono state allenate in maniera supervisionata su 600 immagini provenienti dal database ad accesso libero ISIC.

Parallelamente è stata valutata la validità in campo clinico della lente Nurugo Derma, un dispositivo da smartphone che permette di ricavare immagini cutanee a livello micro. Essendo pensato per un utilizzo amatoriale presenta conseguentemente diverse limitazioni: per ottenere immagini in epiluminescenza è necessario ricorrere ad un vetrino da laboratorio trasparente, da interporre tra cute e Nurugo al fine di applicare del liquido di interfaccia. Le immagini così ottenute sono però caratterizzate da un artefatto di riflessione che ne limita il FOV sulla lesione: è stato sviluppato un algoritmo di



segmentazione in grado di isolare l'area dell'immagine non corrotta. Presso il reparto di dermatologia dell'Azienda Ospedaliera Universitaria Maggiore della Carità di Novara sono state acquisite tramite Nurugo Derma e dermoscopio professionale immagini di 20 nevi benigni, 20 melanomi maligni e 20 di cheratosi seborroiche.

Le immagini così raccolte, insieme ad un test-set proveniente dall'archivio ISIC, sono state utilizzate per testare le tre metodologie proposte, al fine di individuare la struttura più performante. Le prestazioni delle reti sono state poi confrontate con i punteggi ottenuti da alcuni dermatologi esperti, dimostrando un buon *agreement*.

Questo lavoro costituisce una base di partenza per la realizzazione di un sistema di diagnosi assistita intelligente, basata su CNN, utilizzabile dal dermatologo come strumento di supporto durante il processo di diagnosi.

# Capitolo 1

## Introduzione

In questo capitolo viene presentata la cute dal punto di vista anatomico e fisiologico, con una spiegazione dettagliata della composizione e delle molteplici e differenti funzioni che espleta per il corpo umano. Segue un focus sull'importanza e sul ruolo del pigmento melanico. Successivamente vengono descritte le principali lesioni cutanee, differenziando la tipologia non melanocitica dalla melanocitica, con un approfondimento sul melanoma maligno. Ne sono indagati gli aspetti epidemiologici (diffusione, incidenza e mortalità a livello locale e globale), i fattori di rischio e le differenti sottotipologie. Infine viene presentato il mondo della dermoscopia, soffermandosi sugli strumenti, sulle diverse tecniche esistenti e sui principi fisici correlati, e sui differenti modelli di valutazione utilizzati dal dermatologo nel percorso diagnostico. Il capitolo si chiude con una descrizione della visita dermatologica standard e sull'importanza del follow-up nei pazienti a rischio.

### 1.1 Anatomia e fisiologia della cute

La cute, o pelle, è un involucro continuo che riveste, con funzione prevalentemente protettiva e sensitiva, tutto il corpo umano [1]. È l'organo più pesante ed esteso del corpo: in un adulto può occupare una superficie di 1.5-2 m<sup>2</sup> per un peso totale di 15 kg. Si estende inoltre anche a livello degli orifizi (mucose degli apparati respiratorio, digerente, urogenitale, e a livello del margine palpebrale e dei punti lacrimali) [2]. Le caratteristiche macroscopiche principali della cute sono:

- spessore
- colore

- aspetto
- elasticità
- estensibilità

Queste proprietà variano notevolmente a seconda delle zone corporee, dell'età, del peso, dello stato nutrizionale, della razza e del fototipo.

### 1.1.1 Composizione della cute

La pelle è caratterizzata da una struttura stratificata articolata in:

1. **epidermide**: è lo strato più superficiale. Consiste in un epitelio pluri stratificato altamente differenziato. A sua volta è composto, in successione, dall'interno verso l'esterno, da strato basale, strato spinoso, strato granuloso, strato lucido e strato corneo.
2. **derma**: si trova al di sotto dell'epidermide. E' un'impalcatura connettivale ricca di strutture vascolo-nervose. E' possibile distinguere tra derma papillare, reticolare, perianessiale e avventiziale, non è presente però nessuna distinzione netta tra gli strati.
3. **ipoderma**: è lo strato più profondo. E' costituito da tessuto adiposo organizzato in lobuli. All'interno dell'ipoderma si distinguono tre strati di tessuto connettivo: lamina superficiale, lamina intermedia e lamina profonda.

Sono considerati parte della cute anche gli annessi cutanei, ovvero strutture epiteliali e cornee come le unità pilo-sebacee, le ghiandole sudoripare e le unghie.

### 1.1.2 Funzioni della cute

La cute è un organo multifunzionale, in grado di espletare contemporaneamente diverse funzioni.

- **Funzione di barriera**: innanzitutto costituisce un ostacolo per sostanze potenzialmente dannose per l'organismo, non permettendo il passaggio di acqua e sostanze idrosolubili. Previene inoltre la dispersione di liquidi corporei in eccesso.

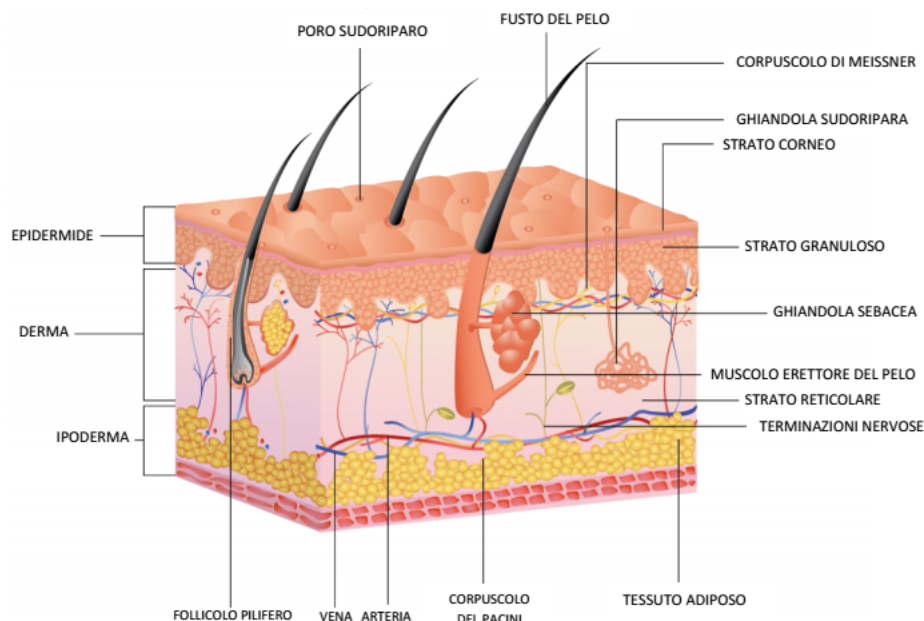


Figura 1.1: Struttura della cute [3].

- **Funzione termoregolatrice:** contribuisce al mantenimento della temperatura corporea dissipando il calore in eccesso tramite la sudorazione.
- **Funzione sensoriale:** è dotata di terminazioni sensoriali che garantiscono sensibilità tattile e dolorifica e in grado di captare stimoli meccanici e termici.
- **Funzione metabolica:** essendo dotata di gran parte dei sistemi enzimatici dei più importanti percorsi metabolici, è in grado di sintetizzare glicogeno, mucopolisaccaridi, lipidi e proteine.
- **Funzione endocrina:** sintetizza la vitamina D<sub>3</sub> e alcuni ormoni sessuali.
- **Funzione secretoria:** assolve a diverse funzioni secretorie, tra le quali cheratinopoesi, pigmentogenesi, secrezione sebacea e sudorale.
- **Funzione di mantenimento dell'omeostasi pressoria:** il tegumento è dotato di un'ampia rete vascolare in grado di fungere da serbatoio per il sangue circolante, mantenendo così un livello di pressione arteriosa sufficiente a soddisfare le esigenze metaboliche di alcuni organi quali cervello, reni e fegato.

- **Funzione immunitaria:** la presenza di diverse popolazioni cellulari permette di liberare dei mediatori solubili che esercitano effetti biologici a livello cutaneo e a distanza.

### 1.1.3 La melanina

La melanina è un pigmento marrone scuro che determina il colore di pelle, capelli e occhi e protegge dagli effetti del sole.

La sintesi della melanina, detta *melanogenesi*, avviene ad opera dei melanociti, cellule ospiti derivanti dalla cresta neurale, localizzate nella giunzione dermo-epidermica. Quest'ultime contengono al loro interno i melanosomi, organuli enzimaticamente attivi in cui viene sintetizzato il pigmento melanico. Una volta prodotti, i granuli di pigmento vengono trasferiti ai cheratinociti (cellule che sintetizzano la cheratina) epidermici circostanti.

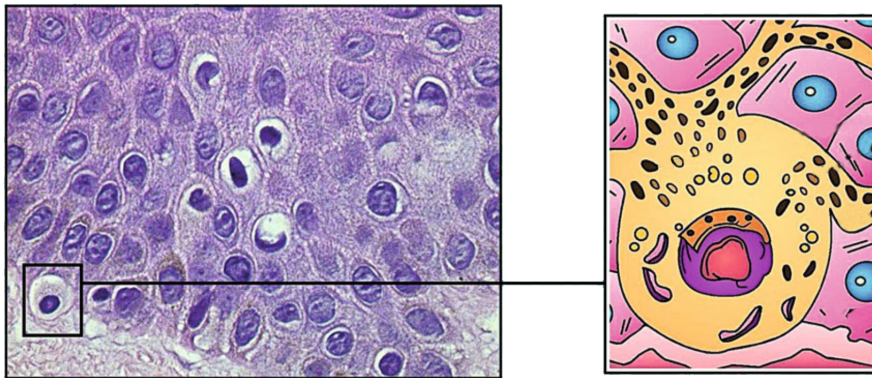


Figura 1.2: A sinistra immagine al microscopio ottico di melanociti. E' possibile osservare i pigmenti di melanina rilasciati dai melanociti. A destra raffigurazione schematica di un melanocita [2].

La melanina presente nell'organismo è suddivisibile in:

- *Costituzionale:* la cui quantità è determinata geneticamente;
- *Facoltativa:* la cui quantità è legata a fattori ambientali (esposizione solare) o ormonali.

Il ruolo principale della melanina è proteggere il genoma dai raggi UV, particolarmente nocivi, formando un scudo, che avvolge il nucleo dei cheratinociti, in grado di assorbire le radiazioni solari e neutralizzare la formazione di radicali liberi.

## 1.2 Lesioni melanocitiche

Un'iper proliferazione dei melanociti, congenita o indotta, porta alla formazione di lesioni cutanee di tipo *melanocitico*. Le lesioni di tipo benigno prendono il nome di *nevi* o *nei*, sono particolarmente frequenti nei soggetti di razza caucasica e presentano caratteristiche macroscopiche variegata. I melanociti possono tuttavia proliferare in maniera più violenta, invadendo e distruggendo le cellule sane circostanti, formando una massa di tessuto maligno che prende il nome di *melanoma maligno*.

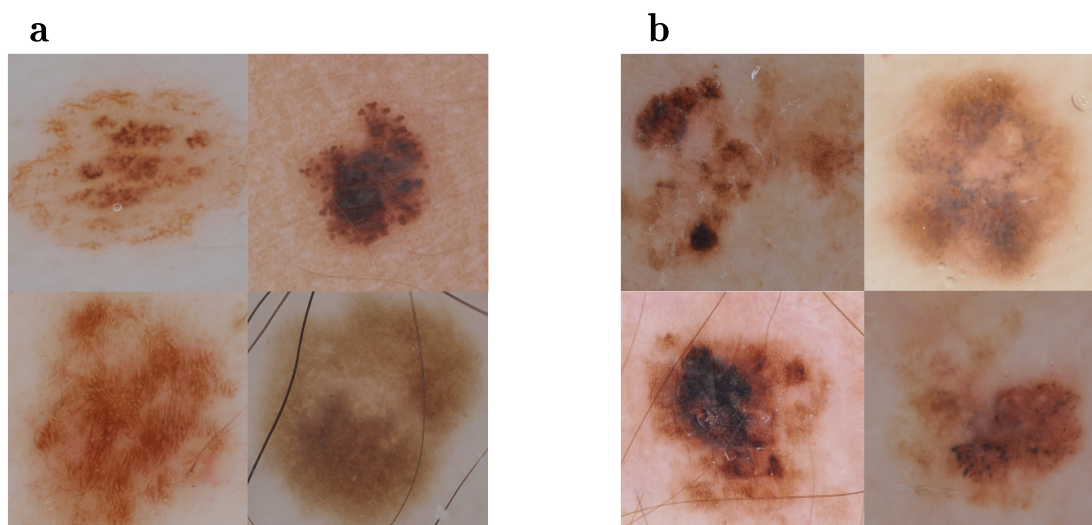


Figura 1.3: Esempio di lesioni melanocitiche osservate in epiluminescenza. **a**, Immagini di nevo benigno. **b**, Immagini di melanoma maligno. [4].

### 1.2.1 Nevi melanocitici

I nevi melanocitici sono determinati da una proliferazione benigna di melanociti, normalmente presenti nello strato basale dell'epidermide. In relazione al periodo di incidenza possiamo suddividere i nevi in:

- **Connatali:** presenti già alla nascita;
- **Congeniti:** presenti nel patrimonio genetico, compaiono al massimo entro il secondo o terzo anno di età;
- **Acquisiti:** compaiono nel corso della vita.

Considerando invece la localizzazione della melanina, è possibile suddividere i nevi in:

- **Giunzionali:** la melanina si trova principalmente a livello della membrana basale;
- **Dermici:** la melanina è distribuita nel derma, al di sotto della membrana basale;
- **Composti:** la melanina si trova sia a livello del derma che della membrana basale.

Particolare attenzione meritano i *nevi displastici*, o aticipi. Sono nevi benigni che presentano somiglianza con i melanomi. La presenza in un individuo di nevi displastici aumenta il rischio di sviluppare melanomi, in maniera proporzionale alla numerosità. Generalmente i nevi displastici sono presenti più in pazienti con melanoma che in pazienti sani. Dal punto di vista epidemiologico circa il 2-8% della popolazione di razza Caucasica presenta nevi displastici [5].

Dal punto di vista clinico il nevo displastico è caratterizzato da grandezza variabile, bordo irregolare e colore mutevole dal marrone scuro al rossiccio. E' localizzato principalmente su tronco, addome e braccia, difficilmente in zone non foto esposte. Pazienti con più di 100 nevi displastici sono considerati affetti dalla *sindrome del nevo displastico* e presentano un rischio maggiore di sviluppare melanoma. I nevi displastici devono essere costantemente sorvegliati e/o escissi per evitare che degenerino in forma maligne [6].

## 1.2.2 Melanoma maligno

Il melanoma è una neoplasia maligna che origina dai melanociti. Data la natura embrionale dei melanociti, che originano dalla cresta neurale, i melanomi possono insorgere anche su mucose, meningi e uvea. Spesso il melanoma rimane non diagnosticato o erroneamente confuso con lesioni benigne della pelle.

### Epidemiologia

Negli ultimi cinquanta anni l'incidenza del melanoma è aumentata a livello globale, rappresentando un importante problema socio-economico. In particolare negli USA, è aumentata del 270% dal 1973 al 2002 rendendo il melanoma il quinto tumore più comune nell'uomo e il sesto nelle donne. Attualmente

un americano su 63 sviluppa il melanoma durante la propria vita. In Italia l'incidenza del melanoma è pari a 5-7 casi su 100.000 abitanti per anno, anche se le popolazioni Mediterranee sono considerate a minor rischio. In Europa infatti i tassi di incidenza sono maggiori per le regioni del Nord rispetto a quelle del Sud, probabilmente a causa della maggiore protezione contro i raggi UV garantite dalla cute più pigmentata e del differente pattern solare. Parallelamente è stato osservato un aumento del tasso di mortalità dovuto al melanoma: annualmente circa 55.500 morti sono dovute al melanoma cutaneo [7]. E' stato osservato come, negli USA, a parità di età, il tasso di mortalità sia maggiore negli uomini che nelle donne, a parità di età. In figura 1.5 i tassi di mortalità normalizzati stimati a livello globale nel 2018. A differenza di altri tipi di tumori, il melanoma affligge principalmente individui giovani e di media età: l'età media di diagnosi è di 57 anni, mentre l'incidenza aumenta linearmente dopo i 25 anni fino ai 50 anni, per poi decrescere. Per quanto riguarda l'incidenza in relazione al sesso esistono dati contrastanti. Secondo alcuni studi gli uomini hanno una probabilità 1.5 volte maggiore delle donne di sviluppare melanoma cutaneo, secondo altri il tasso di incidenza tra i due sessi è correlato all'età. E' stato osservato infine come negli uomini il melanoma si diffonda principalmente sulla schiena, mentre nelle donne sulle braccia e sulle gambe.

Il tasso di incidenza inoltre è differente tra le diverse razze: le popolazioni bianche hanno un rischio 10 volte maggiore di sviluppare il melanoma rispetto alle popolazioni nere, asiatiche o ispaniche: secondo i dati raccolti nel 2007 dal SEER (*The National Cancer Institute Surveillance, Epidemiology, and End Results*) le popolazioni bianche vantano un tasso di incidenza pari a 27.5 su 100.000 abitanti, mentre le popolazioni nere un tasso pari a 1.1 su 100.000 [8]. In figura 1.4 e 1.5 è possibile apprezzare rispettivamente il numero stimato, a livello globale, dei casi di incidenza del melanoma considerando entrambi i sessi e tutte le età, e i tassi di mortalità normalizzati, riportati dal GCO (*Global Cancer Observatory*).

## Fattori di rischio

Il melanoma è considerata una patologia multi-fattoriale derivante dall'interazione tra sensibilità genetica ed esposizione ambientale [8]. I principali fattori di rischio per l'incidenza del melanoma cutaneo sono:

- **Familiarità:** è il fattore di rischio più importante, anche se non sono ancora stati individuati i veri e propri geni coinvolti nella predisposizione allo sviluppo del melanoma.



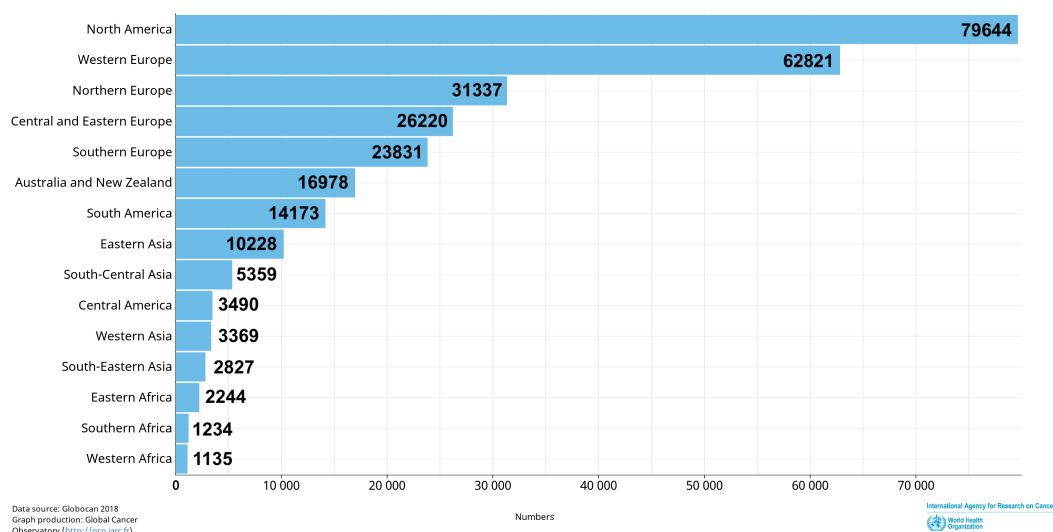


Figura 1.4: Numero stimato dei casi di incidenza del melanoma considerando entrambi i sessi e tutte le età [9].

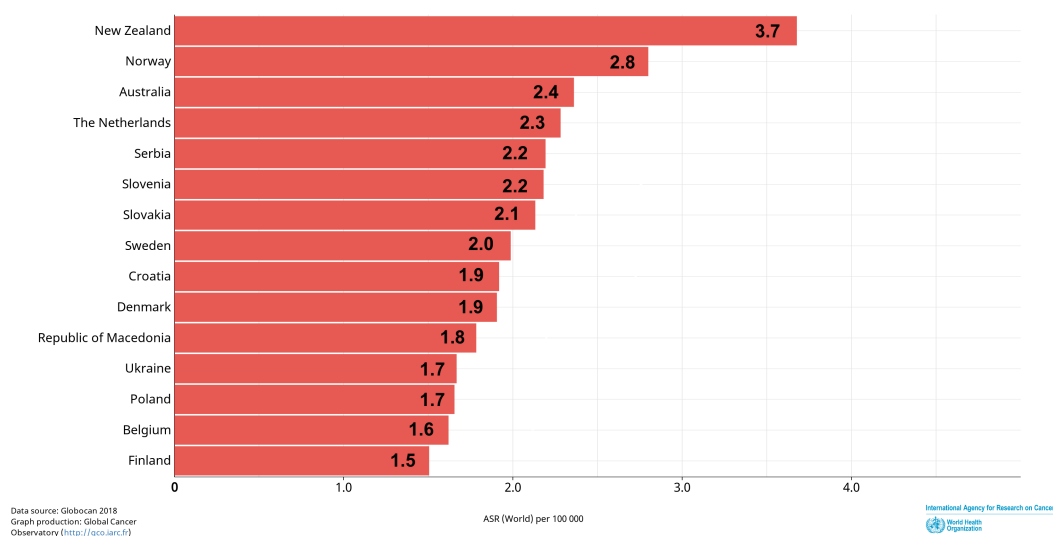


Figura 1.5: Tassi di mortalità normalizzati stimati a livello globale nel 2018. [9].

- **Esposizione ai raggi UV:** sia naturali che artificiali, determinano un *effetto genotossico* per le cellule. L'esposizione intermittente al sole è considerata uno dei fattori più determinanti.
- **Fototipo chiaro:** soggetti di fototipo 1 e 2 sono più suscettibili allo

sviluppo del melanoma.

- **Ustioni solari** in età giovanile-pediatrica.

Il melanoma può insorgere su cute apparentemente sana nel 70% dei casi oppure su precedenti lesioni pigmentate, definite *precursori*, nel restante 30%. Tra le lesioni precursori annoveriamo nevi congeniti o connatali, nevi displastici e lentigo maligna. In particolare soggetti con un numero elevato di nevi e soprattutto di grandi dimensioni ( $> 5$  mm) sono considerati più a rischio.

### Classificazione

Considerando le caratteristiche cliniche e istologiche e l'incidenza, è possibile suddividere il melanoma in tre sotto categorie principali [8]:

1. *Melanoma a diffusione superficiale (SSM)*: è il tipo di melanoma più comune (circa il 70% del casi). E' caratterizzato da una prima fase di crescita radiale orizzontale, seguita da una proliferazione verticale.
2. *Melanoma nodulare (NM)*: neoplasia a progressione rapida. E' caratterizzata da un comportamento biologico particolarmente aggressivo (immediata crescita verticale) con alto potenziale metastatico. E' comunemente privo di caratteristiche dermoscopiche e quindi particolarmente critico da diagnosticare.
3. *Lentigo Maligna (LM)*: è il tumore più comune del viso. E' caratterizzato da una lunga crescita orizzontale all'interno dell'epidermide, seguita da un'invasione del derma sottostante. La diagnosi è resa critica dalla sovrapposizione tipica della lesione con altre macchie pigmentate dovute all'esposizione solare [10].

Esistono anche altre tipologie a minor incidenza:

- *Melanoma acrale lentiginoso (ALM)*: è una forma tumorale con bassissima incidenza (5%) tra le popolazioni bianche ma molto comune tra le popolazioni asiatiche ispaniche e africane. Si sviluppa principalmente su pelle glabra.
- *Melanoma dermoplastico (DM)*: colpisce individui tra i 60 e i 70 anni, diffondendosi sul collo e sulla testa. Può essere *amelanotico*.

## Trattamento

La scelta del metodo di trattamento dipende da numerose variabili quali l'estensione della malattia, la stadiazione ed eventuali patologie concomitanti. Il trattamento di elezione per il melanoma cutaneo è la chirurgia di escissione, particolarmente efficace negli stadi iniziali del tumore: consiste nell'asportazione del tessuto malato e di una porzione di tessuto sano circostante. L'asportazione è seguita da un'analisi istopatologica del campione per verificare la presenza di cellule tumorali. Talvolta vengono rimossi anche i cosiddetti linfonodi "*sentinella*", che potrebbero aver ricevuto linfa direttamente dal tumore. Nel caso di malattia avanzata, con presenza di metastasi, è necessario ricorrere alla chemioterapia combinata con l'utilizzo di radioterapia, il cui utilizzo però negli ultimi anni è sempre più ridotto grazie ai progressi nel campo dell'immunoterapia e della terapia a bersaglio molecolare [11]. La diagnosi precoce del melanoma dunque risulta fondamentale per garantire un trattamento efficace e meno debilitante possibile per il paziente.

## 1.3 Lesioni non melanocitiche

Le lesioni non melanocitiche sono lesioni cutanee che non originano da una proliferazione di melanociti. Nonostante un'apparente somiglianza morfologica con le lesioni melanocitiche, a livello dermoscopico presentano strutture e caratteristiche differenti. Tra le lesioni non melanocitiche più comuni annoveriamo:

- **Cheratosi seborroica:** è uno dei tumori *benigni* cutanei più comuni. Dal punto di vista clinico presenta placche o papule ben circoscritte con cisti pseudocorniche, con colorazioni dal marroni chiaro al marrone scuro o nero. Dal punto di vista dermoscopico è caratterizzata da bordi di tipo "*moth-eaten*", cisti multiple simili a milia, aperture simili a comedoni, giri e solchi e strutture simili ad impronte digitali [12].
- **Carcinoma a cellule basali:** è il tumore *maligno* della cute più comune e anche il più comune tra tutte le forme di tumore. Ha origine da proliferazioni neoplastiche di cellule basali localizzate nello strato più profondo dell'epidermide. L'esposizione cronica e cumulativa ai raggi UV rappresenta la causa principale del carcinoma basocellulare. Dal punto di vista clinico si presenta come placche rossastre, ulcerazioni o erosioni non tendenti alla guarigione, escrescenze di colore rosa, noduli e cicatrici [13]. Al dermoscopio invece mostra pattern caratterizzati da

aree e globuli grigio-blu, nidi ovoidali grigio-blu, peppering, strutture a forma di ruota di carro, aree a foglia d'acero, teleangectasie arboriformi, ed ulcerazione [14].

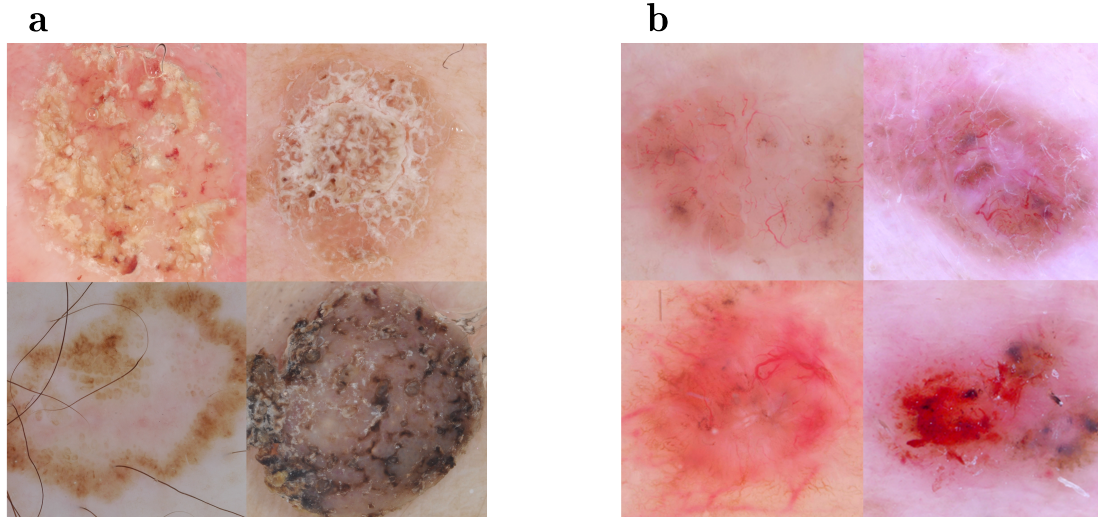


Figura 1.6: Esempio di lesioni non melanocitiche osservate in epiluminescenza. **a**, Immagini di cheratosi seborroica. **b**, Immagini di carcinoma basocellulare. [4].

## 1.4 Dermatoscopia

La dermatoscopia, o dermoscopia o *epiluminescenza*, è una metodica diagnostica *non invasiva* che permette di visualizzare in vivo strutture e pattern cutanei invisibili ad occhio nudo. E' ampiamente utilizzata nella pratica clinica per la diagnosi precoce del melanoma [15]. La tecnica prevede l'utilizzo di un *dermatoscopio*, uno strumento ottico in grado di generare un fascio di luce che incide la superficie cutanea con un angolo di  $20^\circ$  circa[16].

A causa della notevole differenza tra l'indice di rifrazione dello strato corneo (1.55) e l'indice di rifrazione dell'aria (1.0), gran parte della luce incidente viene riflessa. Per poter osservare correttamente le strutture cutanee è necessario ridurre la componente di luce riflessa dalla pelle.

In relazione alla soluzione utilizzata per raggiungere questa condizione è possibile suddividere la dermatoscopia in:

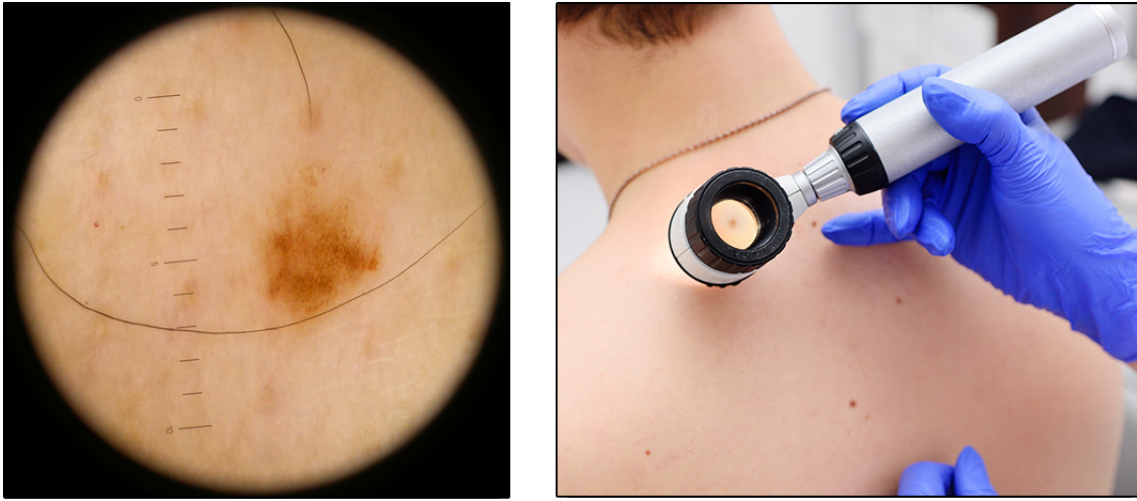


Figura 1.7: A sinistra immagine digitale acquisita al dermatoscopio di nevo benigno. A destra esame dermoscopic [17].

### Dermatoscopia non-polarizzata

Per attenuare la componente di luce riflessa viene interposto un *fluido di interfaccia*, generalmente olio o gel ecografico (considerato il miglior fluido di immersione), tra lo strato corneo e il piatto in vetro del dermatoscopio, in modo da far combaciare i due indici di riflessione e aumentare la trasparenza dello strato corneo. La tecnica richiede dunque il contatto con la superficie della cute ed è indicata per l'analisi delle strutture cutanee più superficiali.

### Dermatoscopia polarizzata

In questo caso la riduzione della componente di luce riflessa viene ottenuta tramite *cross-polarizzazione*: posizionando un filtro polarizzatore lineare a valle della sorgente luminosa e un secondo filtro polarizzatore lineare, ruotato di  $90^\circ$  rispetto al primo, in ingresso all'ottica di acquisizione, è possibile bloccare la componente di luce riflessa, permettendo unicamente il passaggio della componente di luce *back-scattered*, contenente informazioni sulle strutture cutanee più profonde. Il metodo non richiede dunque l'utilizzo di fluidi di interfaccia. Può essere eseguita a contatto con la pelle (**PCD** - *Polarized Contact Dermoscopy*) o non a contatto (**PNC** - *Polarized Non Contact Dermoscopy*). È indicata per l'osservazione più profonda dei pattern cutanei.

Nelle ultime tre decadi la dermatoscopia non polarizzata (**NPD** - *Non Polarized Dermoscopy*) ha costituito lo standard di riferimento in campo

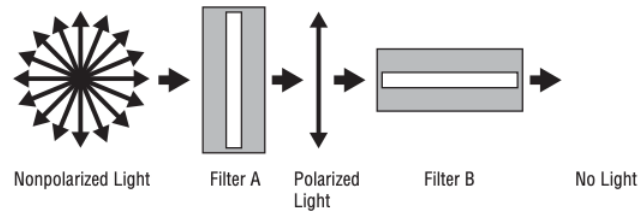


Figura 1.8: Cross polarizzazione[18].

dermatologico, tuttavia l'uso delle tecniche basate sulla polarizzazione sta diventando sempre più frequente.

Per evitare casi di classificazione errata è necessario tener presente come i medesimi colori e le medesime strutture abbiano un aspetto che varia al variare della tecnica utilizzata, come mostrato in figura 1.9, 1.10 e 1.11. In particolare è stato osservato come i colori appaiano più piatti e meno distorti per la NPD, soprattutto il marrone e il blu. Al tempo stesso la PD permette di osservare strutture profonde e un contrasto migliore [18]. La scelta della tecnica da utilizzare dunque dipende principalmente dal tipo di struttura/pattern che si cerca di individuare.

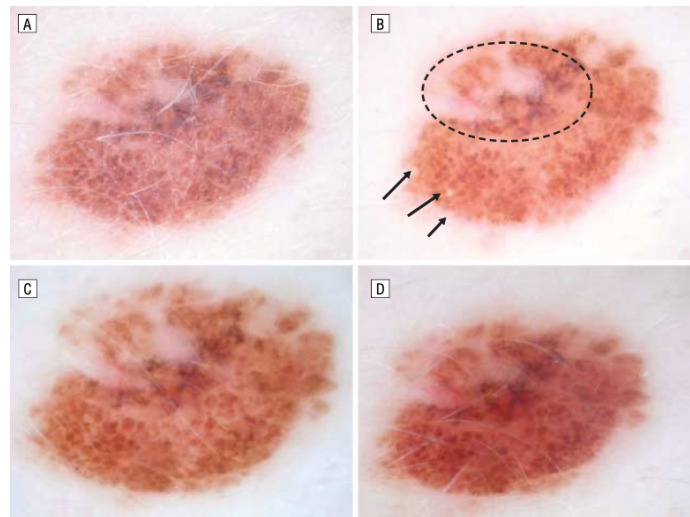


Figura 1.9: Nevo melanocitico osservato tramite *fotografia clinica* (A), *NPD* (B), *PCD* (C) e *PNCD* (D). E' possibile osservare come le strutture di regressione, tra cui il velo blu-bianco e le cisti, siano maggiormente distinguibili se osservate in NPD e quasi inosservabili se viste in PNCD [18].



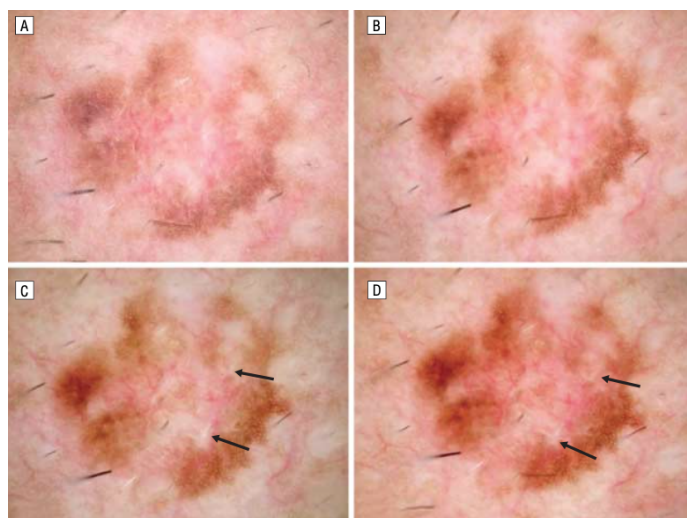


Figura 1.10: Melanoma maligno osservato tramite *fotografia clinica* (A), *NPD* (B), *PCD* (C) e *PNCD* (D). Le zone striate bianco-lucide sono osservabili unicamente tramite tecniche che sfruttano la polarizzazione della luce [18].

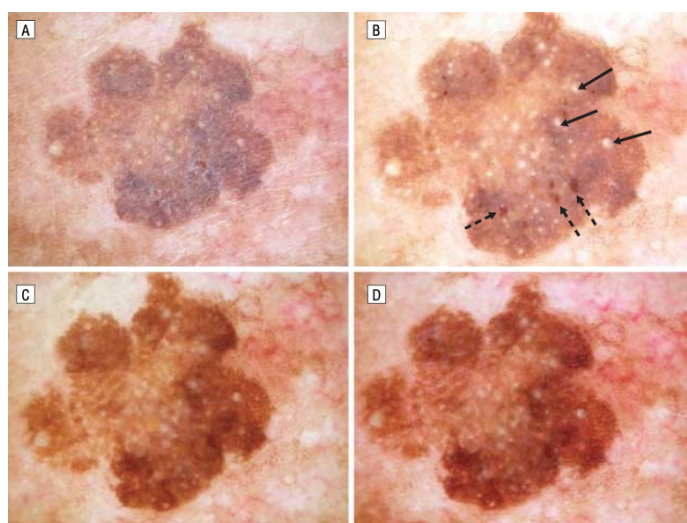


Figura 1.11: Cheratosi seborroica osservata tramite *fotografia clinica* (A), *NPD* (B), *PCD* (C) e *PNCD* (D). L'acquisizione in NPD, al contrario delle restanti, permette di apprezzare maggiormente le aperture "comedolike" e le cisti. [18].

### 1.4.1 Criteri dermoscopic

All'inizio del percorso diagnostico è di fondamentale importanza l'individuazione di criteri morfologici standard all'interno della lesione, in grado di indirizzare l'interpretazione della stessa. I diversi criteri morfologici si diversificano tra loro per:

- **Colore:** dipende dalla posizione della melanina tra i diversi strati della cute.
- **Struttura:** dipende dalla distribuzione spaziale della melanina e dell'emoglobina.

In tabella 1.1 sono mostrati i criteri più utilizzati.

Criterio dermoscopic	Tipo di lesione
Rete pigmentata, globuli aggregati, striature ramificate	Lesione melanocitica
Area blu omogenea	Nevo blu
Pseudocisti cornee, pseudoaperture follicolari o fessure e cripte	Cheratosi seborroica
Chiazze rosse aggregate	Emangioma
Aree che assomigliano a una foglia d'acero, aree radiali, telangiectasia, ulcerazione	Carcinoma basocellulare
Altre strutture differenti	Lesione melanocitica atipica

Tabella 1.1: Principali criteri dermoscopic [16]

### 1.4.2 Modelli diagnostici

Nella pratica diagnostica vengono sfruttati modelli semiquantitativi o qualitativi in grado di guidare la diagnosi. Di seguito i modelli di maggior rilievo.

#### Regola ABCD

Il metodo consiste nella valutazione di determinate caratteristiche morfologiche e l'assegnazione di un punteggio numerico complessivo alla lesione,



definito TDS (*Total Dermoscopic Score*). Come mostrato dalla figura 1.12, le caratteristiche osservate sono:

- **Asimmetria:** suddivisa la lesione con due assi perpendicolari, viene assegnato un punteggio pari a 2.6 se asimmetrica rispetto ad entrambi gli assi, 1.3 se rispetto ad un solo asse, 0 se totalmente simmetrica.
- **Bordi:** suddivisa la lesione con quattro assi in modo da valutare 8 bordi, viene assegnato un punteggio pari a 0 se il bordo "sfuma"; 0.1 se il bordo si interrompe in maniera netta.
- **Colore:** per ogni colore (nero, marrone scuro, marrone chiaro, grigio-bluastro, bianco e rosso) individuato viene assegnato un punteggio pari a 0.5.
- **Dimensioni:** viene valutata la dimensione della lesione e la presenza di strutture caratteristiche (rete pigmentata, globuli clusterizzati, striature ramificate, aree amorphe, puntini). Per ogni struttura individuata viene assegnato un punteggio pari a 0.5.

Se il punteggio totale è maggiore di 5.45 la lesione è considerata come un melanoma altamente probabile, se è maggiore di 4.75 è considerata sospetta, se minore di 4.75 è considerata benigna [16].

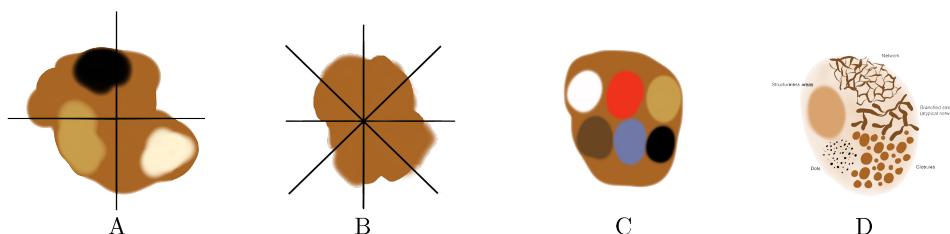


Figura 1.12: Schematizzazione del metodo ABCD [19].

### Regola dei sette punti

In questo caso vengono tenuti in considerazione sette caratteristiche dermo-scopiche, suddivise in più rilevanti, alle quali vengono assegnati due punti, e meno rilevanti, alle quali viene assegnato un punto. Tra le caratteristiche più rilevanti annoveriamo: rete pigmentata atipica, velo blu-biancastro, pattern vascolare atipico. Tra le meno rilevanti invece: striature radiali e pseudopodi radiali, pigmentazione irregolare, globuli e macchie irregolari, pattern di regressione [16].

### **Regola dei tre punti**

Consiste in una versione semplificata della regola dei setti punti, pensata appositamente per i novizi dermatologi. Il metodo consiste nella valutazioni di tre criteri fondamentali:

- **asimmetria:** dal punto di vista strutturale e del colore.
- **rete atipica:** rete pigmentata con fori irregolari e tratti spessi.
- **strutture blu-bianche.**

Se la lesione si dimostra positiva ai tre criteri deve essere asportata [16].

### **Metodo di Menzie**

Il metodo valuta la presenza di undici caratteristiche dermoscopiche, suddivise in *negative* (caratteristica benigna) e *positive* (caratteristica maligna). Le caratteristiche positive comprendono: velo blu-biancastro, puntini marroni multipli, pseudopodi, striature irradiate, aree cicatriziali di depigmentazione, puntini e globuli neri nella periferia della lesione, colori multipli (cinque o sei), puntini multipli e globuli blu, rete pigmentaria allargata. Simmetria e singolo colore sono considerate caratteristiche negative.

La presenza di caratteristiche positive, contemporaneamente all'assenza di quelle negative, è sufficiente per la diagnosi di melanoma cutaneo [16].

### **Analisi dei patterns**

La lesione viene in questo caso analizzata dal punto di vista dei pattern generali e locali, tenendo in considerazione i seguenti criteri [16]:

- rete pigmentaria regolare e irregolare
- pseudopodi
- striature radiali
- puntini e globuli
- velo blu-bianco
- aree di regressione
- ipopigmentazione

- aree amorphe
- pattern vascolare

Tra i diversi criteri l'analisi del pattern mostra la specificità più elevata, ma è il metodo che richiede maggiore esperienza dell'operatore.

### 1.4.3 Esame dermoscopic

Il protocollo dell'esame dermoscopic standardizzato consta di *due step fondamentali*:

1. **Indagine sulla natura melanocitica:** l'obiettivo del primo step consiste nel differenziare la natura melanocitica o non melanocitica della lesione. L'indagine si basa sull'identificazione dei criteri dermoscopic tipici al fine di definire la caratterizzazione del pattern della lesione.
2. **Differenziazione tra lesione benigna e maligna:** diagnosticata la natura melanocitica è necessario verificare la benignità della lesione. Possono essere utilizzati i differenti algoritmi diagnostici illustrati precedentemente [15].

Una volta completato l'esame dermoscopic è necessario integrare i dati con la storia clinica della lesione, con l'obiettivo di una diagnosi il più precoce possibile. La diagnosi finale viene poi confermata da un esame istologico su un campione prelevato con biopsia cutanea mediante asportazione completa o parziale. [20]. Secondo studi di valutazione dell'accuratezza diagnostica, tramite l'esame dermoscopic il dermatologo è in grado di diagnosticare il melanoma in una casistica tra il 65% e l'80%.

Tuttavia l'interpretazione delle immagini dermoscopiche è un processo che richiede allenamento ed esperienza: se utilizzato da dermatologi esperti infatti riesce ad aumentare la sensibilità diagnostica del melanoma del 35% rispetto al solo esame clinico, ma se eseguito da un operatore non esperto può portare unicamente ad un peggioramento dell'accuratezza della diagnosi: un'adeguata preparazione è fondamentale ai fini di un'applicazione diagnostica realmente efficace [15]. Da qui dunque la necessità, sempre maggiore, di affiancare al dermatologo strumenti in grado di assisterlo nella diagnosi.

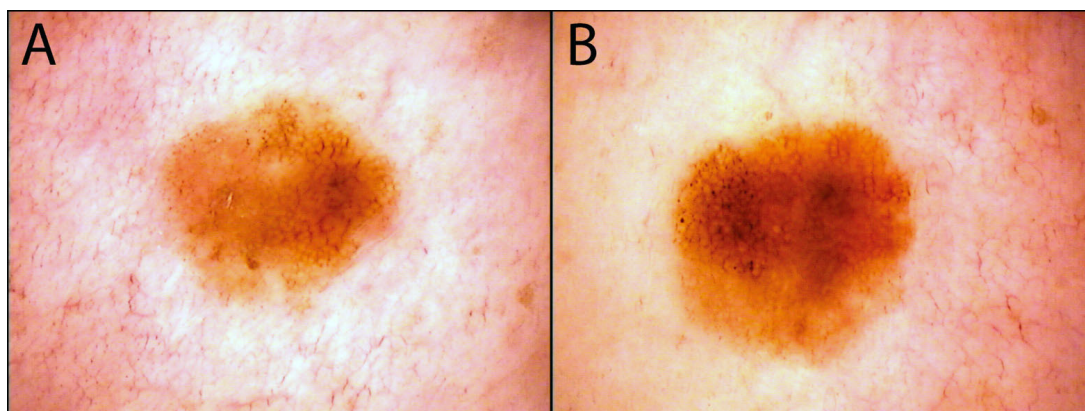


Figura 1.13: Evoluzione nel tempo di un melanoma a crescita superficiale [21]

#### 1.4.4 Follow up dermoscopic

Il monitoraggio nel tempo del paziente risulta fondamentale soprattutto per la diagnosi di lesioni melanocitiche atipiche: i melanomi in stadio iniziale possono presentarsi dal punto di vista dermoscopic come "featureless", ovvero privi di criteri dermoscopic atipici, e possono essere diagnosticati unicamente osservando l'evoluzione della lesione in un intervallo di tempo ristretto. In figura 1.13 viene mostrata l'evoluzione di una melanoma a diffusione superficiale inizialmente privo di strutture sospette che, in un lasso temporale di 6 mesi circa, comincia a mostrare pattern maligni (puntini neri e crescita asimmetrica). Il follow up è inoltre indispensabile per i soggetti caratterizzati da elevati fattori di rischio per lo sviluppo del melanoma (storia familiare di melanoma, elevato numero di nevi, soprattutto atipici) [15].

# Capitolo 2

## Materiali e Metodi

Nella prima parte del capitolo vengono presentati i diversi dataset utilizzati per la realizzazione di reti neurali convoluzionali (CNN) in grado di effettuare classificazioni ternarie di lesioni cutanee, riuscendo a distinguere nevi benigni, melanomi maligni e cheratosi seborroiche: sono descritte nel dettaglio la numerosità, le caratteristiche delle immagini, gli strumenti e i protocolli utilizzati in fase di acquisizione. Segue una panoramica dello strumento NuruGo Derma, una lente da smartphone che permette di acquisire immagini della cute a livello micro in maniera amatoriale, con un focus sui limiti e sull'artefatto di riflessione che ne affligge le immagini. Viene quindi descritto l'algoritmo di segmentazione sviluppato per processarle e la strategia di pre-processing del dataset utilizzato per allenare e testare le diverse CNN. Successivamente viene spiegata nel dettaglio l'architettura di rete scelta e le diverse metodologie utilizzate per affrontare il problema di classificazione.

### 2.1 Archivio ISIC 2019

Le immagini utilizzate come dataset per il training della CNN fanno parte di una raccolta *open source* di immagini dermatologiche, l'*ISIC Archive*, curata dalla *ISIC (Internetial Skin Imaging Collaboration)*, un'organizzazione accademica creata con lo scopo di fornire standard in grado di guidare le tecnologie, le tecniche e la terminologia usata nello *skin imaging* [4]. L'interno archivio consta di 23801 immagini dermoscopiche e 100 immagini cliniche, provenienti da differenti dataset. Le immagini dermoscopiche sono classificate in 8 differenti categorie diagnostiche:

- Melanoma maligno

- Nevo Melanocitico
- Cheratosi seborroica
- Carcinoma basocellulare
- Dermatofibroma
- Carcinoma a cellule squamose
- Lesione vascolare

Con lo scopo di effettuare una classificazione *ternaria* e *binaria* è stato raccolto un dataset contenente 225 immagini di *melanoma maligno*, 237 immagini di *nevo melanocitico* e 235 immagini di *cheratosi seborroica*. Le immagini provengono dalla raccolta *MSK*, caratterizzata da proprietà simili in quanto a risoluzione e dimensione, e soprattutto dotate di righello dermoscopic, vincolo principale nella scelta delle immagini.

La presenza del righello infatti ha permesso, tramite script MatLab, di misurare manualmente i valori di risoluzione, espressi in *pixel/mm*, delle singole immagini.

## 2.2 Dataset Novara

Presso il reparto di dermatologia dell’Azienda Ospedaliera Universitaria Maggiore della Carità di Novara sono state acquisite, da personale specializzato, 39 immagini dermoscopiche in epiluminescenza di nevo benigno, 18 di melanoma maligno e 21 di cheratosi seborroica. Le acquisizioni sono state effettuate in sala operatoria prima dell’escissione delle lesioni (principalmente per quelle maligne e atipiche) e in ambulatorio durante le visite di controllo.

### 2.2.1 Strumento utilizzato

Per le acquisizioni è stato utilizzato un dermoscopio da contatto **HEINE Delta 20T**, tra i migliori in commercio. Il dispositivo vanta le seguenti caratteristiche:

- ottica di precisione ad alta risoluzione con lenti acromatiche.
- ingrandimento 10x.
- illuminazione totale a 6 LED<sup>HQ</sup>.

- funzione "toggle" per passare dalla modalità di illuminazione polarizzata a quella non polarizzata.
- campo visivo di 23 mm di diametro.
- range di correzione da -6 a +6 diottrie.

Il dermoscopio è inoltre dotato di una piastrina di contatto da 23 mm con scala millimetrata. Tramite il fotoadattatore SLR HEINE è possibile agganciare il dispositivo ad una camera reflex professionale per digitalizzare l'acquisizione delle immagini [23].



Figura 2.1: Panoramica del dermoscopio HEINE Delta 20T [23].

Per digitalizzare le immagini è stata utilizzata una camera reflex **NIKON E4500**, le cui specifiche, illustrate nella tabella 2.1, hanno determinato le caratteristiche delle immagini acquisite.

Dimensioni (pixel)	Spazio colore	Profondità (bit)	Risoluzione (pixel/mm)	Formato
2272 x 1704	RGB	24 bit	848	JPEG

Tabella 2.1: Caratteristiche NIKON E4500.

## Pre-processing

Le immagini acquisite con dermoscopio presentano un'area centrale circolare utile, che contiene l'immagine della lesione, e una vasta area di pixel neri. Tutte le immagini sono state così croppate in maniera tale da ridurre l'area nera e rendere il rapporto d'aspetto (*aspect ratio*) quadrato. L'algoritmo realizzato procede per passi:

1. Trasformazione dell'immagine in greyscale.
2. Binarizzazione tramite global thresholding.
3. Individuazione delle coordinate e delle dimensioni dell'area centrale.
4. Cropping.

In figura 2.2 un esempio di crop su immagine dermoscopic di cheratosi seborroica.

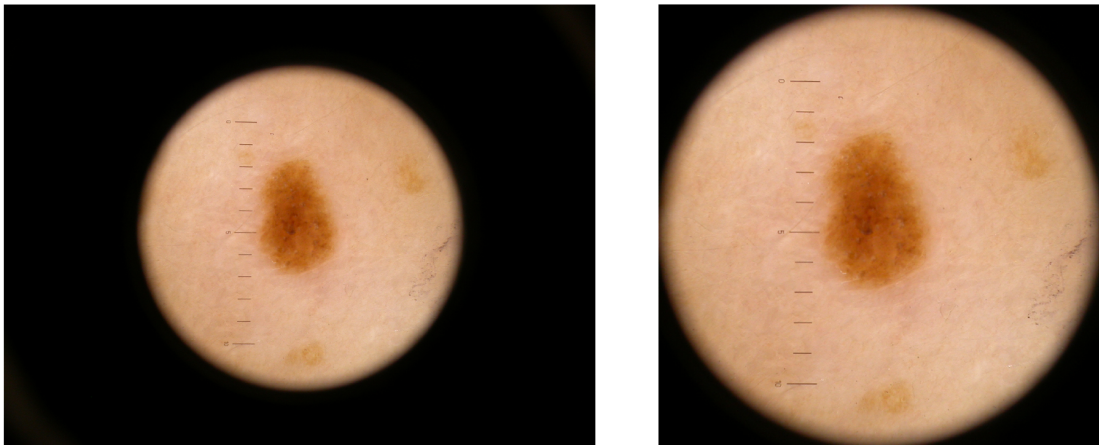


Figura 2.2: Esempio di crop su immagine dermoscopic.

## 2.3 Nurugo Derma

Il *Nurugo Derma* è una lente da smartphone che permette, convogliando tramite un sistema di prismi riflettenti la luce emessa dal flash del dispositivo e un'apposita app di acquisizione per iOS e Android, "*Nurugo Box*",



di esaminare la cute a livello micro in maniera amatoriale. Il dispositivo è compatibile con la maggior parte degli smartphone in commercio e vanta le seguenti caratteristiche:

- Dimensioni (mm): senza clip 24.7 x 21.57 x 33.5, con clip 24.7 x 21.57 x 33.5;
- Peso: circa 10 g;
- Ingrandimento della lente: 38x.

Il dispositivo viene agganciato allo smartphone tramite un'apposita clip in plastica, avendo cura di allineare correttamente la fotocamera e il flash.



Figura 2.3: Nurugo Derma montato su iPhone 6s.

Essendo un dispositivo pensato per un utilizzo amatoriale presenta diversi limiti:

- **Riflessioni:** le immagini acquisite sono affette dal cosiddetto *glare effect*, dovuto alla luce riflessa dalla pelle, che determina artefatti di riflessioni.
- **Effetto ombra:** i margini sono caratterizzati da un effetto ombra dovuto alla conformazione della camera del Nurugo.
- Impossibilità di comprimere la lesione.

Il limite fondamentale del Nurugo risiede tuttavia nella tipologia di immagine, poco utile dal punto di vista diagnostico, non essendo acquisita in epiluminescenza. Il led del flash dei comuni smartphone infatti non produce luce polarizzata, rendendo impossibile ricavare informazioni sulle strutture al di sotto dell'epidermide. Inoltre la struttura del nurugo, priva di un supporto trasparente a valle della lente, non permette di apporre direttamente olio o gel per annullare la riflessione della luce. Per aggirare il problema è stato utilizzato un vetrino da laboratorio trasparente, posizionato tra Nurugo e cute, che permettesse di applicare sulla cute del liquido di interfaccia. Le immagini così ricavate, in epiluminescenza, mostrano le strutture cutanee sottostanti ma sono limitate dalla presenza di un artefatto di riflessione che riduce il campo utile di visione (FOV - *Field Of View*), limitandolo ad un'area *circolare* di circa  $4 \text{ mm}^2$ . Conseguentemente da una singola acquisizione è possibile ricavare generalmente solo una piccola porzione della lesione.

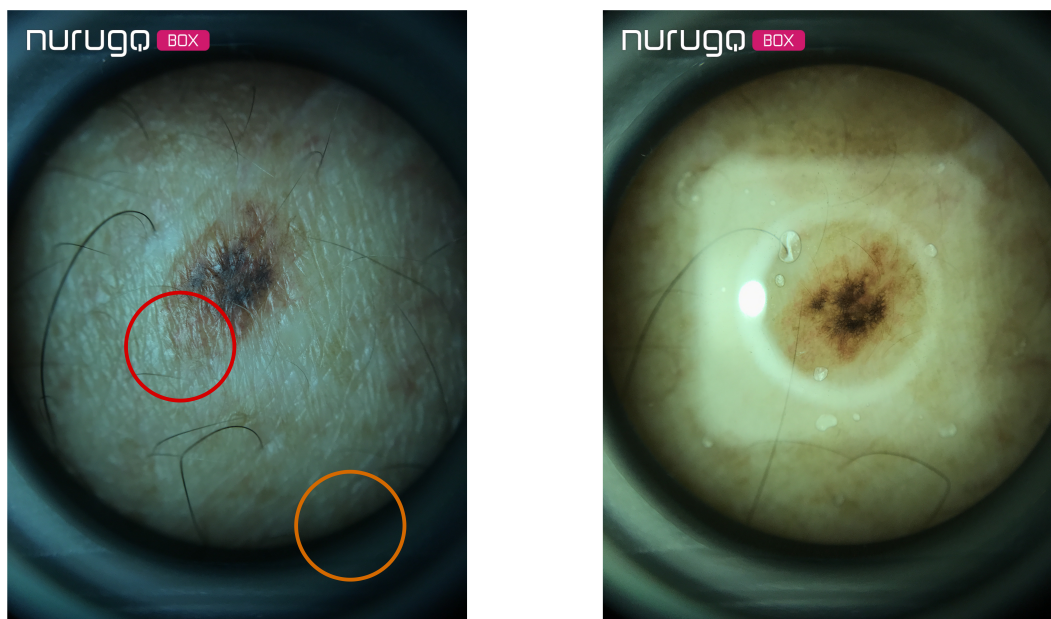


Figura 2.4: A sinistra immagine di nevo benigno acquisita con Nurugo: in arancione l'effetto ombra, in rosso il glare effect. A destra stessa lesione cutanea acquisita con Nurugo con vetrino in condizione di epiluminescenza. E' possibile notare l'artefatto di riflessione che ricopre gran parte dell'immagine.

### 2.3.1 Protocollo di acquisizione

Presso il reparto di dermatologia dell’Azienda Ospedaliera Universitaria Maggiore della Carità di Novara sono state acquisite con il *Nurugo Derma* 39 immagini di nevo benigno, 18 immagini di melanoma maligno e 21 immagini di cheratosi seborroica, corredate con le corrispettive versioni dermoscopiche. Le immagini sono state acquisite sia in che non in condizioni di epiluminescenza. Le acquisizioni in epiluminescenza in particolare sono state effettuate seguendo un determinato protocollo:

1. Applicazione di olio o gel ecografico;
2. Posizionamento di un vetrino da laboratorio trasparente;
3. Posizionamento del nurugo in contatto con il vetrino;
4. Acquisizione dell’immagine.

A causa della presenza dell’artefatto che limita il FOV, per lesioni di grandi dimensioni sono state effettuate più acquisizioni modificando il posizionamento del vetrino rispetto alla lesione. Le immagini sono state acquisite nel tempo utilizzando tre modelli di smartphone differenti e vantano le caratteristiche mostrate nella tabella 2.2.

Modello	Dimensione (pixel)	Spazio colore	Risoluzione ( <i>pixel/mm</i> )	Formato
iPhone 6	2448 x 3264	RGB	1726	JPEG
iPhone 6s	3024 x 4032	RGB	2166	JPEG
iPhone 7	3024 x 4032	RGB	2143	JPEG

Tabella 2.2: Caratteristiche immagini acquisite con Nurugo.

### 2.3.2 Algoritmo di segmentazione

L’artefatto di riflessione che affligge le immagini acquisite con Nurugo in epiluminescenza, determina la saturazione di una vasta area di pixel: un’algoritmo di denoising dovrebbe dunque creare nuova informazione per "pulire" l’immagine. E’ stato scelto quindi di isolare il campo di visione utile, ovvero

l'area circondata dall'artefatto. Per estrarre l'area non contaminata dall'immagine, considerando l'artefatto ripetibile in tutte le immagini secondo determinate caratteristiche (dimensione, forma e posizione), è stato sviluppato un algoritmo di segmentazione basato sulla *trasformata di Hough* [24] (CHT - Circular Hough Transform). La CHT procede per step:

1. *Costruzione di un vettore di accumulo*: innanzitutto i pixel ad alto valore di gradiente vengono selezionati come possibili candidati. Questi pixel scandagliano un intorno circolare con raggio fisso e salvano le loro "preferenze" all'interno di un vettore di accumulo.
2. *Stima del centro*: se i pixel candidati appartengono effettivamente ad un elemento circolare, le preferenze dei diversi candidati andranno ad accumularsi in un punto corrispondente al centro del cerchio.
3. *Stima del raggio*.

In figura 2.5 sono rappresentati i primi due step della trasformata di Hough.

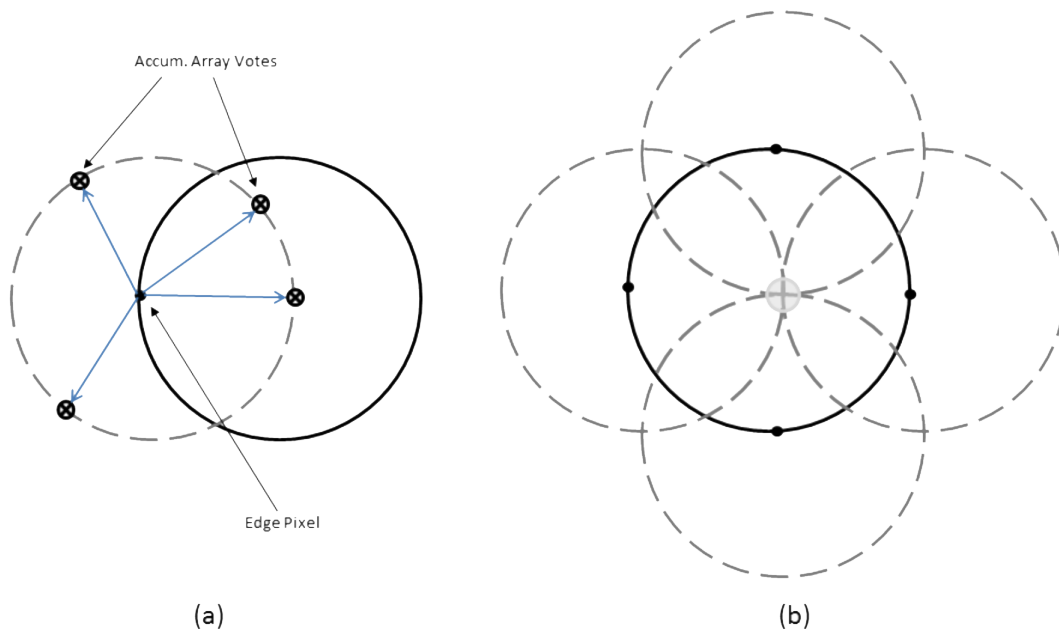


Figura 2.5: Rappresentazione dei primi due step della trasformata di Hough: a) costruzione del vettore di accumulo b) stima del centro [24].

In una prima fase dunque l'algoritmo implementato individua tutti gli elementi circolari all'interno dell'immagine sfruttando la CHT. Per riuscire ad

identificare l'artefatto nella totalità del dataset è stato impostato un valore di soglia elevato. Conseguentemente l'algoritmo individua cerchi aggiuntivi che dovranno essere rimossi tramite una fase di cleaning basata su più controlli:

- **Controllo su area:** innanzitutto vengono scartati i cerchi al di fuori di un determinato range di diametro.
- **Controllo su posizione:** partendo dal presupposto che il cerchio utile si trovi nell'intorno centrale dell'immagine, non vengono mantenuti i cerchi troppo esterni.
- **Controllo su interno:** basandosi sulla considerazione secondo cui il cerchio di interesse è circondato da pixel molto luminosi, viene effettuato un controllo sull'intorno dei cerchi, escludendo quelli con un valore di intensità media dell'intorno sotto soglia.
- **Controllo su robustezza:** viene valutata la robustezza dei singoli cerchi, escludendo i più deboli.

Individuato il cerchio finale l'algoritmo esegue un crop dell'immagine salvando la porzione utile non contaminata dall'artefatto. In figura 2.7 è mostrato un esempio dei diversi step.

Le immagini così ottenute sono state raggruppate in un dataset per la fase di testing delle reti, con l'obiettivo di verificare la validità diagnostica delle immagini acquisite con Nurugo in epiluminescenza con campo di visione ristretto. I diversi step dell'algoritmo sono riassunti dal diagramma di flusso in figura 2.6.

## 2.4 Cropping delle immagini

Con il fine di realizzare una CNN in grado di riconoscere anche le lesioni acquisite con Nurugo è stato scelto di riprodurre le limitazioni date dall'artefatto sul dataset pubblico ISIC, eseguendo un crop circolare, centrato sulla lesione, che ne riducesse il campo di visione. Ne è conseguita la necessità di selezionare dal database ISIC i sotto dataset MSK, contenenti immagini dotate di righello dermoscopico millimetrico: per ogni immagine, tramite script MATLAB, è stato calcolato il numero di pixel contenuti in un mm di pelle, ricavando quindi valori di risoluzione in *pixel/mm*. Ottenuti manualmente i singoli valori di risoluzione sono stati composti più dataset, con crop di raggio differenti:

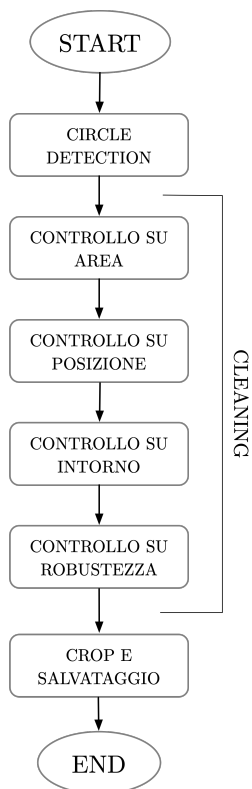


Figura 2.6: Diagramma di flusso dell’algoritmo di segmentazione.

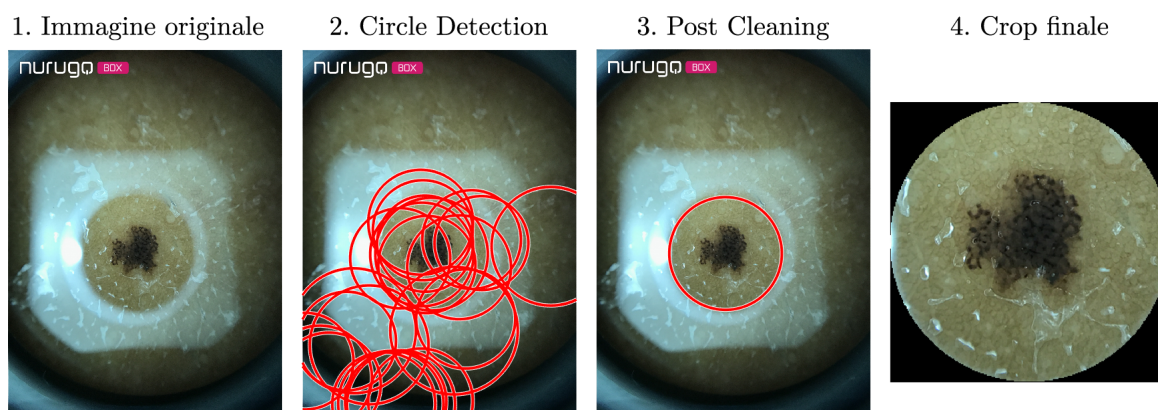


Figura 2.7: Applicazione dell’algoritmo ad immagine di nevo benigno acquisita con Nurugo.

- raggio = 1.14 mm (raggio medio dell’area estratta dalle immagini Nurugo);

- raggio = 3 mm;
- raggio = 5 mm;

Il crop con raggio pari a 1.14 mm, mostrato in figura 2.8, cerca di riprodurre le condizioni del Nurugo.

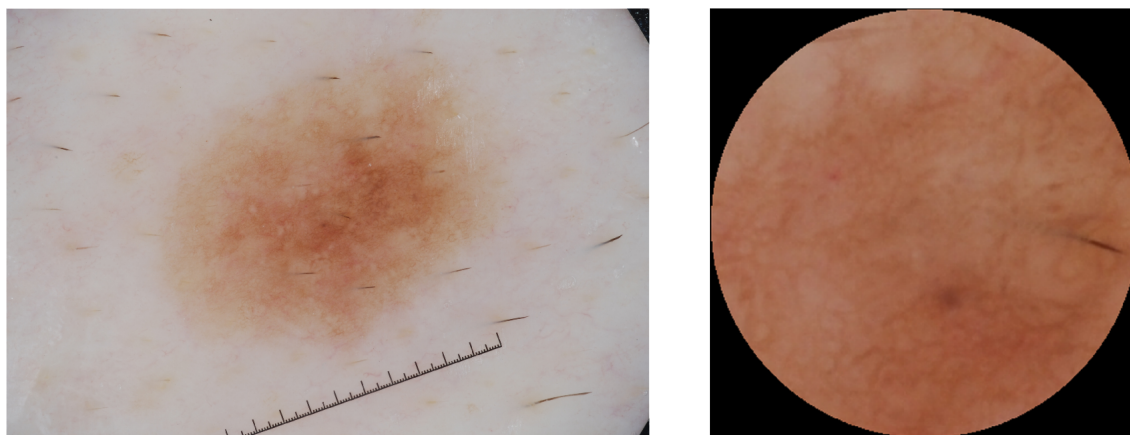


Figura 2.8: Esempio di crop su immagine di melanoma maligno con raggio pari a 1.14 mm

### 2.4.1 MultiCrop

Per riprodurre il protocollo di acquisizione del Nurugo (acquisizioni multiple di lesioni estese) e per non tralasciare porzioni di lesioni che potrebbero risultare preziose per la classificazione, oltre al crop centrale sono stati eseguiti più crop, selezionandone le regioni attorno con un determinato overlap. Per evitare di selezionare aree caratterizzate principalmente da pelle sana o con una percentuale esigua di lesione è stato necessario ricavare le maschere binarie manuali delle singole lesioni.

L'algoritmo di cropping realizzato esegue un controllo sulla maschera binaria, scartando le aree circolari che presentano una percentuale di lesione inferiore ad una determinata soglia, impostata al 50%, e salvando invece le aree che soddisfano i requisiti. In figura 2.9 un esempio di MultiCrop su immagine di cheratosi seborroica.

Questo approccio è risultato vantaggioso dato che ha permesso di aumentare la numerosità del dataset di training e di scartare elementi di disturbo e artefatti presenti nelle zone periferiche delle immagini.



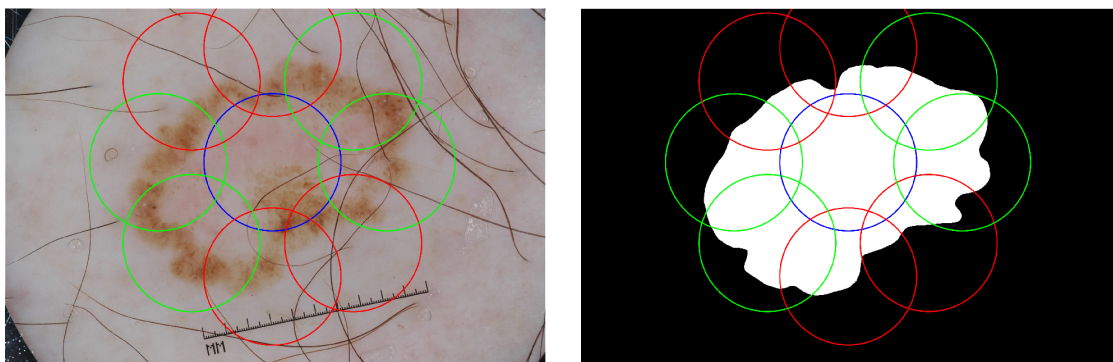


Figura 2.9: Esempio di MultiCrop da 3 mm su immagine di cheratosi seborroica. In blu il crop centrale, in verde i crop considerati idonei, in rosso i crop scartati.

## 2.5 Classificazione di lesioni cutanee

L'obiettivo del lavoro di tesi consiste nella realizzazione di un sistema di classificazione, basato su CNN, in grado di distinguere il melanoma cutaneo da altre lesioni benigne quali nevo e cheratosi seborroica. Le diverse scelte di progetto sono state effettuate con il fine di realizzare un sistema versatile, applicabile su immagini dermoscopiche, ricavate dall'archivio ISIC, immagini dermoscopiche acquisite presso l'Azienda Ospedaliera Universitaria Maggiore della Carità di Novara, e infine su immagini acquisite con il dispositivo Nurugo Derma. Parallelamente dunque il sistema di classificazione ha lo scopo di fornire indicazioni sulla validità diagnostica del suddetto dispositivo.

### 2.5.1 Deep Learning e CNN

Con *deep learning* (apprendimento profondo) si intende una particolare tecnica di *machine learning* in grado di far apprendere ad un computer un determinato compito mimando la tecnica di apprendimento umana: imparare da degli esempi.

Le tecniche di deep learning si prestano particolarmente per creare macchine in grado di classificare oggetti differenti, che siano immagini, testi o suoni. Le applicazioni del deep learning variano infatti dal campo dell'automotive (guida autonoma), fino all'ambito aerospaziale, elettronico e biomedico.

La maggior parte delle tecniche di deep learning si basano sull'utilizzo di reti neurali allenare impiegando set particolarmente numerosi di dati già classificati (*apprendimento supervisionato*). Tramite un processo ricorsivo la rete



neurale impara ad estrarre *features* direttamente dai dati ai quali è sottoposta. In figura 2.10 è possibile osservare la struttura tipica di una rete neurale: un layer di input, seguito da una serie di layers "nascosti" (*hidden layers*) e infine un layer di output.

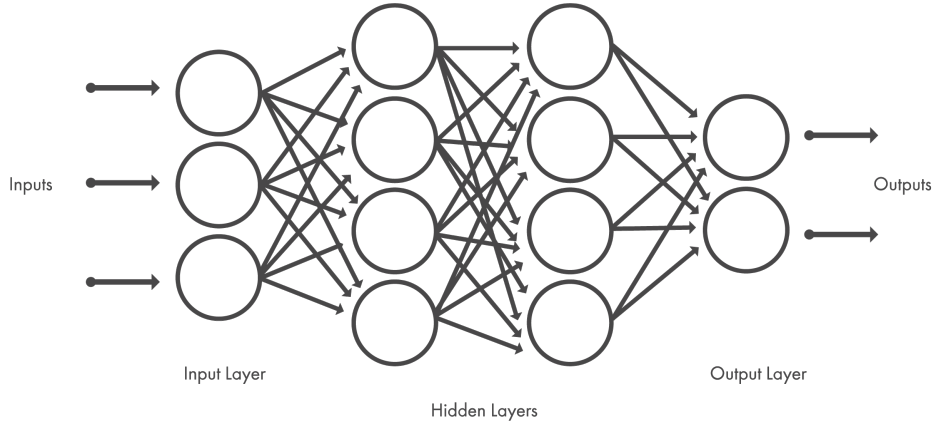


Figura 2.10: Struttura tipica di una rete neurale [25].

Le CNN (*Convolutional Neural Network*), o ConvNet, sono tra i più utilizzati algoritmi di deep learning nell'ambito dell'object recognition e della computer vision. Come le reti neurali feed-forward, anche le CNN sono composte da neuroni, che ricevono informazioni in input e forniscono un determinato output. Ogni neurone è caratterizzato da pesi ( $W$ ) e bias ( $b$ ), parametri aggiornati ed ottimizzati durante la fase di allenamento. Lo scopo dell'allenamento infatti consiste nel minimizzare la cosiddetta funzione di costo, data dall'errore tra l'output atteso e l'output generato:

$$J(W, b) = \frac{1}{m} \sum_{i=1}^m L(\hat{y}_i, y_i)$$

Sottoposte dunque ad un allenamento mirato sono in grado di riconoscere pattern specifici, utilizzati poi per classificare le immagini. E' possibile suddividere la struttura di una CNN in due blocchi: il primo blocco di "*feature learning*" nel quale, tramite operazioni specifiche, vengono estratte informazioni sulle immagini in input, e un secondo blocco di classificazione.

Tra gli elementi più comuni all'interno del blocco di feature learning:

- **Layer di convoluzione:** fa scorrere un filtro bidimensionale sull'immagine di partenza producendo un'immagine d'uscita. E' seguito da una funzione di attivazione, generalmente una ReLu, che mappa a zero

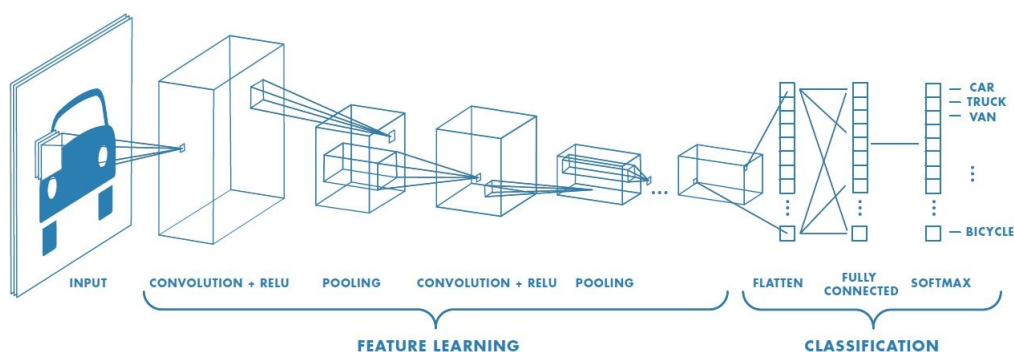


Figura 2.11: Struttura tipica di una CNN [25].

i valori negativi, mantenendo i valori positivi in modo da ottimizzare l'allenamento.

- **Layer di pooling:** semplifica i dati effettuando un *downsampling* non lineare, riducendo la dimensione dell'input. *MaxPooling* (estrae il valore massimo all'interno del kernel) e *AveragePooling* (estrae la tra gli elementi all'interno del kernel) sono i più utilizzati.

Il blocco di classificazione invece è caratterizzato da layers *fully connected*, nei quali ogni neurone è connesso a tutti i neuroni degli altri layers. Le performance di una CNN sono strettamente legate alla struttura della stessa (dimensione e profondità) e alla quantità di dati a disposizione per l'allenamento [25].

## Transfer learning

La fase di allenamento è critica e determinante per la realizzazione di una CNN funzionale ed ottimale. Esistono fondamentalmente due approcci:

1. *Training from Scratch* (allenamento da zero)
2. *Transfer Learning* (trasferimento dell'apprendimento)

La scelta della strategia di allenamento dipende strettamente dalla tipologia di problema da affrontare e soprattutto dalla quantità di dati a disposizione: per allenare una CNN da zero ed ottenere performance accettabili è necessario avere a disposizione un'ingente quantità di dati.

Il Transfer Learning è un metodo che permette di adattare un modello pre allenato ad un problema differente. E' un procedimento che permette di ottenere buoni risultati pur avendo a disposizione un esiguo quantitativo di dati e

soprattutto risparmiando tempo di allenamento e risorse computazionali [25]. Scelto il modello pre allenato da utilizzare è necessario adattarlo al nuovo problema di classificazione sostituendo la softmax e il layer di classificazione. A questo punto dunque è possibile:

1. Riallenare l'intera rete.
2. Allenare solamente alcuni layer del blocco di apprendimento.
3. Allenare unicamente il blocco di classificazione.

Come illustrato dalla figura 2.12 la scelta dovrebbe essere dettata dalle caratteristiche del dataset di allenamento a disposizione.

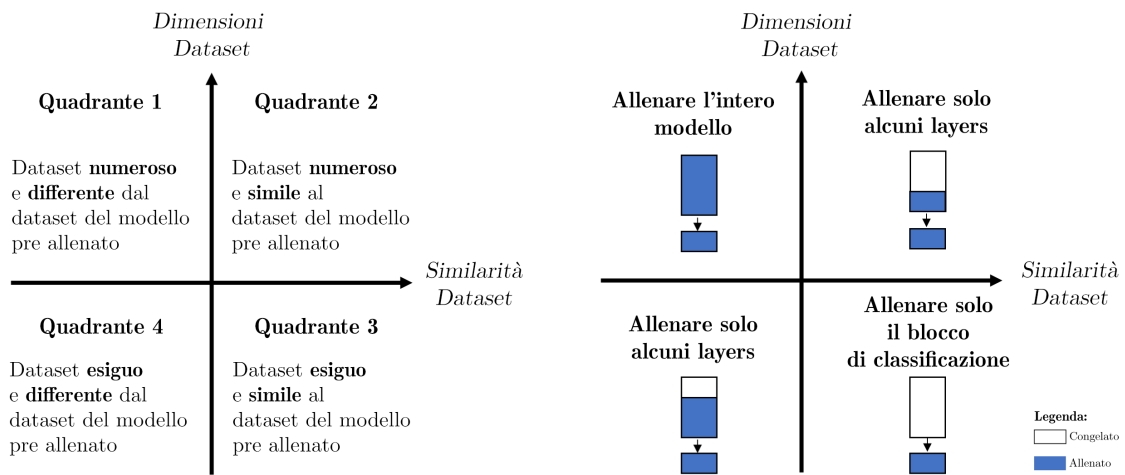


Figura 2.12: Diverse strategie di allenamento [26].

## AlexNet

AlexNet [27] è sicuramente tra i modelli pre allenati più utilizzati nel Transfer Learning. E' una CNN ideata nel 2012 da Alex Krizhevsky, Ilya Sutskever e Geoffrey Hilton nell'ambito della *ImageNet Large Scale Visual Recognition Challenge*, una competizione annuale che consiste nello sviluppo di un modello in grado di classificare il dataset *ImageNet*, composto da più di 1 milione di immagini. Con un errore di predizione del 15.3% AlexNet è risultato il modello vincitore della competizione del 2012.

Dal punto di vista architetturale consta di 650.000 neuroni e 60 milioni di parametri, distribuiti in 5 layer convoluzionali, 5 layer di pooling, 3 layer

totalmente connessi, 2 layer di drop-out, necessari per diminuire l'*overfitting* (adattamento eccessivo della rete agli esempi del training che determina perdita di generalizzazione) durante l'allenamento, una softmax e un layer di classificazione. In figura 2.13 una rappresentazione della struttura di AlexNet.

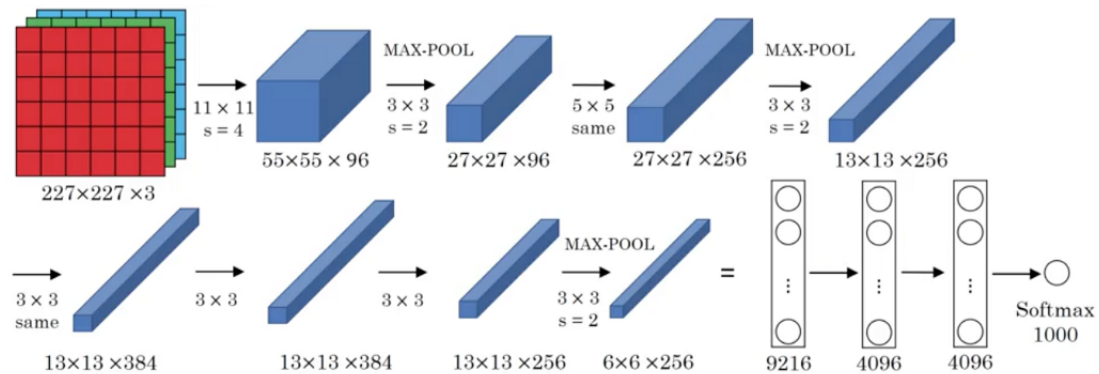


Figura 2.13: Architettura di AlexNet [28].

## Bias, Varianza e Overfitting

Per comprendere il comportamento di una rete neurale convoluzionale è necessario tener presente il concetto di *bias* e *varianza*:

- *Bias elevato*: errore sul training set elevato. La rete non riesce ad apprendere correttamente dagli esempi durante l'allenamento.
- *Varianza elevata*: errore sul validation set elevato. La rete non riesce ad applicare correttamente ciò che ha imparato nella fase di allenamento.

Una CNN ideale è caratterizzata da valori di bias e varianza prossimi allo zero. In condizioni reali però si verifica un "*trade-off*": il miglioramento di un parametro porta al peggioramento dell'altro. In relazione al valore di bias e varianza è possibile definire:

- *Underfitting*: una condizione determinata da elevato bias ed elevata varianza. Le prestazioni della rete sono scadenti sia sul training set che su nuovi esempi.
- *Overfitting*: una condizione determinata da ridotto bias ma elevata varianza. Ne consegue una scarsa capacità di generalizzazione della rete.

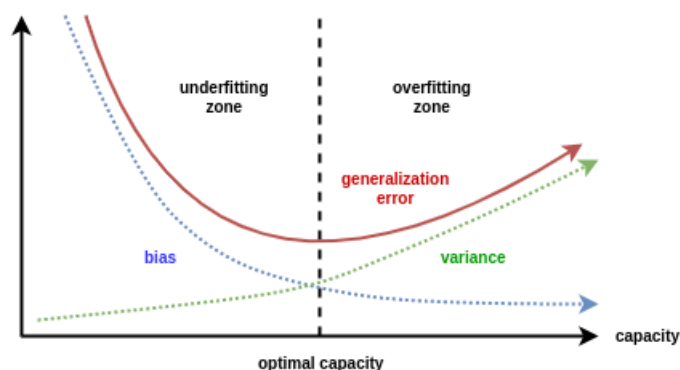


Figura 2.14: Panoramica comportamenti di una rete durante la fase di training.

La condizione di *overfitting* è migliorabile tramite alcuni accorgimenti:

1. Suddivisione del dataset a disposizione in *Training Set*, *Validation Set*, *Test Set*. Il *Validation Set* permette di capire come la rete gestisce esempi diversi dal training, permettendo così di effettuare un tuning dei parametri di allenamento. La suddivisione del dataset dipende dalla numerosità dello stesso, al fine di ottenere informazioni consistenti dal validation e dal test set.
2. Incremento delle dimensioni del *Training Set*. Aumentare la quantità di dati forniti alla rete durante il training diminuisce la probabilità di overfitting. Nel caso in cui non si abbiano a disposizione nuovi dati è possibile operare tecniche di *data augmentation* che effettuano operazioni di rotazione, traslazione, deformazione, aggiunta di rumore, sui dati in input in modo da aumentarne la quantità.
3. Terminazione anticipata dell'allenamento. Un allenamento prolungato potrebbe portare il modello ad adattarsi eccessivamente agli esempi di training, perdendo così capacità di generalizzazione.
4. *Dropout*. La tecnica consiste nel disattivare casualmente durante l'allenamento un certo numero di neuroni in maniera tale da rendere la rete meno sensibile a pesi specifici dei neuroni.
5. *Regolarizzazione L2*. Consiste nell'aggiungere un termine di regolarizzazione  $\lambda$  alla funzione di costo:

$$J(W, b) = \frac{1}{m} \sum_{i=1}^m L(\hat{y}_i, y_i) + \lambda \|w\|^2$$

E' anche definita "weight decay":

$$w^{[l]} = \left[1 - \frac{\alpha\lambda}{m}\right]w^{[l]} - \alpha$$

Se il fattore di regolarizzazione  $\lambda$  aumenta  $w^{[l]}$  diminuisce.

## 2.5.2 Architettura della rete e allenamento

L'utilizzo di AlexNet è però limitato dalle dimensioni del layer di input: la CNN accetta infatti unicamente immagini di dimensione 227 x 227 x 3. Per utilizzare la rete sarebbe dunque necessario effettuare un downsampling delle immagini di training (essendo di dimensioni maggiori), perdendo inevitabilmente informazioni preziose sulla tessitura, considerata caratteristica fondamentale e identificativa, delle diverse lesioni.

Per ovviare a questo vincolo è stata utilizzata una AlexNet modificata che accetta in ingresso immagini di dimensione 512 x 512 x 3. Per adattare il resto della rete ad un input di dimensione differente è stato necessario:

- modificare il primo layer di convoluzione *FC1* aggiungendo un padding con kernel di dimensione 100 x 100.
- modificare i layer fully connected *FC6* e *FC7* in un layer di convoluzione.
- modificare il layer fully connected rimanente adattandolo al numero di classi da identificare.

Scelto il modello pre-allenato da utilizzare è stato deciso di affrontare il problema provando con 3 metodologie differenti:

1. **M1**: singola CNN per classificazione *ternaria* (cheratosi seborroica, nevo benigno, melanoma maligno).
2. **M2**: due CNN in cascata. La prima rete (*CNN 2.1*) effettua una classificazione tra lesione melanocitica e lesione non melanocita; la seconda rete (*CNN 2.2*) classifica in cascata le lesioni melanocitiche in nevi benigni e melanomi maligni. CNN 2.2 dunque "produce" in uscita una confusion matrix nella quale sono contenuti anche i risultati della prima classificazione. Il metodo in questione ricalca dunque lo standard operativo della visita dermatologica.

3. **M3**: due CNN in cascata. La prima rete (*CNN 3.1*) distingue le lesioni benigne dalle lesioni maligne; la seconda (*CNN 3.2*) effettua una classificazione tra nevo benigno e cheratosi seborroica.

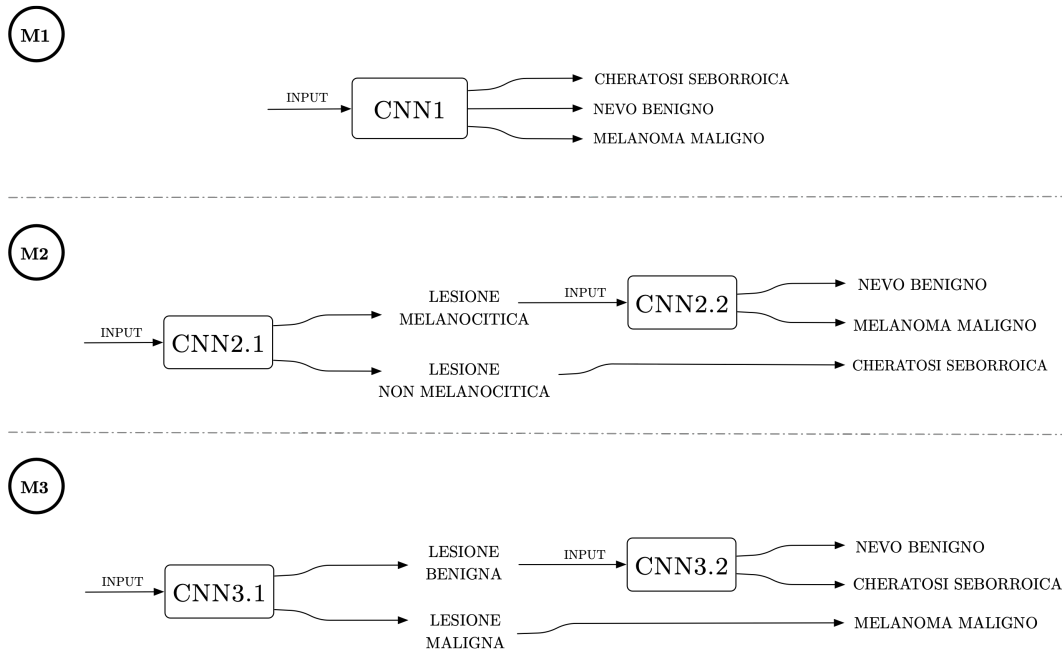


Figura 2.15: Schematizzazione dei tre metodi proposti.

Per le tre diverse strutture sono stati utilizzati dataset di training con numerosità differente, come in tabella 2.3, col fine di effettuare un allenamento bilanciato della rete.

Per ridurre l’overfitting e aumentare la generalizzazione delle diverse CNN, tutti i dataset di training sono stati sottoposti a *Data Augmentation*, in particolare:

- *Rotazioni random*: le immagini vengono ruotate in maniera randomica all’interno di un range di rotazione prestabilito.
- *Riflessioni random*: le immagini vengono riflesse lungo gli assi principali.

In tabella 2.4 sono riportati nel dettaglio i parametri di *Data Augmentation* utilizzati.

Metodo	Rete	Cheratosi seborroica (immagini)	Nevo benigno (immagini)	Melanoma maligno (immagini)
<b>M1</b>	CNN1	200	200	200
<b>M2</b>	CNN2.1	300	150	150
	CNN2.2	0	150	150
<b>M3</b>	CNN3.1	150	150	300
	CNN3.2	150	150	0

Tabella 2.3: Composizione del dataset per il training delle diverse strutture proposte

Rotazione Random (gradi)	Riflessione X Random	Riflessione Y Random	Rescale (fattore di scala)
[0 180]	"true"	"true"	[1 1]

Tabella 2.4: Parametri di Data Augmentation utilizzati

Come base di partenza è stato scelto di iniziare a testare il metodo M1 a differenti valori del raggio di crop, con il fine di individuare la configurazione che permettesse di ottenere le performance migliori. Successivamente, utilizzando il valore del raggio di crop migliore, sono state testate le altre due metodologie, M2 e M3. I parametri di allenamento ottimali delle diverse reti sono stati individuati dopo tentativi reiterati, al fine di ottenere il miglior allenamento possibile in termini di generalizzazione e accuratezza di predizione.

### 2.5.3 Test Sets

Le prestazioni delle diverse CNN sono state testate utilizzando 3 diversi dataset di testing:

1. *Test set 1.* Contenente immagini provenienti dai dataset MSK ISIC. Anch'esse, dotate di righello millimetrato, sono state sottoposte a cropping circolare.



2. *Test set 2*. Contiene immagini di lesioni cutanee acquisite con dermoscopia a contatto.
3. *Test set 3*. Composto dalle immagini acquisite con Nurugo Derma ma sottoposte all'algoritmo di segmentazione descritto al paragrafo 2.3.2.

I test sets 2 e 3 contengono le immagini delle stesse lesioni ma acquisite tramite dispositivi differenti.

In tabella 2.5 è riportata nel dettaglio la numerosità dei singoli dataset.

	<b>Nevo benigno (lesioni)</b>	<b>Melanoma maligno (lesioni)</b>	<b>Cheratosi seborroica (lesioni)</b>
<b>Test Set 1</b>	35	25	37
<b>Test Set 2</b>	39	18	21
<b>Test Set 3</b>	39	18	21

Tabella 2.5: Composizione dei diversi dataset di test.

L'obiettivo è valutare le variazioni di performance tra le diverse metodologie proposte, individuando la struttura più valida e versatile, in grado di riconoscere tipologie di immagini generate con dispositivi differenti.

## 2.6 Metriche di validazione

Per valutare le prestazioni e quantificare la bontà dei diversi metodi proposti sono state utilizzate delle metriche di validazione. I risultati dell'algoritmo sono quindi confrontati con le *ground truth* delle diverse lesioni: ogni immagine appartenente al dataset di test è caratterizzata da un label che ne identifica la tipologia.

### 2.6.1 Confusion Matrix

Per un problema di classificazione la metrica di validazione più utilizzata è la tavola di contingenza, o *confusion matrix*. La confusion matrix permette una visualizzazione immediata delle prestazioni: gli elementi diagonali rappresentano i corretti classificati, gli elementi non diagonali quelli classificati in maniera errata. Le dimensioni della confusion matrix dipendono dal numero di classi da classificare: 2x2 per un problema di classificazione binaria; 3x3 per un problema di classificazione terziaria.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figura 2.16: Esempio di confusion matrix per classificazione binaria.

Si definiscono:

- *TP (True Positive) - Veri Positivi*: elementi classificati correttamente come positivi al test.
- *FP (False Positive) - Falsi Positivi*: elementi classificati erratamente come positivi al test. Costituiscono i cosiddetti falsi allarmi (errore di tipo 1).
- *TN (True Negative) - Veri Negativi*: elementi classificati correttamente come negativi al test.
- *FN (False Negative) - Falsi Negativi*: elementi classificati erratamente come negativi al test. Costituiscono i cosiddetti miss (errore di tipo 2).

Ottenuti i valori della confusion matrix è possibile calcolare i seguenti indicatori statistici:

- **Accuratezza**: fornisce una misura generale delle performance del classificatore.

$$\text{Accuratezza} = \frac{TP + TN}{TP + TN + FN + FP}$$

- **Sensibilità**: TPR (True Positive Rate), definita anche *recall*, misura la percentuale di immagini positive correttamente classificate.

$$\text{Sensibilità} = \frac{TP}{TP + FN}$$

- **Specificità:** TNR (True Negative Rate), misura la percentuale di immagini negative correttamente classificate.

$$\text{Specificità} = \frac{TN}{TN + FP}$$

- **Valore predittivo positivo:**  $VP^+$ , definita anche *precisione*, indica la probabilità che un elemento test-positivo sia effettivamente tale.

$$VP^+ = \frac{TP}{TP + FP}$$

- **Valore predittivo negativo:**  $VP^-$ , indica la probabilità che un elemento test-negativo sia effettivamente tale.

$$VP^- = \frac{TN}{TN + FN}$$

- **F1 Score:** fornisce una misura dell'accuratezza totale del modello combinando le misure di precisione e accuratezza.

$$F1\ Score = 2 * \frac{\text{precisione} * \text{recall}}{\text{precision} + \text{recall}}$$

## 2.6.2 Curva ROC

Per valutare la capacità discriminante di un classificatore è possibile plottare l'andamento del *True Positive Rate* (sensibilità o recall) e del *False Positive Rate* (1-specificità) al variare della soglia scelta, costruendo la cosiddetta **curva ROC** (*Receiver Operating Characteristic*). La curva dunque rappresenta il rapporto tra allarmi veri e falsi allarmi in relazione alle impostazioni del classificatore.

E' uno strumento particolarmente utile dato che permette di individuare la configurazione ottimale del classificatore da utilizzare per un determinato problema, rendendo il modello versatile. All'interno dello spazio della curva ROC, come mostrato in figura 2.17 è possibile individuare la diagonale che congiunge il punto di coordinate (0,0) e il punto di coordinate (1,1), definita *linea di non discriminazione*, che rappresenta una classificazione del tutto casuale. Il punto di coordinate (0,1) rappresenta invece la condizione di classificazione perfetta: più la curva si allontana dalla diagonale e si avvicina

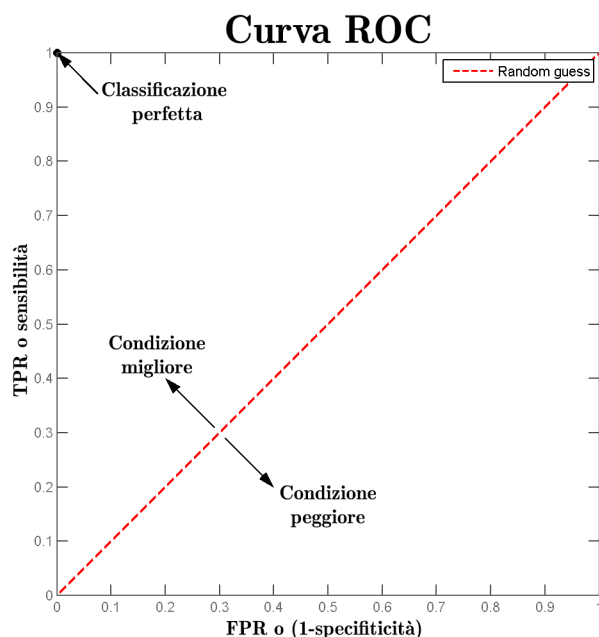


Figura 2.17: Esempio di curva ROC.

al punto (0,1) e migliori saranno le prestazioni del classificatore. Ottenuta la curva ROC è possibile inoltre calcolare l'area sottesa alla curva stessa (**AUC**, *Area Under Curve*), compresa tra 0 e 1, che quantifica il potere discriminante del classificatore:

- $AUC = 0.5$  classificatore non informativo;
- $0.5 \leq AUC \leq 0.7$  classificatore poco accurato;
- $0.7 \leq AUC \leq 0.9$  classificatore moderatamente accurato;
- $0.9 < AUC < 1.0$  classificatore altamente accurato;
- $AUC = 1$  classificatore perfetto [29].

Il piano della curva ROC verrà utilizzato per effettuare un confronto tra le prestazioni dei diversi classificatori utilizzati e tre dermatologi esperti, ai quali è stato sottoposto il medesimo set di immagini da classificare.

### 2.6.3 t-SNE

Il t-SNE (*t-distributed stochastic neighbor embedding*) è un algoritmo di riduzione della dimensionalità utile per visualizzare dati ad alta dimensione in un spazio 3D o 2D, rispettando le informazioni contenute nei dati stessi e le similarità tra i punti [30]. E' una tecnica particolarmente utilizzata nell'ambito dell'apprendimento automatico, dato che permette di capire come un modello artificiale ragioni, fornendone una rappresentazione grafica. Applicando il t-SNE all'ultimo layer nascosto della CNN è possibile ottenere una distribuzione grafica delle features interne apprese. In figura 2.18 un esempio di t-SNE applicato ad una CNN: osservando la distribuzione dei diversi punti all'interno del grafico è possibile capire se la rete stia generalizzando il problema nel modo corretto: la formazione di clusters di colore diverso, ognuno identificativo di una determinata categoria di immagini, è indice di buona generalizzazione. Lo strumento fornisce dunque una rappresentazione del modo di "pensare" della rete.

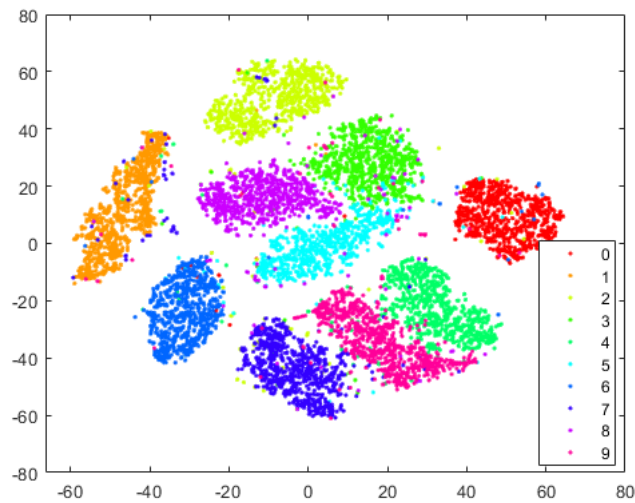


Figura 2.18: Esempio di grafico t-SNE.

### 2.6.4 Occlusion Sensitivity Map

La "*occlusion sensitivity*" è una tecnica che permette di capire quali parti di un'immagine sono più rilevanti per il processo di classificazione. Il metodo consiste nel sostituire piccole aree dell'immagine in input con una maschera

opaca. La maschera scorre sull'immagine come una convoluzione: per ogni posizione della maschera viene calcolato il punteggio di predizione di ogni singola classe. In uscita dunque viene generata una mappa che rappresenta le variazioni dei punteggi di classificazione in funzione della posizione della maschera all'interno dell'immagine [31].

Sostanzialmente viene calcolato il gradiente dell'output rispetto all'immagine in ingresso:

$$\frac{\partial output}{\partial input}$$

E' una tecnica che richiede un tuning di parametri (dimensioni della maschera, "stride" della maschera), permettendo quindi un'analisi più flessibile a diverse scale. Data la presenza all'interno delle immagini dermoscopiche di aree di pixel sani, l'obiettivo è cercare di capire se durante il processo di classificazione questi ultimi non contribuiscano alla classificazione.

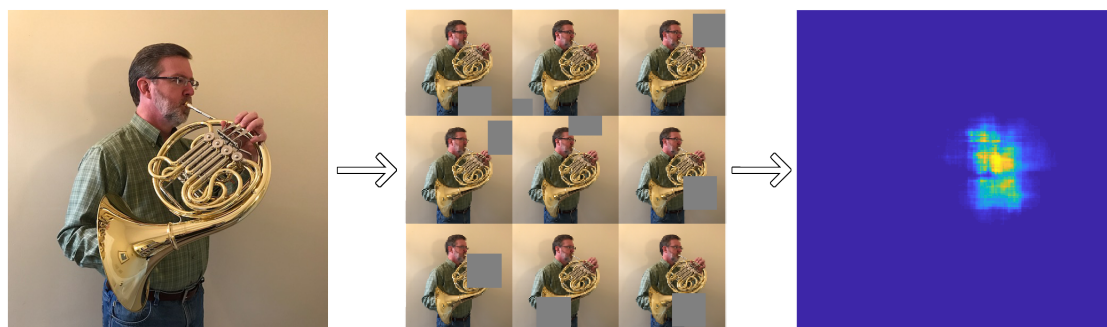


Figura 2.19: Esempio di visualizzazione tramite OSM: in questo caso sono analizzati i pixel che contribuiscono al riconoscimento del corno musicale [31].

# Capitolo 3

## Risultati

All'interno di questo capitolo vengono mostrate e analizzate nel dettaglio le performance dei metodi proposti nel capitolo 2. Innanzitutto viene effettuato un confronto tra i diversi valori di crop al fine di decretarne quello ottimale. Successivamente vengono presentate le architetture a cascata con un focus sull'allenamento e sul modo di "ragionare" delle reti stesse. Segue un confronto complessivo tra i tre metodi proposti, analizzando i punti di forza e i limiti delle diverse strutture. Il capitolo si conclude con una valutazione delle performance sulle acquisizioni con dermoscopio e Nurugo Derma effettuate presso l'Azienda Ospedaliera Universitaria Maggiore della Carità di Novara.

### 3.1 Individuazione del crop ottimale

Inizialmente è stato necessario capire quale fosse la misura del raggio di crop in grado di garantire le performance migliori. Per questo scopo sono state allenate e testate tre diverse CNN seguendo la metodologia M1.

#### 3.1.1 Metodo M1

Per le tre reti è stato utilizzato il medesimo dataset di training, composto da 200 immagini di nevi benigni, 200 immagini di melanomi maligni e 200 immagini di cheratosi seborroiche, sottoposto alle operazioni di Data Augmentation descritte nel paragrafo 2.5.2. Definiamo:

- **M1a** = CNN1 allenata con dataset con crop pari a  $1.14\text{ mm}$ .
- **M1b** = CNN1 allenata con dataset con crop pari a  $3\text{ mm}$ .

- **M1c** = CNN1 allenata con dataset con crop pari a 5 mm.

Sin dalla fase di allenamento la rete M1b ha mostrato performance migliori rispetto alle altre due reti. Osservando le diverse distribuzioni t-SNE in figura 3.1, è possibile evincere come la rete *M1b* riesca a "clusterizzare" in maniera ottimale le tre differenti classi, che appaiono più distanti nello spazio e maggiormente isolate.

Le tre reti sono state poi testate utilizzando il test-set 1, composto da immagini provenienti dall'archivio ISIC. In tabella 3.1 sono riportati i risultati di test delle 3 differenti reti.

	Accuratezza	Sensibilità	Specificità	VP <sup>+</sup>	VP <sup>-</sup>	F1
<b>M1a</b>	72%	70%	67,4%	50%	86%	59%
<b>M1b</b>	73.1%	72%	84.1%	62.1%	89.2%	67%
<b>M1c</b>	68%	75,4%	71%	53.1%	85.2%	59.6%

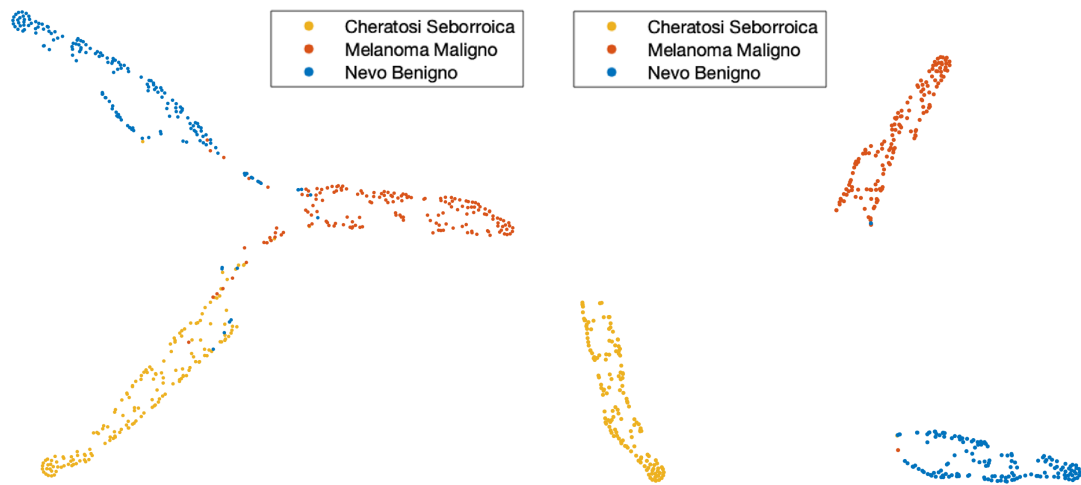
Tabella 3.1: Performance delle tre reti sul testset 1.

A conferma dell'analisi effettuata sul training, la rete M1b offre le prestazioni migliori anche sul testset. La condizione è probabilmente da attribuire alla *FOV* insufficiente delle immagini con cui è stata allenata M1a, e dall'altro lato alla *FOV* eccessiva delle immagini di training della rete M1c, contenenti maggiori elementi di rumore. È stato scelto quindi di utilizzare il crop con valore di raggio pari a 3 mm (M1b), che costituisce il compromesso migliore tra le tre configurazioni, per realizzare le altre due architetture M2 ed M3. Le difficoltà della rete nel classificare immagini provenienti da un dataset differente dal training si palesano osservando la distribuzione t-SNE relativa al testing sul testset 1. La rete, come è possibile osservare in figura 3.2, non riesce a generalizzare in maniera perfetta, come è possibile notare dalla clusterizzazione non perfettamente netta e precisa tra le tre diverse categorie.

## 3.2 Architetture a cascata

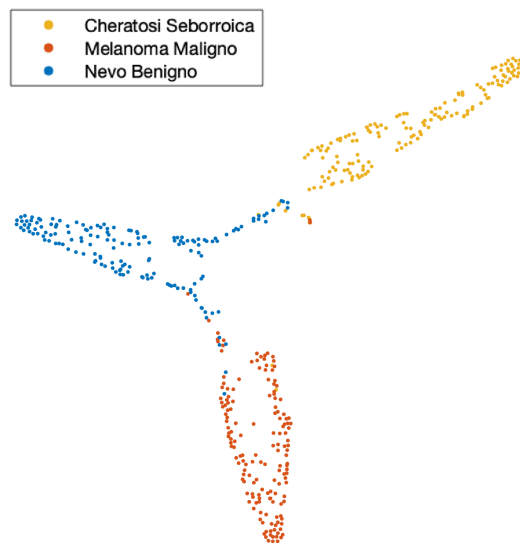
Individuato il valore di crop ottimale sono state sviluppate le due architetture a cascata M2 ed M3. La suddivisione del problema di classificazione ternario in due classificazioni binarie successive dovrebbe determinare un miglioramento delle performance data la maggiore facilità di ottimizzazione di una rete neurale in un problema binario.





(a) *Distribuzione t-SNE del training di M1a*

(b) *Distribuzione t-SNE del training di M1b*



(c) *Distribuzione t-SNE del training di M1c*

Figura 3.1: Distribuzioni t-SNE del training del metodo M1 a diversi valori di crop.



Figura 3.2: Distribuzione t-SNE della rete **M1b** sul testset1. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata.

### 3.2.1 Metodo M2

La metodologia M2 ricalca lo standard procedurale della visita dermatologica: la prima rete valuta la natura melanocitica della lesione, mentre la seconda rete distingue le lesioni melanocitiche in uscita dal primo blocco in nevi benigni e melanomi maligni.

#### CNN 2.1

La prima rete (CNN 2.1) è stato allenata su un dataset contenente 300 immagini di lesioni melanocitiche (150 immagini di nevi benigni e 150 immagini di melanomi maligni) e 300 di lesioni non melanocitiche. La rete raggiunge facilmente un'accuratezza del 100% durante la fase di training. In figura 3.3 è possibile osservare la distribuzione t-SNE relativa all'allenamento: le features relative alle due categorie si raggruppano in clusters ben distanti nello spazio. La rete produce sul testset 1 i risultati mostrati dalla confusion matrix in figura 3.4. Inoltre osservando la distribuzione t-SNE in figura 3.5 è

possibile notare come la rete riesca a raggruppare, seppure in maniera meno precisa rispetto al training, le due classi in clusters separati.

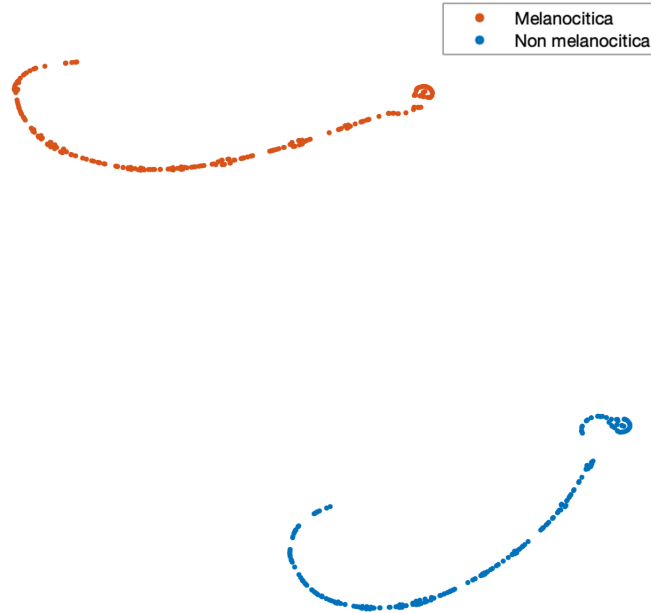


Figura 3.3: Distribuzione t-SNE relativa all'allenamento della CNN 2.1

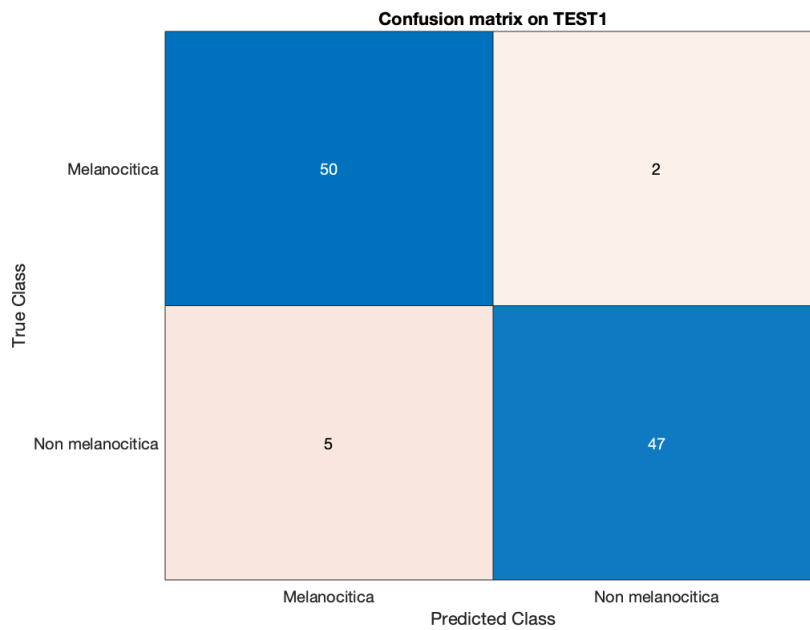


Figura 3.4: Confusion Matrix relativa alla CNN 2.1

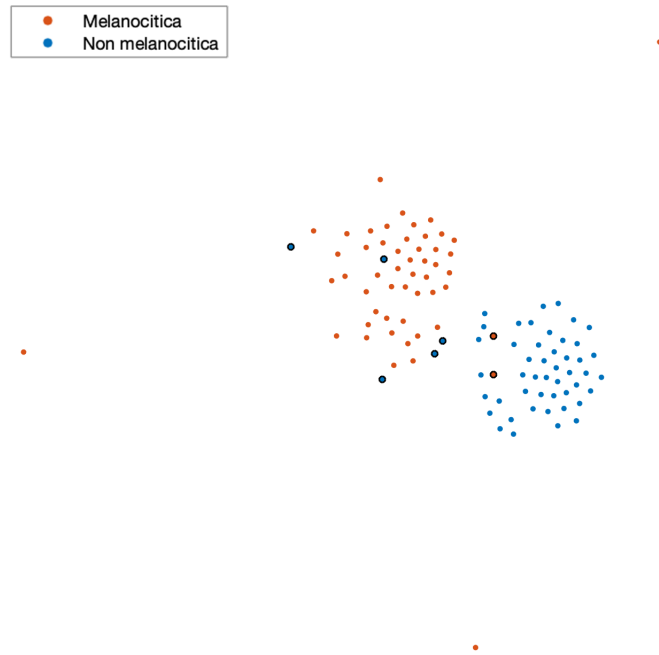


Figura 3.5: Distribuzione t-SNE della rete CNN 2.1 sul testset-1. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata.

## CNN 2.2

La seconda rete (CNN 2.1) invece è stata allenata utilizzando un dataset composto da 150 immagini di nevi benigni e 150 immagini di melanomi maligni, con lo scopo di riconoscere la natura maligna della lesione. Data la somiglianza tra le due tipologie di lesione la fase di allenamento si è rivelata più ostica, come è possibile notare dalla distribuzione t-SNE in figura 3.6.

La rete infatti clusterizza le features relative alle due categorie in due clusters contigui, non isolati tra loro. Per quanto riguarda il test set 1, la rete riceve in ingresso le lesioni classificate dalla CNN 2.1. I risultati sono mostrati dalla confusion matrix in figura 3.7.

Con un'accuratezza sul test set in ingresso pari al 74%, una sensibilità dell'88% e una specificità del 60% costituisce la configurazione che garantisce il miglior compromesso tra sensibilità e specificità: l'obiettivo è infatti quello di individuare il maggior numero di forme maligne possibili, a costo di aumentare il numero di falsi allarmi. Questa polarizzazione è osservabile dalla distribuzione t-SNE in figura 3.8.

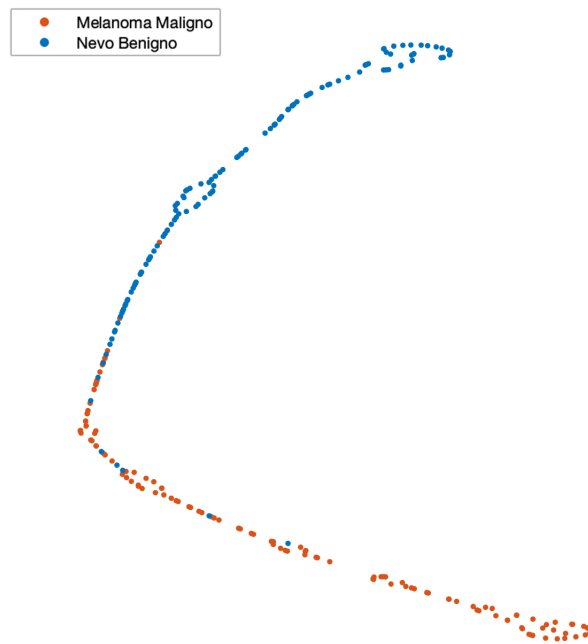


Figura 3.6: Distribuzione t-SNE relativa all'allenamento della CNN 2.2

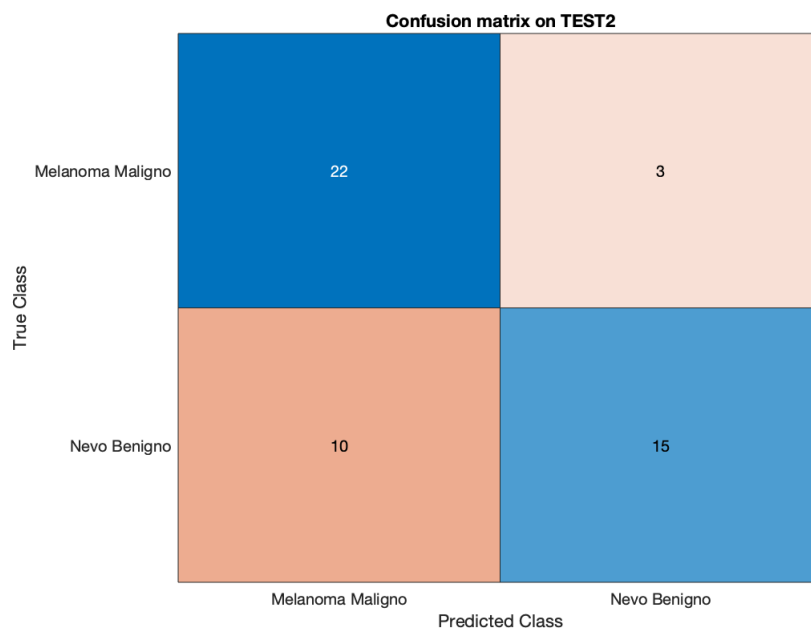


Figura 3.7: Confusion Matrix relativa alla CNN 2.2

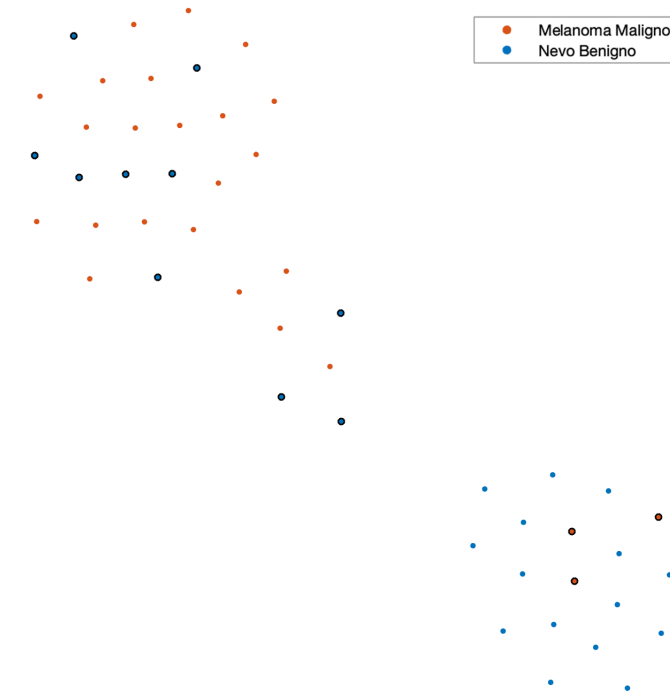


Figura 3.8: Distribuzione t-SNE della rete CNN 2.2 sul testset 1. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata.

### 3.2.2 Metodo M3

L'architettura M3 effettua una classificazione in ordine opposto rispetto a M2: la prima rete (CNN 3.1) riconosce la malignità della lesione, la seconda (CNN 3.2) suddivide le lesioni benigne in nevi (melanocitiche) e cheratosi seborroiche (non melanocitiche).

#### CNN 3.1

La prima rete è allenata su un dataset composto da 300 immagini di lesioni benigne (150 immagini di nevi benigni e 150 immagini di cheratosi seborroica) e 300 di lesioni maligne. Nonostante un'accuratezza prossima al 100% durante la fase di training, la rete non riesce a clusterizzare correttamente le features, come è possibile evincere dalla distribuzione t-SNE in figura 3.9.

In figura 3.10 e 3.11 sono riportate la confusion matrix sul testset 1 e la relativa distribuzione t-SNE.

La CNN 3.1 è caratterizzata da un'accuratezza del 78% una sensibilità dell'80% e una specificità del 77.2%, più bilanciata rispetto alla CNN 2.2.

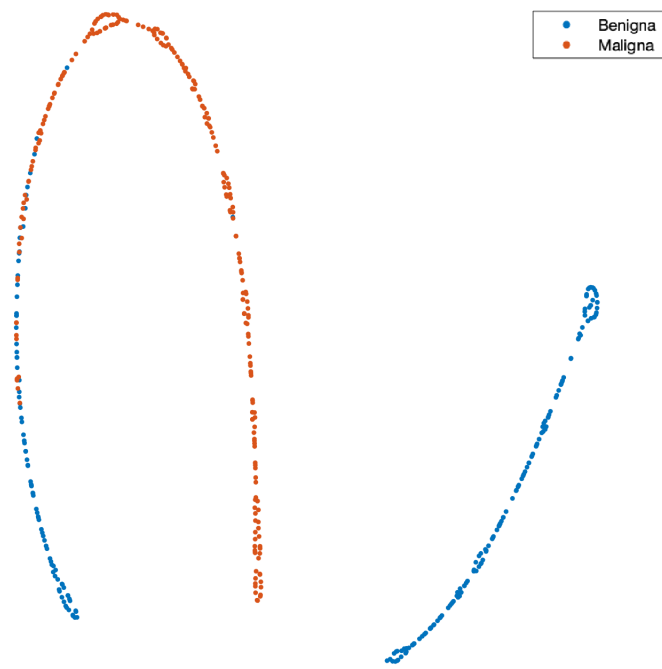


Figura 3.9: Distribuzione t-SNE relativa all’allenamento della CNN 3.1

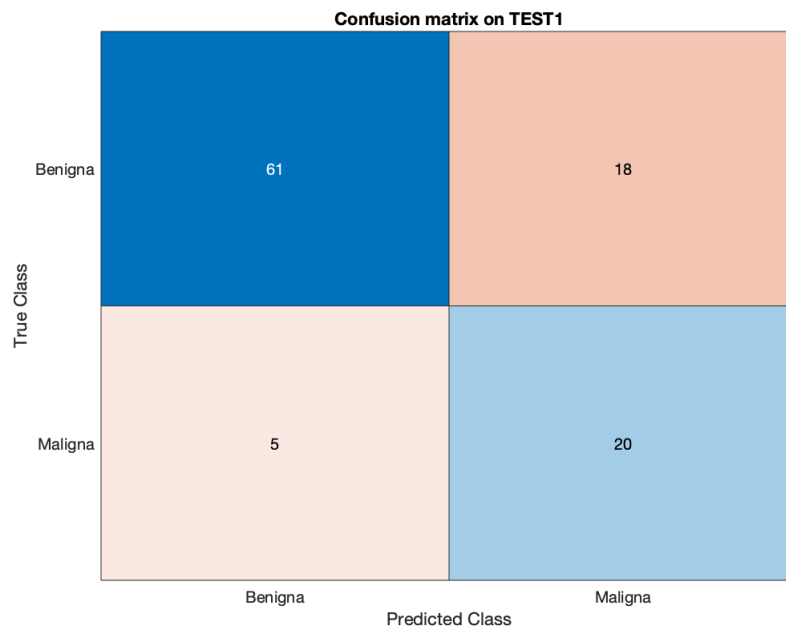


Figura 3.10: Confusion Matrix relativa alla CNN 3.1

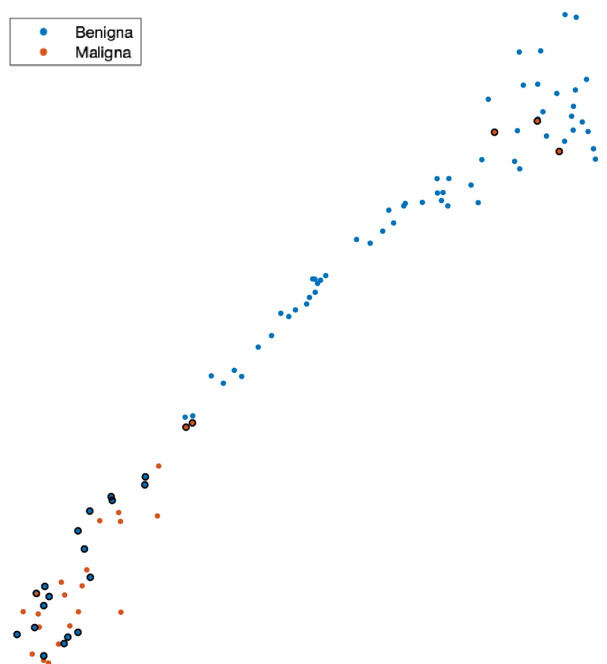


Figura 3.11: Distribuzione t-SNE della rete CNN 3.1 sul testset 1. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata.

## CNN 3.2

Una volta determinata la natura maligna o benigna della lesione, la CNN 3.2 distingue le immagini in ingresso in nevi benigni e cheratosi seborroiche. La rete è stata allenata con un dataset contenente 150 immagini di cheratosi seborroica 150 immagini di nevo benigno. Durante la fase di allenamento la rete riesce ad organizzare le features delle due categorie in clusters separati ma non perfettamente omogenei come osservabile dalla figura 3.12.

Il comportamento della rete sul test-set 1 è descritto dalla distribuzione t-SNE in figura 3.14. La rete raggiunge un'accuratezza del 91.2%, con una sensibilità, rispetto al nevo benigno, pari al 95.7% e una specificità pari al 89.5 %, come è possibile evincere dalla confusion matrix in figura 3.13.



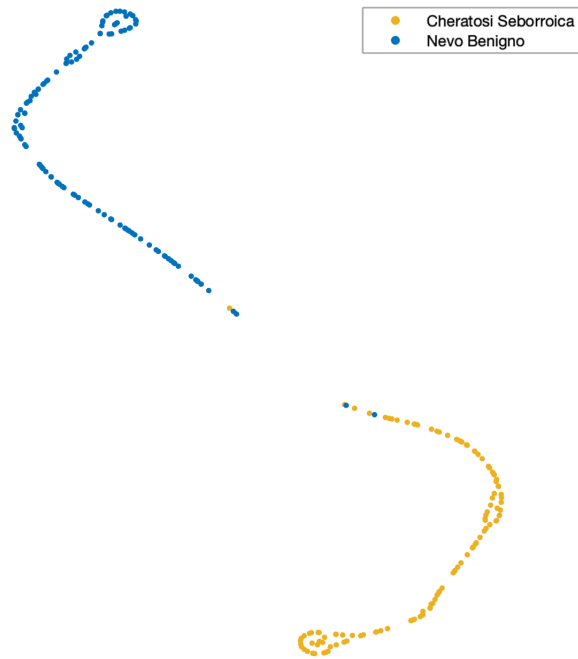


Figura 3.12: Distribuzione t-SNE relativa all’allenamento della CNN 3.2

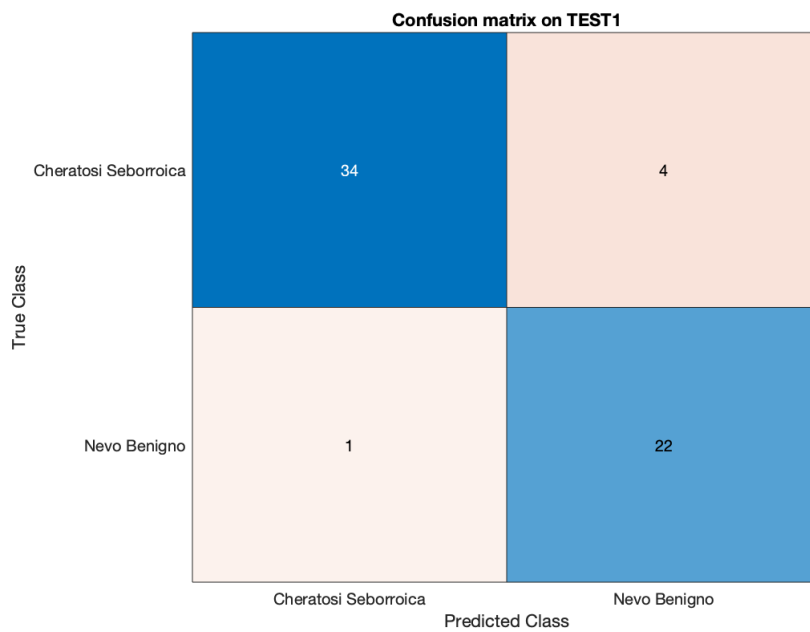


Figura 3.13: Confusion Matrix relativa alla CNN 3.2

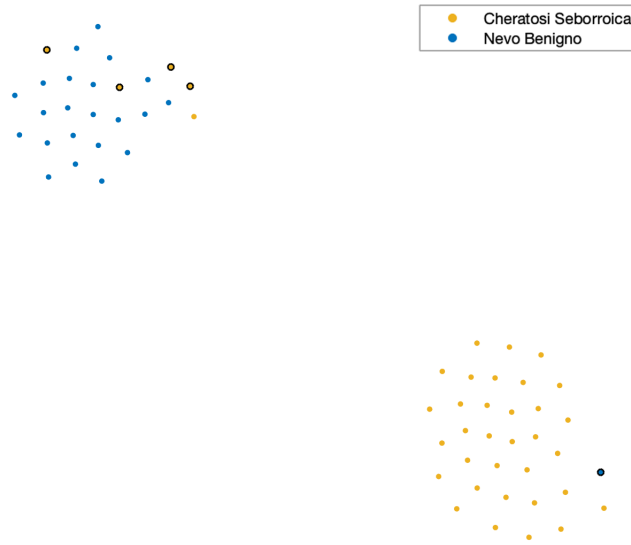


Figura 3.14: Distribuzione t-SNE della rete CNN 3.2 sul testset 1. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata.

### 3.3 Confronto tra le tre metodologie

Nelle figure 3.15, 3.16 e 3.17 sono riportate le confusion matrix relative alle tre metodologie testate utilizzando il test-set 1. Analizzando le performance in tabella 3.2 è possibile osservare come la configurazione a cascata M2 garantisca complessivamente i risultati migliori. Come ci si aspettava infatti una configurazione con due reti binarie raggiunge un'ottimizzazione migliore rispetto ad una rete a classificazione terziaria.

	Accuratezza	Sensibilità	Specificità	VP <sup>+</sup>	VP <sup>-</sup>	F1
<b>M1</b>	73.1%	72%	84.1%	62.1%	89.2%	67%
<b>M2</b>	81%	88%	86%	69%	95.4%	77%
<b>M3</b>	73.1%	80%	76%	53%	92%	63.5%

Tabella 3.2: Performance delle tre metodologie sul testset 1.

Per quantificare la bontà delle performance raggiunte dai diversi metodi è stato effettuato un confronto con le prestazioni ottenute da tre differenti

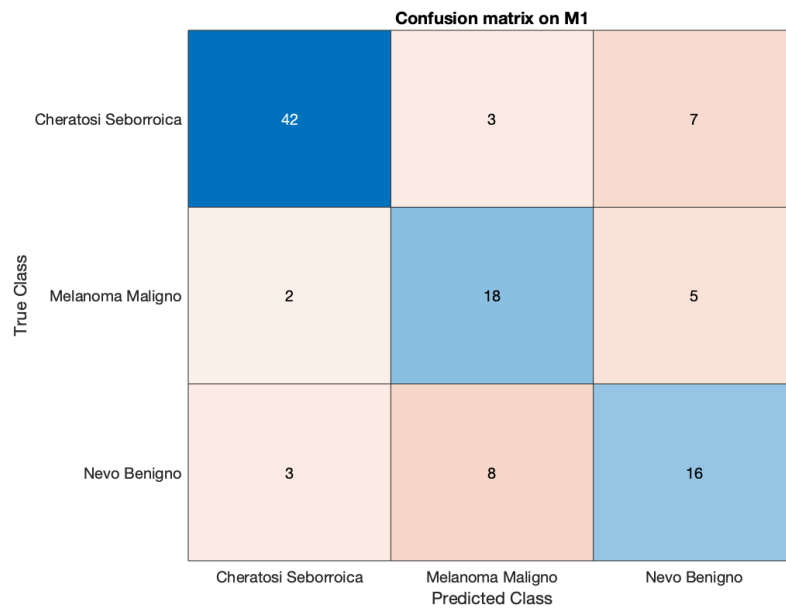


Figura 3.15: Confusion Matrix relativa all'architettura M1, valutata sul test-set 1.

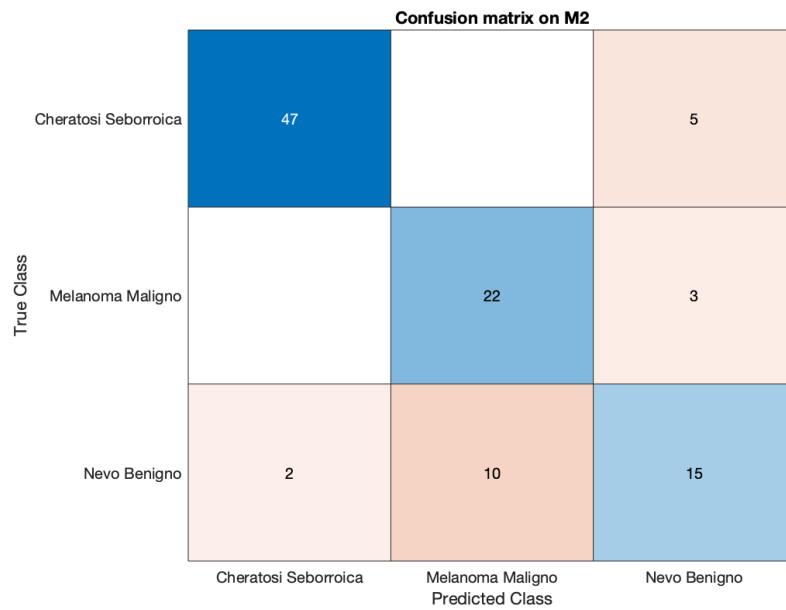


Figura 3.16: Confusion Matrix relativa all'architettura M2, valutata sul test-set 1.

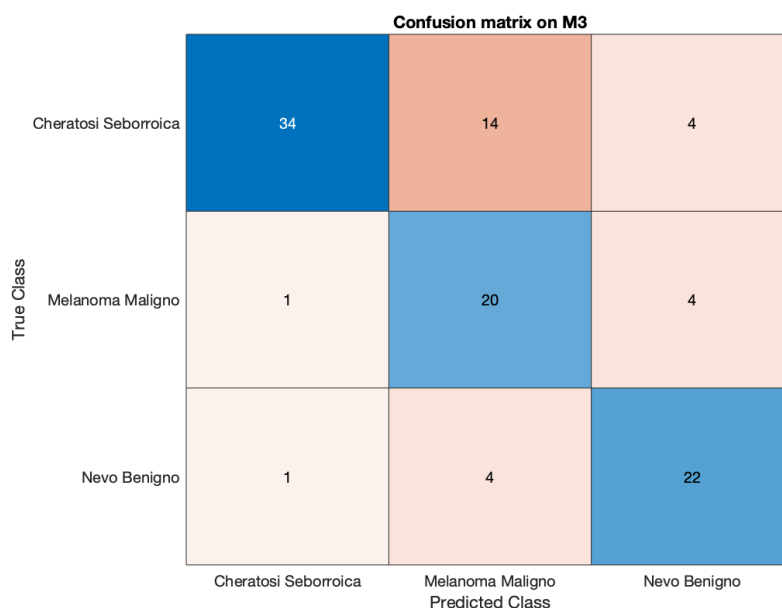


Figura 3.17: Confusion Matrix relativa all’architettura M3, valutata sul test-set 1.

dermatologici esperti (**D1**, **D2** e **D3**), ai quali è stato sottoposto il medesimo dataset di test (Test Set 1). Confrontando i valori in tabella 3.3 relativi agli esperti e i valori in tabella 3.2 relativi alle tre metodologie, è possibile notare come l’architettura M2 vanti livelli di accuratezza, sensibilità e F1 score migliori rispetto alla valutazione degli esperti, mentre ottenga punteggi più bassi per quanto riguarda la specificità.

	Accuratezza	Sensibilità	Specificità	VP <sup>+</sup>	VP <sup>-</sup>	F1
<b>D1</b>	68 %	36%	74.7%	51.3%	93.3%	63.5%
<b>D2</b>	70.1%	32%	97%	80%	78%	46%
<b>D3</b>	75.3%	44%	94%	73.3%	81.6%	55%

Tabella 3.3: Performance dei tre dermatologi sul testset 1.

Per poter infine valutare il potere discriminante dei diversi classificatori sono state generate delle curve ROC che mostrano nello spazio la variazione di sensibilità e specificità al variare della soglia impostata. All’interno dello spazio sono presenti le performance di predizione dei differenti dermatologi in rosso; in verde è rappresentata la performance media tra i dermatologi insieme alla deviazione standard.

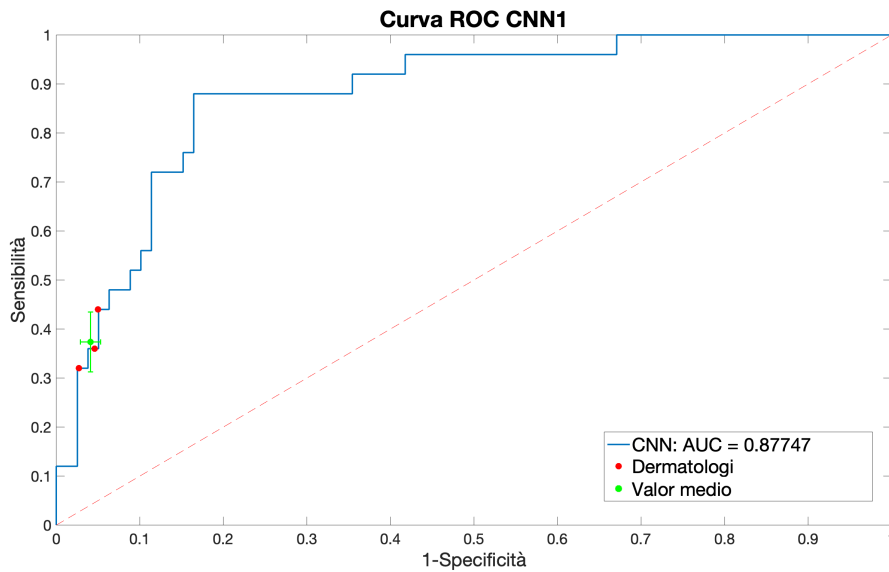


Figura 3.18: Curva ROC relativa all'architettura M1. Viene valutata la capacità della rete di discriminare il melanoma maligno.

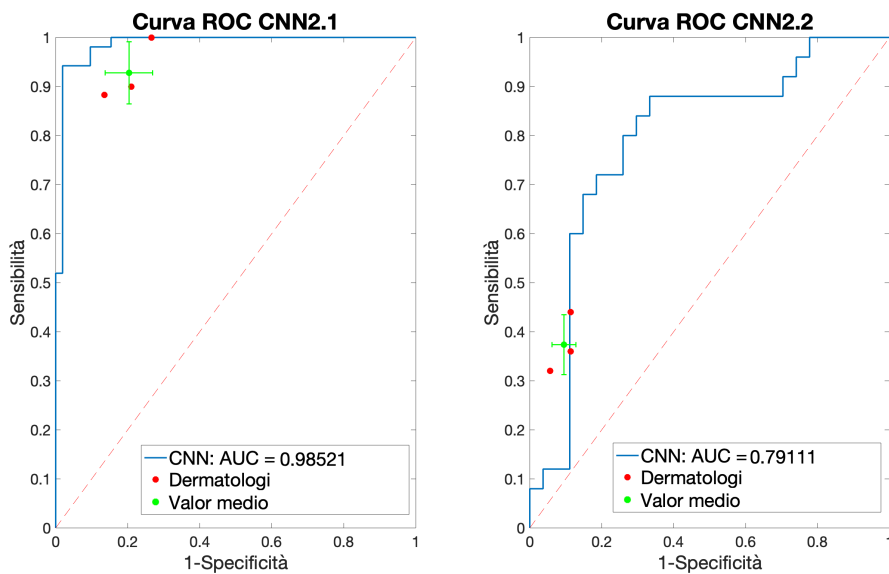


Figura 3.19: Curve ROC relative all'architettura M2. La curva ROC a sinistra valuta la capacità della rete di discriminare le lesioni melanocitiche dalle lesioni non melanocitiche. La curva a destra invece analizza la capacità di differenziare il melanoma maligno dal nevo benigno.

Analizzando le diverse curve riportate nelle figure 3.18, 3.19 e 3.20 è possibile notare come la maggior parte delle predizioni dei dermatologi si trovi al

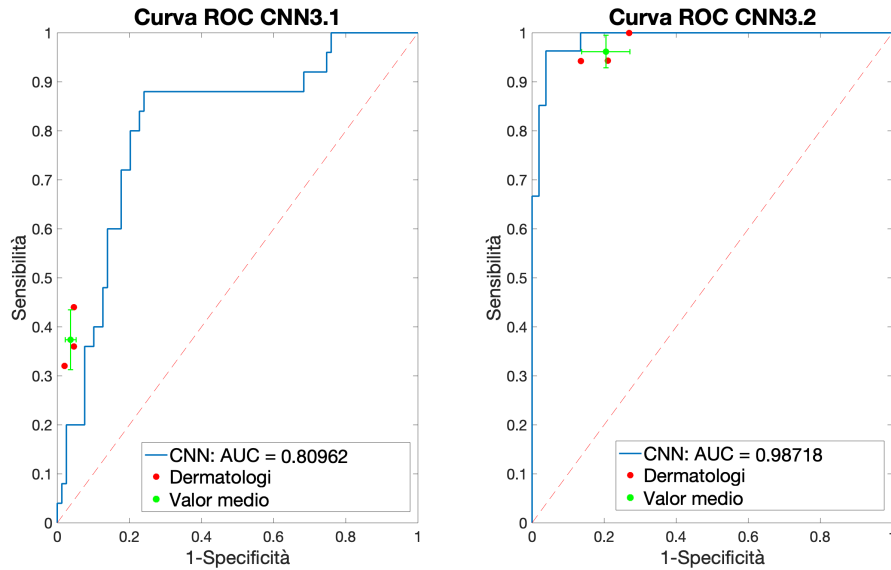


Figura 3.20: Curve ROC relative all’architettura M3. La curva ROC a sinistra valuta la capacità della rete di discriminare le lesioni maligne dalle lesioni benigne. La curva a destra invece analizza la capacità di differenziare il nevo benigno dalla cheratosi seborroica.

di sotto o nei pressi della curva ROC stessa: tutte le CNN si dimostrano alla pari con gli esperti, mostrando un buon *agreement*, o addirittura ne superano le prestazioni in quanto a sensibilità.

	M1		M2		M3	
	<i>CNN1</i>	<i>CNN 2.1</i>	<i>CNN 2.2</i>	<i>CNN 3.1</i>	<i>CNN 3.2</i>	
<b>AUC</b>	88%	98.5%	79.1%	81%	99%	

Tabella 3.4: AUC per le diverse CNN.

Esaminando i valori delle AUC relative alle diverse CNN, riportate in tabella 3.4, è possibile affermare come tutte le CNN realizzate costituiscano dei classificatori moderatamente accurati e in particolare le CNN 2.1 e 3.2 dei classificatori altamente accurati.

Infine per capire se le diverse reti neurali producano una predizione basandosi sulle aree effettivamente costituite da lesione, è stata effettuata un’analisi basata sulle Occlusion Sensitivity Map. In figura sono riportate le mappe di occlusione di alcune immagini d’esempio, nelle quali sono presenti pixel sani

e pixel di lesione. Osservando la figura è possibile notare come le predizioni si basino fortemente sulle zone contenenti pixel di lesione.

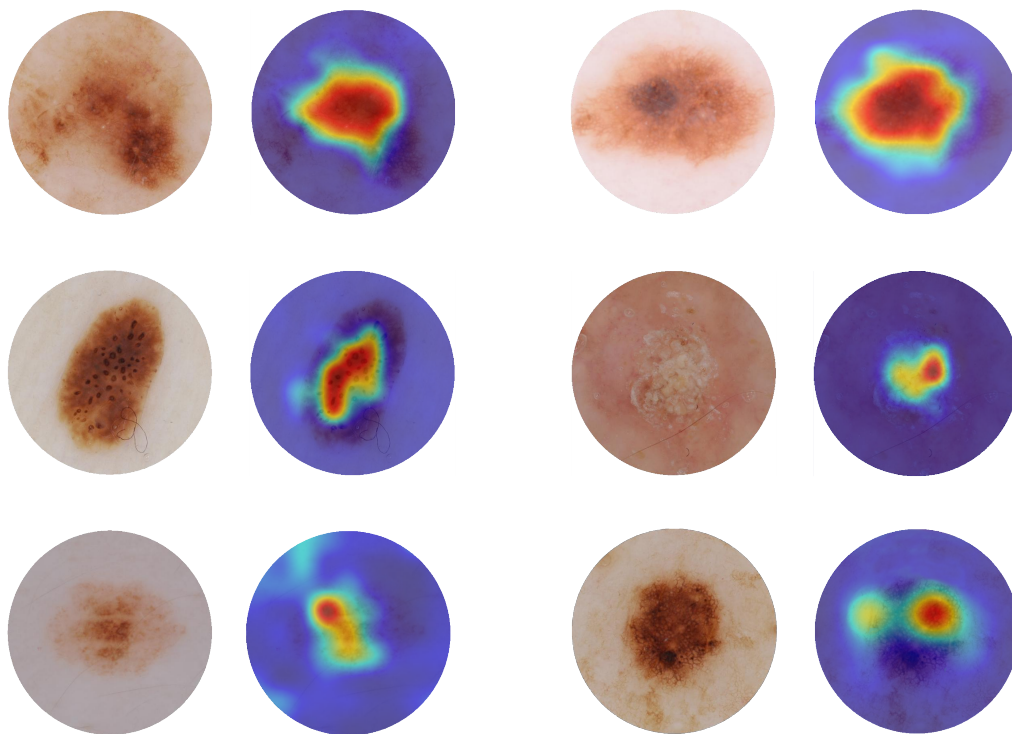


Figura 3.21: Esempio di Occlusion Sensitivity Maps.

### 3.3.1 Analisi delle soglie

Le performance mostrate precedentemente sono state totalizzate considerando come predizione quella avente la probabilità più alta tra le diverse classi. E' stata così effettuata un'analisi delle soglie sulle architettura a cascata M2 e M3 col fine di trovare le quattro coppie di valori in grado di fornire rispettivamente massima sensibilità sul melanoma cutaneo e massima specificità, in maniera tale da rendere il classificatore più versatile ed adattabile a seconda delle esigenze. In tabella 3.5 sono riportate le performance sul test-set 1 delle metodologie M2 e M3 ottenute utilizzando 4 coppie di soglie. Definiamo C1 la configurazione a *probabilità massima*, C2 la configurazione a *sensibilità massima* e C3 la configurazione a *specificità massima*.

E' possibile osservare come per la rete M2 la configurazione a maggiore sensibilità riesca ad ottenere dei valori di performance uguali alla configurazione

	Accuratezza	Sensibilità	Specificità	VP <sup>+</sup>	VP <sup>-</sup>	F1
<b>M2-C1</b>	81%	88%	86%	69%	95.4%	77%
<b>M2-C2</b>	79%	88%	79%	58%	95.25%	70%
<b>M2-C3</b>	82%	72%	90.5%	72%	90.5%	72%
<b>M3-C1</b>	73.1%	80%	76%	53%	92%	63.5%
<b>M3-C2</b>	67.3%	88%	66%	47%	94.1%	61.1%
<b>M3-C3</b>	72.1%	60%	82.2%	53.6%	86%	57%

Tabella 3.5: Confronto tra le prestazioni delle architetture M2 e M3 sul test-set1, utilizzando criteri di calcolo della predizione diversi.

a probabilità massima, mentre è osservabile un incremento della specificità (dall'86% al 90.5%) utilizzando la configurazione a maggiore specificità. Per il metodo M3 invece si osserva un aumento della sensibilità dall'80% all'88% utilizzando la configurazione C2, e un miglioramento della specificità utilizzando la configurazione C3, dal 66% all'82.2%.

### 3.4 Prestazioni sul test-set Dermoscopico

Per quantificare la capacità di generalizzazione su immagini acquisite con dispositivi differenti rispetto a quelle del training, le diverse metodologie proposte sono state sottoposte al test-set 2, contenente 26 immagini di cheratosi seborroica, 33 immagini di melanoma maligno e 39 immagini di nevo benigno, acquisite con dermoscopio a contatto. Per ogni immagine del test-set sono stati misurati manualmente i valori di risoluzione pixel/mm e ricavate le maschere manuali: l'intero test set è stato così sottoposto ad un crop circolare da 3 mm di raggio centrato sulla lesione. Le prestazioni delle diverse metodologie sono descritte in tabella 3.6 e dalle confusion matrix allegate (3.22, 3.23, 3.24).

	Accuratezza	Sensibilità	Specificità	VP <sup>+</sup>	VP <sup>-</sup>	F1
<b>M1</b>	62.2%	79%	62.5%	55.3%	83.3%	65%
<b>M2</b>	59.2%	73%	79.1%	73%	79.1%	73%
<b>M3</b>	57.2%	70%	66%	57.5%	77%	63%

Tabella 3.6: Performance delle tre architetture sul test-set 2



**Confusion matrix on M1**

True Class	Predicted Class		
	Cheratosi Seborroica	Melanoma Maligno	Nevo Benigno
Cheratosi Seborroica	16	6	4
Melanoma Maligno	5	26	2
Nevo Benigno	5	15	19

Figura 3.22: Confusion Matrix relativa all'architettura M1, valutata sul test-set 2

**Confusion matrix on M2**

True Class	Predicted Class		
	Cheratosi Seborroica	Melanoma Maligno	Nevo Benigno
Cheratosi Seborroica	15	2	9
Melanoma Maligno	6	24	3
Nevo Benigno	13	7	19

Figura 3.23: Confusion Matrix relativa all'architettura M2, valutata sul test-set 2

L'architettura M1 garantisce la classificazione più accurata e sensibile, mentre il metodo M2 fornisce una classificazione più bilanciata tra sensibilità

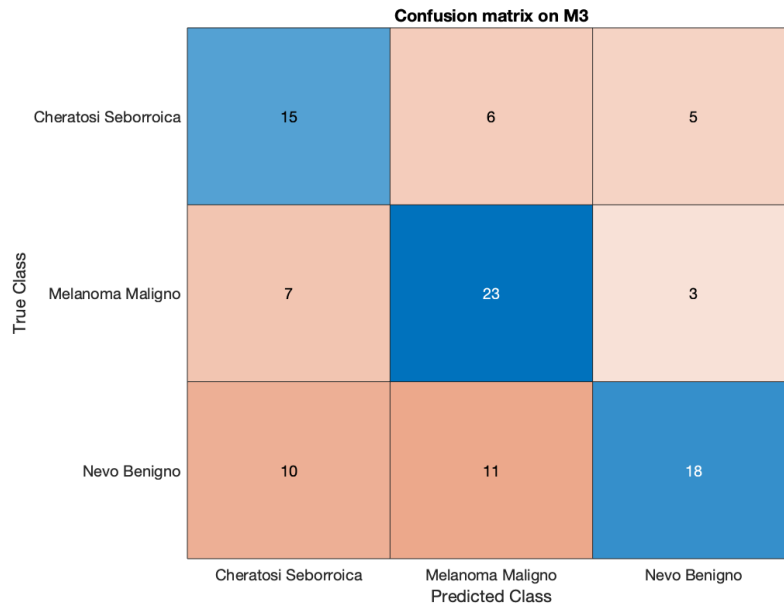


Figura 3.24: Confusion Matrix relativa all’architettura M3, valutata sul test-set 2

e specificità come riportato dal punteggio F1. Il metodo M3 invece presenta prestazioni inferiori ad entrambe le reti.

Come da aspettative, essendo le metodologie allenate su un dataset ristretto, le diverse architetture mostrano difficoltà nel riconoscere lesioni acquisite con un dispositivo differente e avente qualità inferiore in termini di risoluzione pixel/mm. La presenza di artefatti quali righello millimetrico, bolle e la mancanza di messa a fuoco contribuiscono a rendere la classificazione ancora più ardua.

### 3.5 Prestazioni sul test-set Nurugo Derma

Parallelamente l’obiettivo del lavoro di tesi è stato valutare l’utilità diagnostica delle immagini acquisite in epiluminescenza con il Nurugo Derma, cercando di capire se fossero interpretabili da una rete neurale convoluzionale. Le immagini acquisite tramite questo dispositivo presentano tuttavia un elevato numero di problematiche, quali la presenza di numerose bolle d’aria, la mancanza di messa a fuoco, la presenza di striature probabilmente imputabili alla composizione fisica del vetrino, tutte dovute alle modalità obbligate

di acquisizione. Successivamente ad una fase di scrematura, nella quale sono state eliminate le immagini più problematiche, le tre metodologie sono state testate su un test set composto da 40 immagini di cheratosi seborroica, 29 immagini di melanoma maligno e 43 immagini di nevo benigno. In tabella 3.7 sono riportate le prestazioni raggiunte dalle tre differenti architetture, corredate dalle corrispettive confusion matrix in figura 3.25, 3.26 e 3.27.

	<b>Accuratezza</b>	<b>Sensibilità</b>	<b>Specificità</b>	<b>VP<sup>+</sup></b>	<b>VP<sup>-</sup></b>	<b>F1</b>
<b>M1</b>	70 %	62.1%	88%	69.2%	84.5%	65.4%
<b>M2</b>	55.4%	62.1%	85%	69.2%	85%	65.4%
<b>M3</b>	38.4%	86%	23.4%	30%	82%	44.2%

Tabella 3.7: Performance delle tre architetture sul test-set 3

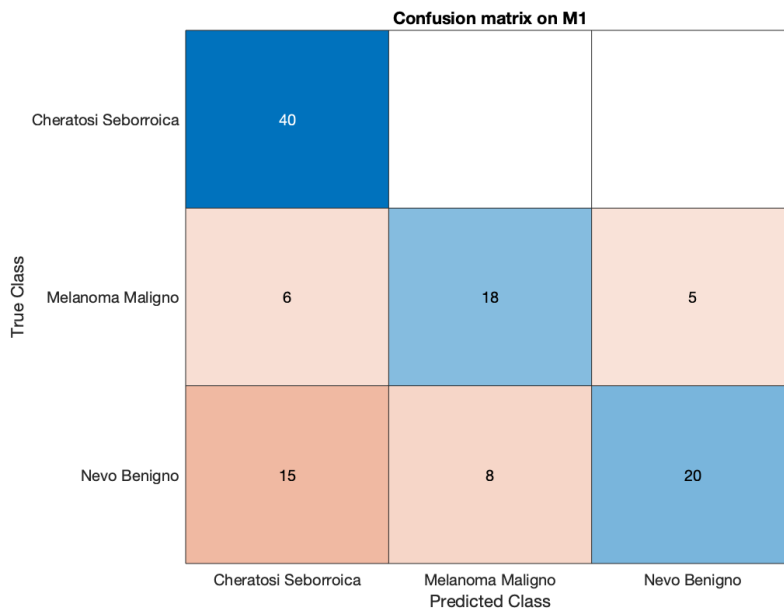


Figura 3.25: Confusion Matrix relativa all'architettura M1, valutata sul test-set 3.

E' possibile notare come a differenza del test-set 1 la configurazione in grado di garantire l'accuratezza migliore sia la M1, che esegue una classificazione one-shot. Le prestazioni della configurazione M2 si avvicinano alla M1 ma sono maggiormente polarizzate sulla cheratosi seborroica. Il metodo M3 invece, che esegue in prima battuta una distinzione tra lesioni benigne e

**Confusion matrix on M2**

True Class	Predicted Class		
	Cheratosi Seborroica	Melanoma Maligno	Nevo Benigno
Cheratosi Seborroica	40		
Melanoma Maligno	11	18	
Nevo Benigno	31	8	4

Figura 3.26: Confusion Matrix relativa all'architettura M2, valutata sul test-set 3.

**Confusion matrix on M3**

True Class	Predicted Class		
	Cheratosi Seborroica	Melanoma Maligno	Nevo Benigno
Cheratosi Seborroica	16	22	2
Melanoma Maligno	3	25	1
Nevo Benigno	4	37	2

Figura 3.27: Confusion Matrix relativa all'architettura M3, valutata sul test-set 3.

maligne e successivamente tra nevo benigno e cheratosi, risulta polarizzata

sul melanoma, non riuscendo a distinguere i caratteri benigni dai maligni, classificando come tali la maggior parte delle lesioni, come testimoniato dalla bassissima specificità (23.4%).

La distribuzione t-SNE, riportata in figura 3.28, mostra la polarizzazione della rete M1 verso le cheratosi seborroiche. Questo errore di classificazione è spiegabile osservando la figura 3.29 nella quale sono riportate degli esempi di immagini di nevo benigno e melanoma maligno classificate come cheratosi: la mancanza di messa a fuoco, le striature dovute al vetrino e soprattutto le bolle d'aria determinate dal liquido di interfaccia, vengono confuse dalla rete con i pattern tipici delle cheratosi seborroiche, causando un errore di classificazione sistematico.

I risultati ottenuti sono particolarmente incoraggianti, a dimostrazione di come un dispositivo ad uso amatoriale, possa prestarsi, con tutti i limiti correlati, ad un'analisi clinica di questo tipo. Nel capitolo seguente sono discusse tutte le possibili migliorie ed accorgimenti attuabili per cercare di aumentare le prestazioni di un dispositivo di questa tipologia.

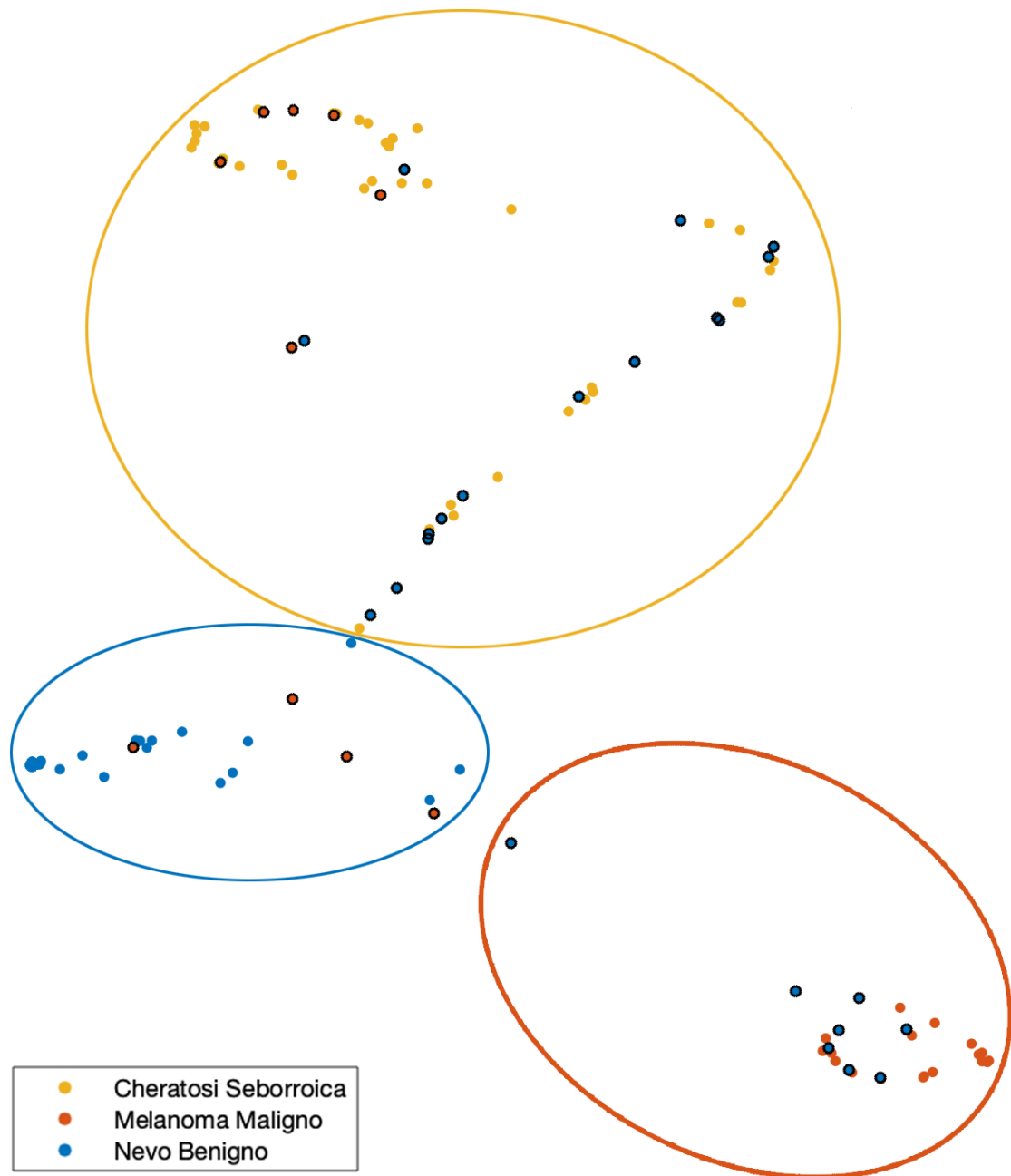
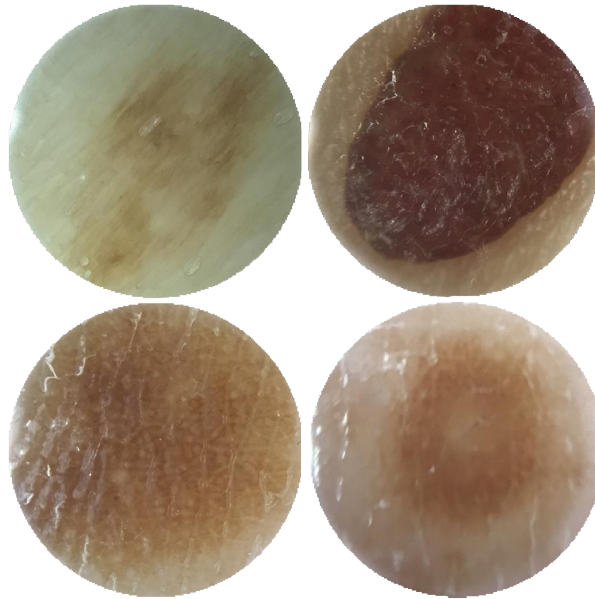
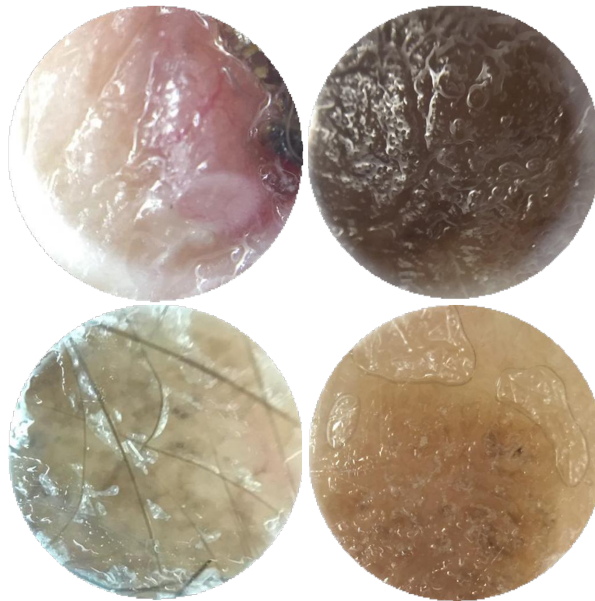


Figura 3.28: Distribuzione t-SNE dell'architettura M1 sul testset 3. I puntini cerchiati in nero rappresentano le immagini classificate in maniera errata. E' possibile osservare come la maggior parte delle misclassificazioni siano collocate all'interno del cluster delle cheratosi seborroiche, in giallo.



(a) Immagini di nevo benigno classificate come cheratosi seborroica



(b) Immagini di melanoma maligno classificate come cheratosi seborroica

Figura 3.29: Esempi di immagini di nevo benigno e melanoma maligno classificate erroneamente come cheratosi seborroica. E' possibile osservare i diversi artefatti, bolle d'aria, mancanza di messa a fuoco e striature, che inficiano sulla qualità dell'immagine.

# Capitolo 4

## Conclusioni e sviluppi futuri

### 4.1 Conclusioni

L'obiettivo di questo lavoro di tesi consiste nella realizzazione di un sistema di classificazione basato su reti neurali convoluzionali in grado di distinguere il melanoma cutaneo da altre due tipologie di lesioni della pelle, il nevo benigno e la cheratosi seborroica. Le scelte di progetto attuate sono state dettate dalla necessità di realizzare un sistema in grado di riconoscere anche lesioni cutanee acquisite con il *Nurugo Derma*, un dispositivo non professionale che, tramite determinati accorgimenti, e con diverse limitazioni, riesce a fornire immagini paragonabili ad un dermoscopio professionale. Parallelamente dunque, il lavoro di tesi propone un "*proof of concept*" del suddetto dispositivo, col fine di decretarne la validità in campo clinico e diagnostico. E' stato dimostrato come le architetture sviluppate forniscano prestazioni incoraggianti sul testset *ISIC* con un'accuratezza di predizione pari all'81% e un *F1 score* massimo pari al 77%, superando le performance ottenute dai dermatologi (accuratezza media del 71% e *F1 score* medio pari al 55%) ai quali sono state sottoposte le medesime immagini. Le diverse metodologie proposte sono state poi testate utilizzando un set di immagini dermoscopiche, aventi caratteristiche differenti rispetto alle immagini *ISIC*, raggiungendo un punteggio *F1* massimo pari al 65.4%. Il riconoscimento di immagini con caratteristiche differenti, presenza di maggiori artefatti e di qualità minore, si è rivelata più difficoltosa, condizione da attribuire principalmente alla numerosità esigua del dataset di allenamento delle diverse reti, che non ha permesso



di raggiungere capacità di generalizzazione elevatissime. Infine, successivamente ad una fase di pre-processing essenziale, le immagini acquisite tramite Nurugo Derma sono state sottoposte alle diverse architetture, raggiungendo un'accuratezza massima pari al 70%. Considerando le diverse condizioni al contorno, particolarmente restrittive, il risultato raggiunto è da considerarsi più che promettente, dimostrando come dispositivi di questo tipo, con i dovuti accorgimenti, possano fornire indicazioni utili ai fini della diagnosi.

Un sistema di questo tipo potrebbe essere utilizzato come base di partenza per la realizzazione di un tool di supporto, o *second opinion*, da affiancare al dermatologo col fine di assisterlo durante l'esame dermoscopico, particolarmente critico.

## 4.2 Sviluppi futuri

Lo studio portato avanti dal lavoro di tesi dimostra dunque come una classificazione basata su CNN costituisca un approccio valido al problema. Parallelamente sono state ottenute performance incoraggianti su un dispositivo amatoriale come il *Nurugo Derma*, fornendo diversi spunti sui possibili miglioramenti atti a by-passarne i limiti. Nelle prossime due sezioni sono discusse idee e suggerimenti da utilizzare come input per il proseguimento del lavoro.

### 4.2.1 Miglioramento del classificatore

Le diverse metodologie realizzate sono state allenate su un dataset di allenamento ristretto contenente 200 immagini per tipologia di lesione. Per ottenere performance migliori il primo passo consiste nel sottoporre la rete, durante la fase di allenamento, ad un numero maggiore di esempi, in modo da aumentarne la capacità di generalizzazione, e ridurre i casi di overfitting. In secondo luogo, potrebbe risultare utile sottoporre le immagini di allenamento ad un *Data Augmentation* più profondo, introducendo magari del rumore artificiale, in grado di emulare le condizioni di mancanza di messa a fuoco, e indagare nuove strutture di rete, verificando se un'architettura ancora più profonda riesca a garantire un miglioramento di prestazioni. Infine, sarebbe utile cercare di allargare il raggio d'azione del classificatore, rendendolo capace di riconoscere altre tipologie di lesione.

### 4.2.2 Progettazione di un dispositivo "Nurugo-like"

Per ottenere delle immagini simili a quelle dermoscopiche, e ricreare dunque le condizioni di epiluminescenza, è stato necessario utilizzare un vetrino da laboratorio. Questo approccio però ha introdotto diverse limitazioni sulle immagini acquisite e sulle modalità di utilizzo del dispositivo stesso. L'artefatto determinato dal vetrino restringe eccessivamente il *FOV* dell'immagine, mentre l'utilizzo di liquido di interfaccia crea delle bolle che interferiscono con l'interpretazione dell'immagine; l'utilizzo del vetrino in sé infine vanifica l'immediatezza di utilizzo pensata del dispositivo.

Dal punto di vista clinico il dispositivo è stato ritenuto valido dagli esperti e una versione migliorata e priva dei difetti sopra menzionati, risulterebbe particolarmente utile principalmente per la facilità e l'immediatezza di utilizzo.

Si potrebbe pensare dunque di realizzare un dispositivo "Nurugo-like", a basso costo, con le seguenti caratteristiche:

- **Sistema di illuminazione LED interno:** l'utilizzo del flash del dispositivo non permette di ottenere una sorgente luminosa uniforme e soprattutto polarizzabile. Si potrebbe realizzare una corona di LED a basso consumo alimentati tramite lo smartphone stesso.
- **Cross polarizzazione:** l'utilizzo di luce polarizzata permetterebbe di evitare l'applicazione di liquido di interfaccia, eliminando le fastidiose bolle d'aria che disturbano l'immagine. Montando due filtri polarizzati lineare in maniera concentrica sarebbe possibile ricreare le condizioni di polarizzazione.
- **Piastrina di contatto:** la possibile di utilizzare una piastrina di contatto trasparente e millimetrata permetterebbe di comprimere le lesioni più spesse durante le acquisizioni.

In figura 4.1 ne è mostrato un primo possibile prototipo.

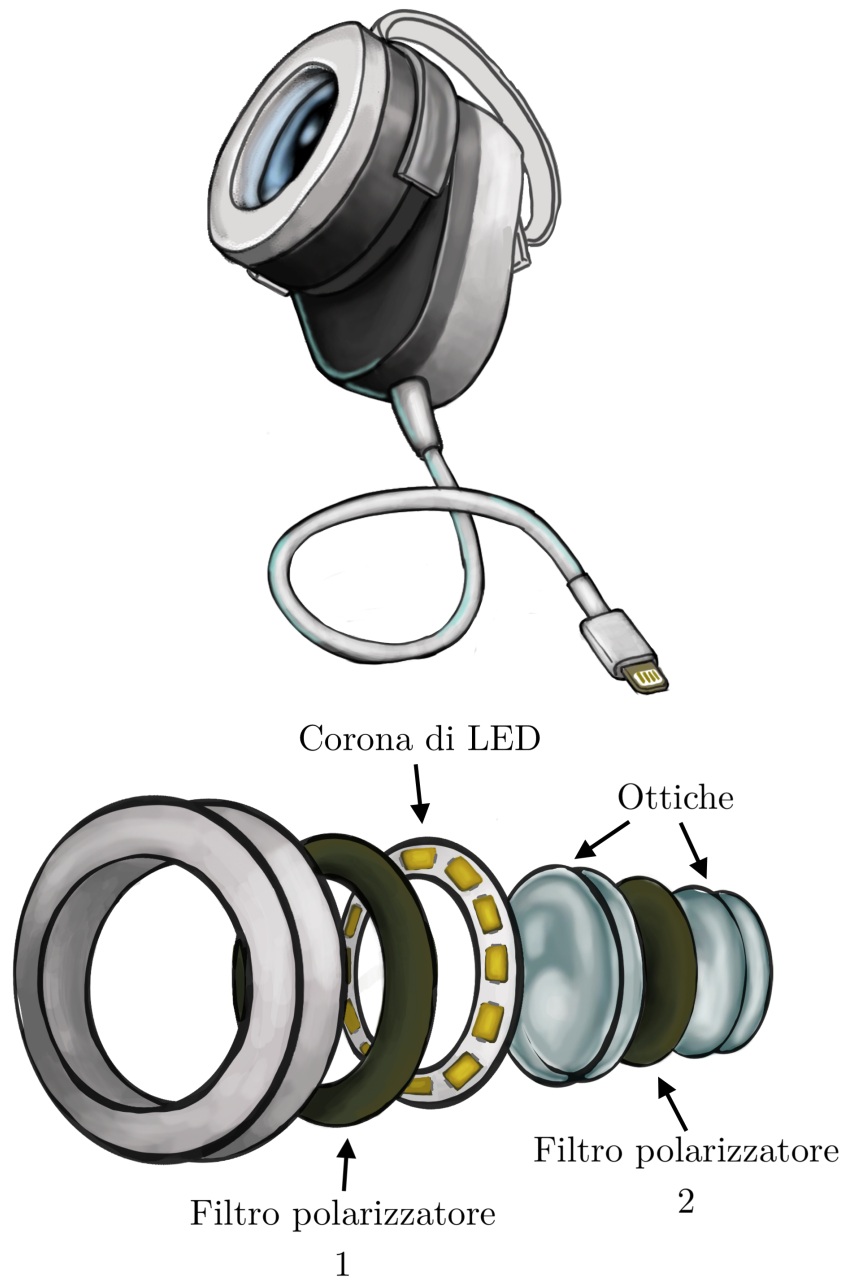


Figura 4.1: Prototipo "Nurugo-like".

# Ringraziamenti

*Una volta raggiunta una vetta ci è concesso un momento di riflessione, un'occasione per guardarsi alle spalle e fare un riassunto di quanto vissuto e imparato. E' stato un percorso arduo, pieno di insidie e difficoltà, che però si è rivelato un'occasione per crescere e imparare, principalmente dai propri errori. Lo sforzo individuale sarebbe risultato vano senza il supporto di tutte le persone che mi hanno accompagnato in questo percorso di crescita.*

*Vorrei ringraziare innanzitutto la professoressa Meiburger e il professore Molinari per avermi dato l'opportunità di affrontare questo lavoro di tesi e permesso di ampliare ulteriormente il mio bagaglio di conoscenze, e soprattutto l'ingegner Salvi per avermi assistito in corsa e per la sua professionalità ineccepibile, grazie per tutti i consigli preziosi.*

*Ringrazio tutti i colleghi che ho avuto la fortuna di conoscere e con cui ho potuto collaborare e legare, ingegneri e persone validissime, e in particolare Umberto, Elena e Federica, senza di voi questo viaggio non sarebbe stato così divertente.*

*Grazie a tutti i miei amici, quelli "nuovi" con cui abbiamo condiviso le vittorie e le cadute universitarie e non solo, e quelli "vecchi", siete sempre un motivo in più per tornare a casa.*

*Ringrazio Ciccio, Ture ed Ernesto, amici di una vita che hanno intrapreso insieme a me questa avventura lontana da casa, con voi è stato tutto più semplice.*

*Grazie a Peppe, coinquilino e amico di sempre, compagno di stanza e di sacrifici, con il tuo supporto si sono rivelati meno ardui da affrontare.*

*Un ringraziamento particolare ad una persona che ho avuto l'immensa fortuna di conoscere nel contesto universitario ma che è diventata molto di più, e che grazie alla sua empatia, umanità e semplicità ha saputo rendermi una persona migliore. Grazie Sara, è ancora più soddisfacente avercela fatta insieme.*

*Questo traguardo però sarebbe stato impossibile da raggiungere senza il supporto costante della persona che ha sempre creduto in me, soprattutto quando*

*ogni ostacolo sembrava insormontabile, grazie Mamma, questa vittoria è tutta nostra. Grazie a Vittorio per avermi insegnato che tramite impegno e dedizione non esiste montagna che non possa essere scalata, e grazie ad Enrica per avermi sempre fatto sentire a casa anche a migliaia di chilometri di distanza.*

# Bibliografia

- [1] *Definizione di cute.* 2020. URL:<http://www.treccani.it/vocabolario/cute/>
- [2] Micali, G., Fabbrocini, G., Monfrecola, G., Tosti, A., Veraldi, S. (2014). "Le basi della dermatologia".
- [3] *Struttura della cute.* 2020. URL:<https://www.istitutoterapiecorporee.it/>
- [4] *Archivio ISIC.* 2020. URL:<https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery>
- [5] *Nevi Displastici.* 2020. URL:<https://www.skincancer.org/international/nevi-displastici-atipici-e-rischio-di-melanoma/>
- [6] *Nevi Displastici.* 2020. URL:<https://www.ilmelanoma.com/nevi-nevo-displastico/>
- [7] Akkooi, A. J. v., Berking, C., Griewank, K. G., Gutzmer R., Hauschild, A., Roesch, A., Schaedendorf, D., Stang, A., Ugurel, S. (2018). "Melanoma" *Lancet*, 392: 971-84.
- [8] Alaibac, M., Rastrelli, M., Rossi, C.R., Rastrelli, M. (2014). "Melanoma: Epidemiology, Risk Factors, Pathogenesis, Diagnosis and Classification". 28: 1005-1012.
- [9] *Epidemiologia del melanoma.* 2020. URL:<http://gco.iarc.fr/>
- [10] Alfano, R., Agozzino, M., Argenziano, G., Ferrara, G., Longo, C., Piccolo, V., Russo, T. (2017) "Dermoscopy pathology correlation in melanoma". *Journal of Dermatology*, 44: 507-514.
- [11] *Trattamento del melanoma.* 2020. URL:<https://www.airc.it/cancro/informazioni-tumori/guida-ai-tumori/melanoma-cutaneo>
- [12] Jaimes, N., Marghoob, A, A., Usatine R. P. (2013) "Dermoscopy for the Family Physician" *American Family Physician*, 88-7.
- [13] *Carcinoma basocellulare.* 2020. URL:<https://www.skincancer.org/international/il-carcinoma-basocellulare/>

- [14] AIOM (Associazione Italiana di Oncologia Medica) (2018) "Linee guida: tumori cutanei non melanoma - carcinoma basocellulare".
- [15] Andreassi, L., Argenziano, G., Carli, P., Chimenti, S., De Giorgi, V., Di Stefano, A., Ferrara, G., Ferrari, A., Gasparini, S., Giannotti, B., Giovane, G.L., Lomuto, M., Mazzocchetti, G., Pellacani, G., Pellicano, R., Peris, K., Piccolo, D., Pizzichetta, M.A., Rubegni, P., Scalvenzi, M., Seidenari, S., Serresi, S., Stanganelli, I. "Linee Guida in Dermoscopia".
- [16] Campos-do-Carmo G., Ramos-e-Silva M. "Dermoscopy: basic concepts".
- [17] *Visita dermoscopia*. 2020. URL:<https://crystalcovedermatology.com/>
- [18] Agero, A. L. C., Benvenuto-Andrade, C., Dusza, S. W., Halpern, A. C., Marghoob, A. A., Rajadhyaksha, M., Scope, A. (2007) "Differences Between Polarized Light Dermoscopy and Immersion Contact Dermoscopy for the Evaluation of Skin Lesions." *Arch Dermatol* 143:329-338.
- [19] *Regola ABCD*. 2020. URL:[https://dermoscopedia.org/ABCD\\_rule](https://dermoscopedia.org/ABCD_rule)
- [20] *Diagnosi del melanoma*. 2020. URL:<https://www.humanitas.it/malattie/melanoma>
- [21] *Evoluzione del melanoma*. 2020. URL:<https://dpcj.org/index.php/dpc/article/view/452>
- [22] Nirmal, B. "Dermatoscopy: Physics and Principles".
- [23] *Dermoscopio HEINE*. 2020. URL:<https://www.heine.com/it>
- [24] *Trasformata di Hough*. 2020. URL:<https://www.mathworks.com/help/images/ref/imfindcircles.html#bti4sim>
- [25] *Matlab for DeepLearning*. 2020. URL:<https://www.mathworks.com/solutions/deep-learning>
- [26] *Transfer learning from pre-trained models*. 2020. URL:<https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>
- [27] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012) «Imagenet classification with deep convolutional neural networks». In: *Advances in neural information processing systems* pp. 1097–1105.
- [28] *CNN Architectures*. 2020. URL:<https://mc.ai/cnn-architectures-lenet-alexnet-vgg-googlenet-and-resnet/>
- [29] D'Arrigo, G., Provenzano, F., Torino, C., Tripepi, G., Zoccali, C. (2011). "I test diagnostici e l'analisi della curva ROC" *G ITAL NEFROL* 28 (6): 642-647.
- [30] L., V. d. Maaten, G., Hinton. (2008) . "Visualizing Data using t-SNE." *Journal of Machine Learning Research* 9, 2008, pp. 2579–2605.

- [31] Occlusion Sensitivity. 2020. URL:<https://blogs.mathworks.com/deep-learning/2017/12/15/network-visualization-based-on-occlusion-sensitivity/>