



**Politecnico  
di Torino**

**Politecnico di Torino**

Ingegneria Biomedica

A.a. 2024/2025

Sessione di laurea Dicembre 2025

# **Utility of Ultraviolet-Enhanced Fluorescence dermoscopy for demarcation of surgical margins in Mohs micrographic surgery**

Relatori:

Kristen M. Meiburger

Massimo Salvi

Candidati:

Alessandro Spina







# Indice

<b>Elenco delle tabelle</b>	V
<b>Elenco delle figure</b>	VII
<b>1 Introduzione</b>	1
1.1 I tumori cutanei . . . . .	1
1.2 La dermoscopia convenzionale . . . . .	2
1.3 Uso della fluorescenza UV . . . . .	3
1.4 La Chirurgia Micrografica di Mohs (MMS) . . . . .	4
1.5 Tecniche di segmentazione delle lesioni cutanee . . . . .	5
1.5.1 Approcci tradizionali di segmentazione . . . . .	6
1.5.2 Deep Learning per la segmentazione dermoscopia . . . . .	6
1.5.3 Limiti e prospettive future . . . . .	7
1.6 Obiettivo della tesi . . . . .	7
<b>2 Materiali e Metodi</b>	9
2.1 Datasets . . . . .	10
2.1.1 Dataset pubblico . . . . .	10
2.1.2 Dataset privato . . . . .	11
2.2 Preprocessing dei dati . . . . .	12
2.2.1 Data augmentation del dataset HAM10000 . . . . .	12
2.2.2 Normalizzazione geometrica . . . . .	13
2.2.3 Divisione del dataset in training, validation e test . . . . .	14
2.2.4 Preprocessing finale: biforcazione dei flussi di elaborazione . . . . .	15
2.3 Architettura del modello . . . . .	18
2.4 Procedura di addestramento . . . . .	20
2.4.1 Impostazione del training . . . . .	20
2.4.2 Funzione di perdita e ottimizzazione . . . . .	20
2.4.3 Strategia di validazione . . . . .	21
2.5 Creazione delle maschere manuali tramite applicazione dedicata . . . . .	21
2.6 Preprocessing del dataset clinico di Novara . . . . .	22

2.7	Post-processing delle predizioni . . . . .	24
2.8	Metriche di valutazione . . . . .	25
2.9	Analisi delle maschere manuali . . . . .	27
2.9.1	Calcolo dell'area e confronto dimensionale delle maschere . .	28
2.9.2	Valutazione della concordanza tra dermatologi . . . . .	28
2.9.3	Confronto tra maschere manuali e predizioni del modello . .	29
2.9.4	Analisi statistica delle maschere manuali e delle predizioni .	29
<b>3</b>	<b>Risultati e discussione</b>	<b>31</b>
3.1	Risultati del modello baseline . . . . .	31
3.2	Risultati dopo l'applicazione della color constancy . . . . .	33
3.3	Risultati dopo l'applicazione della color constancy e del data aug- mentation . . . . .	35
3.4	Risultati dopo l'applicazione del solo canale di luminanza $L^*$ . . . .	36
3.4.1	Analisi dei risultati . . . . .	37
3.5	Risultati dopo data augmentation ed estrazione del canale di lumi- nanza $L^*$ . . . . .	39
3.6	Effetto del post-processing . . . . .	40
3.7	Analisi delle maschere UV e NON-UV . . . . .	41
3.7.1	Esempi qualitativi di accordo e disaccordo . . . . .	43
3.7.2	Analisi statistica delle maschere . . . . .	46
3.7.3	Confronto non paired UV vs NON-UV . . . . .	46
3.7.4	Confronto paired tra dermatologi . . . . .	47
3.7.5	Confronto paired tra modello e dermatologi . . . . .	48
3.7.6	Discussione complessiva dei risultati . . . . .	49
3.8	Confronto con modello presente in letteratura . . . . .	50
3.8.1	Validazione del BA-Transformer sul dataset HAM10000 . . .	51
3.8.2	Confronto sulle immagini non-UV . . . . .	52
3.8.3	Confronto sulle immagini UV . . . . .	52
<b>4</b>	<b>Conclusioni e sviluppi futuri</b>	<b>54</b>
4.1	Sintesi dei risultati principali . . . . .	54
4.2	Contributi del lavoro svolto . . . . .	55
4.3	Limiti dello studio . . . . .	55
4.4	Possibili sviluppi futuri . . . . .	56
	<b>Bibliografia</b>	<b>57</b>

# Elenco delle tabelle

2.1	Distribuzione delle categorie diagnostiche nel dataset HAM10000. . .	10
2.2	Trasformazioni applicate durante la data augmentation dei BCC. . .	13
2.3	Suddivisione del dataset HAM10000 aumentato nei sottoinsiemi di training, validation e test. . . . .	15
3.1	Prestazioni del modello baseline sul dataset HAM10000 (media $\pm$ deviazione standard). . . . .	32
3.2	Prestazioni del modello baseline sulle immagini del dataset di Novara (media $\pm$ deviazione standard, dopo post-processing). . . . .	32
3.3	Prestazioni del modello dopo l'applicazione della color constancy sui tre domini di test (media $\pm$ deviazione standard, dopo post-processing). . .	33
3.4	Prestazioni del modello dopo color constancy + data augmentation sui tre domini (media $\pm$ deviazione standard, dopo post-processing). . .	36
3.5	Prestazioni del modello dopo preprocessing nel solo canale $L^*$ (media $\pm$ deviazione standard, dopo post-processing). . . . .	37
3.6	Prestazioni del modello dopo canale $L^*$ + data augmentation (media $\pm$ deviazione standard, dopo post-processing). . . . .	39
3.7	Confronto tra prestazioni prima e dopo il post-processing per i due modelli finali (media $\pm$ deviazione standard). . . . .	40
3.8	Confronto tra il numero di lesioni con area maggiore in UV e in NON-UV secondo i quattro dermatologi e secondo il modello. . . .	41
3.9	Riepilogo degli accordi tra dermatologi e modello. . . . .	42
3.10	Risultati del test non paired tra aree NON-UV e UV per ciascun operatore. Vengono riportati i p-value dei test di normalità, media e mediana delle due distribuzioni, la statistica del test di Mann-Whitney e il p-value associato. . . . .	47
3.11	P-value del test di Wilcoxon signed-rank tra le differenze UV-NON-UV per ciascuna coppia di dermatologi (triangolo superiore). . . . .	48
3.12	P-value del test di Wilcoxon signed-rank tra il modello e i quattro dermatologi. . . . .	48

3.13	Prestazioni riportate dagli autori del BA-Transformer sui dataset pubblici ISIC 2016 + PH2 e ISIC 2018 (valori tratti dal paper originale). . . . .	51
3.14	Confronto tra BA-Transformer e U-Net sul dataset HAM10000 (media $\pm$ deviazione standard). . . . .	51
3.15	Confronto tra BA-Transformer e U-Net proposta sulle immagini non-UV del dataset di Novara (media $\pm$ deviazione standard). . . .	52
3.16	Confronto tra BA-Transformer e U-Net proposta sulle immagini UV del dataset di Novara (media $\pm$ deviazione standard). . . . .	53

# Elenco delle figure

1.1	Esempio clinico di carcinoma basocellulare (BCC): (a) immagine dermatoscopica e (b) immagine clinica. . . . .	2
1.2	Confronto tra carcinoma basocellulare: (a) luce convenzionle e (b) luce UV. . . . .	3
1.3	Step della chirurgia micrografica di Mohs [9]. . . . .	4
2.1	Schema complessivo della pipeline sviluppata. Il flusso di lavoro si articola in quattro principali fasi: raccolta e analisi dei dataset, preprocessing, addestramento dei due modelli U-Net dedicati ai domini NON-UV e UV, post-processing delle predizioni e valutazione clinica. . . . .	9
2.2	Esempi visivi delle sette categorie diagnostiche del dataset <i>HAM10000</i> . . . . .	11
2.3	Esempi rappresentativi delle immagini del dataset privato di Novara. Riga superiore: immagini non-UV acquisite in dermoscopia convenzionale. Riga inferiore: immagini delle stesse lesioni acquisite in fluorescenza UV. . . . .	12
2.4	Schema complessivo del preprocessing applicato al dataset <i>HAM10000</i> . . . . .	14
2.5	Esempi dell'effetto della normalizzazione cromatica su due immagini dermoscopiche. . . . .	16
2.6	Esempio di conversione in spazio colore CIELUV. In alto: immagine originale in RGB e relativa rappresentazione in spazio LUV. In basso: i tre canali $L^*$ , $u^*$ , $v^*$ . La componente $L^*$ mostra in maniera più chiara la struttura della lesione, risultando la più adatta per il training del modello UV. . . . .	18
2.7	Schema dell'architettura U-Net utilizzata per la segmentazione [18]. . . . .	19
2.8	Immagine originale e corrispondenti segmentazioni manuali ottenute dai quattro dermatologi coinvolti nel progetto. . . . .	22
2.9	Esempio di preprocessing delle immagini dermoscopiche non-UV del dataset clinico. . . . .	23
2.10	Esempio di preprocessing delle immagini UV del dataset clinico. . . . .	23

2.11	Schema della pipeline di post-processing applicata alle predizioni dei modelli. Il processo comprende: rimozione delle componenti minori, selezione della componente principale, riempimento delle cavità interne e applicazione di operazioni morfologiche. . . . .	24
2.12	Esempio dell'effetto del post-processing sulla segmentazione. (a) Predizione grezza; (b) maschera finale binaria dopo post-processing; (c) e (d) mostrano i contorni delle due maschere sovrapposti all'immagine originale. . . . .	26
3.1	Esempi qualitativi dell'effetto della <i>color constancy</i> nelle immagini dermoscopiche non-UV. . . . .	34
3.2	Esempi dell'effetto della <i>color constancy</i> nelle immagini UV. . . . .	35
3.3	Confronto tra la segmentazione ottenuta con la sola <i>color constancy</i> (rosso) e quella prodotta dal modello basato sul canale di luminanza $L^*$ (verde) su quattro immagini UV. . . . .	38
3.4	Percentuali di accordo tra dermatologi e tra dermatologi e modello nella valutazione della maggiore estensione della lesione (UV vs NON-UV). . . . .	42
3.5	Esempi di lesioni in cui tutti i dermatologi e il modello concordano sulla modalità in cui l'area della lesione risulta più estesa. Ogni coppia (a-b) e (c-d) mostra rispettivamente la versione NON-UV e UV della stessa lesione. . . . .	43
3.6	Esempi di lesioni in cui i dermatologi concordano sulla modalità più estesa, mentre il modello fornisce una valutazione discordante. . . . .	44
3.7	Esempi di lesioni per le quali non vi è accordo tra i dermatologi riguardo alla modalità più estesa (UV o NON-UV). . . . .	45

# Capitolo 1

## Introduzione

### 1.1 I tumori cutanei

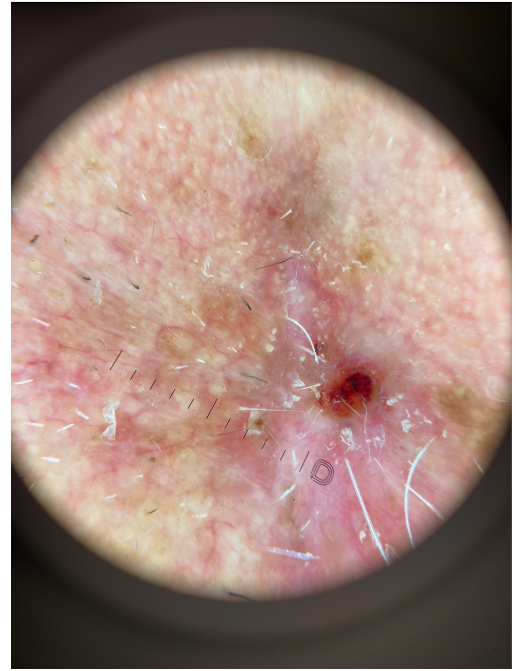
I tumori cutanei non melanoma (NMSC, non-melanoma skin cancer) rappresentano le neoplasie più comuni nella popolazione caucasica, con un'incidenza in costante aumento a livello globale. Tra essi, il carcinoma basocellulare (BCC) costituisce circa l'80% dei casi ed è considerato la forma più frequente di tumore cutaneo maligno [1, 2]. Sebbene il BCC presenti una bassa tendenza a dare metastasi, la sua crescita locale può determinare un'invasione progressiva dei tessuti circostanti, con conseguenti problematiche funzionali ed estetiche, soprattutto nelle aree ad alta rilevanza cosmetica come il volto [3]. La diagnosi precoce e la corretta identificazione dei margini tumorali sono quindi elementi fondamentali per garantire un trattamento efficace. Infatti, se da un lato l'escissione chirurgica rappresenta il trattamento di riferimento, dall'altro la variabilità clinica e istologica del BCC rende complessa la definizione dei margini di sicurezza, con un rischio non trascurabile di recidiva [2, 4]. Dal punto di vista clinico, i BCC possono manifestarsi con aspetti molto eterogenei, includendo forme nodulari, superficiali, infiltrative e sclerodermiformi. Questa eterogeneità complica ulteriormente la valutazione pre-operatoria e rende necessarie tecniche diagnostiche complementari per una corretta pianificazione chirurgica [5]. Nella figura 1.1a è mostrato un esempio di BCC.

In questo contesto, negli ultimi anni, lo sviluppo di metodiche di imaging non invasive ha contribuito a migliorare la capacità di individuare precocemente il tumore e di delimitarne i margini. In particolare, la dermoscopia e le sue evoluzioni tecnologiche hanno assunto un ruolo centrale nella gestione del BCC, integrando l'esame clinico tradizionale e migliorando la sensibilità diagnostica [1, 4].





(a) Immagine clinica.



(b) Immagine dermatoscopica.

**Figura 1.1:** Esempio clinico di carcinoma basocellulare (BCC): (a) immagine dermatoscopica e (b) immagine clinica.

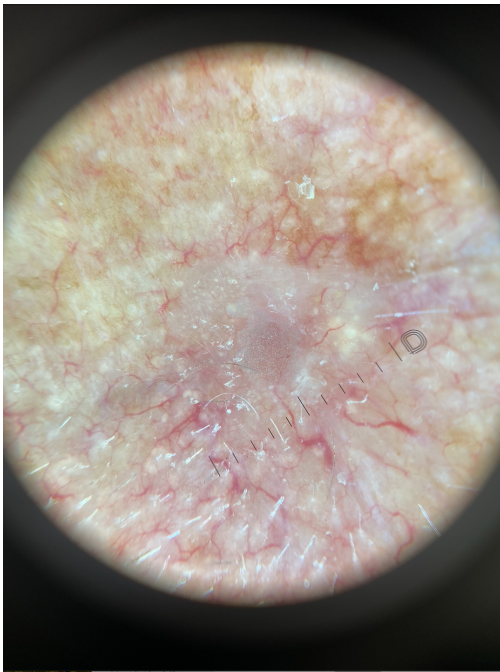
## 1.2 La dermoscopia convenzionale

La dermoscopia è una tecnica diagnostica non invasiva che consente di visualizzare strutture morfologiche della cute non distinguibili a occhio nudo, mediante l'impiego di luce polarizzata o non polarizzata in combinazione con un sistema ottico di ingrandimento [4]. Questa metodica ha progressivamente rivoluzionato la pratica dermatologica, migliorando la sensibilità e la specificità nella diagnosi dei tumori cutanei rispetto al semplice esame clinico visivo [2]. Nel caso del carcinoma basocellulare (BCC), la dermoscopia consente di identificare pattern caratteristici, quali la presenza di arborizing vessels, aree translucide, strutture pigmentate e ulcerazioni. L'analisi di tali elementi permette non solo una diagnosi più accurata, ma anche una migliore definizione preoperatoria dei margini, risultando particolarmente utile nella pianificazione chirurgica [2, 4]. Nonostante i numerosi vantaggi, la dermoscopia presenta alcune limitazioni. In primo luogo, vi è una significativa variabilità inter-osservatore: l'esperienza del dermatologo influenza in modo determinante l'interpretazione delle immagini [4]. Inoltre, la capacità di delimitare i margini tumorali può risultare ridotta in varianti istologiche infiltrative o sclerodermiformi, dove i confini tra tessuto sano e neoplastico non

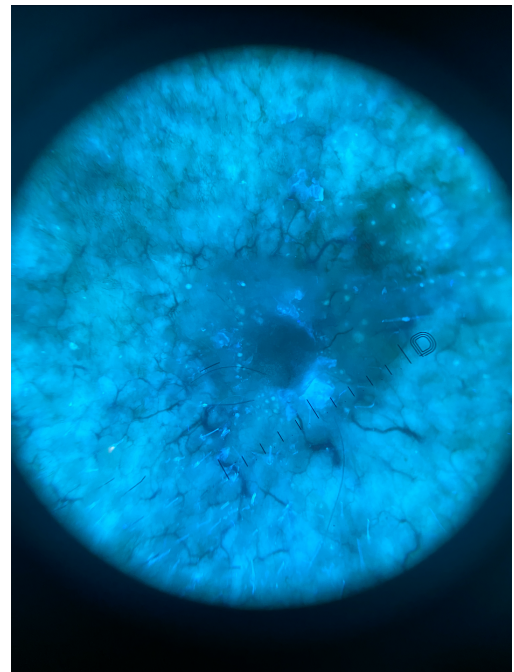
sono chiaramente distinguibili [2]. Per queste ragioni, la dermoscopia convenzionale è oggi considerata uno strumento fondamentale, ma non sempre sufficiente da sola per garantire la completa definizione dei margini chirurgici. Da qui nasce l'interesse crescente verso nuove metodiche di imaging avanzato, come l'imaging a fluorescenza ultravioletta, che possano integrare la dermoscopia tradizionale e incrementarne l'efficacia diagnostica [6].

### 1.3 Uso della fluorescenza UV

L'uso della fluorescenza UV ha mostrato grande potenziale in ambito dermatologico. In particolare, la fluorescenza delle protoporfirine endogene, che tendono ad accumularsi nelle cellule tumorali, consente di delineare con maggiore chiarezza i margini dei tumori cutanei, incluso il carcinoma basocellulare [6] (figura 1.2).



(a) Aspetto clinico del BCC in luce convenzionale.



(b) Aspetto dermoscopico del BCC in luce UV.

**Figura 1.2:** Confronto tra carcinoma basocellulare: (a) luce convenzionale e (b) luce UV.

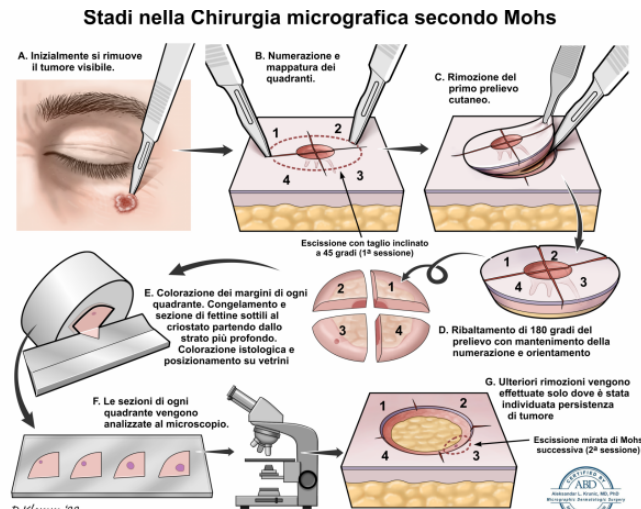
Oltre all'oncologia cutanea, l'imaging UV è stato impiegato nella diagnosi di infezioni fungine, nella valutazione di disordini pigmentari come la vitiligine, e nell'analisi della risposta a trattamenti dermatologici, ad esempio contro la

dermatite seborroica [6]. Negli ultimi anni si è assistito inoltre a una progressiva miniaturizzazione delle tecnologie, con studi che esplorano l'uso di smartphone dotati di sistemi UV per applicazioni diagnostiche rapide e a basso costo [6]. Nonostante i promettenti risultati, uno dei principali limiti dell'imaging UV risiede nella scarsità di dataset pubblici dedicati, che rende difficile lo sviluppo e la validazione di algoritmi automatici di analisi [2]. Per ovviare a questo problema, Kojima e collaboratori hanno proposto UV-photoNet, un modello basato su reti generative avversarie condizionate (CGAN) in grado di generare immagini UV sintetiche a partire da fotografie cutanee standard [7]. In questo studio, condotto su un dataset di 184 individui, le immagini UV sintetiche si sono dimostrate altamente correlate a quelle reali nell'evidenziare macchie pigmentarie, suggerendo un possibile utilizzo di questa tecnologia per lo screening e per la prevenzione del cancro della pelle anche senza dispositivi specialistici [7].

## 1.4 La Chirurgia Micrografica di Mohs

La chirurgia micrografica di Mohs (MMS) rappresenta il gold standard per il trattamento del carcinoma basocellulare ad alto rischio e di altri tumori cutanei localmente invasivi. Introdotta da Frederic Mohs negli anni '30, questa tecnica si basa su una rimozione sequenziale del tumore con analisi istologica intraoperatoria dei margini, consentendo un controllo completo delle cellule residue e il massimo risparmio di tessuto sano [8].

La procedura si articola in più fasi iterative (figura 1.3):



**Figura 1.3:** Step della chirurgia micrografica di Mohs [9].

1. Preparazione del paziente, con marcatura clinica della lesione e anestesia locale.
2. Escissione dello strato superficiale del tumore, comprendente sia la massa tumorale visibile che una piccola porzione di tessuto circostante.
3. Preparazione del campione istologico, mediante congelamento, sezionamento e colorazione del tessuto asportato.
4. Analisi microscopica dei margini, in cui il patologo o il chirurgo stesso osserva l'intero perimetro e la profondità della sezione.
5. Escissione sequenziale, che prevede la rimozione mirata di ulteriore tessuto solo laddove siano presenti cellule neoplastiche residue, ripetendo la procedura fino a ottenere margini completamente liberi [8].

Questa metodica offre diversi vantaggi rispetto alla chirurgia tradizionale:

- Tasso di guarigione superiore al 99 % per BCC primari e tra il 94–97% per BCC recidivi.
- Massima conservazione di tessuto sano, fondamentale nelle aree cosmeticamente o funzionalmente critiche, come volto, orecchie e regioni periorbitali.
- Riduzione del rischio di recidiva, grazie al controllo istologico completo dei margini [8].

Tuttavia, la MMS presenta anche alcuni limiti: si tratta di una procedura complessa, che richiede tempo e risorse significative, oltre alla presenza di operatori altamente specializzati in dermatologia chirurgica e istopatologia [8]. Per questo motivo, l'integrazione di metodiche diagnostiche non invasive, come la dermoscopia convenzionale e, più recentemente, la fluorescenza UV, è di grande interesse per ottimizzare la fase pre-operatoria e rendere la procedura ancora più efficiente.

## 1.5 Tecniche di segmentazione delle lesioni cutanee

La segmentazione delle immagini dermoscopiche rappresenta un passaggio cruciale nell'analisi computerizzata dei tumori cutanei. Essa consiste nell'individuazione e separazione della lesione dal tessuto circostante, con l'obiettivo di fornire una base quantitativa e riproducibile per diagnosi, follow-up e pianificazione chirurgica. Nel caso del carcinoma basocellulare, la corretta delimitazione dei margini assume un ruolo centrale per guidare la chirurgia micrografica di Mohs e ridurre al minimo il sacrificio di tessuto sano [2, 1].

### 1.5.1 Approcci tradizionali di segmentazione

Gli approcci classici si suddividono in diverse categorie:

- **Edge & Region-based methods:** si basano inizialmente sul rilevamento dei bordi, seguito dall'aggregazione dei pixel contigui con caratteristiche simili per formare regioni connesse. Questi metodi risultano rapidi ed efficaci quando applicati a immagini ad alto contrasto, ma soffrono di elevata sensibilità al rumore [1, 4] .
- **Threshold-based methods:** utilizzano soglie di intensità, spesso applicate al canale blu delle immagini dermoscopiche, per distinguere la lesione dal background. Sono semplici da implementare e computazionalmente molto veloci, ma scarsamente robusti in presenza di variabilità di illuminazione o artefatti [4].
- **Intelligence-based methods:** comprendono approcci ispirati all'intelligenza artificiale non profonda, come reti neurali artificiali (ANN), algoritmi genetici e clustering Fuzzy C-Means (FCM). Questi metodi offrono una maggiore capacità di adattamento e di modellazione della complessità delle lesioni, pur richiedendo conoscenze preliminari sul numero e la distribuzione dei cluster [1].

Sebbene queste metodiche abbiano rappresentato un punto di partenza fondamentale, le loro performance risultano oggi inferiori rispetto ai modelli basati sul deep learning, soprattutto in scenari clinici complessi.

### 1.5.2 Deep Learning per la segmentazione dermoscopia

Negli ultimi anni, l'applicazione di reti neurali profonde ha radicalmente trasformato il campo della segmentazione delle lesioni cutanee. Tra le architetture più utilizzate spiccano le reti convoluzionali (CNN) e, in particolare, la U-Net, che si è affermata come standard di riferimento in ambito biomedico grazie alla sua capacità di sfruttare il data augmentation per migliorare le prestazioni anche con dataset di dimensioni limitate [2].

Diversi studi hanno proposto varianti e approcci innovativi:

- **Yu et al.** hanno sviluppato un modello molto profondo basato su residual networks, articolato in due fasi e composto da oltre 50 strati, ottenendo ottime prestazioni ma con elevati costi computazionali [10].
- **Bi et al.** hanno introdotto una CNN multi-stage in grado di apprendere prima le caratteristiche globali della lesione e successivamente i dettagli dei margini [11].

- **Yuan et al.** hanno proposto una rete DCNN a 19 strati, ottimizzata tramite una loss function basata sulla distanza di Jaccard, evidenziando criticità nelle immagini a basso contrasto [12].
- **Goyal e Yap** hanno esplorato approcci multi-classe per distinguere differenti tipologie di lesioni cutanee e modelli ensemble che combinano più reti, migliorando sensibilità e specificità [13].

Dal punto di vista delle performance, i modelli di deep learning raggiungono risultati molto promettenti: valori tipici di Jaccard Index (IoU) tra il 70% e l'85%, Dice Coefficient superiori al 90% e accuratezza oltre il 95% [2]. Tuttavia, restano sfide aperte legate alla disponibilità di dataset pubblici ampi e ben annotati: la variabilità inter-osservatore nella definizione manuale dei margini continua a rappresentare un ostacolo significativo alla creazione di ground truth affidabili [2, 1].

### 1.5.3 Limiti e prospettive future

L'attuale mancanza di dataset specifici per immagini acquisite con luce ultravioletta costituisce un vincolo importante allo sviluppo di modelli dedicati. Alcuni studi preliminari hanno proposto l'uso di reti generative avversarie condizionate (CGAN) per produrre immagini UV sintetiche a partire da fotografie standard, aprendo la strada a un futuro in cui dataset più ricchi potranno supportare lo sviluppo di algoritmi altamente specializzati [7]. In prospettiva, l'integrazione di deep learning e metodiche tradizionali di image processing potrebbe fornire modelli ibridi in grado di coniugare robustezza, interpretabilità e generalizzazione, rappresentando uno strumento di grande utilità nella pratica clinica quotidiana [1].

## 1.6 Obiettivo della tesi

Il razionale di questa tesi si fonda sull'ipotesi che l'integrazione della dermoscopia a fluorescenza UV con algoritmi di segmentazione automatica possa migliorare la precisione nella definizione dei margini tumorali nei pazienti sottoposti a chirurgia micrografica di Mohs. Questo approccio ha il potenziale di ottimizzare il bilancio tra radicalità oncologica e conservazione tissutale, con importanti ricadute in termini di efficacia terapeutica, riduzione delle recidive e miglioramento dell'outcome estetico-funzionale. In particolare, il lavoro si propone di:

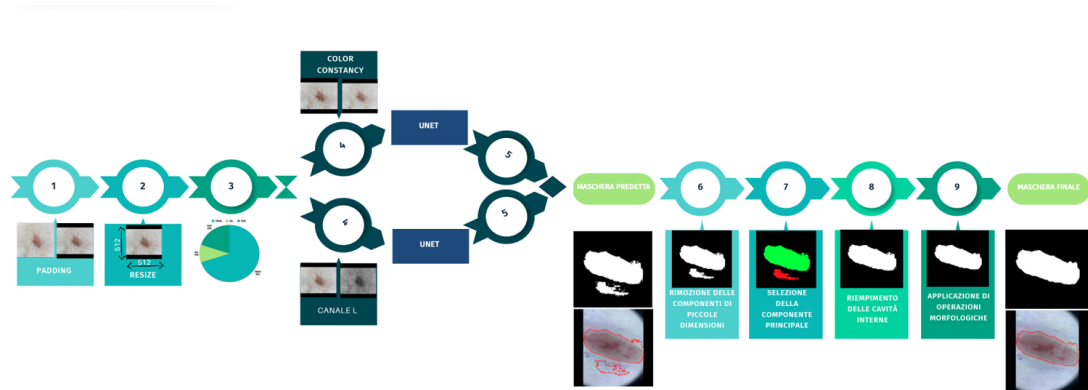
- Analizzare le potenzialità diagnostiche della fluorescenza UV nel migliorare la visualizzazione dei margini tumorali, con l'obiettivo di rendere più efficiente la procedura di Mohs e ridurre la necessità di escissioni sequenziali.

- Valutare le prestazioni del modello proposto mediante metriche di accuratezza consolidate, come Jaccard Index (IoU), Dice Coefficient, sensitivity e specificity, confrontando i risultati con lo stato dell'arte [2, 4, 14].
- Esplorare le implicazioni cliniche di questo approccio, verificando se l'integrazione di dermoscopia UV e analisi automatizzata possa contribuire a un migliore bilancio tra radicalità oncologica e conservazione di tessuto sano.

La prospettiva di questo lavoro è quella di contribuire allo sviluppo di strumenti rapidi, non invasivi e automatizzati, in grado di supportare il dermatologo-chirurgo nella pianificazione e nell'esecuzione della chirurgia di Mohs, migliorando l'accuratezza diagnostica e riducendo i tempi operatori.

## Capitolo 2

# Materiali e Metodi



**Figura 2.1:** Schema complessivo della pipeline sviluppata. Il flusso di lavoro si articola in quattro principali fasi: raccolta e analisi dei dataset, preprocessing, addestramento dei due modelli U-Net dedicati ai domini NON-UV e UV, post-processing delle predizioni e valutazione clinica.

La Figura 2.1 riassume le principali fasi che compongono il workflow sperimentale adottato in questo lavoro. Nei paragrafi successivi ciascun blocco della pipeline verrà descritto nel dettaglio, partendo dalla composizione dei dataset fino ad arrivare alla valutazione delle predizioni e al confronto con l'interpretazione clinica dei dermatologi.



## 2.1 Datasets

### 2.1.1 Dataset pubblico

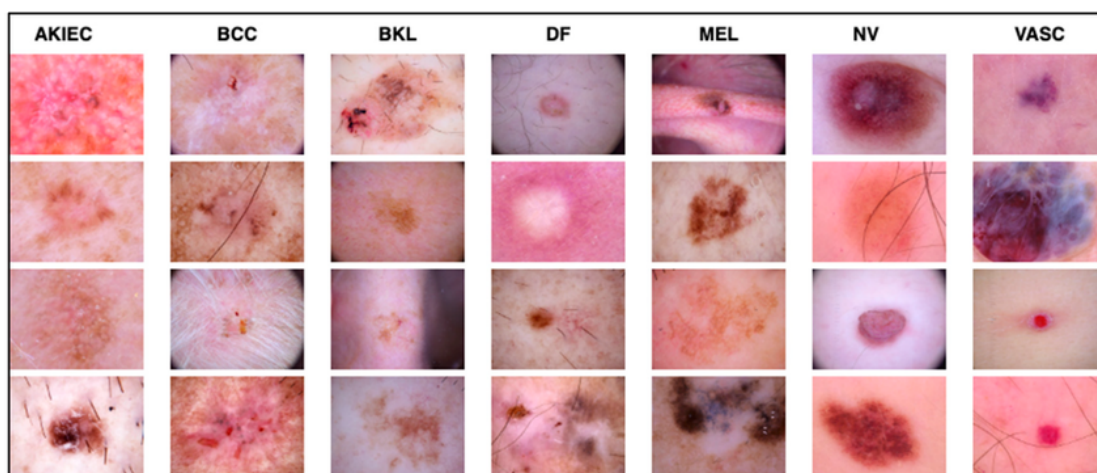
Per l'addestramento del modello è stato utilizzato il dataset pubblico *HAM10000* (Human Against Machine with 10000 training images) [15], costituito da 10 015 immagini dermoscopiche accompagnate dalle rispettive maschere di segmentazione manuale. Il dataset raccoglie immagini provenienti da differenti centri dermatologici e include una varietà di lesioni cutanee di natura sia benigna che maligna.

Le immagini sono suddivise in sette categorie diagnostiche come riportato in Tabella 2.1

**Tabella 2.1:** Distribuzione delle categorie diagnostiche nel dataset HAM10000.

Etichetta	Descrizione	Percentuale approssimata
nv	Nevi melanocitari	~ 67%
mel	Melanoma	~ 11%
bkl	Cheratosi benigne	~ 11%
bcc	Carcinoma basocellulare	~ 5%
akiec	Cheratosi attinica	~ 3%
vasc	Lesioni vascolari	~ 2%
df	Dermatofibroma	~ 1%

La Figura 2.2 mostra alcuni esempi rappresentativi delle sette categorie diagnostiche presenti nel dataset *HAM10000*. L'elevata variabilità delle immagini — in termini di colore, texture, dimensione e condizioni di acquisizione — riflette la natura multicentrica del dataset e rende necessario l'impiego di tecniche di preprocessing dedicate, descritte nelle sezioni successive.



**Figura 2.2:** Esempi visivi delle sette categorie diagnostiche del dataset *HAM10000*.

Le immagini sono fornite in formato `.jpg`, mentre le maschere di segmentazione manuale sono disponibili in formato `.png`. Ai fini della segmentazione, tutte le lesioni sono state considerate come un'unica classe positiva rispetto allo sfondo, adottando un'impostazione binaria lesione vs. non-lesione.

### 2.1.2 Dataset privato

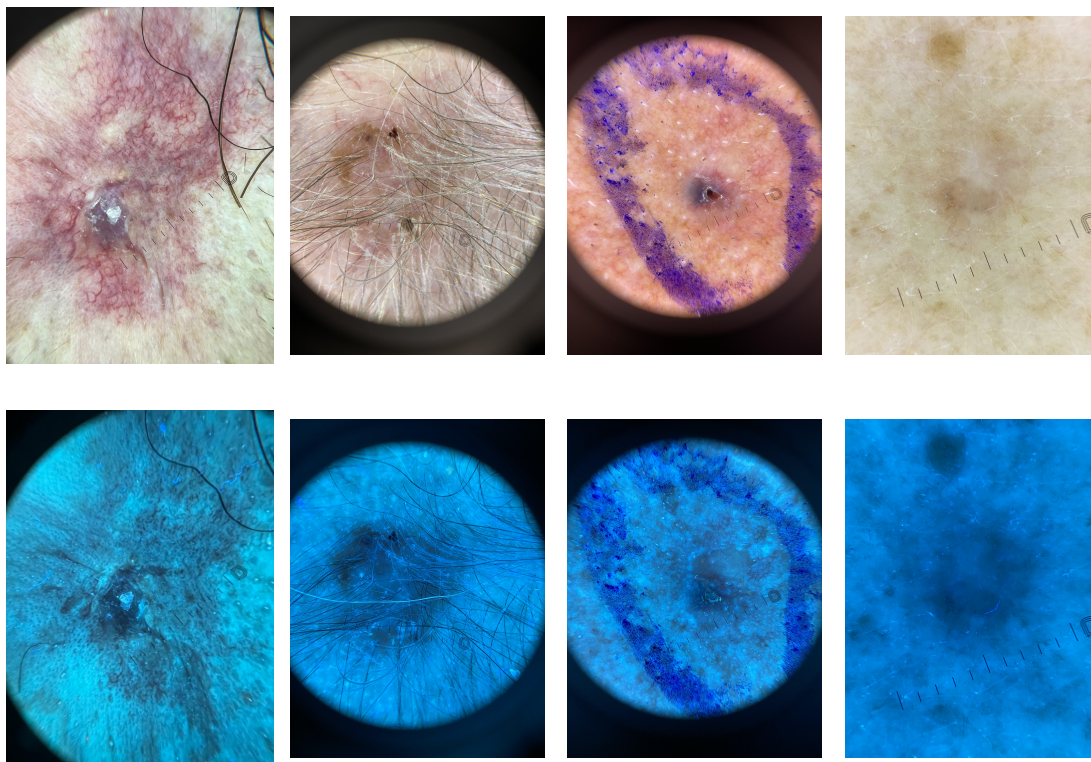
Il secondo dataset utilizzato per la validazione del modello è stato fornito dai dermatologi dell'Ospedale di Novara e consiste in 168 immagini di carcinoma basocellulare (BCC) acquisite in condizioni cliniche reali. Il dataset è composto da:

- 84 immagini acquisite in dermoscopia convenzionale (non-UV),
- 84 immagini acquisite mediante imaging in fluorescenza UV.

Per ciascun paziente sono disponibili entrambe le modalità di acquisizione, formando 84 coppie di immagini raffiguranti la stessa lesione.

Le immagini presentano eterogeneità di formato (`.jpg` e `.heic`), illuminazione, presenza della bordatura del dermatoscopio e vari artefatti clinici (peli, marcature cutanee e riflessi). Esse sono state acquisite tramite il dermatoscopio **DermLite DL5**, un dispositivo portatile dotato di lente da 32 mm con ingrandimento 10×, illuminazione LED multimodale (luce bianca, modalità PigmentBoost® e UV a 365 nm) e polarizzazione variabile[16].

Questa variabilità nei dati riflette la natura non standardizzata del contesto clinico e rende necessario un preprocessing specifico per ciascuna modalità di acquisizione.



**Figura 2.3:** Esempi rappresentativi delle immagini del dataset privato di Novara. Riga superiore: immagini non-UV acquisite in dermoscopia convenzionale. Riga inferiore: immagini delle stesse lesioni acquisite in fluorescenza UV.

Al momento della consegna il dataset non disponeva di maschere di segmentazione; tali annotazioni sono state generate nell’ambito di questo lavoro tramite un’applicazione sviluppata per supportare i dermatologi nella creazione assistita delle maschere (Sezione 2.5).

Questo dataset rappresenta il dominio di interesse principale del presente studio ed è stato utilizzato per valutare la capacità del modello di generalizzare su immagini cliniche reali altamente variabili.

## 2.2 Preprocessing dei dati

### 2.2.1 Data augmentation del dataset HAM10000

Come evidenziato nella Sezione 2.1.1, il dataset pubblico HAM10000 utilizzato per l’addestramento presenta una marcata disomogeneità nella distribuzione delle classi

diagnostiche: le immagini di carcinoma basocellulare (BCC) costituiscono infatti solo una piccola frazione del totale.

Per ridurre tale sbilanciamento e migliorare la capacità di generalizzazione del modello, è stata implementata una procedura di *data augmentation* mirata esclusivamente alla classe BCC. Per identificare in modo univoco tutte le immagini appartenenti a tale classe è stato utilizzato il file `.csv` fornito insieme al dataset originale, contenente la diagnosi associata a ciascun identificativo di immagine.

### Trasformazioni applicate

Le trasformazioni sono state realizzate con la libreria `Albumentations` (v. 2.0.8) in Python e applicate sia alle immagini di BCC sia alle corrispondenti maschere di segmentazione, garantendo la coerenza spaziale tra i due domini. Le operazioni selezionate, di tipo puramente geometrico, sono elencate nella Tabella 2.2.

**Tabella 2.2:** Trasformazioni applicate durante la data augmentation dei BCC.

Tipo di trasformazione	Parametri principali
Flip orizzontale	—
Flip verticale	—
Rotazione casuale	$\pm 20^\circ$
Zoom casuale	$0.9\times - 1.1\times$

### Risultato della procedura

Per ogni immagine BCC sono state generate cinque versioni (originale, flip orizzontale, flip verticale, rotazione, zoom), portando il numero totale di immagini da 514 a 2570. Le maschere binarie sono state trasformate in maniera analoga per mantenere la corrispondenza pixel-wise.

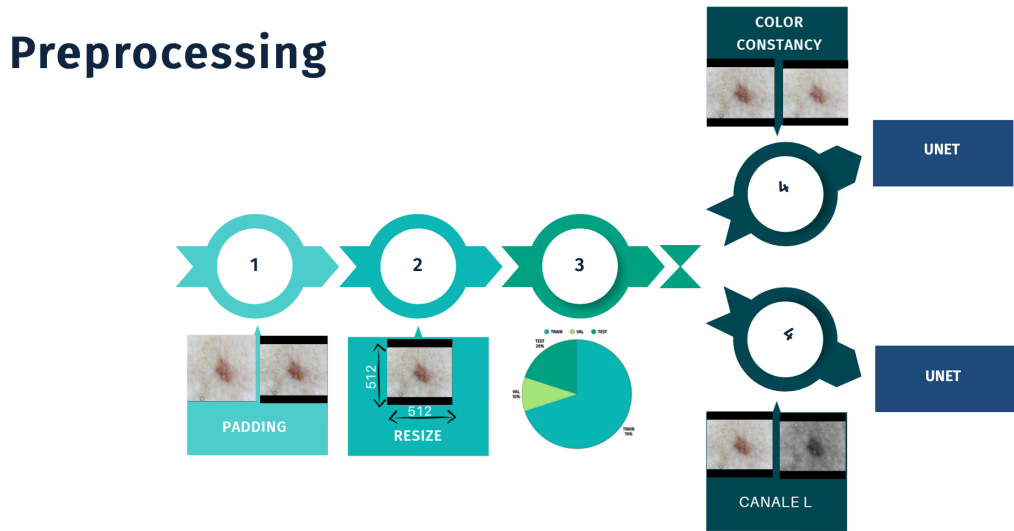
Le immagini aumentate sono state successivamente reintegrate nel dataset di addestramento, incrementando la varietà morfologica e migliorando la rappresentazione statistica dei BCC. Questa procedura ha permesso di ottenere un addestramento più stabile e prestazioni superiori nella segmentazione, come discusso nel Capitolo 3.

### 2.2.2 Normalizzazione geometrica

Le immagini del dataset *HAM10000* presentano risoluzioni e rapporti d'aspetto eterogenei. Per garantire un input uniforme alla rete neurale, è stato applicato un padding simmetrico volto a ottenere un formato quadrato senza distorsione del contenuto anatomico. Successivamente, tutte le immagini sono state ridimensionate

a  $512 \times 512$  pixel. Le maschere di segmentazione associate sono state binarizzate mediante soglia a 128 e salvate in formato `.png` con valori  $\{0, 255\}$ , corrispondenti rispettivamente allo sfondo e alla regione lesionata. Tale rappresentazione è stata mantenuta anche in fase di validazione per assicurare coerenza tra ground truth e predizioni generate dal modello.

La Figura 2.4 mostra una panoramica del workflow completo di preprocessing adottato in questo lavoro. Dopo le fasi comuni di normalizzazione geometrica, il processo prosegue con due percorsi distinti, che verranno descritti nel dettaglio nelle sezioni successive (normalizzazione cromatica tramite GCC-GAN e conversione nel canale  $L^*$  dello spazio CIELUV).



**Figura 2.4:** Schema complessivo del preprocessing applicato al dataset *HAM10000*.

### 2.2.3 Divisione del dataset in training, validation e test

Al termine della normalizzazione geometrica, il dataset *HAM10000* è stato suddiviso in tre sottoinsiemi secondo la proporzione 70/10/20, rispettivamente destinati alle fasi di training, validazione e test:

**Tabella 2.3:** Suddivisione del dataset HAM10000 aumentato nei sottoinsiemi di training, validation e test.

Set	Percentuale	Numero di immagini
Training	70%	8 449
Validation	10%	1 207
Test	20%	2 415

Particolare attenzione è stata posta nel garantire l'assenza di *data leakage* tra i diversi insiemi. In particolare, poiché il dataset include versioni aumentate della stessa immagine originale, tutte le copie di una medesima lesione (immagine originale e corrispondenti varianti generate tramite *data augmentation*) sono state allocate all'interno dello stesso sottoinsieme. In questo modo si evita che il modello venga esposto durante l'addestramento a campioni derivati dalla stessa immagine presenti anche nei set di validazione o di test, preservando così l'indipendenza statistica tra le fasi di training e valutazione.

## 2.2.4 Preprocessing finale: biforcazione dei flussi di elaborazione

A valle della divisione del dataset, il flusso di preprocessing si divide in due percorsi distinti, corrispondenti ai due domini applicativi del modello:

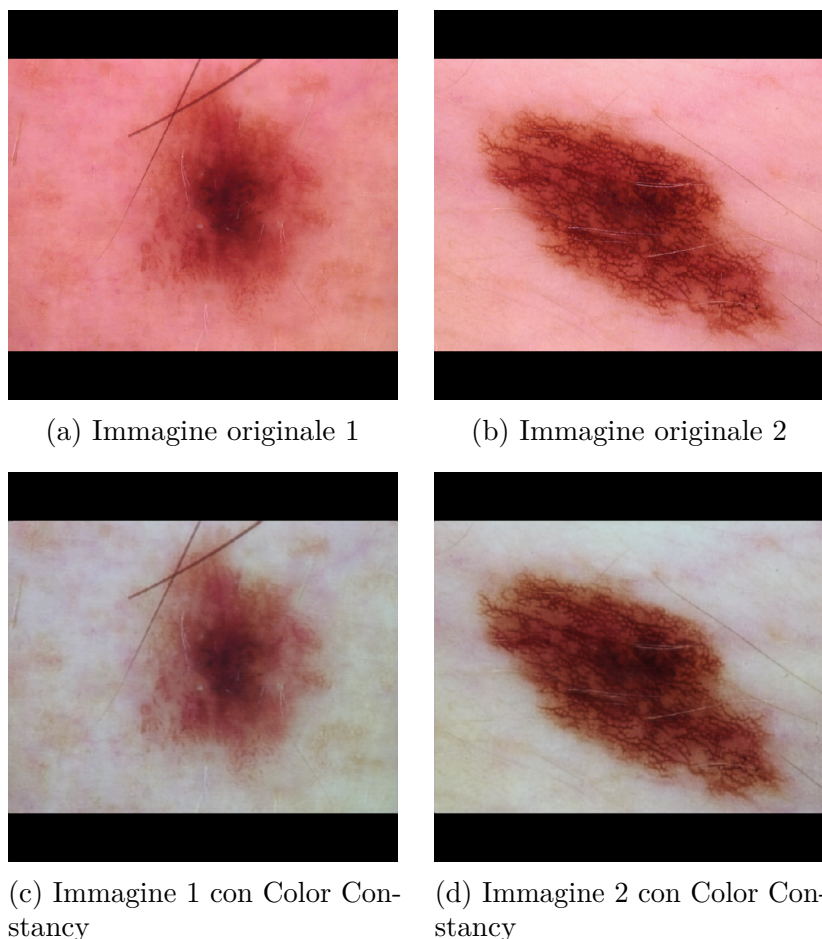
1. **preprocessing cromatico mediante GCC-GAN**, utilizzato per addestrare il modello destinato alla segmentazione delle immagini dermoscopiche convenzionali;
2. **estrazione della componente di luminanza  $L^*$  dallo spazio colore CIELUV**, utilizzata per addestrare il modello dedicato alle immagini acquisite in fluorescenza UV.

La scelta di adottare due flussi separati nasce dalla marcata differenza tra i due domini di acquisizione: le immagini non-UV sono fortemente influenzate dalle condizioni di illuminazione e traggono beneficio dalla normalizzazione cromatica, mentre le immagini UV evidenziano pattern di fluorescenza la cui informazione principale è veicolata dalla componente di luminanza.

Nelle sottosezioni seguenti vengono descritti nel dettaglio i due approcci di preprocessing.

### Normalizzazione cromatica tramite GCC-GAN

Oltre alla normalizzazione geometrica, è stata applicata una procedura di normalizzazione cromatica tramite il metodo *GCC-GAN* (Generative Color Constancy), recentemente proposto in letteratura per ridurre la variabilità cromatica nelle immagini dermatologiche [17].



**Figura 2.5:** Esempi dell'effetto della normalizzazione cromatica su due immagini dermoscopiche.

La color constancy affronta il problema delle differenti condizioni di illuminazione, delle variazioni di esposizione e del bilanciamento del bianco, che possono alterare la percezione del colore e influenzare negativamente l'addestramento dei modelli di segmentazione.

A titolo esemplificativo, la Figura 2.5 mostra una coppia di immagini prima e dopo l'applicazione della correzione cromatica, evidenziando la riduzione delle



dominanti di colore e la maggiore uniformità dell'illuminazione.

Il metodo si basa sull'impiego di una rete generativa avversaria (GAN) che apprende una trasformazione *immagine-originale*  $\rightarrow$  *immagine-normalizzata*, preservando il contenuto strutturale ma standardizzando la componente cromatica. All'interno dell'architettura, il generatore produce una versione cromaticamente normalizzata dell'immagine in ingresso, mentre il discriminatore valuta se il risultato appartenga alla distribuzione cromatica attesa. L'ottimizzazione avviene tramite competizione tra le due reti, permettendo di ottenere una correzione del colore stabile e coerente con le immagini di riferimento.

Nel presente lavoro è stato utilizzato il modello *GCC-GAN pretrained*, come rilasciato dagli autori nell'ambito dello studio originale, senza ulteriore fase di riaddestramento.

Le immagini normalizzate dei vari sets sono state quindi utilizzate come input per l'addestramento del modello di segmentazione finale.

### Estrazione del canale $L^*$ in spazio CIELUV

Per simulare in maniera più realistica il dominio fotometrico delle immagini acquisite in fluorescenza UV, è stato adottato lo spazio colore *CIELUV*, una rappresentazione percettivamente uniforme definita dalla Commission Internationale de l'Éclairage (CIE). A differenza dello spazio RGB, che rispecchia la risposta dei sensori della camera, il modello CIELUV separa la componente di *luminanza* dalle componenti di *crominanza*, permettendo di analizzare l'informazione strutturale dell'immagine in modo più indipendente dal colore.

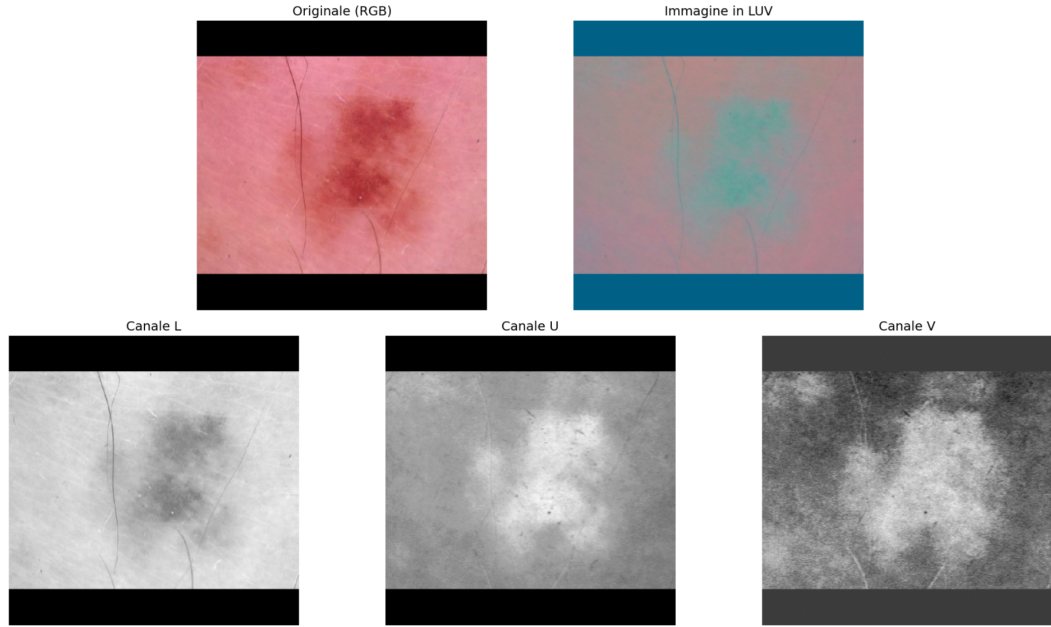
Lo spazio CIELUV è composto da tre canali:

- $L^*$  (**Luminanza**) — rappresenta la brillantezza percepita e contiene la maggior parte delle informazioni strutturali della lesione (bordi, texture, gradienti di intensità);
- $u^*$  — codifica una delle due componenti della crominanza, associata principalmente alle variazioni lungo l'asse blu-giallo;
- $v^*$  — codifica la seconda componente di crominanza, correlata alle variazioni lungo l'asse rosso-verde.

Nel contesto delle immagini dermatologiche UV, la fluorescenza tende a saturare la componente cromatica, mentre la struttura della lesione rimane ben rappresentata nella sola luminanza. Per questo motivo, in questo lavoro è stato utilizzato unicamente il canale  $L^*$ , più stabile e maggiormente informativo per la segmentazione nel dominio UV.



La Figura 2.6 mostra un esempio di conversione di un'immagine dermoscópica dallo spazio RGB allo spazio CIELUV e la corrispondente decomposizione nei tre canali  $L^*$ ,  $u^*$ ,  $v^*$ .



**Figura 2.6:** Esempio di conversione in spazio colore CIELUV. In alto: immagine originale in RGB e relativa rappresentazione in spazio LUV. In basso: i tre canali  $L^*$ ,  $u^*$ ,  $v^*$ . La componente  $L^*$  mostra in maniera più chiara la struttura della lesione, risultando la più adatta per il training del modello UV.

Poiché l'architettura U-Net richiede un input a tre canali, la mappa di luminanza estratta è stata replicata su tre livelli ( $L^*$ ,  $L^*$ ,  $L^*$ ), così da mantenere la compatibilità con il modello senza introdurre informazione cromatica non utile nel dominio UV. Questa trasformazione è stata applicata in modo uniforme a tutte le immagini del dataset preprocessato—training, validation e test—garantendo coerenza tra i diversi sottoinsiemi e correttezza nella fase di valutazione del modello.

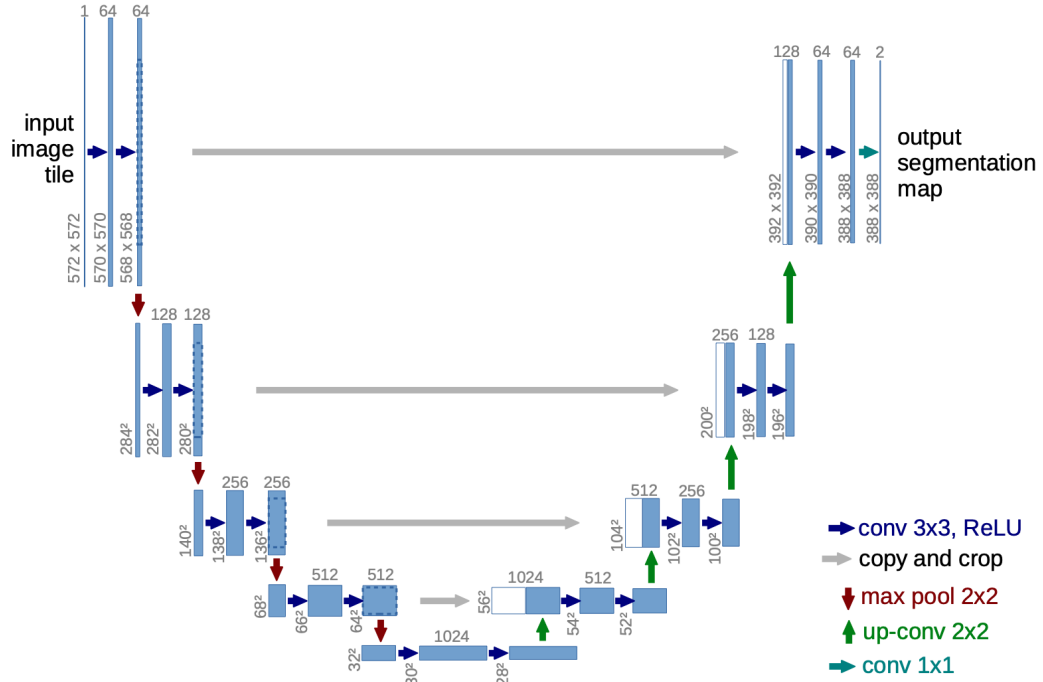
## 2.3 Architettura del modello

### Struttura generale

Il modello di segmentazione adottato in questo lavoro è basato su una rete neurale convoluzionale di tipo U-Net, scelta ampiamente consolidata nel campo dell'analisi

di immagini biomediche grazie alla sua capacità di combinare informazioni contestuali globali con dettagli locali. L'architettura segue uno schema encoder-decoder simmetrico, in cui la fase discendente estrae progressivamente caratteristiche di livello crescente, mentre la fase ascendente ricostruisce la mappa di segmentazione alla risoluzione originale.

A titolo illustrativo, la Figura 2.7 mostra una rappresentazione schematica della U-Net impiegata in questo lavoro



**Figura 2.7:** Schema dell'architettura U-Net utilizzata per la segmentazione [18].

## Encoder-Decoder e skip connections

Nel percorso di encoding l'immagine attraversa una sequenza di blocchi convoluzionali (*DoubleConv*) composti da due convoluzioni 3x3 seguite rispettivamente da Batch Normalization e funzione di attivazione ReLU. Dopo ogni blocco viene applicata un'operazione di sottocampionamento, che consente alla rete di apprendere rappresentazioni di ordine superiore.

La fase di decoding esegue l'operazione inversa, ripristinando progressivamente la risoluzione spaziale. A ogni livello, le feature map ricostruite vengono concatenate con le corrispondenti mappe dell'encoder tramite *skip connections*, che hanno il ruolo di preservare l'informazione morfologica fine (ad esempio i contorni della lesione) che andrebbe altrimenti persa durante il downsampling.

## Testa di segmentazione

Completata la fase di decoding, un livello convoluzionale finale a kernel  $1 \times 1$  produce una mappa a singolo canale, corrispondente alla probabilità pixel-wise di appartenenza alla lesione. Poiché il problema è formulato come segmentazione binaria (lesione vs. sfondo), l'output finale viene interpretato tramite funzione sigmoide in fase di inferenza. L'architettura è stata implementata in PyTorch e parametrizzata con tre livelli di profondità e un numero iniziale di 32 filtri, senza utilizzo di pesi pre-addestrati, consentendo un addestramento supervisionato direttamente sul dominio dermatologico oggetto di studio.

## 2.4 Procedura di addestramento

### 2.4.1 Impostazione del training

Sono stati addestrati due modelli distinti, entrambi basati sulla stessa architettura U-Net descritta nella Sezione 2.3, ma differenziati esclusivamente per il tipo di preprocessing applicato alle immagini:

- **Modello non-UV**: addestrato sulle immagini preprocessate tramite GCC-GAN;
- **Modello UV**: addestrato sulle immagini convertite nel canale  $L^*$  dello spazio CIELUV.

Entrambi i modelli sono stati addestrati per 50 epoche con `batch size` pari a 8, inizializzazione casuale dei pesi e salvataggio dei *checkpoint* al termine di ogni epoca.

### 2.4.2 Funzione di perdita e ottimizzazione

Nella fase iniziale dell'addestramento è stata adottata la funzione *BCEWithLogitsLoss*, particolarmente indicata per compiti di segmentazione binaria poiché combina in un unico operatore la sigmoid e la funzione di entropia binaria. Successivamente, per migliorare la sensibilità ai contorni e la coerenza delle regioni segmentate, è stata utilizzata la *Dice Loss*, che risulta più adatta in presenza di classi con estensione spaziale ridotta rispetto allo sfondo.

L'ottimizzazione dei pesi è stata effettuata mediante l'algoritmo *Adam* con learning rate iniziale pari a  $10^{-3}$ . Non è stato applicato alcun scheduler adattativo del tasso di apprendimento, né meccanismi di early stopping, così da permettere alla rete di convergere stabilmente nel corso delle 50 epoche previste.

### 2.4.3 Strategia di validazione

Per monitorare l'andamento dell'addestramento, a ogni epoca è stata condotta una fase di validazione sul 10% del dataset, mantenuto separato rispetto al training. Oltre alla loss di validazione, è stato calcolato l'Intersection over Union (IoU), adottato come metrica principale di valutazione in quanto più informativa in termini di sovrapposizione spaziale tra predizione e maschera di riferimento. La soglia di binarizzazione delle probabilità in output è stata fissata a 0.5.

Al termine dell'addestramento, il modello è stato selezionato sulla base delle prestazioni medie ottenute sul validation set e successivamente impiegato per la fase di test e per l'inference sulle immagini del dataset clinico fornito dai dermatologi di Novara.

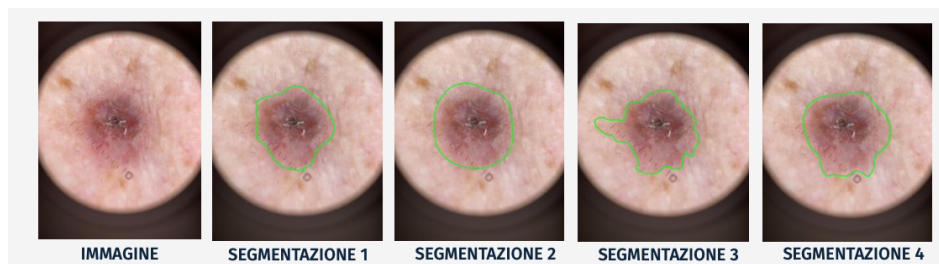
## 2.5 Creazione delle maschere manuali tramite applicazione dedicata

Il dataset fornito dai dermatologi di Novara non includeva le corrispondenti maschere di segmentazione, rendendo necessario un processo di annotazione manuale per la definizione del ground truth. A tal fine è stata sviluppata un'applicazione dedicata in ambiente *MATLAB*, progettata per consentire ai dermatologi la creazione assistita delle maschere in modo semplice, controllato e riproducibile. L'applicazione permette di caricare automaticamente un insieme di immagini da una cartella selezionata dall'utente e le visualizza in ordine casuale, riducendo possibili bias derivanti dalla sequenza di presentazione. Per ciascuna immagine è possibile delineare manualmente il contorno della lesione tramite strumenti grafici dedicati. Una volta confermata la segmentazione, la maschera viene salvata automaticamente in formato `.png` con valori binari  $\{0,255\}$  e mantenendo lo stesso nome dell'immagine originale per garantire l'associazione univoca.

Il software integra inoltre un meccanismo di tracciamento tramite file `.json`, che registra le immagini già segmentate e quelle escluse. Le immagini non idonee (ad esempio fotografie puramente localizzative o prive di valore informativo utile) possono essere contrassegnate tramite l'opzione "*Salta immagine*" e vengono escluse dal dataset finale. Il sistema consente di sospendere e riprendere il lavoro in più sessioni senza perdita dei progressi e senza ripresentazione di immagini già annotate o scartate. L'applicazione è stata distribuita come eseguibile sia per Windows (`.exe`) sia per macOS (`.app`), grazie alla compilazione tramite MATLAB Runtime, rendendo possibile l'utilizzo anche in assenza di MATLAB installato sulla macchina dell'operatore clinico.

La segmentazione manuale è stata eseguita in parallelo da quattro dermatologi coinvolti nel progetto, producendo quattro annotazioni indipendenti per ciascuna

immagine.



**Figura 2.8:** Immagine originale e corrispondenti segmentazioni manuali ottenute dai quattro dermatologi coinvolti nel progetto.

A partire da queste quattro maschere è stata generata la maschera finale di riferimento mediante *majority voting* pixel-wise, ovvero classificando come appartenente alla lesione ogni pixel presente in almeno due delle quattro annotazioni. Questa procedura consente di ottenere un ground truth validato inter-osservatore, riducendo l’effetto della variabilità soggettiva e aumentando l’affidabilità clinica del dataset. In questo modo, la maschera finale rappresenta il consenso clinico aggregato tra i quattro annotatori.

## 2.6 Preprocessing del dataset clinico di Novara

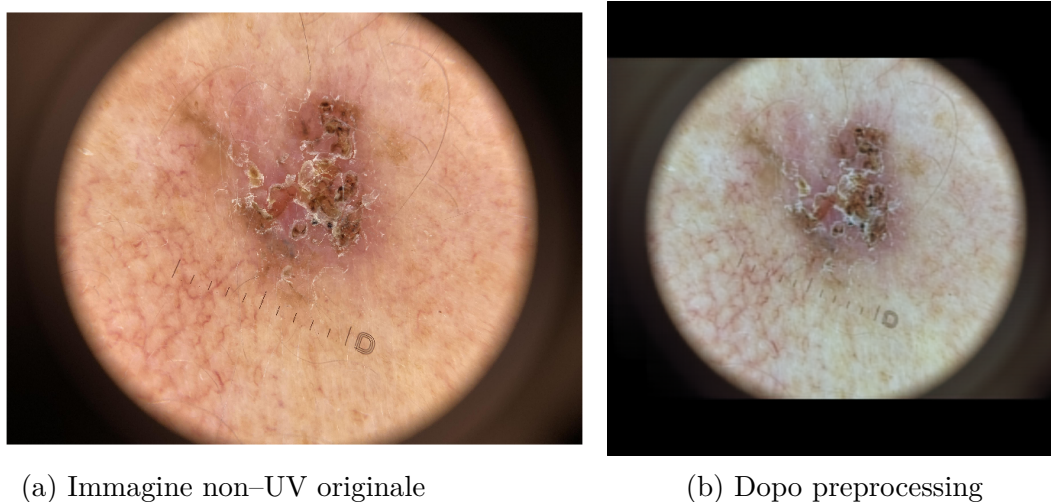
Il preprocessing applicato alle immagini cliniche di Novara è stato progettato per rispecchiare esattamente le trasformazioni utilizzate nella fase di addestramento dei due modelli (non-UV e UV), così da mantenere coerenza di dominio tra training e inference. Poiché la maggior parte delle procedure è già stata descritta nella Sezione 2.2,

**Normalizzazione geometrica** Tutte le immagini, indipendentemente dalla modalità di acquisizione, sono state:

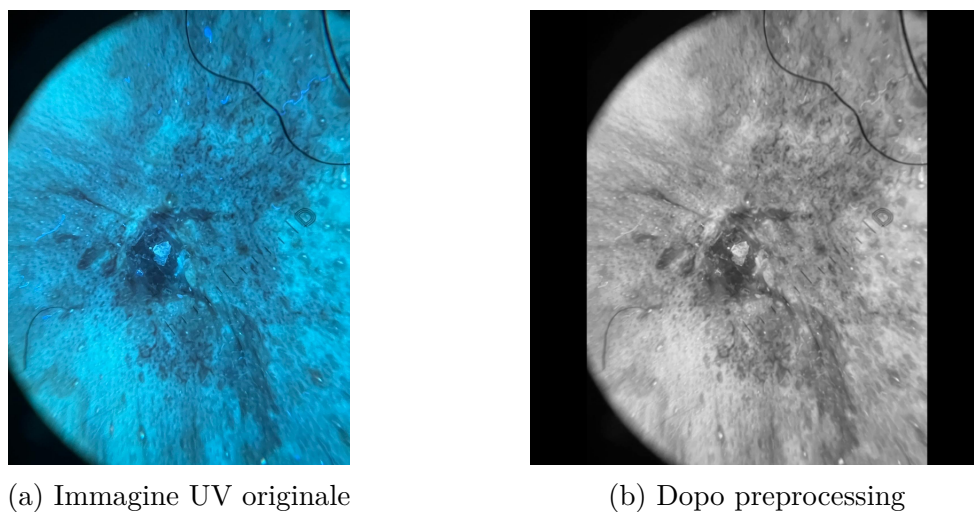
- convertite in `.png`,
- rese quadrate tramite padding simmetrico,
- ridimensionate a  $512 \times 512$  pixel.

**Immagini dermoscopiche non-UV** Per le immagini non-UV è stata applicata la stessa normalizzazione cromatica impiegata durante l’addestramento del modello dedicato a questo dominio (Sezione 2.2), ottenendo un input perfettamente allineato

al preprocessing del dataset HAM10000. Un esempio del risultato di questa trasformazione è mostrato in Figura 2.9.



**Figura 2.9:** Esempio di preprocessing delle immagini dermoscopiche non-UV del dataset clinico.



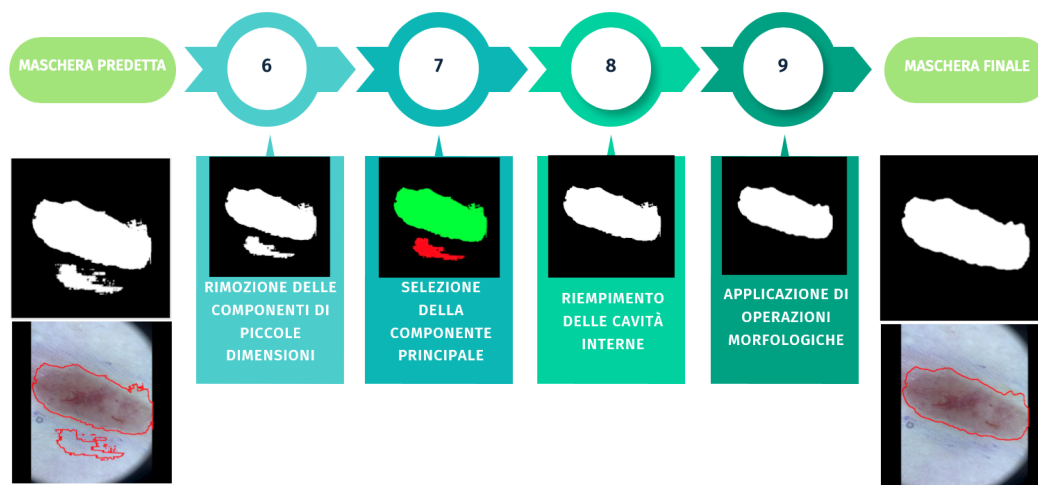
**Figura 2.10:** Esempio di preprocessing delle immagini UV del dataset clinico.

**Immagini acquisite in fluorescenza UV** Per le immagini UV è stato applicato lo stesso flusso utilizzato in fase di training: conversione nello spazio colore CIELUV, estrazione del solo canale di luminanza  $L^*$  e replicazione in tre canali per garantire la compatibilità con l'architettura U-Net.

Un esempio del preprocessing applicato alle immagini UV è riportato in Figura 2.10.

## 2.7 Post-processing delle predizioni

La Figura 2.11 illustra in maniera schematica il flusso di post-processing applicato in maniera uniforme a tutte le maschere prodotte dai due modelli.



**Figura 2.11:** Schema della pipeline di post-processing applicata alle predizioni dei modelli. Il processo comprende: rimozione delle componenti minori, selezione della componente principale, riempimento delle cavità interne e applicazione di operazioni morfologiche.

Una volta generate le maschere di segmentazione dai due modelli (dedicato alle immagini non-UV e dedicato alle immagini UV), tutte le predizioni sono state sottoposte alla stessa pipeline di post-processing. L'obiettivo è quello di eliminare piccoli artefatti residui, regolarizzare i contorni e ottenere una maschera finale più coerente e adatta sia alla valutazione quantitativa sia alla rappresentazione grafica dei risultati.

Il processo è articolato nelle seguenti operazioni:

1. **Rimozione delle componenti di piccole dimensioni**, tipicamente generate da rumore o falsi positivi locali.
2. **Selezione della componente principale**, identificata come quella più vicina al centro geometrico dell'immagine. Questa scelta deriva dall'osservazione che, nelle acquisizioni dermatoscopiche, la lesione di interesse è quasi sempre

posizionata in corrispondenza dell'asse ottico del dispositivo, rendendo tale criterio più affidabile rispetto al semplice mantenimento della componente di area massima.

3. **Riempimento delle cavità interne** (hole filling), al fine di migliorare la coerenza spaziale della regione segmentata ed evitare discontinuità nel corpo centrale della lesione.
4. **Applicazione di operazioni morfologiche** (apertura e chiusura) per eliminare irregolarità residue e regolarizzare il bordo della lesione.

Al termine del processo viene prodotta una maschera binaria con valori  $\{0,255\}$  pronta per l'utilizzo sia nella valutazione quantitativa delle performance sia nella visualizzazione qualitativa delle predizioni. Un esempio del risultato complessivo del post-processing è mostrato in Figura 2.12.

## 2.8 Metriche di valutazione

Per valutare le prestazioni del modello di segmentazione sono state utilizzate metriche standard derivate dal confronto pixel-wise tra la maschera predetta e la maschera manuale di riferimento (*ground truth*). Ciascuna metrica è calcolata a partire dai valori della *confusion matrix* binaria, che distingue tra:

- **True Positive (TP)**: pixel correttamente identificati come appartenenti alla lesione;
- **False Positive (FP)**: pixel erroneamente classificati come lesione;
- **False Negative (FN)**: pixel della lesione non individuati dal modello;
- **True Negative (TN)**: pixel correttamente classificati come sfondo.

Le principali metriche di valutazione adottate sono descritte di seguito.

### Intersection over Union (IoU)

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (2.1)$$

Misura la sovrapposizione tra la maschera predetta e quella reale: valori prossimi a 1 indicano un'elevata corrispondenza. È una delle metriche più utilizzate per la segmentazione semantica.

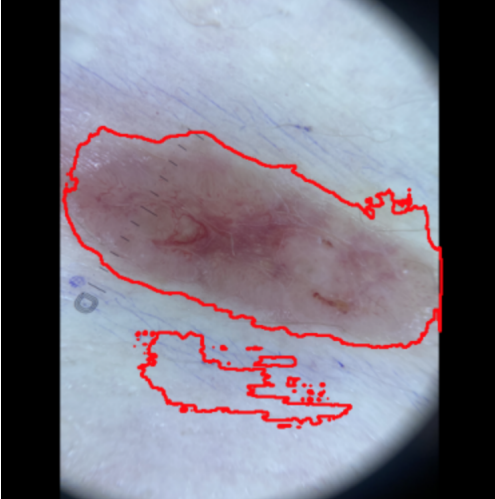




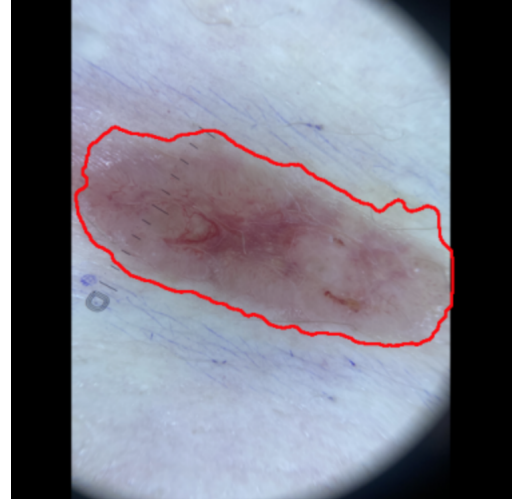
(a) Maschera predetta grezza



(b) Maschera dopo post-processing



(c) Contorno della maschera iniziale sovrapposto all'immagine



(d) Contorno della maschera finale sovrapposto all'immagine

**Figura 2.12:** Esempio dell'effetto del post-processing sulla segmentazione. (a) Predizione grezza; (b) maschera finale binaria dopo post-processing; (c) e (d) mostrano i contorni delle due maschere sovrapposti all'immagine originale.

### Dice Coefficient (Dice)

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \quad (2.2)$$

Quantifica la similarità tra due insiemi di pixel (predetti e reali). Rispetto all'IoU

risulta più sensibile alla presenza di piccole regioni, caratteristica particolarmente utile per la segmentazione di lesioni cutanee.

### Precision

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.3)$$

Indica la proporzione di pixel predetti come lesione che sono effettivamente corretti. Un'elevata *precision* corrisponde a una bassa incidenza di falsi positivi.

### Recall (Sensibilità)

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.4)$$

Misura la capacità del modello di individuare tutti i pixel appartenenti alla lesione. Un valore elevato di *recall* implica una bassa probabilità di omissione di aree patologiche.

### Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.5)$$

Rappresenta la proporzione complessiva di pixel classificati correttamente. Nel caso di segmentazione binaria con forte squilibrio tra classi, questa metrica può risultare meno rappresentativa rispetto a IoU e Dice.

L'utilizzo di queste metriche è ampiamente consolidato nella valutazione di sistemi di segmentazione medica, come discusso in [19]. Tutte le metriche sono state calcolate a livello di singola immagine e successivamente riportate come **media** e **deviazione standard** sull'intero insieme di test, al fine di descrivere in modo accurato la variabilità delle prestazioni del modello.

## 2.9 Analisi delle maschere manuali

Dopo la creazione delle maschere manuali descritta nella Sezione 2.5, sono state condotte diverse analisi finalizzate a:

- caratterizzare quantitativamente le annotazioni cliniche;
- confrontare la dimensione delle lesioni in luce convenzionale e in fluorescenza UV;
- valutare la coerenza tra i dermatologi e il modello di segmentazione sviluppato.

L'obiettivo di questa sezione è quindi analizzare le annotazioni ottenute e confrontarle con le predizioni automatiche, allo scopo di comprendere come la fluorescenza UV influenzi la percezione dei margini e in che misura il modello riesca a riprodurre i criteri decisionali adottati dai dermatologi.

### 2.9.1 Calcolo dell'area e confronto dimensionale delle maschere

Per ogni immagine sono state considerate le due maschere manuali distinte: una ottenuta da immagini dermoscopiche convenzionali (NON-UV) e una ricavata dalle corrispondenti immagini acquisite in fluorescenza UV. L'area della lesione è stata calcolata come numero totale di pixel appartenenti alla regione segmentata:

$$A = \sum_{x,y} M(x, y), \quad (2.6)$$

dove  $M(x, y) = 1$  indica i pixel classificati come lesione. Il confronto dimensionale tra le due acquisizioni è stato espresso come differenza:

$$\Delta A = A_{UV} - A_{NONUV}, \quad (2.7)$$

da cui è possibile distinguere tre casi:

- $\Delta A > 0$ : la lesione risulta più estesa in UV;
- $\Delta A < 0$ : la lesione appare più ampia nella versione NON-UV;
- $\Delta A = 0$ : aree equivalenti nelle due modalità.

Queste analisi sono state ripetute per ciascuno dei quattro dermatologi coinvolti e successivamente sulle predizioni del modello automatico, al fine di confrontarne il comportamento con quello osservato nella pratica clinica.

### 2.9.2 Valutazione della concordanza tra dermatologi

Per ogni coppia di immagini, le valutazioni dei dermatologi sono state confrontate per determinare il livello di accordo clinico sulla maggiore estensione della lesione tra UV e NON-UV. È stata adottata la seguente classificazione:

- **Accordo totale:** tutti i dermatologi esprimono la stessa valutazione riguardo quale delle due modalità (UV o NON-UV) mostri una lesione di dimensioni maggiori.

- **Disaccordo:** non emerge alcuna decisione predominante, indicando un'elevata ambiguità interpretativa dovuta a margini poco definiti o a fluorescenza non uniforme. In questi casi i dermatologi esprimono valutazioni discordanti (ad esempio distribuzioni 2-2 o 3-1), senza convergere su quale modalità mostri una lesione più estesa.

Questa classificazione permette di suddividere il dataset clinico in livelli di *complessità interpretativa*, evidenziando i casi in cui la segmentazione manuale risulta coerente e quelli in cui la variabilità inter-osservatore è più elevata.

### 2.9.3 Confronto tra maschere manuali e predizioni del modello

Per valutare l'aderenza del modello al comportamento clinico, è stato effettuato un confronto diretto tra la decisione dei dermatologi e quella prodotta dal modello riguardo la maggiore estensione della lesione in UV o in NON-UV.

La valutazione del modello si basa sulle maschere predette sulle immagini UV e NON-UV, calcolando la differenza di area come descritto nella sezione 2.9.1. Da tale confronto emergono tre categorie di casi:

- **Accordo modello-clinici:** il modello fornisce la stessa valutazione condivisa dalla totalità dei dermatologi;
- **Disaccordo modello-clinici:** i dermatologi concordano su una valutazione ma il modello fornisce il risultato opposto;
- **Casi ambigui:** i dermatologi non convergono su una decisione univoca, rendendo la valutazione del modello non interpretabile in termini di correttezza.

Questa analisi consente di comprendere la capacità del modello di riprodurre il criterio di interpretazione utilizzato nella pratica clinica, un aspetto fondamentale per un'eventuale applicazione in supporto alla chirurgia micrografica o alla valutazione pre-operatoria.

### 2.9.4 Analisi statistica delle maschere manuali e delle predizioni

Oltre al confronto diretto tra le maschere UV e NON-UV, è stata condotta un'analisi statistica finalizzata a quantificare in modo oggettivo le differenze dimensionali tra le due modalità di acquisizione e a valutare la coerenza tra i dermatologi e il modello di segmentazione sviluppato.

L'analisi si articola in tre parti:

- confronto non paired tra le aree UV e NON-UV per ciascun operatore;
- confronto paired tra le differenze UV–NON-UV dei dermatologi;
- confronto paired tra le differenze UV–NON-UV del modello e quelle dei dermatologi.

### Confronto non paired tra UV e NON-UV

Per ciascun dermatologo e per il modello, sono state considerate due distribuzioni di area: una relativa alle immagini NON-UV e una relativa alle immagini UV. L'obiettivo era verificare se la fluorescenza UV comportasse un ampliamento statisticamente significativo dell'area segmentata.

La normalità delle distribuzioni è stata verificata tramite test di Shapiro–Wilk. In assenza di normalità, condizione riscontrata sistematicamente in tutte le distribuzioni analizzate, il confronto tra le due modalità è stato effettuato tramite il test di *Mann–Whitney U*, equivalente al test di Wilcoxon rank–sum, mentre il test *t* di Student non paired sarebbe stato utilizzato in presenza di distribuzioni Gaussiane.

Questo approccio permette di valutare, per ciascun operatore, se le immagini in fluorescenza UV producano una differenza significativa nell'estensione della lesione rispetto alla dermoscopia convenzionale.

### Confronto paired tra dermatologi

Poiché tutti gli operatori segmentano le stesse immagini, il confronto tra le loro differenze  $\Delta A$  configura un problema di tipo *paired*. La normalità delle differenze è stata verificata tramite test di Shapiro–Wilk; in assenza di normalità, è stato applicato il test di *Wilcoxon signed-rank*.

Questa procedura consente di identificare eventuali divergenze sistematiche tra operatori e di evidenziare gruppi omogenei in base al criterio adottato per la valutazione dei margini UV.

### Confronto paired tra modello e dermatologi

La stessa procedura basata sul confronto paired delle differenze  $\Delta A$  è stata estesa al modello di segmentazione. Le differenze UV–NON-UV prodotte dal modello sono state confrontate, sempre tramite test di Wilcoxon signed-rank, con quelle di ciascun dermatologo.

Tale analisi permette di quantificare in che misura il modello riproduca il comportamento clinico dei dermatologi e di identificare eventuali similarità o divergenze nel criterio decisionale adottato sulle immagini UV rispetto alla dermoscopia convenzionale.

## Capitolo 3

# Risultati e discussione

In questo capitolo vengono presentati i risultati sperimentali ottenuti dall'applicazione dei modelli di segmentazione sviluppati. L'obiettivo principale è valutare l'impatto delle diverse fasi di preprocessing e post-processing descritte nel Capitolo 2 sulle prestazioni del modello U-Net proposto.

I risultati vengono analizzati sia in termini quantitativi, mediante le metriche presentate nella Sezione 2.8, sia in termini qualitativi attraverso esempi di segmentazione. Particolare attenzione è stata posta al confronto tra i due domini di acquisizione (non-UV e UV), trattati come dataset distinti.

### 3.1 Risultati del modello baseline

La prima fase di analisi ha riguardato la valutazione del modello *baseline*, ovvero la versione della rete U-Net addestrata sul dataset *HAM10000* preprocessato ma priva di *data augmentation* e *color constancy*. Questo esperimento ha lo scopo di fornire un punto di riferimento iniziale per quantificare l'efficacia dei successivi interventi di miglioramento.

#### Prestazioni sul dataset pubblico HAM10000

Sul dataset di test del *HAM10000*, il modello baseline ha mostrato prestazioni complessivamente buone, confermando la capacità della rete di apprendere la morfologia delle lesioni nel dominio di addestramento. Le metriche medie ottenute sono riportate in Tabella 3.1.

**Tabella 3.1:** Prestazioni del modello baseline sul dataset HAM10000 (media  $\pm$  deviazione standard).

Metrica	Valore medio
IoU	$0.756 \pm 0.211$
Dice	$0.841 \pm 0.170$
Precision	$0.877 \pm 0.188$
Recall	$0.854 \pm 0.169$
Accuracy	$0.942 \pm 0.071$

Questi risultati evidenziano una buona capacità di generalizzazione all’interno del dataset stesso, con un equilibrio tra *precision* e *recall* che suggerisce un comportamento stabile della rete nelle diverse tipologie di lesione.

### Prestazioni sulle immagini dermatologiche di Novara

Il modello baseline è stato successivamente testato sulle immagini cliniche provenienti dal dataset di Novara, comprendente sia acquisizioni dermoscopiche non-UV sia immagini in fluorescenza UV. Poiché tali immagini appartengono a un dominio completamente diverso rispetto al dataset di addestramento, si osserva un netto calo di prestazioni in entrambi i casi, come riportato in Tabella 3.2.

**Tabella 3.2:** Prestazioni del modello baseline sulle immagini del dataset di Novara (media  $\pm$  deviazione standard, dopo post-processing).

Metrica	Novara non-UV	Novara UV
IoU	$0.420 \pm 0.194$	$0.011 \pm 0.054$
Dice	$0.565 \pm 0.200$	$0.015 \pm 0.081$
Precision	$0.731 \pm 0.222$	$0.039 \pm 0.186$
Recall	$0.559 \pm 0.267$	$0.010 \pm 0.055$
Accuracy	$0.868 \pm 0.095$	$0.768 \pm 0.108$

Come si osserva, le prestazioni risultano significativamente inferiori rispetto a quelle ottenute sul dataset di addestramento, con un calo particolarmente marcato nel caso delle immagini in fluorescenza UV. Anche le immagini dermoscopiche non-UV mostrano valori più bassi rispetto a quanto atteso, a causa di differenze sostanziali tra i due domini.

Le immagini cliniche di Novara presentano infatti condizioni di illuminazione più variabili e una marcata vignettatura dovuta alla presenza del cerchio del dermatoscopio, elemento assente nel dataset pubblico *HAM10000*. Inoltre, molte lesioni

mostrano caratteristiche morfologiche e cromatiche diverse rispetto a quelle tipiche delle immagini di addestramento, rappresentando quindi casi *out-of-distribution*.

Tali differenze, combinate con l'assenza di dataset pubblici aventi caratteristiche simili a quelle delle immagini cliniche di Novara, limitano la possibilità di addestrare reti neurali su dati realmente comparabili. Queste considerazioni spiegano le basse prestazioni osservate per il modello *baseline* e motivano l'introduzione, nelle sezioni successive, di strategie di *color constancy* e *data augmentation* per migliorare la robustezza del modello ai cambiamenti di dominio.

## 3.2 Risultati dopo l'applicazione della color constancy

Il modello allenato sul dataset *HAM10000* preprocessato tramite *color constancy* è stato valutato sia sul relativo test set pubblico sia sui due sottoinsiemi del dataset clinico di Novara (immagini non-UV e UV). I risultati quantitativi ottenuti nei tre domini sono riportati in Tabella 3.3.

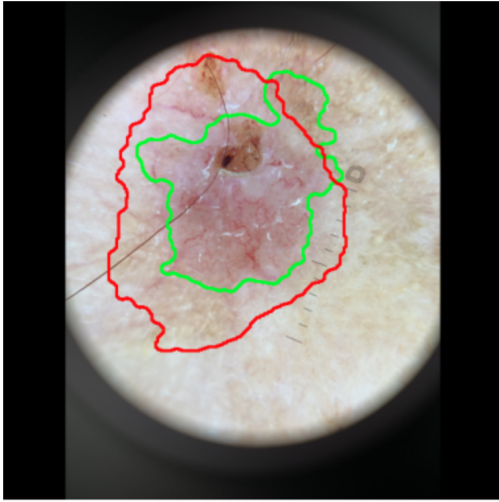
**Tabella 3.3:** Prestazioni del modello dopo l'applicazione della color constancy sui tre domini di test (media  $\pm$  deviazione standard, dopo post-processing).

Metrica	HAM10000	Novara non-UV	Novara UV
IoU	$0.860 \pm 0.154$	$0.549 \pm 0.176$	$0.479 \pm 0.175$
Dice	$0.915 \pm 0.118$	$0.690 \pm 0.174$	$0.618 \pm 0.186$
Precision	$0.914 \pm 0.141$	$0.651 \pm 0.217$	$0.568 \pm 0.230$
Recall	$0.936 \pm 0.106$	$0.834 \pm 0.187$	$0.804 \pm 0.211$
Accuracy	$0.969 \pm 0.044$	$0.895 \pm 0.068$	$0.813 \pm 0.090$

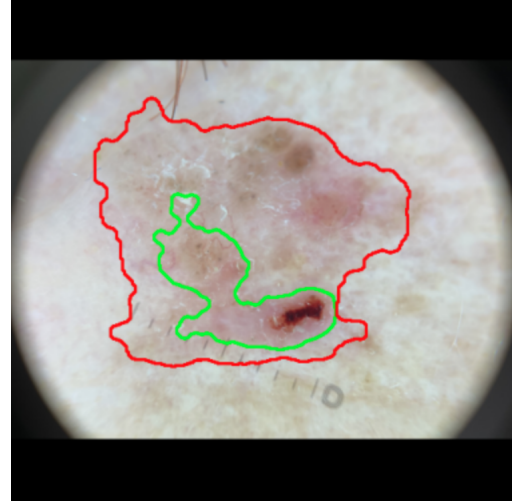
L'applicazione della *color constancy* ha determinato un miglioramento significativo delle prestazioni rispetto al modello *baseline* in tutti i domini. Nelle immagini dermoscopiche non-UV del dataset di Novara, l'IoU medio passa da 0.42 a 0.55 e il *Dice coefficient* da 0.56 a 0.69, indicando una segmentazione più coerente dei margini periferici e una migliore separazione della lesione dal tessuto cutaneo circostante.

Il beneficio della *color constancy* è ben visibile anche sul piano qualitativo. In Figura 3.1 sono riportati alcuni esempi tratti dalle immagini non-UV del dataset di Novara. Il contorno verde rappresenta la maschera ottenuta senza preprocessing, mentre il contorno rosso mostra la segmentazione prodotta dopo l'applicazione della normalizzazione cromatica.

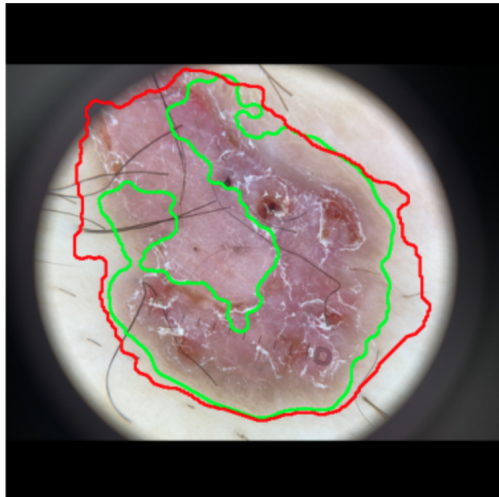




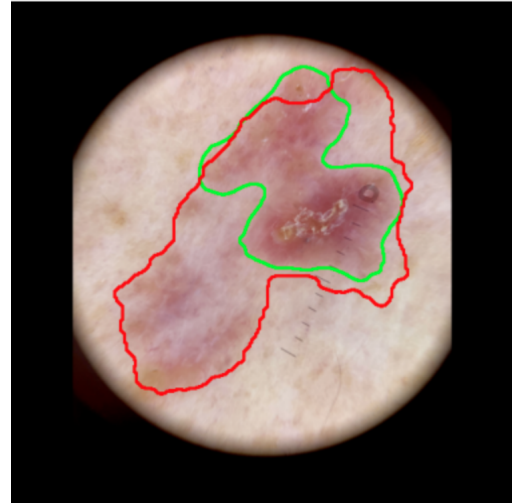
(a) Esempio 1



(b) Esempio 2



(c) Esempio 3

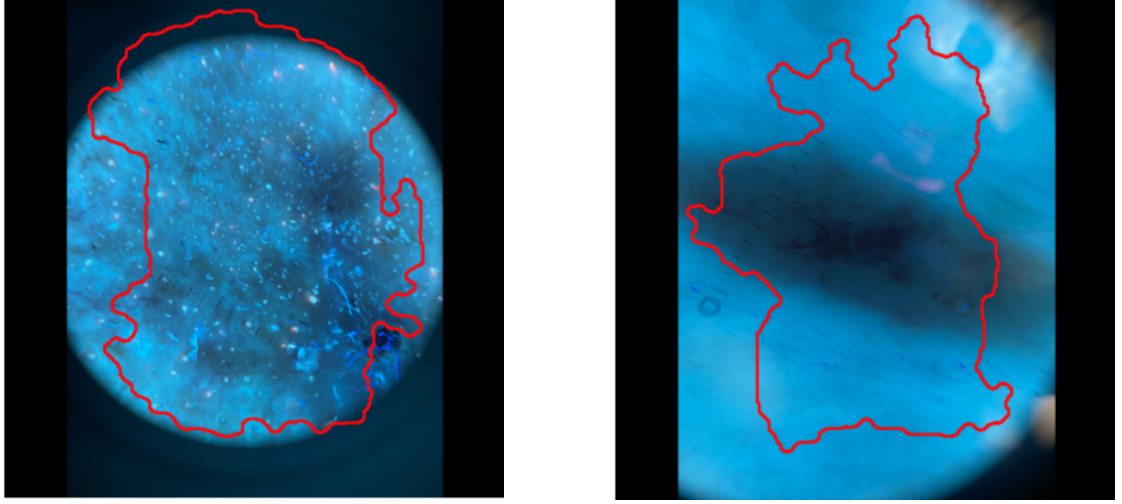


(d) Esempio 4

**Figura 3.1:** Esempi qualitativi dell'effetto della *color constancy* nelle immagini dermoscopiche non-UV.

Il comportamento del modello sulle immagini acquisite in fluorescenza UV è invece più complesso da interpretare. Come riportato in Tabella 3.3, si osserva un incremento numerico delle metriche (IoU da 0.01 a 0.48 e Dice da 0.02 a 0.62), ma tale miglioramento non corrisponde a un reale aumento della qualità della segmentazione. Il modello tende infatti a sovrasegmentare la lesione, includendo aree non patologiche o, in alcuni casi, l'intero campo visivo del dermatoscopio. Ciò porta a maschere molto ampie che, pur aumentando la sensibilità in termini

numerici, risultano clinicamente inutilizzabili.



(a) Esemplio UV 1

(b) Esemplio UV 2

**Figura 3.2:** Esempi dell'effetto della *color constancy* nelle immagini UV.

Gli esempi in Figura 3.2 mostrano chiaramente questo comportamento: la maschera generata dopo *color constancy* (in rosso) non segue la distribuzione reale della fluorescenza tumorale, ma tende ad espandersi oltre i margini veri della lesione fino a coprire quasi tutta l'area illuminata.

### 3.3 Risultati dopo l'applicazione della *color constancy* e del *data augmentation*

#### Prestazioni sui diversi domini

Dopo aver valutato l'efficacia della normalizzazione cromatica, è stato introdotto un ulteriore passaggio di *data augmentation* al fine di incrementare la variabilità morfologica del dataset di addestramento.

Il modello allenato sul dataset aumentato e normalizzato cromaticamente è stato valutato sia sul test set del *HAM10000*, sia sui due sottoinsiemi del dataset clinico di Novara (immagini non-UV e UV). Tutti i valori riportati includono il post-processing morfologico descritto in Sezione 2.7. I risultati sono riassunti in Tabella 3.4.

**Tabella 3.4:** Prestazioni del modello dopo color constancy + data augmentation sui tre domini (media  $\pm$  deviazione standard, dopo post-processing).

Metrica	HAM10000	Novara non-UV	Novara UV
IoU	$0.820 \pm 0.185$	$0.589 \pm 0.177$	$0.518 \pm 0.191$
Dice	$0.886 \pm 0.154$	$0.722 \pm 0.172$	$0.659 \pm 0.188$
Precision	$0.916 \pm 0.161$	$0.720 \pm 0.215$	$0.773 \pm 0.223$
Recall	$0.880 \pm 0.167$	$0.814 \pm 0.197$	$0.671 \pm 0.247$
Accuracy	$0.959 \pm 0.061$	$0.910 \pm 0.064$	$0.863 \pm 0.077$

### Analisi dei risultati

L'aggiunta del *data augmentation* ha migliorato ulteriormente le prestazioni rispetto al modello addestrato con la sola *color constancy*.

Sul dataset pubblico *HAM10000*, le metriche rimangono elevate e stabili (IoU = 0.82, Dice = 0.89), indicando un apprendimento solido senza fenomeni evidenti di overfitting.

L'effetto più rilevante si osserva sul dataset clinico di Novara:

- **Immagini non-UV:** l'IoU cresce da 0.55 a 0.59 e il Dice da 0.69 a 0.72, con un ulteriore incremento della *recall* (0.81), che riflette una migliore individuazione delle regioni periferiche.
- **Immagini UV:** l'IoU aumenta da 0.48 a 0.52 e il Dice da 0.62 a 0.66, suggerendo un miglioramento numerico rispetto al modello con la sola *color constancy*. Tuttavia, come osservato nella sezione precedente, questi valori non riflettono un reale incremento della qualità della segmentazione.

Per questo motivo, nonostante un miglioramento quantitativo, la qualità clinica delle segmentazioni UV rimane insufficiente. Ciò ha motivato lo sviluppo di un'ulteriore pipeline basata sull'utilizzo del solo canale di luminanza  $L^*$  dello spazio colore LUV, più adatto a rappresentare le variazioni reali di intensità osservabili nelle immagini in fluorescenza UV.

## 3.4 Risultati dopo l'applicazione del solo canale di luminanza $L^*$

### Prestazioni sui diversi domini

Il modello allenato sul dataset contenente esclusivamente il canale di luminanza  $L^*$  è stato valutato sul test set del *HAM10000* e sui due sottoinsiemi del dataset clinico

di Novara (immagini non-UV e UV). Tutti i valori includono il post-processing morfologico descritto in Sezione 2.7. I risultati sono riportati in Tabella 3.5.

**Tabella 3.5:** Prestazioni del modello dopo preprocessing nel solo canale  $L^*$  (media  $\pm$  deviazione standard, dopo post-processing).

Metrica	HAM10000	Novara non-UV	Novara UV
IoU	$0.763 \pm 0.185$	$0.521 \pm 0.177$	$0.533 \pm 0.174$
Dice	$0.842 \pm 0.133$	$0.685 \pm 0.172$	$0.676 \pm 0.171$
Precision	$0.905 \pm 0.142$	$0.703 \pm 0.215$	$0.684 \pm 0.235$
Recall	$0.875 \pm 0.156$	$0.814 \pm 0.197$	$0.766 \pm 0.192$
Accuracy	$0.942 \pm 0.062$	$0.910 \pm 0.064$	$0.855 \pm 0.069$

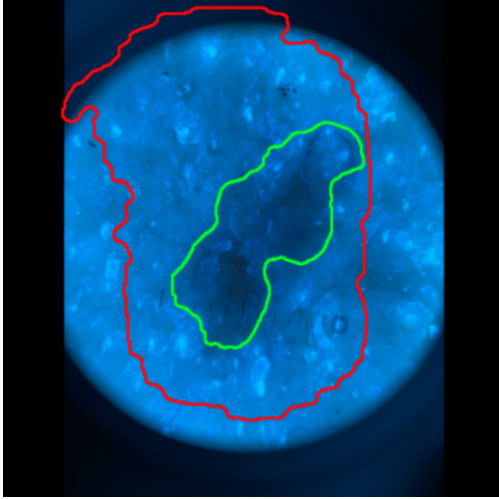
### 3.4.1 Analisi dei risultati

L'estrazione del solo canale  $L^*$  consente al modello di concentrarsi sulle informazioni di luminanza eliminando le dominanti cromatiche tipiche della fluorescenza UV, riducendo così la variabilità spettrale e gli artefatti dovuti alla riflessione superficiale. I risultati mostrano chiaramente che questo approccio è particolarmente efficace nel dominio UV:

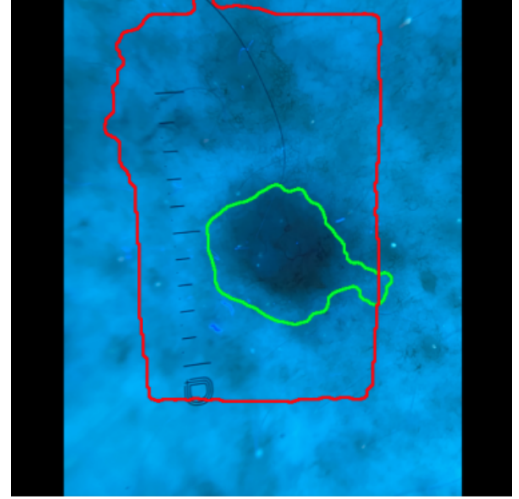
- **Immagini UV:** l'*IoU* raggiunge 0.53 (contro 0.48 con la sola *color constancy*) e il *Dice* sale a 0.68. Il canale  $L^*$  fornisce una rappresentazione più stabile dell'intensità di fluorescenza, permettendo al modello di seguire meglio i margini reali della lesione e riducendo le sovrasegmentazioni osservate nella sezione precedente.
- **Immagini non-UV:** le prestazioni risultano leggermente inferiori a quelle ottenute con la *color constancy* (*IoU* = 0.52 contro 0.55). Nel dominio visibile, infatti, l'informazione cromatica contribuisce in modo rilevante alla definizione dei margini e la sola luminanza non è sufficiente a catturare tutte le variazioni strutturali utili alla segmentazione.
- **Dataset HAM10000:** le metriche rimangono buone (*IoU* = 0.76), dimostrando che la morfologia delle lesioni rimane ben rappresentata anche in assenza di informazione cromatica.

Dal punto di vista qualitativo, la Figura 3.3 confronta la segmentazione ottenuta con la sola *color constancy* (contorno rosso) e quella prodotta dal modello basato sul canale  $L^*$  (contorno verde) su quattro immagini UV del dataset di Novara. Gli esempi mostrano come la segmentazione basata su luminanza sia generalmente più

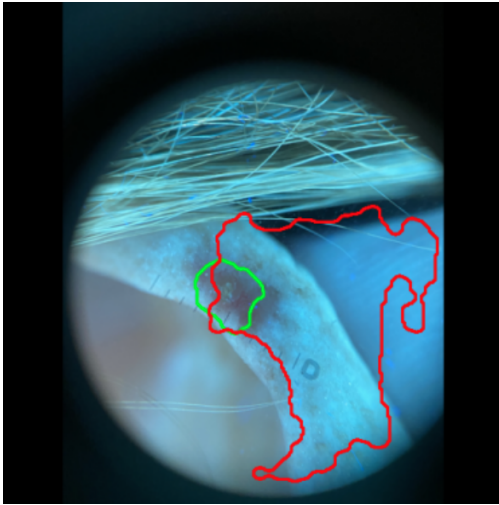
contenuta e meglio aderente alla reale distribuzione della fluorescenza tumorale, mentre la *color constancy* tende spesso a sovrasegmentare coprendo porzioni del campo visivo non associate alla lesione.



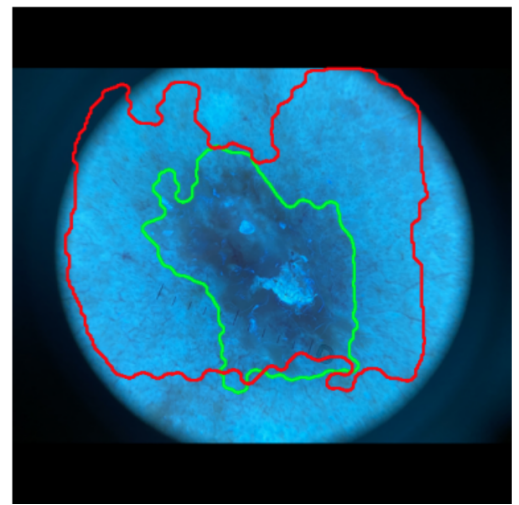
(a) Esempio 1



(b) Esempio 2



(c) Esempio 3



(d) Esempio 4

**Figura 3.3:** Confronto tra la segmentazione ottenuta con la sola *color constancy* (rosso) e quella prodotta dal modello basato sul canale di luminanza  $L^*$  (verde) su quattro immagini UV.

Nel complesso, il preprocessing nel solo canale  $L^*$  rappresenta una soluzione più adatta per la segmentazione delle immagini in fluorescenza UV, mentre risulta

meno efficace sulle immagini dermoscopiche convenzionali. Nella prossima sezione verrà analizzato l'effetto del *data augmentation* in combinazione con il canale  $L^*$ , con l'obiettivo di incrementare ulteriormente la stabilità e la robustezza del modello nel dominio UV.

### 3.5 Risultati dopo data augmentation ed estrazione del canale di luminanza $L^*$

#### Prestazioni sui diversi domini

Di seguito vengono riportate le prestazioni del modello precedente con aggiunto lo step di data augmentation sul test set di *HAM10000* e sui due sottoinsiemi del dataset di Novara. Tutti i valori includono il post-processing morfologico descritto nella Sezione 2.7.

**Tabella 3.6:** Prestazioni del modello dopo canale  $L^*$  + data augmentation (media  $\pm$  deviazione standard, dopo post-processing).

Metrica	HAM10000	Novara non-UV	Novara UV
IoU	$0.793 \pm 0.175$	$0.541 \pm 0.177$	$0.555 \pm 0.181$
Dice	$0.872 \pm 0.133$	$0.686 \pm 0.172$	$0.693 \pm 0.178$
Precision	$0.895 \pm 0.142$	$0.704 \pm 0.215$	$0.648 \pm 0.224$
Recall	$0.885 \pm 0.156$	$0.814 \pm 0.197$	$0.859 \pm 0.147$
Accuracy	$0.952 \pm 0.062$	$0.910 \pm 0.064$	$0.853 \pm 0.071$

#### Analisi dei risultati

I risultati confermano che la combinazione tra canale di luminanza  $L^*$  e *data augmentation* è particolarmente vantaggiosa per la segmentazione delle immagini in fluorescenza UV.

- **Immagini UV:** si registra un ulteriore miglioramento rispetto al modello con solo canale  $L^*$  (IoU da 0.53 a 0.56, Dice da 0.68 a 0.69). L'aumento della *recall* (0.86) evidenzia una migliore capacità di catturare le regioni periferiche della fluorescenza, spesso caratterizzate da bassi livelli di intensità.
- **Immagini non-UV:** le prestazioni rimangono inferiori a quelle ottenute applicando la *color constancy*.
- **Dataset HAM10000:** le metriche restano elevate (IoU = 0.79, Dice = 0.87), dimostrando che il modello mantiene una buona capacità di apprendimento anche quando privato delle informazioni cromatiche.

Nel complesso, il preprocessing basato sul canale di luminanza  $L^*$  combinato con il *data augmentation* rappresenta la soluzione più efficace per la segmentazione delle immagini in fluorescenza UV, migliorandone la qualità dei contorni e la sensibilità.

### 3.6 Effetto del post-processing

#### Confronto quantitativo

Per quantificare il contributo del post-processing morfologico descritto nella Sezione 2.7, è stato effettuato un confronto diretto tra le prestazioni dei due modelli finali prima e dopo l'applicazione delle operazioni di filtraggio.

Sono stati analizzati i due modelli migliori per ciascun dominio:

- **Color Constancy + Data Augmentation** per le immagini non-UV;
- **Canale  $L^*$  + Data Augmentation** per le immagini UV.

**Tabella 3.7:** Confronto tra prestazioni prima e dopo il post-processing per i due modelli finali (media  $\pm$  deviazione standard).

Dominio	IoU (pre)	IoU (post)	Dice (pre)	Dice (post)
Non-UV (CC + DA)	$0.533 \pm 0.176$	$0.589 \pm 0.177$	$0.676 \pm 0.174$	$0.722 \pm 0.172$
UV (L + DA)	$0.543 \pm 0.180$	$0.555 \pm 0.181$	$0.683 \pm 0.178$	$0.693 \pm 0.178$

#### Analisi dei risultati

L'effetto del post-processing risulta chiaramente positivo in entrambi i domini di test, seppur con entità diversa.

**Immagini non-UV.** Nel dominio dermoscopedico convenzionale, l'aumento dovuto al post-processing è più marcato: l'*IoU* cresce da 0.53 a 0.59 e il *Dice coefficient* da 0.68 a 0.72. Ciò è dovuto al fatto che le immagini non-UV presentano margini lesionati complessi e talvolta irregolari; la rimozione delle componenti spurie e la regolarizzazione del contorno migliorano significativamente la consistenza della segmentazione.

**Immagini UV.** Nel dominio della fluorescenza UV, il beneficio è più contenuto ma comunque presente, con un incremento dell'*IoU* ( $0.54 \rightarrow 0.56$ ) e del *Dice* ( $0.68 \rightarrow 0.69$ ). La fluorescenza presenta già bordi più morbidi e distribuzioni di intensità omogenee; il post-processing agisce principalmente eliminando piccole aree rumorose e consolidando la componente principale della lesione.

### 3.7 Analisi delle maschere UV e NON-UV

Per ciascuna delle 84 lesioni del dataset clinico sono state confrontate le aree delle maschere manuali ottenute dai quattro dermatologi, con l'obiettivo di valutare in quale modalità di acquisizione (dermoscopia convenzionale o fluorescenza UV) la lesione risultasse più estesa. La Tabella 3.8 riassume il numero di casi in cui la maschera UV presenta un'area maggiore rispetto alla corrispondente NON-UV e viceversa.

Osservatore	UV > NON-UV	NON-UV > UV
Dermatologo 1	65	19
Dermatologo 2	67	17
Dermatologo 3	63	21
Dermatologo 4	65	19
Rete	68	16

**Tabella 3.8:** Confronto tra il numero di lesioni con area maggiore in UV e in NON-UV secondo i quattro dermatologi e secondo il modello.

Tutti gli osservatori mostrano una tendenza coerente: nella maggior parte dei casi la lesione appare più estesa in fluorescenza UV. Tale comportamento riflette la maggiore evidenza dei margini osservata in UV, in condizioni in cui la fluorescenza mette in risalto le aree periferiche non sempre distinguibili in luce convenzionale. Il modello automatico riproduce lo stesso andamento, con una distribuzione (68 vs 16) molto simile a quella dei dermatologi, suggerendo che la rete apprende e generalizza tale caratteristica morfologica anche sulle immagini cliniche.

#### Concordanza tra osservatori e confronto con il modello

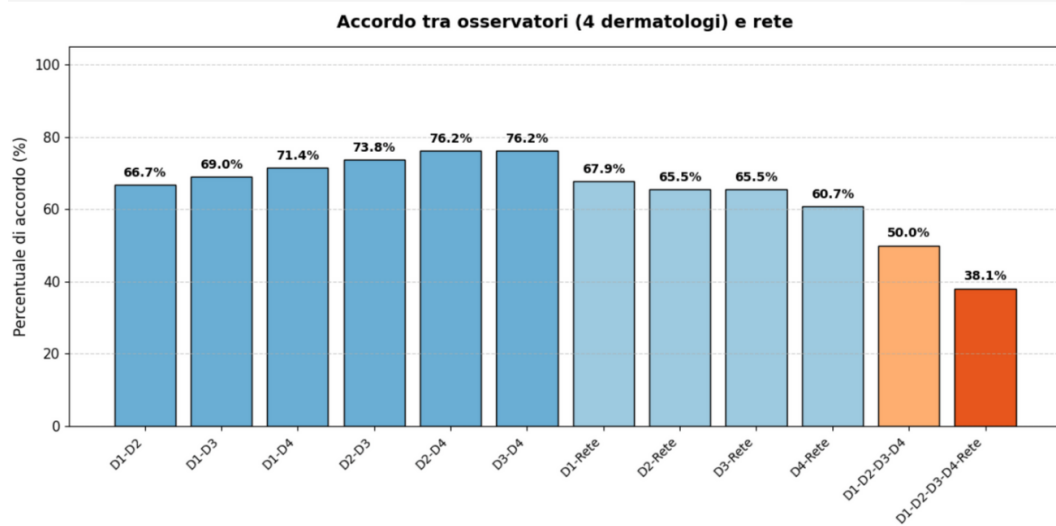
Oltre alla valutazione delle differenze di area, è stata analizzata la coerenza tra le decisioni degli osservatori riguardo alla maggiore estensione della lesione (UV o NON-UV). La Figura 3.4 riporta le percentuali di accordo per ciascuna coppia di dermatologi, per ciascun dermatologo confrontato con il modello e per i gruppi composti da tutti gli osservatori.

La concordanza tra le coppie di dermatologi varia dal 66.7% al 76.2%, valori che indicano un buon livello di consistenza inter-osservatore. Le concordanze tra ciascun dermatologo e il modello risultano comprese tra il 60.7% e il 67.9%, valori paragonabili a quelli osservati tra clinici, e che suggeriscono come il comportamento della rete sia in linea con il criterio decisionale adottato nella pratica clinica.

Considerando il gruppo composto dai quattro dermatologi, l'accordo completo è raggiunto nel 50% dei casi. Quando si include anche il modello, il consenso comune



scende al 38%, evidenziando che una quota delle lesioni presenta caratteristiche intrinsecamente ambigue, per le quali né gli esperti né il modello convergono su una valutazione univoca.



**Figura 3.4:** Percentuali di accordo tra dermatologi e tra dermatologi e modello nella valutazione della maggiore estensione della lesione (UV vs NON-UV).

### Sintesi dei risultati

La Tabella 3.9 riassume i principali scenari osservati.

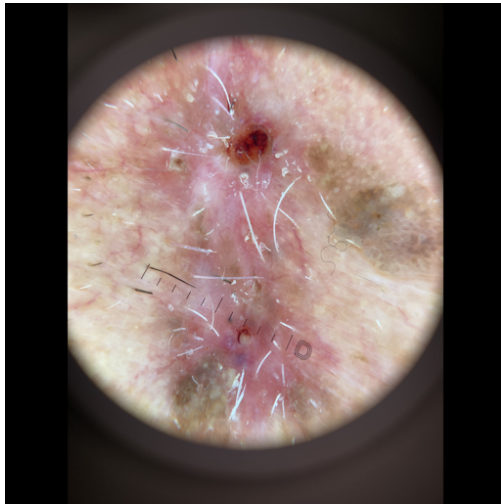
Scenario	Numero immagini	Percentuale
Dermatologi tutti d'accordo	42	50.0%
Tutti (4 Dermatologi + Rete) d'accordo	32	38.1%
Dermatologi d'accordo, rete in disaccordo	10	11.9%

**Tabella 3.9:** Riepilogo degli accordi tra dermatologi e modello.

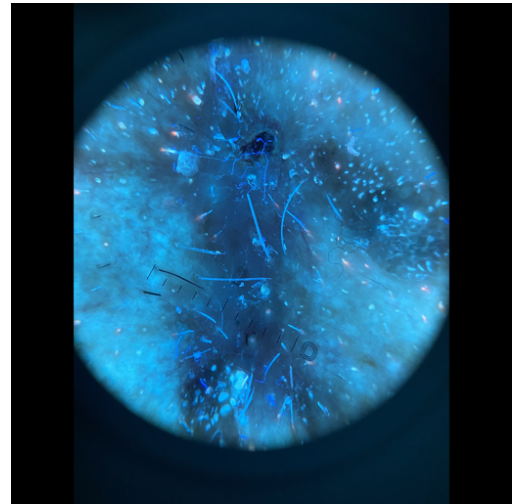
Nel complesso, i risultati evidenziano come il modello mostri un comportamento coerente con quello degli osservatori clinici: quando la valutazione specialistica raggiunge un consenso elevato, anche la rete tende a fornire la stessa classificazione. I casi di disaccordo, che rappresentano il 10% del totale, rappresentano situazioni intrinsecamente più ambigue dal punto di vista interpretativo, confermando che la complessità del segnale dermoscopico e fluorescente si riflette sia nella variabilità umana sia nelle predizioni automatiche.

### 3.7.1 Esempi qualitativi di accordo e disaccordo

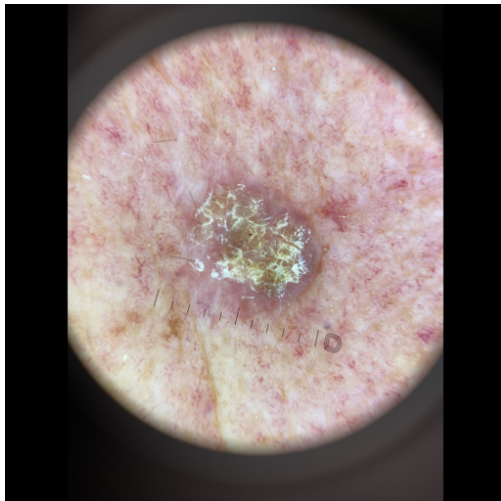
Per meglio comprendere il comportamento dei dermatologi e del modello nella valutazione dell'estensione delle lesioni tra le due modalità di acquisizione (UV e NON-UV), sono stati analizzati esempi rappresentativi dei diversi scenari osservativi emersi dall'analisi quantitativa.



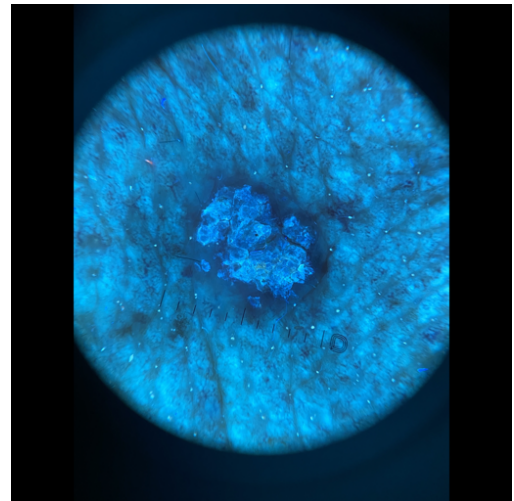
(a) NON-UV — Esempio 1



(b) UV — Esempio 1



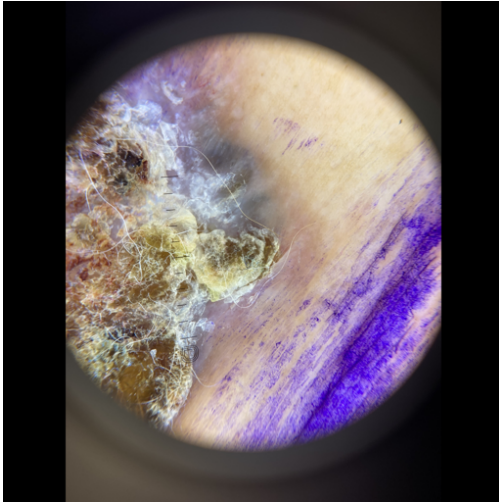
(c) NON-UV — Esempio 2



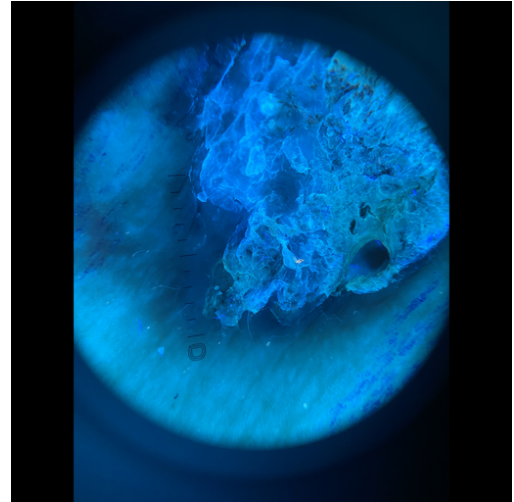
(d) UV — Esempio 2

**Figura 3.5:** Esempi di lesioni in cui tutti i dermatologi e il modello concordano sulla modalità in cui l'area della lesione risulta più estesa. Ogni coppia (a-b) e (c-d) mostra rispettivamente la versione NON-UV e UV della stessa lesione.

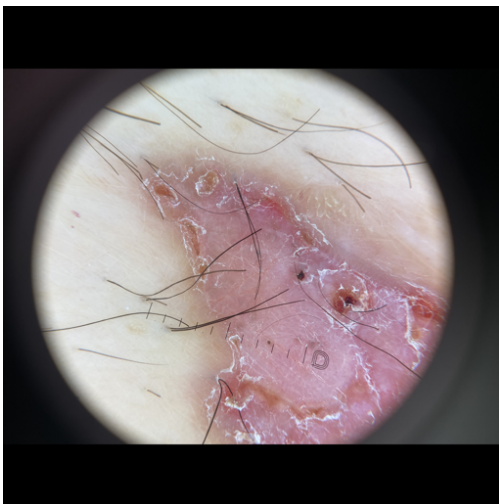
Un primo insieme di casi è costituito dalle lesioni per le quali **tutti i dermatologi e il modello sono concordi** nell'indicare la stessa modalità come più estesa. Questi esempi rappresentano le situazioni più chiare dal punto di vista clinico: i margini appaiono ben definiti e la fluorescenza UV, quando presente, evidenzia in modo netto l'area periferica della lesione. In Figura 3.5 vengono riportati quattro esempi di questa tipologia, che mostrano un'elevata coerenza tra osservatori e modello.



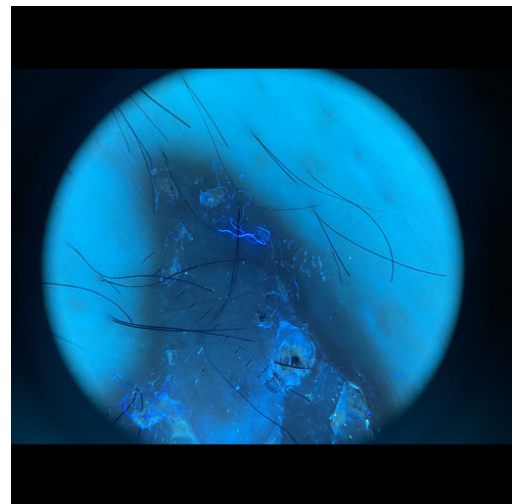
(a) NON-UV — Esempio 1



(b) UV — Esempio 1



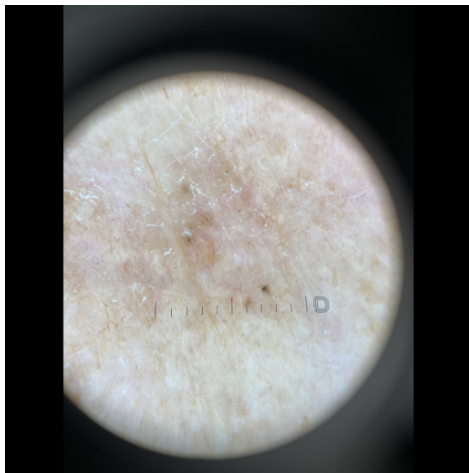
(c) NON-UV — Esempio 2



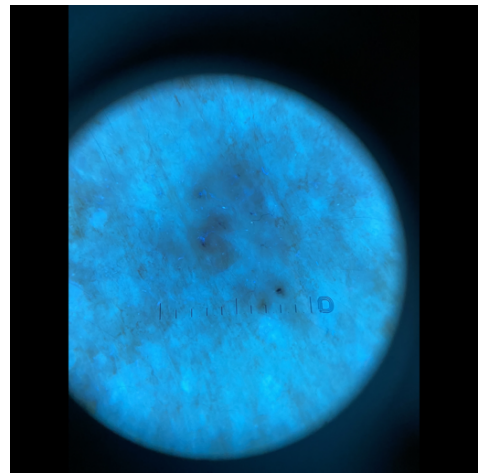
(d) UV — Esempio 2

**Figura 3.6:** Esempi di lesioni in cui i dermatologi concordano sulla modalità più estesa, mentre il modello fornisce una valutazione discordante.

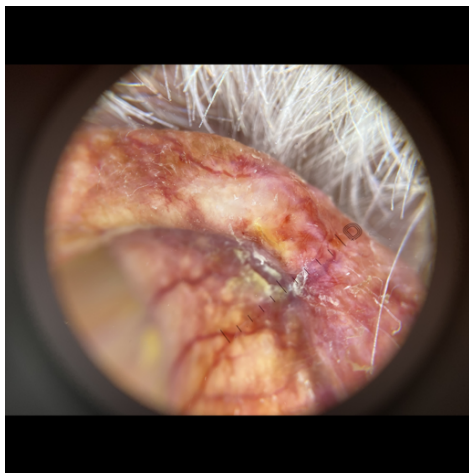
Un secondo insieme di casi è rappresentato dalle situazioni in cui i **dermatologi concordano tra loro**, ma la **valutazione del modello è discordante**. Si tratta spesso di lesioni caratterizzate da pattern irregolari di fluorescenza, da variazioni di luminosità o da artefatti (riflessi, peli, marcature con pennarello), che possono indurre la rete a sovrastimare o sottostimare le aree periferiche rispetto al giudizio clinico. Questi esempi sono particolarmente utili per identificare i limiti del modello e per comprendere quali caratteristiche visive risultano più difficili da apprendere. La Figura 3.6 mostra quattro esempi di questa categoria.



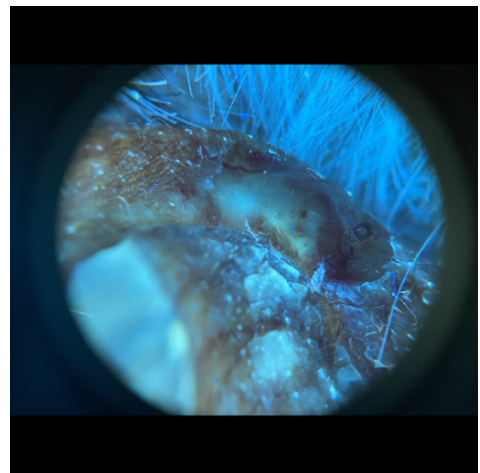
(a) NON-UV — Esempio 1



(b) UV — Esempio 1



(c) NON-UV — Esempio 2



(d) UV — Esempio 2

**Figura 3.7:** Esempi di lesioni per le quali non vi è accordo tra i dermatologi riguardo alla modalità più estesa (UV o NON-UV).

Infine, un terzo gruppo comprende le lesioni per le quali **non vi è accordo tra i dermatologi stessi**. In questi casi la definizione dei margini risulta complessa: la fluorescenza può essere irregolare o disomogenea, il contrasto con lo sfondo ridotto, oppure la lesione può presentare morfologie atipiche che rendono difficile determinare quale modalità ne mostri una porzione maggiore. Si tratta delle situazioni più ambigue dal punto di vista clinico, in cui la variabilità inter-osservatore è intrinsecamente elevata e nessun modello automatico può aspirare a una risposta univoca. La Figura 3.7 riporta alcuni esempi di queste condizioni.

Nel complesso, questi esempi qualitativi permettono di visualizzare in modo diretto la variabilità interpretativa delle lesioni e di comprendere meglio in quali scenari il modello riesce ad allinearsi al comportamento clinico e in quali invece permangono criticità legate alla natura stessa delle immagini o ai limiti del processo di apprendimento.

### 3.7.2 Analisi statistica delle maschere

In questa sezione vengono presentati i risultati dei test statistici descritti nel Capitolo 2.9.4, con l'obiettivo di valutare in modo quantitativo:

- l'eventuale differenza tra le aree segmentate in UV e NON-UV per ciascun operatore;
- il grado di coerenza inter-osservatore;
- la similarità tra il modello di segmentazione e i dermatologi.

I risultati sono organizzati in tre parti: analisi non paired UV vs NON-UV, analisi paired tra operatori e analisi paired modello-clinici.

#### 3.7.3 Confronto non paired UV vs NON-UV

La Tabella 3.10 riporta i risultati del test di Mann-Whitney U per ciascun operatore. In tutti i casi le distribuzioni NON-UV e UV non rispettavano l'ipotesi di normalità, rendendo il test non parametrico l'approccio più appropriato.

**Tabella 3.10:** Risultati del test non paired tra aree NON-UV e UV per ciascun operatore. Vengono riportati i p-value dei test di normalità, media e mediana delle due distribuzioni, la statistica del test di Mann–Whitney e il p-value associato.

Operatore	$p_{\text{norm,nonuv}}$	$p_{\text{norm,uv}}$	Media NON-UV	Media UV	p-value
Derm1	0.000816	0.214982	$4.22 \times 10^4$	$5.91 \times 10^4$	<b>0.000072</b>
Derm2	0.000265	0.007027	$2.29 \times 10^6$	$2.78 \times 10^6$	<b>0.044639</b>
Derm3	0.001733	0.012813	$2.35 \times 10^6$	$2.89 \times 10^6$	<b>0.027923</b>
Derm4	0.000166	0.001340	$2.69 \times 10^6$	$3.01 \times 10^6$	0.211348
Rete	0.011760	0.576932	$4.97 \times 10^4$	$6.58 \times 10^4$	<b>0.000086</b>

### Interpretazione

I risultati indicano che:

- per i **Dermatologi 1, 2 e 3** la differenza tra UV e NON-UV è statisticamente significativa ( $p < 0.05$ ), con la modalità UV che mostra un ampliamento dell'area segmentata;
- il **Dermatologo 4** presenta un ampliamento non statisticamente significativo, a causa dell'elevata variabilità delle differenze UV–NON-UV;
- anche la **rete** mostra una differenza significativa  $UV > \text{NON-UV}$ , con una magnitudine simile a quella osservata per Derm1.

Nel complesso, la fluorescenza UV tende a evidenziare un'estensione maggiore della lesione per la maggior parte degli operatori, pur con differenze nel grado di consistenza.

### 3.7.4 Confronto paired tra dermatologi

La Tabella 3.11 riporta i p-value ottenuti tramite test di Wilcoxon signed-rank applicati alle differenze  $\Delta A$  dei diversi operatori. La matrice è simmetrica e la diagonale principale è posta a zero.

**Tabella 3.11:** P-value del test di Wilcoxon signed-rank tra le differenze UV–NON-UV per ciascuna coppia di dermatologi (triangolo superiore).

	D1	D2	D3	D4
D1	–	4.5e-10	3.9e-07	4.3e-05
D2	–	–	0.893	0.823
D3	–	–	–	0.363
D4	–	–	–	–

### Interpretazione

Dai confronti emergono due comportamenti distinti:

- il **Dermatologo 1** risulta significativamente diverso da tutti gli altri operatori ( $p < 0.001$ ), adottando un criterio più conservativo nell’interpretazione dell’estensione UV.
- I **Dermatologi 2, 3 e 4** non presentano differenze significative tra loro ( $p > 0.36$ ), formando un gruppo omogeneo con un comportamento più estensivo.

Questi risultati confermano una variabilità inter-osservatore marcata e una distinzione tra un approccio più prudente (Dermatologo 1) e uno più espansivo (Dermatologo 2–Dermatologo 4).

### 3.7.5 Confronto paired tra modello e dermatologi

La Tabella 3.12 riassume i risultati del test di Wilcoxon between model and clinicians.

**Tabella 3.12:** P-value del test di Wilcoxon signed-rank tra il modello e i quattro dermatologi.

	D1	D2	D3	D4
Rete	0.535	6.5e-10	4.6e-07	4.6e-05

### Interpretazione

- la rete è statisticamente indistinguibile dal Dermatologo 1 ( $p = 0.535$ );
- la rete è invece **significativamente diversa** da Dermatologi 2, 3 e 4 ( $p < 0.001$ ).



Questi risultati indicano che il modello ha appreso un criterio di segmentazione maggiormente allineato a quello di Derm1, caratterizzato da un ampliamento UV più contenuto rispetto al gruppo estensivo D2–D4.

### 3.7.6 Discussione complessiva dei risultati

L'analisi statistica mette in evidenza come la fluorescenza UV tenda a mostrare un'estensione maggiore della lesione per la maggior parte degli operatori, pur con variabilità individuale. I dermatologi si suddividono in due cluster: un gruppo conservativo (*Derm1*) e un gruppo estensivo (*Derm2–Derm4*). Il modello di segmentazione riproduce il comportamento del gruppo conservativo, suggerendo una maggiore cautela nel seguire segnali di fluorescenza meno marcati.

Questo risultato è coerente con l'elevata variabilità delle annotazioni UV e indica che il modello, anziché ampliare eccessivamente i margini UV, tende a privilegiare un'interpretazione più prudente, potenzialmente vantaggiosa in un contesto clinico dove la sovra-segmentazione potrebbe non essere desiderata.

L'analisi delle maschere manuali e delle predizioni automatiche ha permesso di caratterizzare in modo approfondito come la fluorescenza UV influenzi la percezione dei margini e come il modello di segmentazione si comporti rispetto ai dermatologi. In una prima fase, il confronto qualitativo tra UV e NON-UV ha evidenziato che, nella maggior parte dei casi, la fluorescenza tende a mostrare un'estensione maggiore della lesione, fenomeno osservato in tutte le annotazioni manuali ma con variabilità tra gli operatori. La classificazione in accordo totale o disaccordo ha inoltre messo in luce la presenza di casi a elevata ambiguità interpretativa, attribuibili a margini poco definiti o a fluorescenza non uniforme, nei quali i dermatologi non convergono verso la stessa valutazione.

I test statistici non paired hanno confermato e quantificato queste osservazioni. Per tre dermatologi su quattro (*Derm1*, *Derm2* e *Derm3*) le differenze tra le aree UV e NON-UV risultano statisticamente significative, indicando una tendenza sistematica della fluorescenza UV ad ampliare la superficie percepita della lesione. Solo per *Derm4* tale differenza non è risultata significativa, coerentemente con la maggiore variabilità riscontrata nelle sue annotazioni manuali. Anche la rete mostra una differenza  $UV > NON-UV$  statisticamente significativa, suggerendo che il modello riconosca un'informazione addizionale nelle immagini UV.

L'analisi paired tra dermatologi, condotta confrontando le differenze  $\Delta A = A_{UV} - A_{NONUV}$ , ha permesso di valutare il grado di coerenza inter-osservatore. I risultati rivelano l'esistenza di due comportamenti distinti: *Derm1* presenta un profilo significativamente diverso da tutti gli altri operatori, adottando un criterio più conservativo nel valutare l'estensione UV; al contrario, *Derm2*, *Derm3* e *Derm4* formano un gruppo omogeneo, caratterizzato da un ampliamento UV più marcato e statisticamente simile tra loro. Questi due cluster riflettono una variabilità



clinica reale nell'interpretazione del segnale UV, già evidenziata dalla classificazione qualitativa ma ora confermata da una valutazione statistica formale.

Il confronto paired tra il modello e i dermatologi completa il quadro. La rete risulta statisticamente indistinguibile da Derm1, mentre differisce in modo significativo da Derm2, Derm3 e Derm4. Questo risultato suggerisce che il modello abbia appreso un criterio più prudente, allineato all'approccio conservativo rappresentato da Derm1, evitando di ampliare eccessivamente i margini nelle immagini UV. Tale comportamento appare coerente con la filosofia del modello, che tende a generalizzare in modo stabile su dati a elevata variabilità e potrebbe risultare vantaggioso in contesti clinici dove un'eccessiva espansione dei margini non è desiderabile.

Nel loro insieme, queste analisi mostrano come la fluorescenza UV fornisca informazioni aggiuntive sulla morfologia della lesione, ma anche come la sua interpretazione sia soggetta a variabilità clinica. La rete, pur riconoscendo la tendenza all'ampliamento UV, si posiziona su un comportamento più conservativo e riproduce in modo coerente uno dei criteri adottati manualmente. Questi risultati offrono indicazioni importanti per una futura integrazione dell'informazione UV nei percorsi diagnostici e pre-chirurgici, suggerendo un ruolo complementare tra segmentazione automatica e valutazione clinica.

### 3.8 Confronto con modello presente in letteratura

Per valutare in modo più approfondito la qualità delle segmentazioni ottenute dalla pipeline proposta, è stato eseguito un confronto con un modello presente in letteratura: il *Boundary-aware Transformer* (BA-Transformer), introdotto da Wang et al. nel lavoro "Boundary-aware Transformers for Skin Lesion Segmentation" [20].

Il BA-Transformer rappresenta una delle architetture più recenti dedicate alla segmentazione dermoscopica e combina tre componenti principali:

- un backbone **ResNet50** per l'estrazione delle feature locali;
- un **modulo Transformer globale** (Trans.), in grado di modellare relazioni spaziali a lungo raggio e di catturare la struttura complessiva della lesione;
- un **boundary-refinement module** (BAG), progettato per migliorare in modo esplicito la definizione dei bordi.

Nel lavoro originale, l'architettura viene validata su tre dataset pubblici ISIC 2016 [21], ISIC 2018 [22] e PH2 [23] ottenendo risultati *state-of-the-art* su metriche come IoU, Dice, HD95 e ASSD. È importante sottolineare che tali dataset sono

costituiti esclusivamente da immagini dermoscopiche convenzionali, acquisite in condizioni controllate e prive di fluorescenza UV.

**Tabella 3.13:** Prestazioni riportate dagli autori del BA-Transformer sui dataset pubblici ISIC 2016 + PH2 e ISIC 2018 (valori tratti dal paper originale).

Trans.	BAG	ISIC 2016 + PH2		ISIC 2018	
		Dice $\uparrow$	IoU $\uparrow$	Dice $\uparrow$	IoU $\uparrow$
$\checkmark$		0.884	0.805	0.879	0.810
	$\checkmark$	0.900	0.827	0.890	0.821
$\checkmark$	$\checkmark$	<b>0.921</b>	<b>0.858</b>	<b>0.912</b>	<b>0.843</b>

Il confronto effettuato in questo lavoro permette quindi di valutare quanto un modello molto potente e ottimizzato per il dominio ISIC sia in grado di generalizzare anche su un dataset clinico reale caratterizzato da un marcato *domain shift* sia cromatico sia morfologico.

### 3.8.1 Validazione del BA-Transformer sul dataset HAM10000

Prima di applicare il BA-Transformer al dataset clinico di Novara, è stato eseguito un passaggio di validazione per verificare che il modello, allenato tramite il codice messo a disposizione dagli autori, fosse in grado di raggiungere prestazioni comparabili a quelle riportate nel lavoro originale.

Il modello è stato addestrato sul dataset HAM10000 utilizzando la stessa pipeline di preprocessing descritta nella Sezione 2: ridimensionamento, *color constancy* e *data augmentation*. Il fine di questa fase non era ottenere prestazioni massime, ma confermare la correttezza dell’implementazione e la capacità del modello di apprendere correttamente la morfologia delle lesioni.

La Tabella 3.14 riassume il confronto diretto tra le prestazioni del BA-Transformer e quelle della U-Net proposta, includendo le due metriche principali di segmentazione (IoU e Dice).

**Tabella 3.14:** Confronto tra BA-Transformer e U-Net sul dataset HAM10000 (media  $\pm$  deviazione standard).

Modello	IoU	Dice
BA-Transformer	0.8440 $\pm$ 0.1401	0.9078 $\pm$ 0.1012
U-Net (proposta)	0.8200 $\pm$ 0.1850	0.8860 $\pm$ 0.1540

Le prestazioni ottenute risultano coerenti con le performance riportate nel paper originale, confermando la correttezza dell’implementazione. Il BA-Transformer mostra metriche leggermente superiori nel dominio ISIC, come atteso, grazie alla capacità dei moduli Transformer di modellare relazioni globali e migliorare la delineazione dei bordi.

Nel complesso, questi risultati confermano che il modello è correttamente implementato e pienamente funzionale, ponendo le basi per un confronto affidabile con la U-Net sul dataset clinico di Novara.

### 3.8.2 Confronto sulle immagini non–UV

Per valutare la capacità dei modelli di generalizzare sul dominio clinico reale, è stato eseguito un confronto diretto tra il BA-Transformer e la U-Net proposta sulle 84 immagini non–UV del dataset di Novara. Le performance riportate in tabella 3.15 sono quelle ottenute dopo il post–processing morfologico per entrambe le predizioni effettuate dai due modelli, così da garantire una valutazione coerente e comparabile.

**Tabella 3.15:** Confronto tra BA-Transformer e U-Net proposta sulle immagini non–UV del dataset di Novara (media  $\pm$  deviazione standard).

Modello	IoU	Dice
BA-Transformer (originale)	$0.4256 \pm 0.1758$	$0.5747 \pm 0.1825$
U-Net (CC + DA)	<b><math>0.5890 \pm 0.1770</math></b>	<b><math>0.7224 \pm 0.1720</math></b>

Nonostante entrambi i modelli siano stati addestrati e testati sulle stesse immagini preprocessate, la U-Net mostra prestazioni nettamente superiori, con un incremento di oltre **16 punti percentuali** in IoU e circa **15 punti percentuali** nel Dice rispetto al BA-Transformer.

Questo risultato indica che, la U-Net si dimostra più stabile e più capace di generalizzare in presenza di marcato *domain shift*, mentre la maggiore complessità del BA-Transformer non si traduce in un miglioramento operativo sulle immagini cliniche reali.

### 3.8.3 Confronto sulle immagini UV

Per garantire un confronto equo nel dominio UV, il BA-Transformer è stato riaddestrato utilizzando lo stesso preprocessing impiegato per la U-Net, basato sull’estrazione del solo canale di luminanza  $L^*$  dello spazio colore LUV, seguito dalle stesse tecniche di *data augmentation*. Entrambi i modelli sono quindi valutati

sulle stesse 84 immagini UV e con lo stesso post-processing morfologico, così da assicurare una comparazione coerente.

I risultati sono riportati in Tabella 3.16.

**Tabella 3.16:** Confronto tra BA-Transformer e U-Net proposta sulle immagini UV del dataset di Novara (media  $\pm$  deviazione standard).

Modello	IoU	Dice
BA-Transformer ( $L^* + DA$ )	$0.5250 \pm 0.1566$	$0.6725 \pm 0.1557$
<b>U-Net (<math>L^* + DA</math>)</b>	<b><math>0.5326 \pm 0.1741</math></b>	<b><math>0.6930 \pm 0.1780</math></b>

Nonostante le immagini UV presentino caratteristiche molto differenti rispetto ai dataset dermoscospici tradizionali — tra cui distribuzioni di luminanza irregolari, fluorescenza eterogenea e una maggiore presenza di artefatti — la U-Net ottiene prestazioni leggermente superiori a quelle del BA-Transformer, con un incremento di circa 1 punto percentuale in IoU e oltre 2 punti percentuali nel Dice.

Nel complesso, la U-Net preprocessata in luminanza si dimostra più affidabile per la segmentazione delle immagini UV reali, confermando la necessità di adattare il preprocessing al dominio piuttosto che aumentare la complessità architetturale.

## Capitolo 4

# Conclusioni e sviluppi futuri

### 4.1 Sintesi dei risultati principali

In questo lavoro è stata analizzata l'utilità della dermoscopia in fluorescenza UV per la demarcazione dei margini del carcinoma basocellulare (BCC) e valutata la possibilità di integrare tale informazione con un sistema di segmentazione automatica basato su architettura U-Net. Il modello è stato addestrato sul dataset pubblico HAM10000, opportunamente preprocessato tramite normalizzazione geometrica, color constancy e data augmentation mirata alla classe BCC, e successivamente validato su un dataset clinico reale costituito da 84 coppie di immagini NON-UV e UV.

I risultati sperimentali mostrano che:

- la fluorescenza UV evidenzia in numerosi casi un'estensione maggiore della lesione rispetto alla dermoscopia convenzionale, suggerendo un potenziale valore nella fase pre-chirurgica;
- il modello proposto ottiene prestazioni competitive sul dataset di test pubblico e una buona capacità di generalizzazione, pur in presenza di condizioni di acquisizione cliniche altamente variabili;
- l'introduzione della color constancy e del canale di luminanza  $L^*$  migliora sensibilmente la robustezza della segmentazione nei due domini di acquisizione;
- il confronto con il modello di letteratura BA-Transformer mostra come la U-Net personalizzata risulti più stabile nelle immagini NON-UV e comparabile nelle immagini UV.

Complessivamente, i risultati confermano il potenziale dell'imaging UV e del modello proposto come strumenti utili per supportare la valutazione dei margini tumorali.

## 4.2 Contributi del lavoro svolto

Il contributo di questa tesi può essere sintetizzato nei seguenti punti principali:

- sviluppo di una pipeline completa per la segmentazione dermoscópica, comprendente preprocessing avanzato, addestramento, validazione e post-processing delle predizioni;
- introduzione di due percorsi di preprocessing indipendenti (NON-UV e UV), basati rispettivamente su normalizzazione cromatica tramite GCC-GAN e su estrazione della luminanza nello spazio CIELUV;
- creazione di un dataset clinico annotato, ottenuto tramite un'applicazione dedicata che ha permesso la generazione di maschere manuali da parte di quattro dermatologi e la produzione di un ground truth tramite majority voting;
- analisi sistematica della concordanza tra dermatologi e confronto con il comportamento del modello, evidenziando casi di accordo, disaccordo e ambiguità clinica;
- confronto diretto con un modello di letteratura (BA-Transformer) per valutare il posizionamento della rete proposta rispetto allo stato dell'arte.

## 4.3 Limiti dello studio

Nonostante i risultati incoraggianti, il presente lavoro presenta alcuni limiti:

- il dataset clinico è di dimensioni ridotte e comprende esclusivamente immagini acquisite con un unico dermatoscopio, limitando la variabilità inter-dispositivo;
- la fluorescenza UV, pur evidenziando in diversi casi un'estensione maggiore, risulta in alcune immagini difficilmente interpretabile a causa di artefatti, riflessi, peli o scarso contrasto;
- il modello non è stato addestrato direttamente su immagini reali UV, poiché non esistono dataset pubblici sufficientemente ampi, costringendo a un adattamento tramite estrazione della luminanza;
- la mancanza di una validazione istologica dei margini limita la possibilità di stabilire un collegamento diretto tra estensione delle maschere e margini chirurgici reali;
- il confronto con altri modelli deep learning è stato limitato dal vincolo computazionale e dalle differenze tra domini di acquisizione.

## 4.4 Possibili sviluppi futuri

Alla luce dei risultati ottenuti, diverse linee di sviluppo futuro risultano promettenti:

- ampliare il dataset clinico includendo nuove acquisizioni, diversi dispositivi e casi con validazione istologica per una valutazione più accurata dell'utilità clinica;
- esplorare tecniche di domain adaptation o transfer learning per ridurre la distanza tra immagini pubbliche e immagini cliniche UV reali;
- integrare modelli più recenti basati su architetture transformer o modelli ibridi CNN-Transformer, ottimizzati per la segmentazione ad alta sensibilità dei margini;
- sviluppare sistemi multimodali che combinino simultaneamente informazioni NON-UV e UV, sfruttando la complementarità dei due domini;
- implementare un prototipo software utilizzabile in sala operatoria per la visualizzazione in tempo reale delle maschere segmentate e del confronto UV/NON-UV;
- valutare la possibilità di utilizzare le predizioni automatiche per stimare margini chirurgici suggeriti, come ausilio alla pianificazione della chirurgia micrografica di Mohs.

In conclusione, il lavoro pone le basi per l'integrazione di imaging UV e segmentazione automatica nella pratica clinica, aprendo la strada a strumenti di supporto che possano migliorare la precisione nella definizione dei margini tumorali e contribuire a ottimizzare le procedure di chirurgia dermatologica.

# Bibliografia

- [1] K. M. Hosny, D. Elshoura, E. R. Mohamed, E. Vrochidou e G. A. Papakostas. «Deep Learning and Optimization-Based Methods for Skin Lesions Segmentation: A Review». In: *IEEE Access* 11 (2023), pp. 85467–85488 (cit. alle pp. 1, 5–7).
- [2] Z. Mirikharaji, K. Abhishek, A. Bissoto, C. Barata, S. Avila, E. Valle, M. Celebi e G. Hamarneh. «A survey on deep learning for skin lesion segmentation». In: *Medical Image Analysis* 88 (2023) (cit. alle pp. 1–8).
- [3] Baba P.U.F., Hassan A.u. e Khurshid J.and Wani A.H. «Basal Cell Carcinoma: Diagnosis, Management and Prevention.» In: *Journal of Molecular Pathology* 5.2 (2024), pp. 153–170. DOI: <https://doi.org/10.3390/jmp5020010> (cit. a p. 1).
- [4] M. M. Stofa, M. A. Zulkifley e M. M. Zainuri. «Skin Lesions Classification and Segmentation: A Review». In: *International Journal of Advanced Computer Science and Applications* 12 (gen. 2021) (cit. alle pp. 1, 2, 6, 8).
- [5] H. Husein-ElAhmed, M. T. Gutiérrez-Salmerón, R. Naranjo-Sintes e R. Aneiros-Cachaza. «Sclerodermiform basal cell carcinoma: How much can we rely on dermatoscopy to differentiate from non-aggressive basal cell carcinomas? Analysis of 1 256 cases». In: *Anais Brasileiros de Dermatologia* 93.2 (2018), pp. 229–232. DOI: 10.1590/abd1806-4841.20186699 (cit. a p. 1).
- [6] J. A. Mojeski, M. Almashali, P. Jowdy, M. E. Fitzgerald, K. L. Brady, N. C. Zeitouni, O. R. Colegio e G. Paragh. «Ultraviolet imaging in dermatology». In: *Photodiagnosis and Photodynamic Therapy* 30 (mar. 2020) (cit. alle pp. 3, 4).
- [7] K. Kojima, K. Shido e G. Tamiya. «Facial UV photo imaging for skin pigmentation assessment using conditional generative adversarial networks». In: *Sci Rep* 11 (gen. 2021) (cit. alle pp. 4, 7).
- [8] Skin Cancer Foundation. «Mohs Surgery». In: (jun, 2021). URL: <https://www.skincancer.org/it/treatment-resources/mohs-surgery/> (cit. alle pp. 4, 5).



- [9] Skin Doctors' Center. *Chirurgia micrografica di Mohs*. Accessed: 21-Nov-2025. s.d. URL: <https://skindoctors.it/it-IT/servizio/chirurgia-dermatologica/chirurgia-micrografica-di-mohs-11.html> (cit. a p. 4).
- [10] L. Yu, H. Chen, Q. Dou, J. Qin e P. A. Heng. «Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks». In: *IEEE Transactions on Medical Imaging*. Vol. 36. 4. 2017, pp. 994–1004. DOI: 10.1109/TMI.2016.2642839 (cit. a p. 6).
- [11] L. Bi, J. Kim, E. Ahn e D. Feng. «Dermoscopic Image Segmentation via Multi-Stage Fully Convolutional Networks». In: *IEEE Trans Biomed Eng.* 2017, pp. 2065–2074. DOI: 10.1109/TBME.2017.2712771. Epub2017Jun7. PMID: 28600236 (cit. a p. 6).
- [12] Y. Yuan, M. Chao e Y. Lo. «Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance». In: *IEEE Transactions on Medical Imaging* 36.9 (2017), pp. 1876–1886. DOI: doi:10.1109/TMI.2017.2695227. Epub2017Apr18. PMID: 28436853 (cit. a p. 7).
- [13] M. Goyal e M. H. Yap. «Multi-Class Semantic Segmentation of Skin Lesions via Fully Convolutional Networks». In: *arXiv preprint arXiv:1711.10449* (2018) (cit. a p. 7).
- [14] M. Goyal, A. Oakley, P. Bansal, D. Dancey e M.H. Yap. «Skin Lesion Segmentation in Dermoscopic Images With Ensemble Deep Learning Methods». In: *IEEE Access* 8 (gen. 2020), pp. 4171–4181 (cit. a p. 8).
- [15] P. Tschandl, C. Rosendahl e H. Kittler. *HAM10000: A large collection of multi-source dermatoscopic images of common pigmented skin lesions*. Kaggle dataset: Skin Cancer MNIST – HAM10000. Accessed: 15-Nov-2025. 2018. URL: <https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000> (cit. a p. 10).
- [16] AdoMedical. *Dermatoscopio DermLite DL5 Wood 365nm UV*. <https://adomedical.it/it/dermatoscopi-foto-e-video-dermatoscopi/1126-dermatoscopio-dermlite-dl5-wood-365nm-UV>. Accessed: 22-Nov-2025. s.d. (Cit. a p. 11).
- [17] M. Salvi, F. Branciforti, F. Molinari e K.M. Meiburger. «GCC-GAN and GSN-GAN script and dataset». In: *Mendeley Data V2* (2024), doi: 10.17632/32bvf6xhj.2 (cit. a p. 16).
- [18] O. Ronneberger, P. Fischer e T. Brox. «U-Net: Convolutional Networks for Biomedical Image Segmentation». In: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4\_28 (cit. a p. 19).

- [19] A. A. Taha e A. Hanbury. «Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool». In: *BMC Medical Imaging* 15.29 (2015). DOI: <https://doi.org/10.1186/s12880-015-0068-x> (cit. a p. 27).
- [20] Jiacheng Wang, Lan Wei, Liansheng Wang, Qichao Zhou, Lei Zhu e Jing Qin. «Boundary-Aware Transformers for Skin Lesion Segmentation». In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Springer International Publishing, 2021, pp. 206–216. ISBN: 9783030871932. DOI: 10.1007/978-3-030-87193-2\_20. URL: [http://dx.doi.org/10.1007/978-3-030-87193-2\\_20](http://dx.doi.org/10.1007/978-3-030-87193-2_20) (cit. a p. 50).
- [21] M. H. Tasin. *ISIC-2016 Original Dataset*. Kaggle dataset: ISIC 2016 Original Dataset. Accessed: 21-Nov-2025. 2016. URL: <https://www.kaggle.com/datasets/mahmudulhasantasin/isic-2016-original-dataset> (cit. a p. 50).
- [22] T. Tschandl, C. Rosendahl e H. Kittler. *ISIC2018 Challenge Task 1 Data: Segmentation*. Kaggle dataset: ISIC2018 Challenge Task1 – Segmentation. Accessed: 21-Nov-2025. 2018. URL: <https://www.kaggle.com/datasets/tschandl/isic2018-challenge-task1-data-segmentation> (cit. a p. 50).
- [23] T. Mendonça, P. M. Ferreira, J. Marques, A. R. S. Marçal e J. Rozeira. *PH<sup>2</sup>: A Dermoscopic Image Database for Research and Benchmarking*. Kaggle dataset: PH2 Dataset. Accessed: 21-Nov-2025. 2013. URL: <https://www.kaggle.com/datasets/spacesurfer/ph2-dataset> (cit. a p. 50).