

POLITECNICO DI TORINO

Collegio di Ingegneria Biomedica

**Corso di Laurea Magistrale
in Ingegneria Biomedica-Curriculum “Strumentazione
Biomedica”**

Tesi di Laurea Magistrale

**Facial Muscle Segmentation in Ultrasound
Imaging: Evaluating Image Preprocessing
and Medical Foundation Models**



Relatori

prof.ssa Kristen Meiburger

prof. Francesco Marzola

Candidato

Fabrizio Fiorello

20 Luglio 2026

ABSTRACT

Background. La distrofia muscolare facio-scapolo-omeroale (Facioscapulohumeral Muscular Dystrophy, FSHD) è una patologia neuromuscolare ereditaria caratterizzata da una progressiva degenerazione del tessuto muscolare e da un coinvolgimento precoce della muscolatura facciale. L'ecografia muscolare rappresenta una metodica non invasiva, accessibile e ripetibile per la valutazione del tessuto muscolare. Tuttavia, la segmentazione automatica dei muscoli facciali rimane complessa a causa delle ridotte dimensioni anatomiche, della variabilità morfologica, del basso contrasto rispetto ai tessuti circostanti e della presenza di rumore speckle.

Obiettivo. Il presente lavoro ha avuto l'obiettivo di analizzare possibili strategie per migliorare la segmentazione dei muscoli facciali in immagini ecografiche, valutando l'effetto di diverse tecniche di preprocessing e il comportamento di un foundation model medico applicato alla segmentazione guidata da prompt.

Metodi. Lo studio è stato condotto su un dataset di immagini ecografiche muscolari acquisito presso il Radboud University Medical Center. Sono state considerate diverse tecniche di miglioramento del contrasto e riduzione del rumore, tra cui Contrast Limited Adaptive Histogram Equalization (CLAHE), una configurazione più conservativa denominata CLAHE_LOW, Guided Filtering, Speckle Reducing Anisotropic Diffusion, Adaptive Non-Local Means e Stationary Wavelet Transform. L'effetto del preprocessing è stato inizialmente valutato sui quattro muscoli facciali principali, Depressor Anguli Oris, Mentalis, Orbicularis Oris e Zygomaticus, mediante una pipeline basata su MMSegmentation, Swin Transformer, UPerNet e K-Net. Successivamente, le configurazioni più rilevanti sono state applicate a MedSAM, un foundation model per la segmentazione di immagini mediche, utilizzato a partire da un checkpoint pre-addestrato e senza fine-tuning sul dataset considerato. Per MedSAM, la segmentazione è stata guidata da bounding box derivate dalle maschere manuali di riferimento. Le prestazioni sono state valutate mediante Dice Similarity Coefficient, Intersection over Union e segmentation failure, definita come Dice Similarity Coefficient inferiore a 0,5.

Risultati. Gli esperimenti condotti con la pipeline supervisionata implementata in MMSegmentation hanno confermato la difficoltà della segmentazione dei muscoli facciali. Sulle immagini originali, la pipeline basata su Swin Transformer, UPerNet e K-Net ha ottenuto un Dice medio pari a $0,504 \pm 0,289$ e una IoU media pari a $0,383 \pm 0,244$, con una segmentation failure pari al 38,28%. L'applicazione del preprocessing non ha prodotto un miglioramento globale uniforme: CLAHE_LOW ha ottenuto valori vicini alle immagini originali, con Dice medio pari a $0,500 \pm 0,291$, IoU media pari a $0,381 \pm 0,246$ e segmentation failure pari al 38,80%. MedSAM ha mostrato prestazioni più elevate nella segmentazione guidata da bounding box dei quattro muscoli facciali principali, ottenendo un Dice medio pari a $0,741 \pm 0,124$ sulle immagini originali, $0,733 \pm 0,125$ con CLAHE e $0,737 \pm 0,124$ con CLAHE_LOW. Le corrispondenti IoU medie sono state pari a $0,602 \pm 0,143$, $0,593 \pm 0,143$ e $0,597 \pm 0,142$. La segmentation failure si è ridotta rispettivamente al 5,91%, 6,18% e 6,01%. Considerando l'intero dataset dei 16 muscoli, MedSAM ha ottenuto valori medi globali di Dice superiori a 0,80 in tutte le condizioni analizzate, con differenze limitate tra immagini originali, CLAHE e CLAHE_LOW.

Conclusioni. I risultati indicano che il solo preprocessing non è sufficiente a risolvere in modo stabile le difficoltà associate alla segmentazione automatica dei muscoli facciali.

CLAHE_LOW rappresenta una configurazione moderata e complessivamente equilibrata, ma il suo vantaggio rimane limitato e dipendente dal muscolo considerato. MedSAM ha mostrato buone prestazioni quando la regione di interesse viene fornita tramite bounding-box prompt, ma nella configurazione valutata non costituisce una pipeline completamente automatica end-to-end. Futuri sviluppi dovranno quindi concentrarsi sulla generazione automatica dei prompt, sull'adattamento dei foundation models al dominio ecografico e sulla validazione esterna su dati acquisiti presso centri e dispositivi differenti.

Abstract (english)

Background. Facioscapulohumeral muscular dystrophy (FSHD) is an inherited neuromuscular disorder characterized by progressive skeletal muscle degeneration and early involvement of the facial muscles. Muscle ultrasound is a non-invasive, accessible, and repeatable imaging technique for the assessment of muscle tissue. However, automatic segmentation of facial muscles remains challenging because of their small anatomical size, morphological variability, low contrast with surrounding tissues, and the presence of speckle noise.

Objective. This thesis aimed to investigate possible strategies to improve facial muscle segmentation in ultrasound images by evaluating the effect of different preprocessing techniques and the behavior of a medical foundation model applied to prompt-guided segmentation.

Methods. The study was conducted on a dataset of muscle ultrasound images acquired at Radboud University Medical Center. Several contrast-enhancement and denoising techniques were considered, including Contrast Limited Adaptive Histogram Equalization (CLAHE), a more conservative configuration termed CLAHE_LOW, Guided Filtering, Speckle Reducing Anisotropic Diffusion, Adaptive Non-Local Means, and Stationary Wavelet Transform. The effect of preprocessing was initially evaluated on the four main facial muscles, Depressor Anguli Oris, Mentalis, Orbicularis Oris, and Zygomaticus, using a segmentation pipeline based on MMSegmentation, Swin Transformer, UPerNet, and K-Net. The most relevant configurations were then applied to MedSAM, a foundation model for medical image segmentation, used from a pre-trained checkpoint without fine-tuning on the dataset considered. For MedSAM, segmentation was guided by bounding boxes derived from the manual reference masks. Performance was assessed using the Dice Similarity Coefficient, Intersection over Union, and segmentation failure, defined as a Dice Similarity Coefficient below 0.5.

Results. Experiments performed with the supervised pipeline implemented in MMSegmentation confirmed the difficulty of segmenting facial muscles. On original images, the Swin Transformer, UPerNet and K-Net-based pipeline achieved a mean Dice score of 0.504 ± 0.289 and a mean IoU of 0.383 ± 0.244 , with a segmentation failure rate of 38.28%. Preprocessing did not lead to a consistent global improvement: CLAHE_LOW achieved values close to those obtained on original images, with a mean Dice score of 0.500 ± 0.291 , a mean IoU of 0.381 ± 0.246 , and a segmentation failure rate of 38.80%. MedSAM achieved higher performance in bounding-box-guided segmentation of the four main facial muscles, with mean Dice scores of 0.741 ± 0.124 on original images, 0.733 ± 0.125 with CLAHE, and 0.737 ± 0.124 with CLAHE_LOW. The corresponding mean IoU values were 0.602 ± 0.143 , 0.593 ± 0.143 , and 0.597 ± 0.142 . Segmentation failure decreased to 5.91%, 6.18%, and 6.01%, respectively. When considering the full dataset of 16 muscles, MedSAM achieved global mean Dice scores above 0.80 across all analyzed conditions, with limited differences between original, CLAHE, and CLAHE_LOW images.

Conclusion. The results indicate that preprocessing alone is not sufficient to consistently overcome the challenges associated with automatic segmentation of facial muscles. CLAHE_LOW represents a moderate and overall balanced configuration, but its advantage remains limited and muscle-dependent. MedSAM showed good performance when the region of interest was provided through a bounding-box prompt; however, in the evaluated configuration, it does not represent a fully automatic end-to-end pipeline. Future work should

therefore focus on automatic prompt generation, adaptation of foundation models to the ultrasound domain, and external validation on data acquired from different centers and devices.

Indice

Indice figure	ix
Indice tabelle	xi
1. Introduzione	1
2. Facioscapulohumeral Muscular Dystrophy (FSHD).....	5
3. Ecografia Muscolare Diagnostica	8
4. Dataset.....	12
4.1 Organizzazione del dataset per le analisi sperimentali.....	15
4.2 Identificazione del Problema: la Segmentazione dei Muscoli Facciali.....	16
4.3 Organizzazione dello studio sperimentale.....	18
5. Tecniche di preprocessing delle Immagini Ecografiche	22
5.1 Contrast Limited Adaptive Histogram Equalization (CLAHE).....	22
5.1.1 CLAHE_LOW	23
5.1.2 CLAHE_HIGH.....	25
5.2 Guided Filtering	25
5.2.1 Guided_LOW	26
5.3 Speckle Reducing Anisotropic Diffusion (SRAD)	27
5.4 Adaptive Non-Local Means (A-NLM)	28
5.5 Stationary Wavelet Transform (SWT)	29
5.6 Confronto Qualitativo delle Tecniche di Preprocessing	31
6. Metodi	34
6.1 MMSegmentation.....	34
6.1.1 Architettura Utilizzata	36
6.1.2 Swin Transformer.....	36
6.1.3 UPerNet.....	37
6.1.4 K-Net.....	38
6.1.5 Organizzazione degli esperimenti	38
6.2 Foundation Models per la segmentazione di Immagini Mediche	39
6.2.1 Segmentazione e Foundation Models	39
6.2.2 Segment Anything Model (SAM).....	40
6.2.3 Limiti di SAM nell'Imaging Medico.....	41
6.2.4 MedSAM.....	42
6.2.5 Motivazioni della Scelta di MedSAM.....	43
6.3 Metriche di Valutazione	44
7. Risultati sperimentali sui muscoli facciali	47
7.1 MMSegmentation.....	47

7.1.1 Confronto quantitativo globale.....	47
7.1.2 Confronto per singolo muscolo	49
7.1.3 Limiti del solo preprocessing	50
7.1.4 Considerazioni.....	51
7.2 MedSAM.....	52
7.2.1 Risultati sui quattro muscoli facciali principali.....	53
7.2.2 Risultati estesi sull'intero dataset - Original	55
7.2.3 Risultati estesi sull'intero dataset – CLAHE.....	57
7.2.4 Risultati estesi sull'intero dataset – CLAHE_LOW	58
7.2.5 Confronto collettivo tra le tecniche	60
7.3 Confronto tra MMSegmentation e MedSAM	63
7.3.1 Differenze metodologiche tra i due approcci	63
7.3.2 Confronto delle prestazioni	64
7.3.3 Sensibilità al preprocessing	65
7.3.4 Limiti del confronto diretto	65
7.3.5 Considerazioni finali	66
8. Discussione.....	67
8.1 Interpretazione complessiva dei risultati.....	67
8.2 Ruolo e limiti del preprocessing.....	67
8.3 Interpretazione dei risultati ottenuti con MedSAM.....	68
8.4 Limiti dello studio	68
8.5 Sintesi della discussione	69
9. Conclusioni e sviluppi futuri	71
9.1 Conclusioni.....	71
9.2 Sviluppi futuri	72
Bibliografia.....	74

Indice figure

Figura 2.1: Esempi di immagini ecografiche relative a quattro muscoli facciali analizzati	7
Figura 3.1: I quattro gradi della scala di Heckmatt utilizzata per classificare l'alterazione ecografica del muscolo	9
Figura 4.1: Un esempio di tre acquisizioni bilaterali per ciascun muscolo	12
Figura 4.2: Un esempio di immagine, la corrispondente maschera manuale e la sovrapposizione tra immagine e maschera	14
Figura 4.3: Schema della pipeline seguita in questo lavoro di tesi	15
Figura 5.1: Confronto di una stessa immagine prima e dopo l'applicazione del preprocessing tramite la tecnica CLAHE	23
Figura 5.2: Comparazione di un'immagine originale con la stessa immagine dopo aver utilizzato la tecnica CLAHE con i parametri meno aggressivi	24
Figura 5.3: La stessa immagine ecografica viene riportata nelle tre versioni per mostrare l'effetto del preprocessing	24
Figura 5.4: Immagini ottenute con CLAHE standard, CLAHE_LOW e CLAHE_HIGH.....	25
Figura 5.5: Comparazione di un'immagine originale con la stessa immagine dopo aver utilizzato la tecnica del Guided Filter	26
Figura 5.6: Immagine di confronto tra l'immagine originale e la tecnica del Guided Filtering utilizzando parametri meno aggressivi.....	27
Figura 5.7: L'effetto della diffusione anisotropa su un'immagine del muscolo zigomatico ..	28
Figura 5.8: Un esempio che mostra l'effetto del filtro A-NLM.....	29
Figura 5.9: Un esempio che mostra l'effetto della SWT	30
Figura 5.10: Confronto tra un'immagine originale e le versioni preprocessate mediante CLAHE, CLAHE_HIGH, SRAD, CLAHE_LOW, Non-Local Means, Stationary Wavelet Transform, Guided Filter e Guided Filter_LOW	31
Figura 6.1: Schema generale del flusso dati in un modello di segmentazione con MMSegmentation.....	34
Figura 6.2: Differenza tra Swin Transformer e Vision Transformer classico (ViT)	37
Figura 6.3: Schema generale del Segment Anything Model.....	40
Figura 6.4: Un esempio di applicazione di MedSAM alla segmentazione ecografica dei muscoli scheletrici	43
Figura 7.1: Dice medio generale ottenuto mediante MMSegmentation sui quattro muscoli facciali nelle quattro condizioni sperimentali considerate: ORIGINAL, CLAHE, CLAHE_LOW e GUIDED_LOW	48
Figura 7.2: Percentuale di segmentation failure ottenuta mediante MMSegmentation sui quattro muscoli facciali	48
Figura 7.3: Dice medio per ciascuno dei quattro muscoli facciali analizzati mediante MMSegmentation.....	49
Figura 7.4: Confronto qualitativo tra immagine originale, ground truth e maschere predette mediante MMSegmentation per un caso di Orbicularis oris caratterizzato da buona sovrapposizione tra predizione e annotazione manuale	50
Figura 7.5: Confronto qualitativo tra immagine originale, ground truth e maschere predette mediante MMSegmentation per un caso critico di Orbicularis oris.....	51
Figura 7.6: Un esempio di un'immagine ecografica originale, la maschera manuale di riferimento, la maschera predetta da MedSAM e l'overlay	56
Figura 7.7: Un esempio di un'immagine ecografica preprocessata mediante CLAHE, la maschera manuale di riferimento, la maschera predetta da MedSAM e l'overlay	58
Figura 7.8: Un esempio di un'immagine ecografica preprocessata mediante CLAHE_LOW, la maschera manuale di riferimento, la maschera predetta da MedSAM e l'overlay	60

Figura 7.9: Confronto complessivo delle prestazioni di MedSAM nelle tre condizioni analizzate.....	62
Figura 7.10: Confronto qualitativo delle segmentazioni ottenute da MedSAM su due casi appartenenti ai muscoli facciali.....	62

Indice tabelle

Tabella 4.1: Attribuzione ad ogni muscolo il proprio codice, la propria regione anatomica e la presenza o meno nel dataset finale	13
Tabella 4.3: Valori per ciascun muscolo di: accuratezza di classificazione, segmentation failure e il valore medio di Intersection over Union.....	16
Tabella 6.1: Principali componenti della pipeline MMSegmentation utilizzata, con i casi specifici utilizzati nel progetto	35
Tabella 7.1: Prestazioni complessive di MedSAM sui quattro muscoli facciali principali nelle tre condizioni analizzate.....	53
Tabella 7.2: Prestazioni di MedSAM sui quattro muscoli facciali principali, suddivise per muscolo e per tecnica di preprocessing.....	54
Tabella 7.3: Risultati quantitativi ottenuti da MedSAM sulle immagini originali.....	55
Tabella 7.4: Risultati quantitativi ottenuti da MedSAM dopo preprocessing CLAHE.....	57
Tabella 7.5: Risultati quantitativi ottenuti da MedSAM dopo preprocessing CLAHE_LOW	59
Tabella 7.6: Confronto complessivo delle prestazioni di MedSAM nelle tre condizioni analizzate	61

1. Introduzione

Le patologie neuromuscolari costituiscono un ampio e complesso gruppo di malattie che interessano il sistema nervoso periferico, la giunzione neuromuscolare e il tessuto muscolare scheletrico. Tali condizioni comprendono numerose entità cliniche caratterizzate da differenti eziologie, manifestazioni e prognosi, ma accomunate dalla progressiva compromissione della funzione muscolare. La degenerazione delle fibre muscolari e la conseguente perdita di forza rappresentano infatti una delle principali cause di disabilità nei pazienti affetti da queste patologie, con importanti ripercussioni sulla qualità della vita, sull'autonomia personale e sulla partecipazione alle attività quotidiane.

Negli ultimi decenni, il crescente interesse verso le malattie neuromuscolari ha favorito lo sviluppo di nuove metodologie diagnostiche e di monitoraggio, finalizzate a migliorare la comprensione dei meccanismi patologici e a supportare l'introduzione di strategie terapeutiche sempre più efficaci. In questo contesto, la disponibilità di strumenti in grado di valutare in maniera oggettiva e quantitativa lo stato del muscolo assume un ruolo fondamentale sia nella pratica clinica sia nell'ambito della ricerca scientifica [1].

Tra le diverse patologie neuromuscolari, la distrofia muscolare facio-scapolo-omeroale (Facioscapulohumeral Muscular Dystrophy, FSHD) [2] rappresenta una delle forme ereditarie più diffuse nell'età adulta. La malattia è caratterizzata da una progressiva degenerazione muscolare che interessa prevalentemente la muscolatura facciale, scapolare e degli arti superiori, sebbene il coinvolgimento possa estendersi anche ad altri distretti anatomici [3]. Una delle caratteristiche distintive della FSHD è l'elevata variabilità clinica, sia tra individui differenti sia all'interno dello stesso soggetto nel corso della progressione della malattia. Tale eterogeneità rende particolarmente complessa la valutazione dell'evoluzione patologica e richiede l'impiego di strumenti diagnostici sensibili e riproducibili.

L'imaging medico riveste un ruolo centrale nello studio delle patologie neuromuscolari. Tecniche quali la risonanza magnetica (Magnetic Resonance Imaging, MRI), la tomografia computerizzata (Computed Tomography, CT) e l'ecografia muscolare consentono di osservare le alterazioni strutturali associate alla degenerazione del tessuto muscolare e di ottenere informazioni utili ai fini diagnostici e prognostici [4]. Tra queste metodologie, l'ecografia muscolare ha acquisito negli ultimi anni una crescente rilevanza grazie alla combinazione di numerosi vantaggi pratici e clinici.

L'ecografia rappresenta infatti una tecnica di imaging non invasiva, priva di radiazioni ionizzanti, relativamente economica e facilmente accessibile. A differenza di altre metodologie diagnostiche, consente inoltre di effettuare esami dinamici in tempo reale e di ripetere le acquisizioni senza particolari limitazioni, caratteristica particolarmente importante nel monitoraggio longitudinalmente ripetuto di pazienti affetti da malattie croniche. La portabilità delle apparecchiature ecografiche permette inoltre di eseguire esami anche in contesti ambulatoriali o domiciliari, ampliando significativamente le possibilità di utilizzo clinico.

Nonostante tali vantaggi, l'interpretazione delle immagini ecografiche rimane fortemente dipendente dall'esperienza dell'operatore. La qualità dell'acquisizione, la scelta delle impostazioni dello strumento e l'identificazione delle strutture anatomiche possono infatti influenzare significativamente il risultato finale dell'esame. Inoltre, la valutazione visiva delle

immagini è intrinsecamente soggettiva e può determinare una variabilità inter-operatore e intra-operatore non trascurabile. Questa limitazione rappresenta uno dei principali ostacoli alla diffusione di metodologie ecografiche completamente standardizzate per la valutazione quantitativa delle patologie neuromuscolari.

Per superare tali problematiche, negli ultimi anni è emerso un crescente interesse verso lo sviluppo di sistemi automatici di analisi delle immagini mediche. L'integrazione di algoritmi di elaborazione digitale e tecniche di intelligenza artificiale ha infatti aperto nuove prospettive nell'ambito della diagnostica assistita da computer, consentendo di estrarre informazioni quantitative direttamente dalle immagini acquisite. In particolare, l'applicazione di metodologie basate sul Machine Learning e sul Deep Learning ha mostrato risultati estremamente promettenti in numerosi settori dell'imaging biomedico.

L'introduzione delle reti neurali convoluzionali (Convolutional Neural Networks, CNN) ha rappresentato una svolta significativa nel campo della computer vision applicata alla medicina. Tali modelli sono in grado di apprendere automaticamente caratteristiche rilevanti a partire dai dati di input, eliminando la necessità di definire manualmente descrittori e parametri specifici. Grazie alla disponibilità di dataset sempre più estesi e all'aumento della potenza computazionale delle moderne GPU, le tecniche di Deep Learning hanno progressivamente raggiunto prestazioni elevate in numerose applicazioni di imaging biomedico, soprattutto in compiti di classificazione, rilevamento e segmentazione.

Nell'ambito dell'ecografia muscolare, una delle operazioni più importanti è rappresentata dalla segmentazione anatomica delle strutture di interesse. La segmentazione consiste nell'identificazione automatica dei confini del muscolo all'interno dell'immagine ecografica e costituisce il passaggio preliminare necessario per l'esecuzione di analisi quantitative avanzate. Una segmentazione accurata permette infatti di estrarre parametri morfologici, caratteristiche tessutali e descrittori radiomici utilizzabili per la classificazione delle patologie, la valutazione della severità clinica e il monitoraggio della progressione della malattia.

Negli ultimi anni, le reti neurali profonde hanno mostrato un forte potenziale nella segmentazione di immagini biomediche. Architetture encoder-decoder come U-Net hanno contribuito in modo rilevante alla diffusione del deep learning in questo ambito, grazie alla capacità di combinare informazioni contestuali e localizzazione spaziale precisa [5]. Anche nell'ecografia muscoloscheletrica sono stati proposti sistemi automatici basati su deep learning, con l'obiettivo di ridurre il tempo necessario all'annotazione manuale e rendere più riproducibile l'estrazione di misure quantitative.

Inoltre, recentemente, l'evoluzione dell'intelligenza artificiale ha portato alla nascita dei cosiddetti foundation models [6], modelli addestrati su enormi quantità di dati e successivamente adattabili a differenti compiti specifici mediante tecniche di trasferimento dell'apprendimento. Nell'ambito della segmentazione di immagini mediche, tali approcci stanno attirando un interesse crescente grazie alla loro capacità di essere adattati a compiti differenti anche in presenza di dataset relativamente limitati. Tra le soluzioni più promettenti rientrano le architetture derivate dal framework Segment Anything Model (SAM) e le sue varianti specificamente sviluppate per applicazioni biomedicali, come MedSAM.

Il presente lavoro di tesi si inserisce all'interno di un progetto di ricerca sviluppato in collaborazione con il Radboud University Medical Center e le Molinette di Torino, prendendo come punto di partenza uno studio recentemente pubblicato relativo alla segmentazione

automatica di immagini ecografiche muscolari in pazienti affetti da FSHD [7]. Lo studio originale ha dimostrato come l'impiego di tecniche avanzate di Machine Learning sia in grado di ottenere prestazioni elevate nella segmentazione della maggior parte dei muscoli scheletrici analizzati. Tuttavia, i risultati hanno evidenziato prestazioni significativamente inferiori nel caso dei muscoli facciali, identificando questa regione anatomica come uno dei principali limiti del sistema.

Alla luce di tali considerazioni, si propone di investigare possibili strategie per migliorare la segmentazione automatica dei muscoli facciali mediante l'utilizzo combinato di tecniche di preprocessing delle immagini e moderni approcci basati sull'intelligenza artificiale. In particolare, verranno analizzati differenti metodi di miglioramento della qualità delle immagini ecografiche e verranno confrontate le prestazioni di modelli di segmentazione tradizionali e foundation models di ultima generazione.

L'obiettivo finale dello studio consiste nel valutare l'impatto delle diverse strategie adottate sulla qualità della segmentazione automatica e nel contribuire allo sviluppo di strumenti più affidabili e potenzialmente utili per l'analisi quantitativa delle immagini ecografiche muscolari.

I risultati possono contribuire a chiarire il ruolo del preprocessing e dei foundation models nell'analisi ecografica muscolare, fornendo indicazioni utili per futuri sviluppi di sistemi automatici più affidabili. In prospettiva, una segmentazione più accurata dei muscoli facciali potrebbe facilitare l'estrazione di misure quantitative e supportare studi longitudinali sulla progressione della FSHD, soprattutto in un distretto anatomico clinicamente rilevante ma ancora difficile da valutare in modo standardizzato [8].

Alla luce di queste considerazioni, l'ipotesi alla base del presente lavoro è che le prestazioni di segmentazione dei muscoli facciali possano essere influenzate sia dalla qualità dell'immagine ecografica sia dal tipo di modello utilizzato. L'obiettivo primario della tesi è quindi valutare se l'applicazione di tecniche di preprocessing possa migliorare la segmentazione automatica dei quattro principali muscoli facciali considerati: Depressor Anguli Oris, Mentalis, Orbicularis Oris e Zygomaticus.

Gli obiettivi secondari sono: confrontare diverse configurazioni di preprocessing su una pipeline basata su MMSegmentation; analizzare il comportamento di MedSAM nella segmentazione guidata da bounding box; valutare la stabilità delle prestazioni rispetto alle condizioni ORIGINAL, CLAHE e CLAHE_LOW; discutere i limiti metodologici del confronto tra una pipeline automatica supervisionata e un modello prompt-guided.

Il contributo principale del lavoro consiste quindi nell'analisi comparativa di preprocessing e foundation models nel contesto specifico della segmentazione ecografica dei muscoli facciali, evidenziando sia le potenzialità sia i limiti di tali approcci.

2. Facioscapulohumeral Muscular Dystrophy (FSHD)

La distrofia muscolare facio-scapolo-omeroale (Facioscapulohumeral Muscular Dystrophy, FSHD) rappresenta una delle più comuni forme di distrofia muscolare ereditaria dell'età adulta [3]. La prevalenza stimata varia tra 1 e 12 casi ogni 100.000 individui, rendendola una delle distrofie muscolari più diffuse. La malattia è caratterizzata da una progressiva degenerazione del muscolo scheletrico associata a un marcato grado di eterogeneità clinica, sia in termini di distribuzione anatomica del coinvolgimento muscolare sia in termini di severità della sintomatologia.

Ciò che distingue la FSHD è il particolare pattern di distribuzione della debolezza muscolare. Nelle fasi iniziali della malattia, il coinvolgimento interessa prevalentemente i muscoli facciali, determinando difficoltà nell'esecuzione di movimenti mimici quali il sorriso, la chiusura completa delle palpebre o il mantenimento dell'espressione facciale. La FSHD è una malattia lentamente progressiva, ma non uniforme: la distribuzione dei muscoli coinvolti, l'età di esordio e la severità clinica possono variare in modo marcato da un paziente all'altro [2], [3]. Con il progredire della patologia, la debolezza si estende generalmente ai muscoli della cintura scapolare e degli arti superiori, causando limitazioni funzionali sempre più significative. Nelle fasi più avanzate possono essere coinvolti anche i muscoli addominali, gli arti inferiori e altri gruppi muscolari scheletrici.

Un aspetto particolarmente importante della FSHD è la sua eterogeneità. Alcuni pazienti presentano sintomi lievi e una progressione molto lenta, mantenendo una buona autonomia per molti anni; altri sviluppano un quadro più severo, con limitazioni funzionali rilevanti e, in una quota minoritaria di casi, necessità di ausili per la deambulazione. Questa variabilità non riguarda solo il confronto tra pazienti diversi, ma può emergere anche all'interno dello stesso individuo, con muscoli vicini tra loro che mostrano livelli differenti di compromissione. Per questo motivo, la semplice osservazione clinica non sempre è sufficiente a descrivere con precisione lo stato del tessuto muscolare e la sua evoluzione nel tempo [3].

Dal punto di vista genetico, la FSHD è una patologia complessa la cui origine è stata chiarita soltanto negli ultimi decenni grazie ai progressi della genetica molecolare. La forma più comune, denominata FSHD di tipo 1 (FSHD1), è associata a una contrazione della regione ripetitiva D4Z4 localizzata sul cromosoma 4q35. Nei soggetti sani questa regione presenta generalmente un numero maggiore di unità, mentre nei pazienti con FSHD1 il numero di D4Z4 è ridotto, tipicamente fino a 1–10 ripetizioni [9].

Una seconda forma della malattia, denominata FSHD di tipo 2 (FSHD2), presenta manifestazioni cliniche sostanzialmente sovrapponibili alla FSHD1 ma è associata a mutazioni in geni coinvolti nei meccanismi di regolazione epigenetica della cromatina. Nonostante questa differenza iniziale, FSHD1 e FSHD2 convergono verso un meccanismo patogenetico comune: l'espressione inappropriata del gene DUX4 nel muscolo scheletrico adulto [9].

La presenza della proteina DUX4 induce l'attivazione di numerosi geni normalmente silenziati nel tessuto muscolare adulto, alterando profondamente i processi cellulari fisiologici. Tra gli effetti osservati vi sono l'aumento dello stress ossidativo, la disregolazione dei meccanismi di riparazione cellulare, l'attivazione di processi infiammatori e la compromissione della rigenerazione muscolare. Nel tempo, tali alterazioni conducono alla progressiva perdita delle fibre muscolari funzionali e alla loro sostituzione con tessuto adiposo e fibroso [10].

Dal punto di vista istopatologico, la degenerazione muscolare associata alla FSHD si manifesta attraverso una combinazione di fenomeni che includono atrofia delle fibre, infiltrazione adiposa, fibrosi interstiziale e variazioni nella struttura interna del muscolo. Questi cambiamenti modificano significativamente le proprietà meccaniche e acustiche del tessuto muscolare, rendendoli osservabili mediante tecniche di imaging diagnostico.

Questi cambiamenti strutturali hanno conseguenze dirette anche dal punto di vista dell'imaging. La perdita della normale architettura muscolare, l'infiltrazione adiposa e la fibrosi modificano infatti le proprietà del tessuto e possono essere osservate mediante tecniche di imaging muscolare. La risonanza magnetica rappresenta uno degli strumenti più utilizzati nello studio della FSHD, poiché consente di valutare in modo accurato l'infiltrazione adiposa, l'edema e la distribuzione anatomica del coinvolgimento muscolare. Tuttavia, i costi, i tempi di acquisizione e la disponibilità non sempre immediata possono limitarne l'impiego in contesti di monitoraggio frequente [11].

In questo scenario, l'ecografia muscolare rappresenta una tecnica particolarmente interessante. È rapida, non invasiva, priva di radiazioni ionizzanti e più facilmente accessibile rispetto alla risonanza magnetica. Inoltre, può essere ripetuta nel tempo e applicata anche in contesti ambulatoriali, rendendola adatta al follow-up di pazienti con patologie croniche. Nella FSHD, studi quantitativi hanno mostrato che l'ecografia e la risonanza magnetica possono fornire informazioni complementari: la MRI è più adatta a valutare infiltrazione adiposa avanzata ed edema, mentre l'ecografia può risultare sensibile a modificazioni strutturali più sottili, soprattutto nei muscoli superficiali [11], [12].

Dal punto di vista ecografico, la degenerazione muscolare si associa spesso a un aumento dell'ecogenicità. In altre parole, il muscolo patologico tende ad apparire più chiaro rispetto al muscolo sano, a causa delle modificazioni tissutali legate a fibrosi, infiltrazione adiposa e perdita della normale organizzazione delle fibre. L'analisi quantitativa dei livelli di grigio e della texture dell'immagine può quindi fornire informazioni utili sul grado di compromissione muscolare. Tuttavia, questi parametri sono affidabili solo se vengono calcolati su una regione anatomica correttamente delimitata [11].

Per questo motivo, la segmentazione del muscolo rappresenta un passaggio centrale nell'intera pipeline di analisi. Identificare con precisione i contorni del muscolo permette di estrarre misure morfologiche e tessutali più attendibili, come area, spessore, intensità media dei pixel e caratteristiche di texture. Al contrario, una segmentazione imprecisa può introdurre tessuti non appartenenti al muscolo o escludere porzioni rilevanti della struttura anatomica, alterando le misure successive e riducendo l'affidabilità dell'analisi quantitativa.

Nel caso della FSHD, la segmentazione dei muscoli facciali assume un'importanza particolarmente rilevante, come mostrato nella Figura 2.1. Proprio questi rappresentano infatti uno dei primi distretti interessati dalla malattia e costituiscono un potenziale supporto per la caratterizzazione e il monitoraggio del coinvolgimento muscolare nella progressione clinica. Tuttavia, la segmentazione manuale richiede l'intervento di un operatore esperto e può quindi essere influenzata da variabilità intra- e inter-operatore. Inoltre, i muscoli del viso per via delle ridotte dimensioni anatomiche, della complessità della morfologia e del limitato contrasto rispetto ai tessuti circostanti, rendono la segmentazione automatica particolarmente difficile [8].

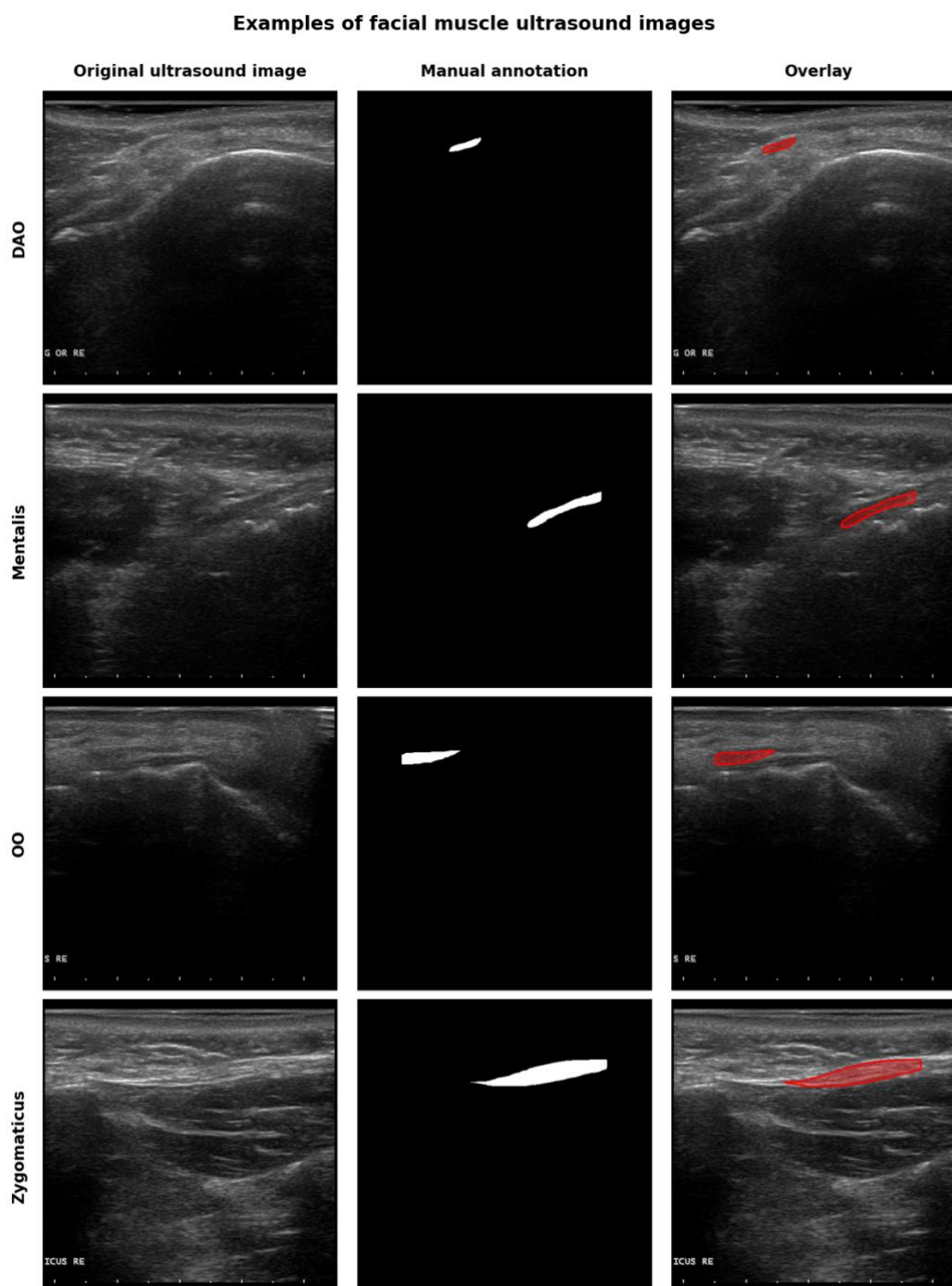


Figura 2.1: Esempi di immagini ecografiche relative a quattro muscoli facciali analizzati nel presente lavoro: Depressor Anguli Oris (DAO), Mentalis, Orbicularis Oris (OO) e Zygomaticus. La figura evidenzia le ridotte dimensioni anatomiche dei muscoli facciali e il basso contrasto rispetto ai tessuti circostanti, due aspetti che rendono complessa sia la segmentazione manuale sia quella automatica. Elaborazione originale a partire dalle immagini ecografiche del dataset utilizzato nello studio.

Per tali ragioni, lo sviluppo di algoritmi sempre più accurati per l'identificazione automatica dei muscoli facciali nelle immagini ecografiche rappresenta una delle principali sfide attuali nell'ambito dell'imaging neuromuscolare assistito dall'intelligenza artificiale e costituisce il punto di partenza del presente lavoro di ricerca.

3. Ecografia Muscolare Diagnostica

L'ecografia muscolare diagnostica (Diagnostic Muscle Ultrasound, MUS) rappresenta una delle metodologie di imaging migliori per la valutazione delle patologie neuromuscolari. Grazie alla sua natura non invasiva, all'assenza di radiazioni ionizzanti, alla ridotta complessità operativa e ai costi relativamente contenuti, essa ha acquisito un ruolo sempre più importante sia nella pratica clinica quotidiana sia nell'ambito della ricerca scientifica.

Negli ultimi anni, il miglioramento delle sonde ecografiche, della qualità dei sistemi di acquisizione e delle tecniche di elaborazione digitale delle immagini ha ampliato ulteriormente le potenzialità della MUS [13].

Oggi l'ecografia non viene utilizzata soltanto per una valutazione qualitativa del muscolo, ma anche come possibile strumento quantitativo per analizzare la struttura, la composizione e l'evoluzione del tessuto muscolare nel tempo. In questo senso, essa può contribuire alla definizione di biomarcatori utili per la diagnosi, il monitoraggio della progressione di malattia e la valutazione della risposta a eventuali trattamenti .

Il principio fisico alla base dell'ecografia consiste nell'emissione di onde ultrasonore ad alta frequenza da parte di una sonda piezoelettrica. Quando tali onde attraversano i tessuti biologici, una parte dell'energia viene riflessa in corrispondenza delle interfacce caratterizzate da differenti proprietà acustiche. I segnali riflessi vengono successivamente raccolti dalla stessa sonda e convertiti in immagini bidimensionali che rappresentano la distribuzione spaziale delle strutture anatomiche esaminate.

Nel muscolo scheletrico sano, l'aspetto ecografico è generalmente riconoscibile e relativamente costante. In sezione longitudinale, le fibre muscolari appaiono come strutture orientate in modo parallelo, separate da sottili bande iperecogene che corrispondono principalmente al tessuto connettivo intramuscolare. In sezione trasversale, invece, il muscolo assume un aspetto più puntinato o reticolare, dovuto alla disposizione tridimensionale dei fasci muscolari e delle strutture connettivali. Questo insieme di caratteristiche definisce la normale echotexture del muscolo, cioè il pattern ecografico fisiologico che riflette l'organizzazione interna del tessuto.

Quando il muscolo è interessato da processi patologici, tale organizzazione tende progressivamente ad alterarsi. Fenomeni come degenerazione delle fibre, infiltrazione adiposa, fibrosi e processi infiammatori modificano la composizione del tessuto e, di conseguenza, anche il modo in cui esso interagisce con gli ultrasuoni. In particolare, l'aumento della componente fibroadiposa determina una maggiore riflessione del fascio ultrasonoro, rendendo il muscolo più brillante, cioè più iperecogeno, rispetto alla condizione fisiologica. Parallelamente, la normale architettura delle fibre muscolari può diventare meno distinguibile, con perdita progressiva dei pattern anatomici tipici del muscolo sano.

Queste modificazioni strutturali rendono l'ecografia uno strumento particolarmente utile per la valutazione delle malattie neuromuscolari. Numerosi studi hanno infatti dimostrato come l'ecogenicità muscolare sia correlata al grado di degenerazione del tessuto e possa essere utilizzata come biomarcatore della progressione patologica. Inoltre, l'ecografia consente di valutare simultaneamente diversi gruppi muscolari durante una singola sessione di acquisizione, caratteristica particolarmente vantaggiosa nel caso di patologie caratterizzate da un coinvolgimento anatomico eterogeneo, come la FSHD.

Uno degli strumenti clinici più utilizzati per interpretare le immagini ecografiche muscolari è la scala di Heckmatt. Proposta originariamente negli anni Ottanta per la valutazione dei bambini affetti da distrofia muscolare di Duchenne, questa classificazione continua ancora oggi a essere impiegata nella pratica clinica e nella ricerca neuromuscolare come sistema semplice e immediato per descrivere il grado di alterazione ecografica del muscolo.

La scala di Heckmatt, come mostrato nella Figura 3.1, suddivide le immagini ecografiche in quattro differenti livelli di gravità sulla base dell'aspetto visivo del muscolo e della visibilità delle strutture profonde. Nel grado 1 il muscolo presenta una normale ecogenicità e una chiara visualizzazione dell'osso sottostante. Nel grado 2 si osserva un lieve aumento dell'ecogenicità muscolare, pur mantenendo una buona visibilità delle strutture profonde. Nel grado 3 l'aumento dell'ecogenicità diventa più marcato e la visualizzazione dell'osso risulta parzialmente compromessa. Infine, nel grado 4 il muscolo appare fortemente iperecogeno e le strutture profonde risultano difficilmente distinguibili o completamente invisibili [14].

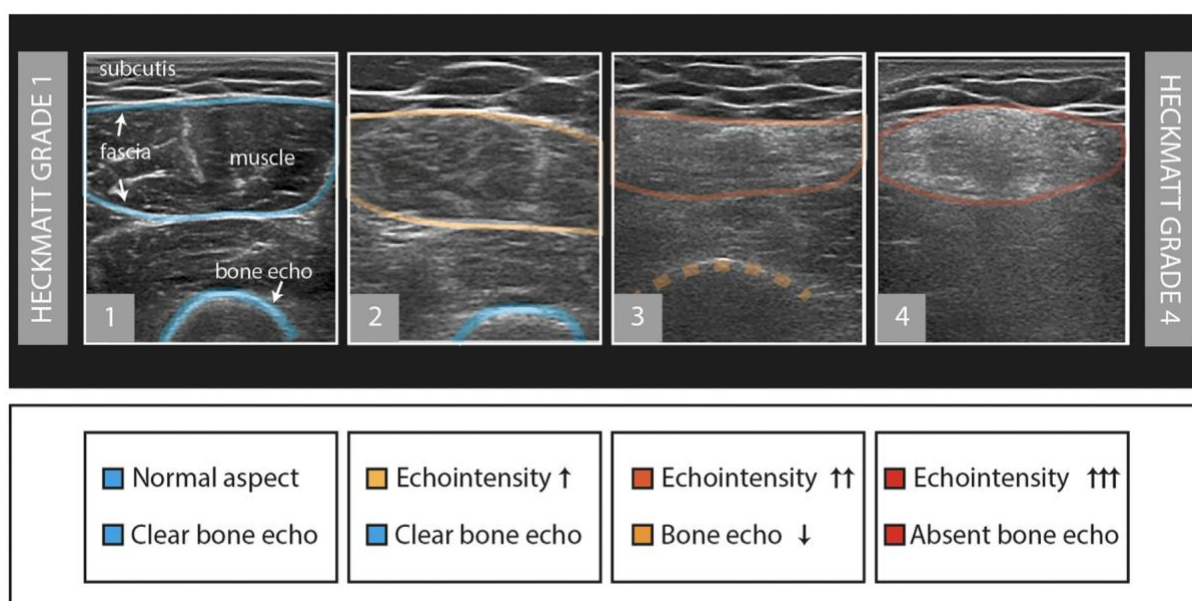


Figura 3.1: La figura mostra i quattro gradi della scala di Heckmatt utilizzata per classificare l'alterazione ecografica del muscolo. Nel grado 1 il muscolo presenta un aspetto normale, con la tipica struttura “a cielo stellato”. Nel grado 2 si osserva un aumento dell'ecogenicità muscolare rispetto allo strato adiposo sovrastante, mantenendo comunque una buona visibilità dell'osso. Nel grado 3 l'ecogenicità del muscolo risulta marcatamente aumentata e l'eco osseo appare ridotto. Nel grado 4, infine, il muscolo presenta un'ecogenicità molto elevata e l'eco osseo non è più visibile a causa dell'attenuazione del fascio ultrasonoro all'interno del tessuto muscolare. Figura tratta da J. Wijntjes et al., 2024 [14].

Nonostante la sua ampia diffusione e la relativa semplicità di utilizzo, la classificazione di Heckmatt presenta alcune limitazioni intrinseche. Il sistema si basa infatti sulla valutazione visiva da parte dell'operatore e pertanto risulta influenzato dall'esperienza clinica dell'esaminatore. Differenti specialisti possono attribuire valutazioni diverse alla medesima immagine, generando una variabilità inter-operatore che può ridurre la riproducibilità delle analisi. Analogamente, lo stesso operatore potrebbe fornire giudizi differenti in momenti diversi, introducendo una variabilità intra-operatore non trascurabile.

Un ulteriore limite delle metodologie basate esclusivamente sulla valutazione visiva è la loro natura qualitativa o semi-quantitativa. Sebbene tali approcci siano utili nella pratica clinica, essi

non permettono di sfruttare pienamente le informazioni contenute nelle immagini digitali moderne. Le immagini ecografiche, infatti, non rappresentano soltanto un supporto visivo per l'interpretazione dell'operatore, ma contengono dati numerici relativi all'intensità dei pixel, alla distribuzione dei livelli di grigio e alle caratteristiche spaziali della texture. Per questo motivo, negli ultimi anni è cresciuto l'interesse verso approcci di analisi quantitativa applicati all'imaging muscolare.

L'analisi quantitativa consente infatti di estrarre parametri oggettivi direttamente dai dati digitali, riducendo la dipendenza dall'interpretazione soggettiva dell'operatore. Tra le caratteristiche più frequentemente analizzate rientrano l'intensità media dei pixel, la distribuzione statistica dei livelli di grigio, le caratteristiche di texture e numerosi descrittori radiomici avanzati. Tali parametri possono essere successivamente utilizzati per sviluppare modelli predittivi in grado di classificare automaticamente il grado di coinvolgimento muscolare o di identificare specifiche condizioni patologiche, per poi effettuarne la segmentazione.

Tradizionalmente, la segmentazione dei muscoli nelle immagini ecografiche viene eseguita manualmente da operatori esperti. Sebbene questo approccio possa garantire un'elevata accuratezza, esso richiede tempi considerevoli ed è soggetto alla variabilità associata all'intervento umano. L'aumento delle dimensioni dei dataset utilizzati nella ricerca moderna rende inoltre sempre meno praticabile l'approccio manuale, soprattutto quando si devono analizzare migliaia di immagini.

Per queste ragioni, la segmentazione automatica basata su algoritmi di intelligenza artificiale è diventata uno dei principali ambiti di ricerca nell'analisi delle immagini ecografiche muscolari. Le tecniche di Deep Learning, in particolare, hanno mostrato un'elevata capacità di apprendere automaticamente caratteristiche rilevanti dalle immagini e di identificare le strutture anatomiche di interesse. L'impiego di modelli automatici può ridurre drasticamente il tempo necessario per l'elaborazione dei dati, migliorare la standardizzazione delle analisi e rendere più praticabile l'utilizzo dell'ecografia quantitativa in contesti clinici e sperimentali su larga scala [15].

L'integrazione tra ecografia muscolare, segmentazione automatica e analisi quantitativa rappresenta oggi una delle prospettive più promettenti per la realizzazione di sistemi di supporto decisionale clinico. Tali sistemi potrebbero consentire una valutazione più oggettiva della progressione delle patologie neuromuscolari, facilitare il monitoraggio longitudinale dei pazienti e contribuire allo sviluppo di biomarcatori digitali utilizzabili sia nella pratica clinica sia negli studi sperimentali.

Alla luce di queste considerazioni, il miglioramento delle tecniche di segmentazione automatica delle immagini ecografiche muscolari assume un'importanza strategica per l'intero settore dell'imaging neuromuscolare. Il presente lavoro si colloca in questo contesto, con particolare attenzione alle problematiche associate alla segmentazione dei muscoli facciali, che rappresentano uno dei distretti anatomici più complessi da analizzare mediante algoritmi automatici.

4. Dataset

Il dataset rappresenta uno degli elementi centrali, poiché la qualità delle immagini e delle annotazioni disponibili condiziona direttamente l'affidabilità dei modelli di segmentazione automatica. Questo aspetto è particolarmente importante nel caso dell'ecografia muscolare, dove la variabilità dell'immagine può dipendere da numerosi fattori, tra cui il muscolo acquisito, l'orientamento della sonda, la profondità della struttura anatomica, la presenza di tessuto adiposo o fibrotico e le caratteristiche del paziente. Per questo motivo, la disponibilità di un dataset ampio, ben organizzato e acquisito secondo un protocollo standardizzato costituisce una condizione essenziale per sviluppare sistemi di intelligenza artificiale realmente utili in ambito clinico.

Il dataset utilizzato nel presente lavoro deriva dalla banca dati sviluppata e descritta da Marzola et al. nell'ambito di uno studio dedicato all'analisi quantitativa dell'ecografia muscolare in pazienti affetti da distrofia muscolare facio-scapolo-omeroale (FSHD) [7].

Le immagini sono state ottenute presso il Radboud University Medical Center nell'ambito di protocolli clinici standardizzati finalizzati alla valutazione di pazienti affetti da patologie neuromuscolari. Le acquisizioni sono state effettuate mediante uno scanner ecografico Esaote MyLabTwice equipaggiato con sonda lineare LA533 da 3–13 MHz, caratterizzata da una risoluzione assiale di circa 0,2 mm. Un aspetto rilevante del dataset è rappresentato dal mantenimento di impostazioni ecografiche il più possibile costanti durante le acquisizioni. Questa scelta ha permesso di ridurre la variabilità tecnica legata allo strumento e di ottenere immagini più confrontabili tra loro.

Il dataset originale comprende 25.005 immagini .png (ognuna 512x512, a 8-bit grayscale) ecografiche B-mode acquisite da 290 partecipanti, di cui 180 soggetti sani e 110 pazienti appartenenti a una coorte FSHD. Le immagini riguardano inizialmente 21 muscoli differenti, con tre acquisizioni per ciascun muscolo, come schematicamente rappresentato nella Figura 4.1.

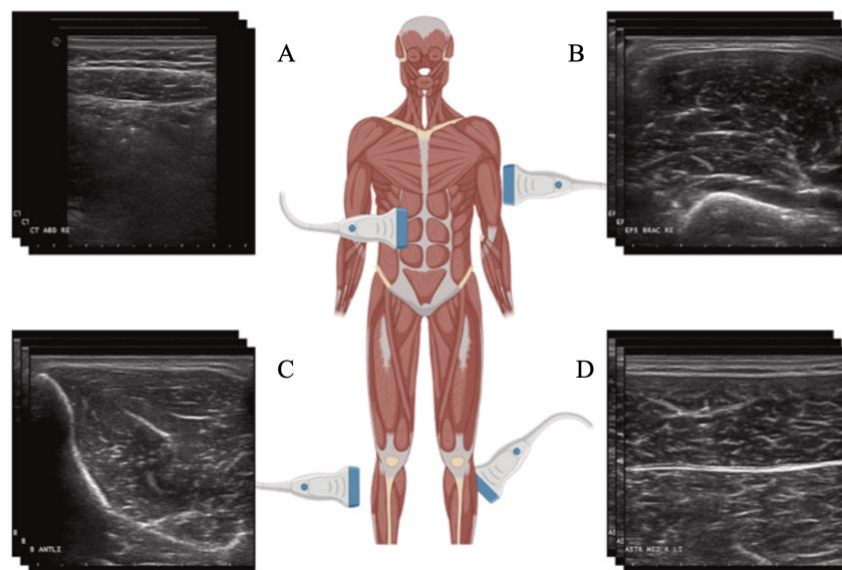


Figura 4.1: Le tre acquisizioni bilaterali per ciascun muscolo, in figura sono riportati gli esempi dei muscoli Rectus Abdominis (A), Biceps Brachii (B), Tibialis Anterior (C) e Gastrocnemius (D).

Dopo una fase di selezione, sono stati esclusi i muscoli rappresentati da meno di 200 immagini, così da ridurre lo sbilanciamento del dataset e garantire una quantità minima di esempi per ciascuna classe anatomica. A seguito di questa procedura, il dataset utilizzato per la segmentazione e la classificazione muscolare comprende 24.756 immagini relative a 16 muscoli.

Per favorire una migliore organizzazione del dataset, ogni immagine è identificata da un nome file strutturato secondo la convenzione “A_B_C_D”. In questa convenzione, “A” rappresenta il codice numerico del paziente, “B” indica il muscolo, secondo la codifica riportata in Tabella 4.1, “C” specifica il lato di acquisizione, con “00” per la scansione sinistra e “01” per la scansione destra, mentre “D” rappresenta il grado di Heckmatt, compreso tra 1 e 4.

Tabella 4.1: Tabella schematica per attribuzione ad ogni muscolo il proprio codice, la propria regione anatomica e la presenza o meno nel dataset finale.

Muscle type	Code	Anatomic Region	Inclusion in the final dataset
Biceps brachii	001	ARM	Included
Deltoides	002	ARM	Included
Depressor anguli_oris, DAO	003	FACE	Included
Digastricus	004	FACE	Included
Extensor_digitorum_brevis	005	LEG	Not Included
Flexor_carpi_radialis	006	ARM	Not Included
Flexor_digitorum_profundus	007	ARM	Not Included
Gastrocnemius_medial_head	008	LEG	Included
Geniohyoideus	009	FACE	Included
Levator_labii_superior	010	FACE	Not Included
Masseter	011	FACE	Included
Mentalis	012	FACE	Included
Orbicularis_oris, OO	013	FACE	Included
Peroneus_tertius	014	LEG	Not Included
Rectus_abdominis	015	CHEST	Included
Rectus_femoris	016	LEG	Included
Temporalis	017	FACE	Included
Tibialis_anterior	018	LEG	Included
Trapezius	019	ARM	Included
Vastus_lateralis	020	LEG	Included
Zygomaticus	021	FACE	Included

Le immagini sono state raccolte nell’ambito della pratica clinica ordinaria e successivamente anonimizzate prima della condivisione tra i centri coinvolti. Le maschere di segmentazione associate alle immagini sono state ottenute tramite annotazione manuale da parte di operatori esperti nell’ambito dell’ecografia neuromuscolare. Tali annotazioni rappresentano la ground truth, cioè il riferimento utilizzato per addestrare e valutare i modelli automatici.

La selezione di questi muscoli è stata effettuata con l’obiettivo di ottenere una rappresentazione ampia delle differenti regioni anatomiche coinvolte nella FSHD. Alcuni di essi appartengono infatti alla muscolatura facciale, altri alla regione cervicale, mentre altri ancora rappresentano importanti gruppi muscolari degli arti superiori, del tronco e degli arti inferiori.

Particolare attenzione è stata dedicata ai muscoli facciali, che costituiscono l'iniziale principale oggetto di interesse del presente lavoro. Tra questi figurano il Depressor Anguli Oris, il Mentalis, l'Orbicularis Oris e lo Zygomaticus, tutti frequentemente coinvolti nelle fasi iniziali della progressione della FSHD [8]. L'importanza clinica di tali muscoli è particolarmente elevata poiché il loro coinvolgimento può rappresentare uno dei primi segni della patologia e contribuire significativamente alla compromissione funzionale del paziente.

Dal punto di vista dell'analisi automatica delle immagini, i muscoli facciali rappresentano inoltre una delle categorie anatomiche più difficili da segmentare. Le loro dimensioni ridotte, la complessità della morfologia, la presenza di tessuti adiacenti con caratteristiche ecografiche simili e il limitato contrasto rispetto alle strutture circostanti rendono infatti particolarmente impegnativo il riconoscimento accurato dei contorni muscolari. Proprio per questa ragione, il precedente studio [7] aveva evidenziato come tali muscoli costituissero il principale punto debole del sistema di segmentazione automatico sviluppato.

Le immagini presenti nel dataset sono accompagnate, in un'altra cartella, da maschere di segmentazione binarie (512x512, a 8-bit grayscale binarie) con gli stessi nomi delle immagini per favorire un'associazione univoca e facilitata. Tali maschere sono state generate manualmente attraverso procedure di annotazione eseguite da esperti e successivamente verificate per garantire la massima accuratezza possibile (un esempio in Figura 4.2). Nel contesto del machine learning supervisionato, queste maschere costituiscono il riferimento utilizzato per confrontare le predizioni prodotte dagli algoritmi con la segmentazione considerata corretta dal punto di vista clinico.

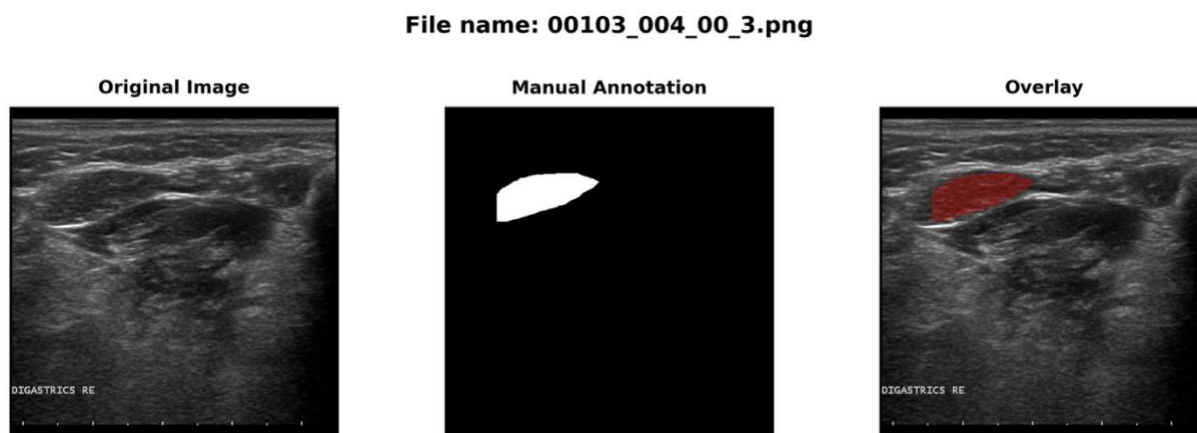


Figura 4.2: Un esempio di immagine, la corrispondente maschera manuale e la sovrapposizione tra immagine e maschera. Il nome del file riportato nella parte superiore segue la convenzione utilizzata per identificare paziente, muscolo (che nel caso specifico è il Digastricus), lato di acquisizione (sinistra) e grado di Heckmatt (3).

La disponibilità di un dataset di grandi dimensioni, caratterizzato da annotazioni esperte e da un'elevata variabilità anatomica, ha reso possibile l'esecuzione di un'analisi approfondita delle prestazioni dei diversi approcci investigati, che si analizzeranno più avanti nella trattazione. In particolare, il dataset ha consentito di valutare sia l'impatto delle tecniche di preprocessing applicate alle immagini ecografiche sia l'efficacia di differenti strategie di segmentazione basate su architetture di intelligenza artificiale tradizionali e foundation models [16].

4.1 Organizzazione del dataset per le analisi sperimentali

A partire dal dataset finale descritto in precedenza, composto da immagini ecografiche e relative maschere manuali di segmentazione, il lavoro di tesi è stato organizzato in due fasi principali. Questa scelta è stata dettata dalla necessità di affrontare inizialmente il problema più critico, rappresentato dalla segmentazione dei muscoli facciali, per poi estendere le procedure sviluppate all'intero dataset.

In una prima fase, sono state separate le immagini e le rispettive maschere relative ai muscoli facciali dagli altri gruppi muscolari. In particolare, l'analisi si è concentrata su Depressor Anguli Oris, Mentalis, Orbicularis Oris e Zygomaticus, che rappresentano i muscoli di maggiore interesse per il presente lavoro.

La scelta di focalizzarsi inizialmente su questi muscoli è stata motivata non solo dalla loro rilevanza clinica nella FSHD, ma anche dalle maggiori difficoltà che essi presentano dal punto di vista dell'analisi automatica. Infatti, nello studio preso come riferimento [7], tali muscoli risultano tra quelli con le prestazioni più critiche in termini di classification accuracy, segmentation failures e Intersection over Union (IoU) rispetto agli altri gruppi muscolari.

In particolare, il Depressor Anguli Oris mostra una classification accuracy pari a 83.83% e una percentuale di segmentation failures del 16.67%; il Mentalis presenta un tasso di fallimento della segmentazione del 17.93%; l'Orbicularis Oris evidenzia la classification accuracy più bassa (75.84%) e segmentation failures pari al 25.21%; infine, lo Zygomaticus risulta il caso più critico, con segmentation failures pari al 32.89% e un valore medio di IoU pari a 0.40 [7].

Questi risultati confermano come i muscoli facciali costituiscano il sottogruppo anatomico più complesso da trattare e giustificano la scelta di utilizzarli come punto di partenza per le prime prove di preprocessing. Come verrà descritto nei capitoli successivi, una volta individuate le strategie più promettenti, il preprocessing è stato successivamente esteso all'intero dataset e valutato anche nell'ambito dell'applicazione di MedSAM. La pipeline appena descritta è schematizzata nella Figura 4.3.

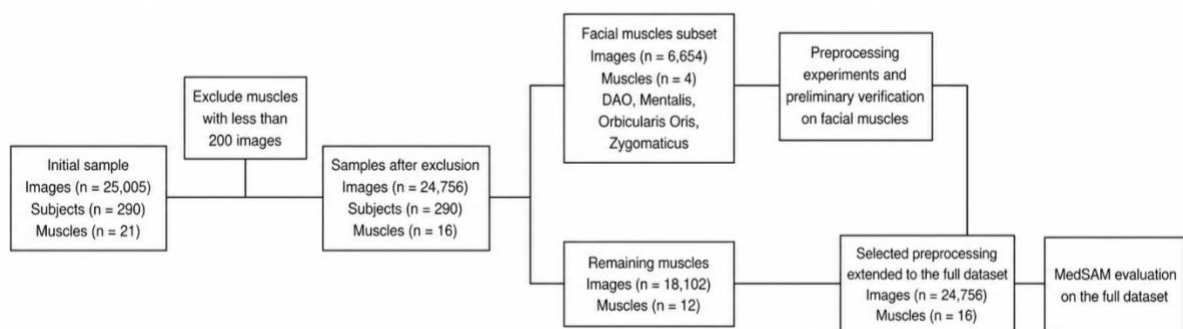


Figura 4.3: Schema della pipeline seguita in questo lavoro di tesi. Dopo l'esclusione dei muscoli con meno di 200 immagini, il dataset finale è stato suddiviso in un sottoinsieme di muscoli facciali, e nei restanti gruppi muscolari. Le prime prove di preprocessing sono state condotte sui muscoli facciali, mentre le strategie selezionate sono state successivamente estese all'intero dataset e valutate mediante MedSAM.

4.2 Identificazione del Problema: la Segmentazione dei Muscoli Facciali

Lo studio sviluppato presso il Radboud University Medical Center [7] ha dimostrato come le moderne tecniche di segmentazione automatica basate su Machine Learning possano raggiungere prestazioni elevate nell'analisi delle immagini ecografiche muscolari. I risultati ottenuti hanno evidenziato valori di accuratezza particolarmente elevati per numerosi gruppi muscolari appartenenti agli arti superiori, agli arti inferiori e al tronco, confermando il potenziale dell'intelligenza artificiale come strumento di supporto per la valutazione quantitativa delle patologie neuromuscolari.

Nonostante questi risultati incoraggianti, l'analisi dettagliata delle metriche di segmentazione ha evidenziato la presenza di un'importante criticità. Alcuni gruppi muscolari hanno infatti mostrato prestazioni inferiori rispetto alla media del dataset, con particolare riferimento ai muscoli facciali. Questa osservazione ha rappresentato il principale punto di partenza del presente progetto di ricerca e ha motivato l'intero percorso di approfondimento sviluppato nei capitoli successivi.

Tra tutti i muscoli analizzati, il Depressor Anguli Oris, il Mentalis, l'Orbicularis Oris e lo Zygomaticus hanno evidenziato le maggiori difficoltà di segmentazione. I valori delle metriche di valutazione, in particolare della Segmentation Failure e della IoU (Intersection over Union) [18], risultavano infatti sensibilmente inferiori rispetto a quelli osservati per altri muscoli scheletrici di maggiori dimensioni. Parallelamente, la frequenza di segmentazioni fallite o parzialmente errate risultava significativamente più elevata, indicando una ridotta capacità del modello di identificare correttamente tali strutture anatomiche (come si evince in Tabella 4.3).

Tabella 4.2: La tabella riporta, per ciascun muscolo, l'accuratezza di classificazione, la percentuale di segmentation failure e il valore medio di Intersection over Union (IoU $\pm$ deviazione standard). I muscoli facciali, in particolare Depressor Anguli Oris, Mentalis, Orbicularis Oris e Zygomaticus, mostrano le prestazioni più critiche, giustificando la loro selezione come sottoinsieme iniziale per le prove di preprocessing. Valori adattati da Marzola et al. [7].

Muscle type	Classification Accuracy (%)	Segmentation Failure (%)	IoU
Biceps brachii	99.54	0	0.85 ± 0.11
Deltoides	99.81	0.29	0.88 ± 0.09
Depressor anguli_oris, DAO	83.83	16.67	0.57 ± 0.28
Digastricus	98.22	1.38	0.78 ± 0.13
Gastrocnemius_medial_head	99.66	0.74	0.91 ± 0.10
Geniohyoideus	98.64	1.23	0.77 ± 0.13
Masseter	99.59	0.41	0.87 ± 0.09
Mentalis	98.95	17.93	0.45 ± 0.23
Orbicularis_oris, OO	75.84	25.21	0.44 ± 0.28
Rectus_abdominis	98.96	0.69	0.85 ± 0.11
Rectus_femoris	97.93	0.29	0.88 ± 0.10
Temporalis	99.80	2.58	0.77 ± 0.22
Tibialis_anterior	99.88	0.23	0.78 ± 0.15
Trapezius	99.87	0.66	0.80 ± 0.14
Vastus_lateralis	98.61	0.58	0.88 ± 0.12
Zygomaticus	99.73	32.89	0.40 ± 0.30

Le cause delle ridotte prestazioni osservate sono molteplici e derivano dalla combinazione di fattori anatomici, ecografici e patologici. Uno dei principali elementi di difficoltà è rappresentato, come già sottolineato in precedenza, dalle dimensioni estremamente ridotte dei muscoli facciali rispetto ai grandi muscoli scheletrici degli arti o del tronco. In un'immagine ecografica, strutture anatomiche di piccole dimensioni occupano infatti una porzione limitata del campo visivo, riducendo la quantità di informazioni disponibili per il modello durante il processo di apprendimento.

Dal punto di vista dell'intelligenza artificiale, questo fenomeno determina un importante squilibrio tra la regione appartenente al muscolo e lo sfondo circostante. In termini di segmentazione semantica, il numero di pixel associati alla classe di interesse diventa notevolmente inferiore rispetto al numero di pixel appartenenti alle altre strutture presenti nell'immagine. Tale situazione può rendere più difficile l'apprendimento delle caratteristiche distintive del muscolo e aumentare la probabilità di errori di classificazione.

Un secondo fattore critico riguarda la definizione dei confini anatomici. Molti muscoli facciali presentano margini poco netti e una continuità anatomica con strutture adiacenti quali tessuto adiposo sottocutaneo, fasce connettivali e altri piccoli gruppi muscolari. Di conseguenza, i contorni osservabili nelle immagini ecografiche risultano spesso meno evidenti rispetto a quelli di muscoli più grandi e meglio delimitati.

La qualità dei bordi rappresenta un aspetto particolarmente importante nei modelli di segmentazione basati su reti neurali profonde. L'identificazione accurata delle transizioni tra muscolo e tessuti circostanti costituisce infatti una delle principali fonti di informazione utilizzate dagli algoritmi per definire la maschera finale. Quando tali transizioni risultano poco marcate, la probabilità di errori aumenta sensibilmente [19].

Ulteriori difficoltà derivano dalla complessità anatomica della regione facciale. I muscoli del volto sono caratterizzati da una disposizione tridimensionale estremamente articolata e da frequenti sovrapposizioni con altre strutture anatomiche. Inoltre, il loro orientamento può variare considerevolmente tra individui differenti, introducendo una significativa variabilità morfologica all'interno del dataset.

A differenza dei grandi muscoli scheletrici, che tendono a presentare una forma relativamente costante e facilmente riconoscibile, i muscoli facciali mostrano una maggiore eterogeneità geometrica. Questa variabilità rende più difficile la costruzione di modelli generalizzabili capaci di mantenere elevate prestazioni su soggetti differenti [20].

Nel contesto specifico della FSHD, la situazione risulta ulteriormente complicata dalle alterazioni patologiche associate alla malattia. La degenerazione progressiva del tessuto muscolare determina infatti modificazioni strutturali che possono alterare significativamente l'aspetto ecografico del muscolo. Fenomeni quali infiltrazione adiposa, fibrosi e atrofia modificano l'ecogenicità dei tessuti e possono rendere ancora più difficile l'identificazione dei confini anatomici.

L'analisi delle prestazioni ottenute nello studio originale ha quindi evidenziato che il problema della segmentazione dei muscoli facciali non può essere attribuito a un singolo fattore, ma deriva dall'interazione di numerose criticità concomitanti.

Di seguito un elenco riassuntivo dei fattori contribuiscono a rendere la categoria muscolare dei muscoli del viso particolarmente difficile da analizzare mediante algoritmi automatici rispetto ai muscoli scheletrici di grandi dimensioni:

- Le ridotte dimensioni anatomiche,
- La scarsa definizione dei bordi,
- La complessità morfologica della regione facciale
- Le alterazioni patologiche indotte dalla FSHD

Per riuscire a contrastare questi attributi si è ipotizzato di applicare adeguate tecniche di preprocessing per incrementare la qualità delle informazioni presenti nelle immagini ecografiche, facilitando il riconoscimento delle strutture anatomiche da parte dei modelli di intelligenza artificiale. Successivamente, si è sfruttata l'introduzione di architetture più recenti e avanzate, inclusi i foundation models sviluppati specificamente per l'imaging medico, per consentire di superare almeno in parte le limitazioni osservate nei modelli precedentemente utilizzati.

4.3 Organizzazione dello studio sperimentale

Al fine di affrontare le criticità evidenziate nella segmentazione automatica dei muscoli facciali, è stato progettato un percorso sperimentale strutturato in più fasi successive. L'obiettivo principale dello studio consisteva nell'identificare strategie in grado di migliorare l'accuratezza della segmentazione automatica nelle immagini ecografiche muscolari, con particolare attenzione ai gruppi muscolari che avevano mostrato le prestazioni più limitate nello studio originale sviluppato presso il Radboud University Medical Center.

La progettazione dell'attività sperimentale è stata guidata dall'ipotesi che le difficoltà osservate nella segmentazione dei muscoli facciali possano dipendere sia dalla qualità intrinseca delle immagini ecografiche sia dalle caratteristiche dei modelli utilizzati per l'analisi automatica. Per questo motivo, il lavoro è stato suddiviso in due macrofasi principali, ciascuna dedicata all'investigazione di uno specifico aspetto della pipeline di elaborazione.

La prima fase è stata dedicata al preprocessing delle immagini ecografiche. In ambito medico, la qualità dell'immagine rappresenta infatti un elemento fondamentale per il corretto funzionamento degli algoritmi di segmentazione. Le immagini ecografiche sono caratterizzate dalla presenza di rumore speckle, variazioni di contrasto e artefatti che possono ridurre la visibilità delle strutture anatomiche e complicare l'identificazione dei contorni muscolari. Pertanto, è stato ipotizzato che l'applicazione di opportune tecniche di miglioramento dell'immagine potesse facilitare il riconoscimento delle strutture anatomiche da parte dei modelli di intelligenza artificiale.

Sono state quindi investigate diverse strategie di preprocessing specificamente selezionate per il contesto dell'ecografia muscolare. L'obiettivo era aumentare il contrasto tra il muscolo e i tessuti circostanti, migliorare la visibilità dei bordi anatomici e ridurre l'influenza del rumore presente nelle acquisizioni originali. Ogni tecnica è stata applicata all'intero dataset e successivamente valutata attraverso esperimenti di segmentazione controllati.

La seconda fase dello studio è stata invece dedicata all'analisi e al confronto di differenti modelli di segmentazione automatica. In questa fase si è cercato di comprendere se le limitazioni osservate nello studio originale fossero attribuibili esclusivamente alla qualità delle immagini

oppure se potessero essere superate mediante l'impiego di architetture di intelligenza artificiale più recenti e avanzate.

Per garantire la massima comparabilità con il lavoro precedentemente sviluppato, la prima serie di esperimenti è stata condotta utilizzando il framework basato su Swin Transformer derivato dallo studio originale. Tale architettura rappresentava il punto di riferimento dell'intero progetto e costituiva la baseline rispetto alla quale confrontare tutte le successive modifiche introdotte.

L'utilizzo iniziale del modello originale ha consentito di valutare in maniera controllata l'impatto delle differenti tecniche di preprocessing, mantenendo invariata l'architettura di segmentazione. In questo modo è stato possibile isolare l'effetto delle modifiche apportate alle immagini e identificare le strategie di preprocessing maggiormente promettenti per il miglioramento delle prestazioni.

Successivamente, una volta individuate le tecniche di preprocessing più efficaci, l'attenzione si è spostata verso l'analisi di approcci alternativi basati sui più recenti sviluppi dell'intelligenza artificiale. In particolare, è stato investigato MedSAM (Medical Segment Anything Model), una versione specializzata per l'imaging medico del foundation model Segment Anything Model (SAM).

L'interesse verso MedSAM deriva dal crescente successo dei foundation models in numerosi ambiti della computer vision. A differenza delle architetture tradizionali addestrate per uno specifico compito, questi modelli vengono inizialmente allenati su enormi quantità di dati e successivamente adattati a differenti applicazioni. Tale approccio ha dimostrato notevoli capacità di generalizzazione e rappresenta attualmente una delle principali direzioni di sviluppo dell'intelligenza artificiale applicata all'imaging medico.

L'intera pipeline sperimentale è stata organizzata secondo una sequenza di operazioni ben definita. La prima fase ha previsto l'estrazione e la preparazione delle immagini provenienti dal dataset originale. Successivamente sono state applicate le diverse tecniche di preprocessing selezionate, generando differenti versioni dello stesso insieme di dati.

Una volta completata la fase di preprocessing, il dataset è stato organizzato secondo una strategia di cross-validation a cinque fold. Tale approccio ha consentito di ottenere una valutazione statisticamente più robusta delle prestazioni dei modelli, riducendo il rischio che i risultati fossero influenzati da particolari caratteristiche di una singola suddivisione dei dati. Inoltre, la suddivisione dei dati in training, validation e test è stata effettuata a livello di paziente, secondo l'organizzazione sperimentale dello studio di riferimento. In questo modo, le immagini appartenenti allo stesso soggetto non sono state distribuite contemporaneamente tra insiemi di addestramento e test, riducendo il rischio di data leakage e rendendo più corretta la valutazione delle prestazioni del modello.

Per ciascun fold sono stati eseguiti gli esperimenti di segmentazione utilizzando sia le immagini originali sia le diverse versioni preprocessate. Le prestazioni sono state valutate mediante metriche quantitative standard nel campo della segmentazione semantica, consentendo un confronto oggettivo tra le differenti configurazioni sperimentali.

I risultati ottenuti nella prima fase hanno permesso di identificare le strategie di preprocessing più promettenti. Solo le configurazioni che hanno mostrato un comportamento più promettente

dal punto di vista qualitativo e numerico sono state successivamente considerate per le analisi più approfondite.

Nella fase finale dello studio è stato effettuato un confronto diretto tra l'approccio basato sul framework Swin Transformer e il foundation model MedSAM. Questo confronto ha avuto lo scopo di valutare non soltanto l'efficacia delle tecniche di preprocessing, ma anche il potenziale contributo offerto dalle più recenti evoluzioni dell'intelligenza artificiale alla risoluzione del problema della segmentazione dei muscoli facciali.

L'organizzazione progressiva dello studio ha quindi consentito di affrontare il problema da due prospettive complementari: da un lato il miglioramento della qualità dell'informazione contenuta nelle immagini, dall'altro l'introduzione di modelli di segmentazione caratterizzati da capacità rappresentative più avanzate. Questa strategia ha permesso di costruire una pipeline sperimentale completa, finalizzata a identificare la combinazione più efficace tra preprocessing e architettura di segmentazione per l'analisi automatica delle immagini ecografiche muscolari.

I capitoli successivi descrivono nel dettaglio le tecniche di preprocessing investigate, le configurazioni sperimentali adottate, i modelli di segmentazione utilizzati e i risultati ottenuti nelle diverse fasi dello studio.

5. Tecniche di preprocessing delle Immagini Ecografiche

Le immagini ecografiche presentano caratteristiche molto diverse rispetto ad altre tecniche di imaging medico, come la tomografia computerizzata o la risonanza magnetica. Sebbene l'ecografia sia una metodica non invasiva, economica, facilmente accessibile e utilizzabile in tempo reale, la qualità dell'immagine ottenuta dipende in modo rilevante da diversi fattori legati sia all'acquisizione sia al paziente. In particolare, parametri come la posizione della sonda, la pressione esercitata dall'operatore, l'angolo di insonazione, il guadagno, la profondità di acquisizione e le proprietà acustiche dei tessuti possono modificare sensibilmente l'aspetto finale dell'immagine [21]

Tra le principali limitazioni delle immagini ecografiche rientrano il rumore speckle, il basso contrasto tra strutture adiacenti, le variazioni locali di intensità e la presenza di artefatti acustici. Lo speckle, in particolare, è una componente tipica dell'imaging ultrasonoro e deriva dall'interferenza degli echi generati dai tessuti. Sebbene faccia parte della natura fisica dell'immagine ecografica, esso può ridurre la visibilità dei dettagli anatomici più fini, rendere meno netti i confini tra le strutture e complicare sia l'interpretazione clinica sia l'elaborazione automatica dell'immagine [22].

Nel contesto dell'ecografia muscolare, tali problematiche risultano particolarmente rilevanti nel caso dei muscoli facciali. Per questo motivo è stato ipotizzato che adeguate tecniche di preprocessing potessero migliorare la qualità delle immagini e facilitare il processo di segmentazione automatica. Nel caso specifico delle immagini ecografiche, il preprocessing può essere utile per aumentare il contrasto locale, ridurre l'effetto del rumore, rendere più omogenea la distribuzione delle intensità e facilitare l'identificazione dei margini anatomici.

L'obiettivo del preprocessing è quello di aumentare la visibilità delle strutture anatomiche rilevanti, ridurre il rumore e migliorare il contrasto locale senza alterare le informazioni clinicamente significative contenute nell'immagine originale.

5.1 Contrast Limited Adaptive Histogram Equalization (CLAHE)

Tra le tecniche di miglioramento del contrasto investigate, una delle più utilizzate in ambito medico è il Contrast Limited Adaptive Histogram Equalization (CLAHE). Questo metodo rappresenta un'evoluzione dell'equalizzazione dell'istogramma tradizionale e consente di migliorare il contrasto locale dell'immagine attraverso l'elaborazione indipendente di piccole regioni denominate tiles.

Una caratteristica fondamentale di CLAHE è la presenza di un parametro di clipping che limita l'amplificazione eccessiva delle intensità, evitando l'introduzione di artefatti indesiderati e contenendo la crescita del rumore [23].

Nel caso delle immagini ecografiche muscolari, CLAHE ha mostrato una significativa capacità di evidenziare i confini muscolari, le fasce connettivali e altre strutture anatomiche di interesse. Tuttavia, un aumento troppo marcato del contrasto può talvolta enfatizzare anche il rumore speckle presente nell'immagine.

Nel presente studio, l'algoritmo CLAHE è stato implementato utilizzando un valore di clipLimit pari a 2.0 (parametro che controlla il limite massimo di amplificazione dell'istogramma locale: valori elevati aumentano il contrasto ma possono enfatizzare anche il rumore, mentre valori inferiori producono un miglioramento più conservativo) e una tileSize di 8×8 questo valore invece la dimensione della griglia in cui l'immagine viene suddivisa; una griglia più fitta migliora il contrasto locale ma può incrementare la variabilità dell'immagine (come mostrato in Figura 5.1).

Preprocessing comparison - Zygomaticus

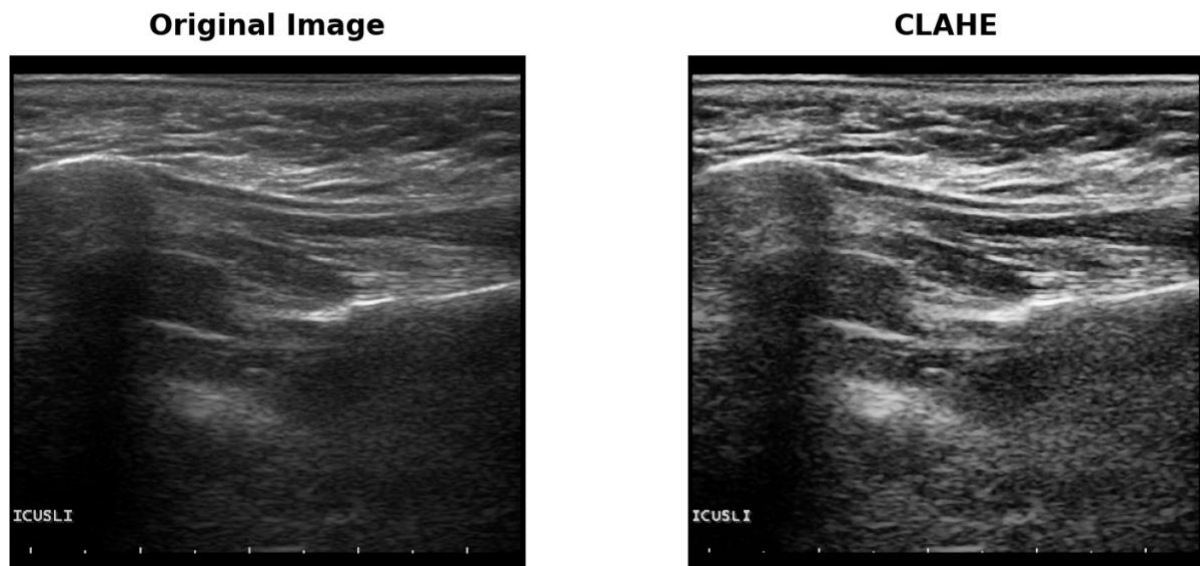


Figura 5.1: Confronto di una stessa immagine (00013_021_01_3) prima e dopo l'applicazione del preprocessing tramite la tecnica CLAHE

Prima dell'applicazione della CLAHE è stato eseguito un filtro mediano con kernel 3×3 per attenuare il rumore impulsivo mantenendo i contorni anatomici. Successivamente l'immagine è stata normalizzata nell'intervallo di intensità [0,255] mediante normalizzazione Min-Max. Questa configurazione è stata scelta come compromesso tra miglioramento del contrasto locale e limitazione dell'amplificazione del rumore speckle.

5.1.1 CLAHE_LOW

Per limitare gli effetti indesiderati associati a un miglioramento troppo aggressivo del contrasto, è stata sviluppata una seconda configurazione denominata CLAHE_LOW.

Questa variante utilizza parametri più conservativi rispetto alla configurazione CLAHE standard, mantenendo i vantaggi dell'equalizzazione adattiva ma riducendo l'intensità dell'elaborazione. L'obiettivo era ottenere un compromesso più equilibrato tra miglioramento del contrasto e preservazione dell'aspetto naturale dell'immagine.

La variante CLAHE_LOW è stata ottenuta riducendo i parametri dell'algoritmo rispetto alla configurazione standard, utilizzando clipLimit = 1.0 e tileSize = 4×4. Tale scelta consente di limitare ulteriormente l'enfatizzazione del rumore speckle mantenendo comunque un miglioramento del contrasto locale, risultando particolarmente adatta alle immagini ecografiche

dei muscoli facciali, caratterizzate da strutture anatomiche di piccole dimensioni (esempio applicativo mostrato nella Figura 5.2).

Preprocessing comparison - Zygomaticus



Figura 5.2: Comparazione di un'immagine originale (00013_021_01_3) con la stessa immagine dopo aver utilizzato la tecnica CLAHE con i parametri meno aggressivi.

Le immagini ottenute mediante CLAHE_LOW (paragone visibile in Figura 5.3) hanno mostrato un contrasto locale migliorato mantenendo al contempo una rappresentazione più realistica delle strutture anatomiche. Come verrà discusso nei capitoli dedicati ai risultati sperimentali, questa tecnica ha prodotto le prestazioni più elevate nella maggior parte degli esperimenti eseguiti.

Preprocessing comparison - Zygomaticus



Figura 5.3: La stessa immagine ecografica (00013_021_01_3) viene riportata nelle tre versioni per mostrare l'effetto del preprocessing sul contrasto locale, sulla visibilità delle fibre muscolari e sulla definizione dei margini.

5.1.2 CLAHE_HIGH

CLAHE_HIGH rappresenta una versione più aggressiva della CLAHE, ottenuta aumentando il limite di contrasto e la dimensione della griglia locale. Questa configurazione enfatizza maggiormente i dettagli e il contrasto delle strutture ecografiche, ma può anche aumentare la presenza di rumore e rendere i margini meno naturali rispetto alle configurazioni più moderate.

La variante CLAHE_HIGH è stata applicata utilizzando parametri più elevati rispetto alla configurazione standard, con $\text{clipLimit} = 4.0$ e $\text{tileGridSize} = 16 \times 16$. Questa impostazione produce un incremento più marcato del contrasto locale e rende maggiormente visibili alcune microstrutture dell'immagine ecografica. Tuttavia, l'aumento dell'enfasi sui dettagli può amplificare anche il rumore e generare un'immagine visivamente più artificiale, motivo per cui questa tecnica è stata valutata principalmente come confronto rispetto alla CLAHE standard e alla CLAHE_LOW (come mostrato in Figura 5.4).



Figura 5.4: La figura mostra le versioni ottenute con CLAHE standard, CLAHE_LOW e CLAHE_HIGH, evidenziando le differenze introdotte dai diversi parametri di equalizzazione adattiva del contrasto.

5.2 Guided Filtering

Il Guided Filtering è una tecnica di filtraggio edge-preserving adottata da He, Sun e Tang, e sviluppata per ridurre il rumore e le variazioni indesiderate dell'immagine mantenendo il più possibile i principali contorni strutturali [24]. Questo aspetto risulta particolarmente importante nelle immagini ecografiche dei muscoli facciali, dove le strutture da segmentare sono di piccole dimensioni e spesso presentano margini poco evidenti.

In ambito ecografico, approcci basati sul Guided Filter sono stati utilizzati per la riduzione dello speckle noise, anche in combinazione con tecniche come lo SRAD, con l'obiettivo di migliorare la qualità dell'immagine preservando i bordi anatomici utili alla successiva segmentazione [25].

Il Guided Filtering è stato valutato all'interno di una pipeline composta da una fase iniziale di miglioramento del contrasto mediante CLAHE, seguita dall'applicazione del filtro guidato. In particolare, l'immagine ecografica è stata prima elaborata con CLAHE, utilizzando clipLimit pari a 2.0 e tileGridSize pari a 8×8 , e successivamente filtrata mediante Guided Filter, impiegando come immagine guida la stessa immagine preprocessata.

Per la configurazione standard, indicata come GUIDED, sono stati utilizzati un raggio pari a 8 ($\text{radius} = 8$) e un parametro di regolarizzazione $\varepsilon = 10^{-2}$. Il raggio controlla l'estensione della finestra locale considerata dal filtro, mentre ε regola il grado di levigatura in funzione delle variazioni di intensità. Rispetto alla configurazione più conservativa, questi parametri producono un filtraggio più marcato, con l'obiettivo di ridurre maggiormente le variazioni locali dell'immagine, pur cercando di preservare i principali bordi anatomici, come mostrato in Figura 5.5.

Preprocessing comparison - Zygomaticus

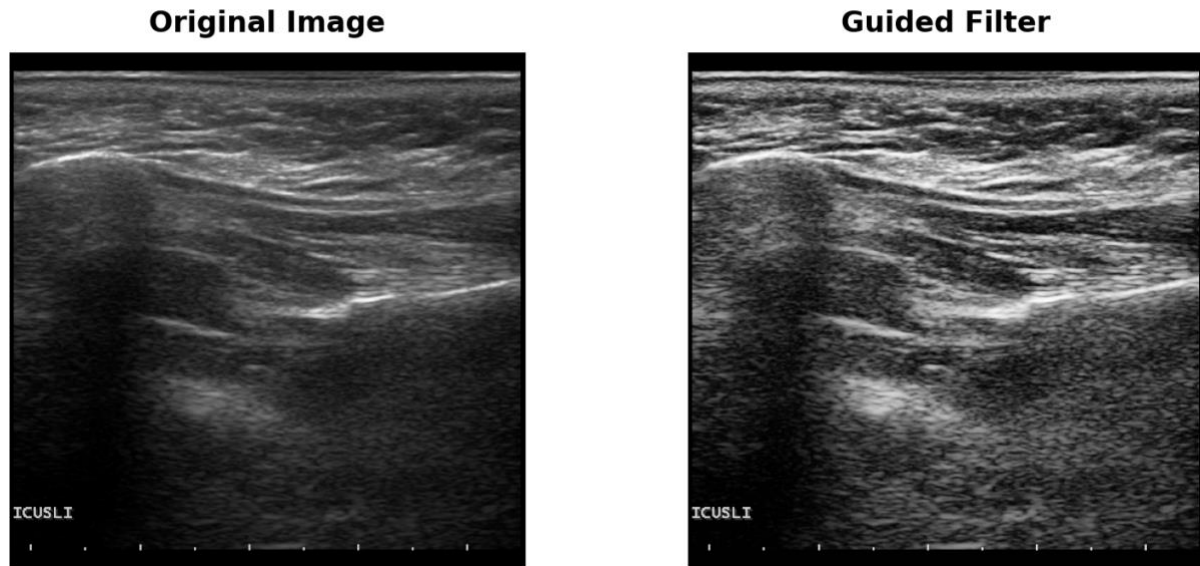


Figura 5.5: Comparazione di un'immagine originale (00013_021_01_3) con la stessa immagine dopo l'applicazione della pipeline GUIDED, costituita da CLAHE seguita da Guided Filtering con configurazione standard.

5.2.1 Guided_LOW

La configurazione GUIDED_LOW è stata introdotta come variante meno aggressiva della pipeline basata su CLAHE e Guided Filtering. Anche in questo caso, l'immagine è stata inizialmente preprocessata mediante CLAHE e successivamente filtrata con Guided Filter, utilizzando come guida la stessa immagine elaborata.

Rispetto alla configurazione GUIDED standard, GUIDED_LOW utilizza parametri più conservativi, con un raggio pari a 4 ($\text{radius} = 4$) e un parametro di regolarizzazione $\varepsilon = 10^{-3}$. Un raggio inferiore limita l'estensione della regione locale considerata dal filtro, mentre un valore più basso di ε rende il filtro più selettivo rispetto alle variazioni di intensità. L'obiettivo di questa configurazione era quindi ridurre parte del rumore e rendere l'immagine più uniforme, preservando al tempo stesso una maggiore quantità di dettagli anatomici fini, particolarmente importanti nella segmentazione dei piccoli muscoli facciali.

Sebbene GUIDED_LOW abbia prodotto immagini visivamente più regolari, come mostrato in Figura 5.6, i risultati sperimentali non hanno evidenziato un vantaggio netto rispetto alle configurazioni basate su CLAHE_LOW. Per questo motivo, questa tecnica è stata mantenuta come confronto nella prima fase sperimentale, ma non è stata selezionata per l'analisi successiva con MedSAM.

Preprocessing comparison - Zygomaticus

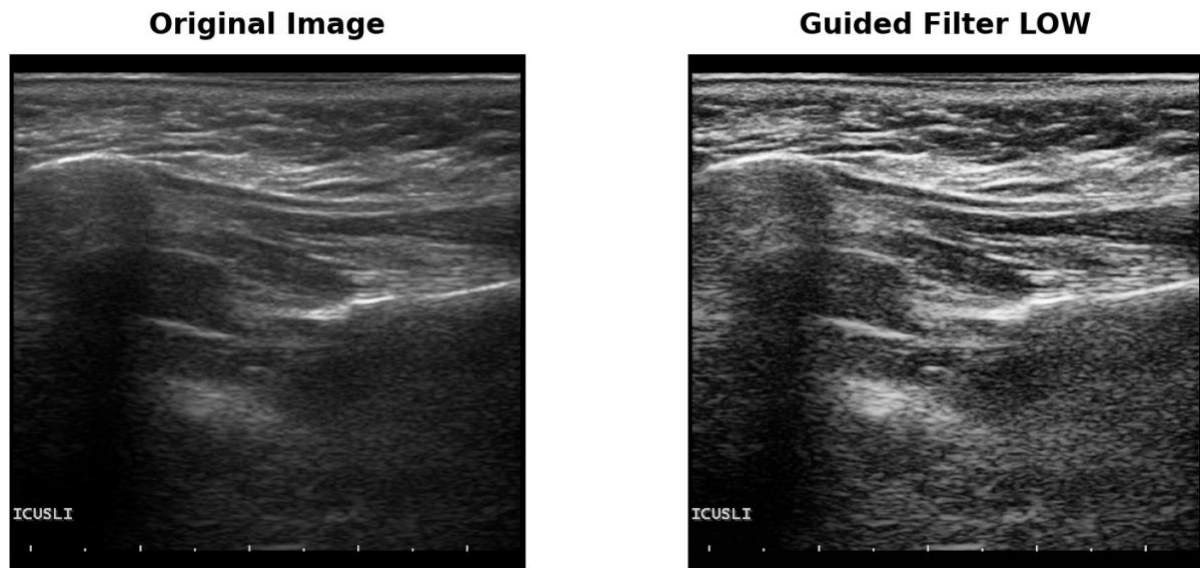


Figura 5.6: Confronto tra l'immagine originale e la versione ottenuta mediante GUIDED_LOW, cioè la pipeline CLAHE + Guided Filtering applicata con parametri più conservativi.

5.3 *Speckle Reducing Anisotropic Diffusion (SRAD)*

Poiché lo speckle noise rappresenta uno degli artefatti più tipici delle immagini ecografiche, è stata valutata anche la tecnica Speckle Reducing Anisotropic Diffusion (SRAD). Questo metodo è pensato per ridurre il rumore preservando, per quanto possibile, i bordi e le principali strutture anatomiche dell'immagine [25].

L'algoritmo agisce attraverso un processo di diffusione anisotropa: le aree più omogenee dell'immagine vengono progressivamente regolarizzate, mentre le zone caratterizzate da variazioni di intensità più marcate, come i margini anatomici, vengono preservate. In questo modo, SRAD cerca di ottenere un compromesso tra riduzione del rumore e mantenimento delle informazioni utili alla segmentazione.

In questo lavoro, SRAD è stato implementato utilizzando 30 iterazioni e un passo temporale $dt = 0.25$. Successivamente, è stata applicata una CLAHE con gli stessi parametri descritti in precedenza ($clipLimit = 2.0$, $tileGridSize = 8 \times 8$), seguita dalla normalizzazione dell'immagine. Le iterazioni multiple permettono una riduzione progressiva dello speckle, mentre il parametro dt controlla la stabilità del processo diffusivo.

Nonostante le buone premesse teoriche, nei test effettuati questa tecnica non ha portato a un miglioramento significativo qualitativamente parlando rispetto alle altre strategie di preprocessing considerate (lo si evince anche dalla Figura 5.7). Per questo motivo, SRAD è stato mantenuto principalmente come tecnica di confronto all'interno dell'analisi sperimentale.

Preprocessing comparison - Zygomaticus

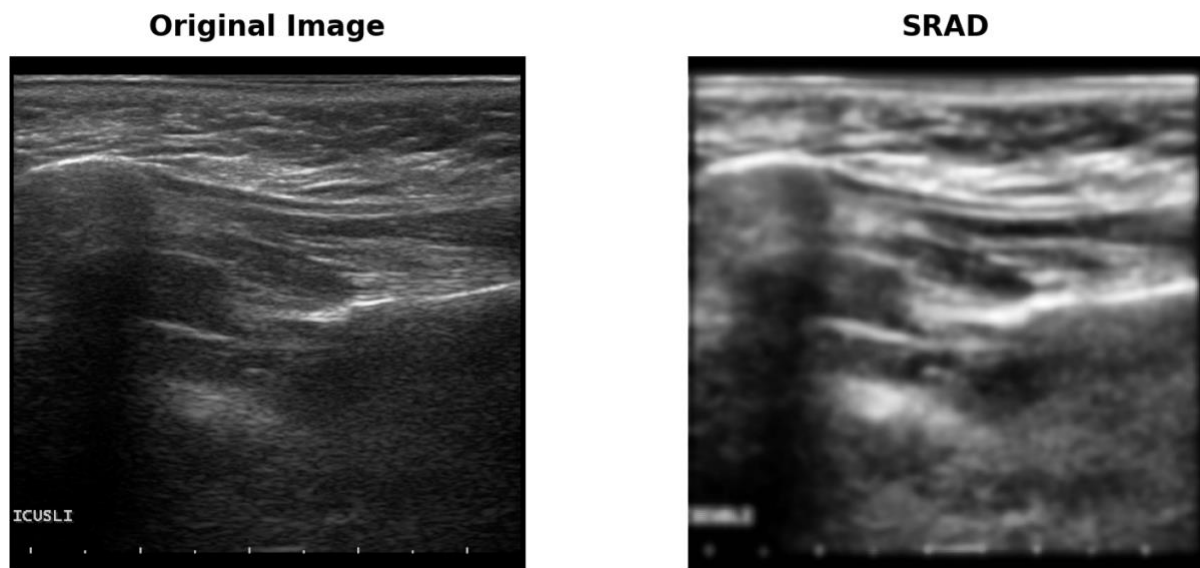


Figura 5.7: L'esempio mostra l'effetto della diffusione anisotropa sulla stessa immagine (00013_021_01_3) del muscolo zigomatico, evidenziando un peggioramento dell'immagine data la difficoltà post-applicazione della tecnica a distinguere i vari limiti anatomici

5.4 Adaptive Non-Local Means (A-NLM)

L'Adaptive Non-Local Means è stato valutato come tecnica di denoising con l'obiettivo di ridurre il rumore senza eliminare completamente le informazioni strutturali dell'immagine. Il principio alla base del metodo è che, all'interno della stessa immagine, possono essere presenti regioni con caratteristiche simili anche se non sono necessariamente vicine tra loro.

A differenza dei filtri più tradizionali, che lavorano soprattutto sul vicinato immediato di ciascun pixel, il Non-Local Means confronta piccole porzioni dell'immagine, chiamate patch, e combina le informazioni provenienti da regioni simili. Questo approccio è particolarmente interessante nelle immagini ecografiche, dove la texture muscolare può presentare pattern ripetitivi e dove la riduzione dello speckle deve essere bilanciata con la preservazione dei dettagli anatomici [26].

Nel presente lavoro, il livello di rumore è stato stimato automaticamente mediante la funzione `estimate_sigma()`. Il parametro di filtraggio è stato poi impostato pari a $h = 1.2\sigma$, utilizzando patch di dimensione 5×5 e una distanza massima di ricerca pari a 7 pixel. In questo modo, il filtro ha potuto confrontare regioni anche non adiacenti dell'immagine, sfruttando la somiglianza tra diverse aree della texture ecografica.

Dal punto di vista visivo, in Figura 5.8, l'Adaptive Non-Local Means ha mostrato una buona capacità di attenuare il rumore e rendere l'immagine più uniforme. Tuttavia, in alcuni casi, anche dettagli anatomici particolarmente fini risultavano meno evidenti dopo il filtraggio. Per questo motivo, pur essendo una tecnica efficace per il denoising, il suo contributo al problema specifico della segmentazione dei piccoli muscoli facciali è risultato limitato.

Preprocessing comparison - Zygomaticus

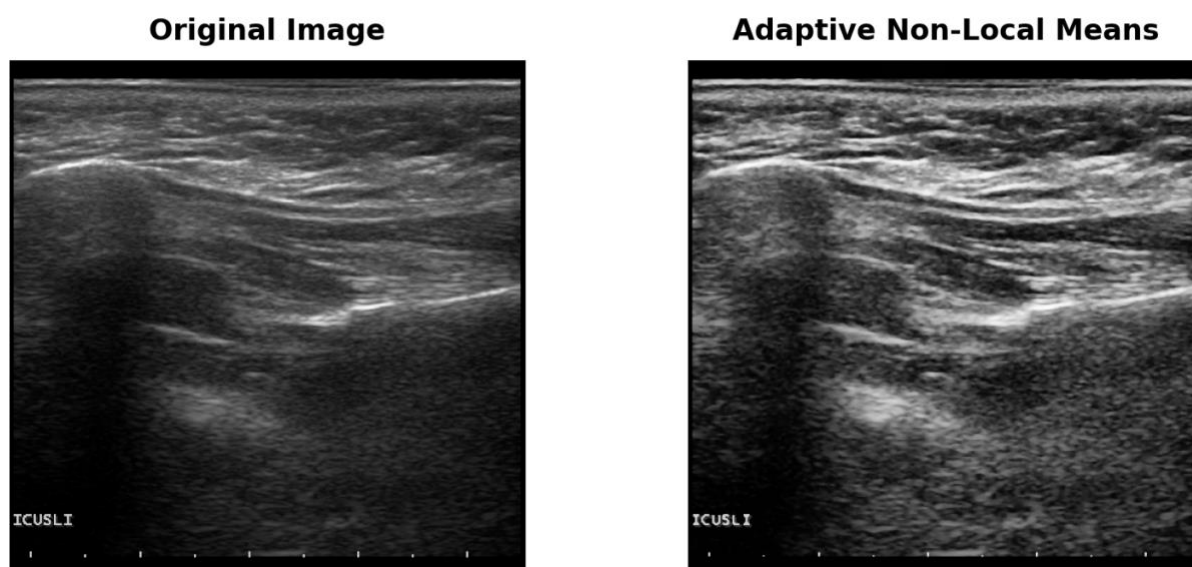


Figura 5.8: L'esempio mostra l'effetto del filtro A-NLM sulla stessa immagine (00013_021_01_3) del muscolo zigomatico, evidenziando una riduzione del rumore e una maggiore uniformità dell'immagine, pur con una possibile attenuazione di alcuni dettagli anatomici più fini.

5.5 Stationary Wavelet Transform (SWT)

Le tecniche basate sulla trasformata wavelet sono state considerate perché permettono di analizzare l'immagine a diverse scale spaziali. Questo è utile nelle immagini ecografiche, dove il rumore speckle può sovrapporsi alle strutture anatomiche e rendere meno evidente il confine del muscolo.

In questo lavoro è stata valutata la Stationary Wavelet Transform (SWT), una variante della trasformata wavelet che ha il vantaggio di essere più stabile rispetto a piccoli spostamenti dell'immagine.

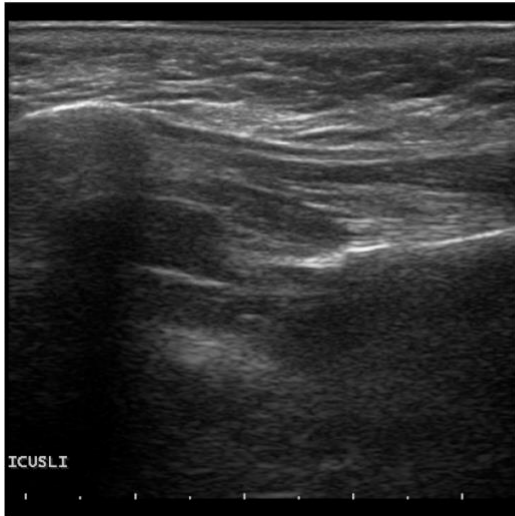
Questa proprietà, detta invarianza alla traslazione, consente di mantenere meglio l'allineamento delle strutture durante le fasi di decomposizione e ricostruzione. Proprio per questo motivo, la SWT è stata utilizzata in diversi lavori sul denoising ecografico, dove è importante ridurre il rumore senza perdere bordi e dettagli anatomici [27].

È stata utilizzata la wavelet biortogonale bior3.5, con tre livelli di decomposizione e un coefficiente di soglia $k = 4$, applicato mediante soft-thresholding. La wavelet biortogonale è stata scelta per favorire la preservazione dei bordi, mentre i tre livelli di decomposizione permettono di separare in modo più efficace le componenti strutturali dell'immagine dal rumore.

Dal punto di vista visivo (come in Figura 5.9), la SWT ha prodotto immagini più pulite e meno disturbate dallo speckle. Tuttavia, come spesso accade nei metodi di denoising, come era successo per la tecnica A-NLM, la riduzione del rumore è stata accompagnata da una lieve perdita di dettagli fini. Questo aspetto è particolarmente rilevante nel caso dei muscoli facciali, dove anche piccole variazioni nei margini possono influenzare la qualità della segmentazione.

Preprocessing comparison - Zygomaticus

Original Image



Stationary Wavelet Transform



Figura 5.9: L'esempio mostra l'effetto della SWT sulla stessa immagine del muscolo zigomatico (00013_021_01_3), evidenziando una riduzione del rumore speckle ma un minore contrasto complessivo che potrebbe portare alla perdita di dettagli anatomici.

5.6 Confronto Qualitativo delle Tecniche di Preprocessing

È stato effettuato un confronto visivo tra i diversi metodi di preprocessing applicati alla stessa immagine ecografica. Come si può osservare in Figura 5.10, alcune tecniche producono un'immagine più contrastata, mentre altre agiscono soprattutto sulla riduzione del rumore e sull'uniformità della texture.

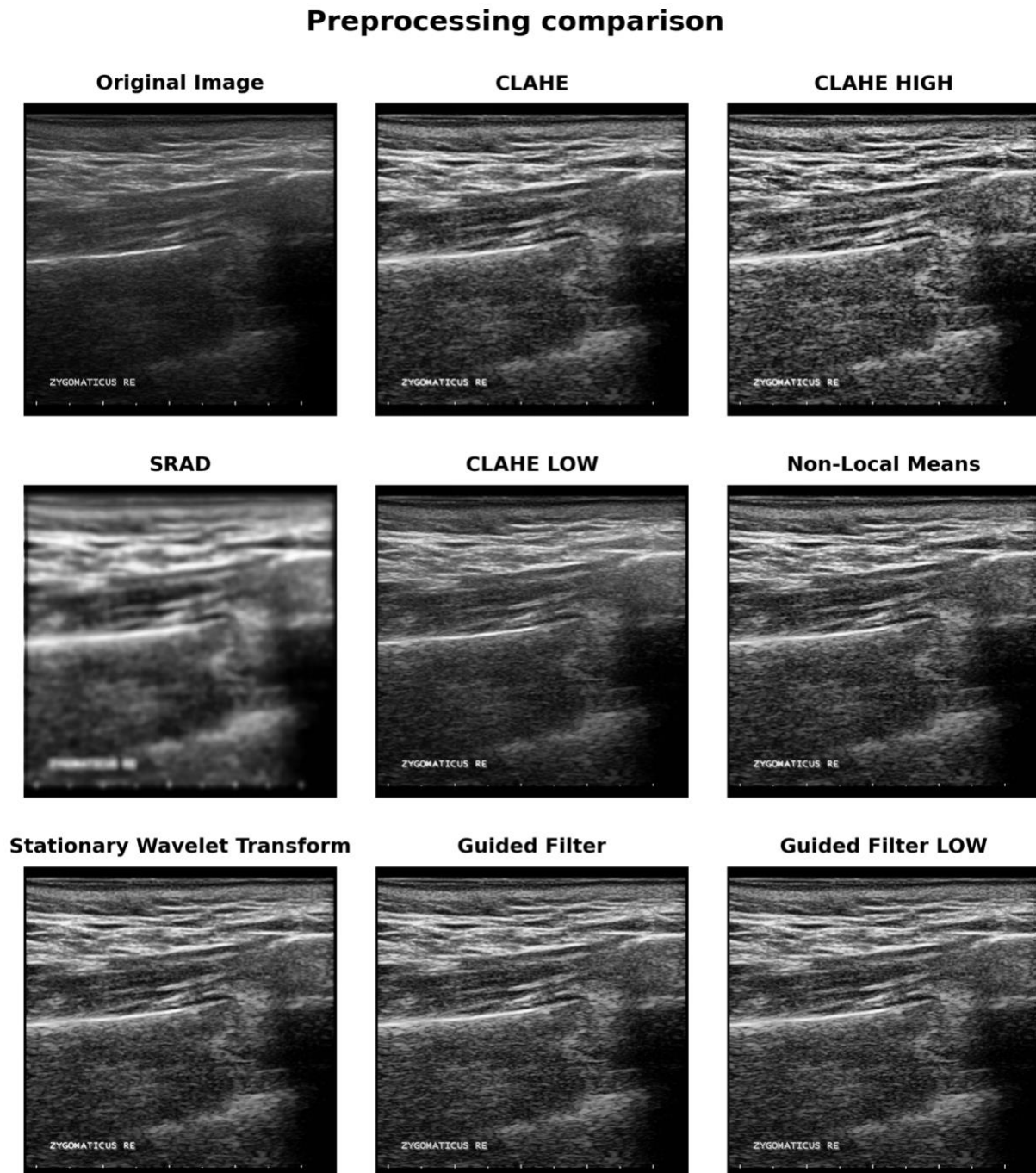


Figura 5.10: La figura mostra l'immagine originale (00195_021_00_3) e le versioni preprocessate mediante CLAHE, CLAHE_HIGH, SRAD, CLAHE_LOW, Non-Local Means, Stationary Wavelet Transform, Guided Filter e Guided Filter LOW. Il confronto evidenzia come ciascuna tecnica modifichi in modo diverso il contrasto, la rumorosità e la visibilità delle strutture muscolari.

Come si evince dalla Figura 5.10, CLAHE ha prodotto il maggiore incremento del contrasto locale, migliorando sensibilmente la visibilità delle strutture anatomiche. CLAHE_LOW ha fornito un risultato più equilibrato, mantenendo gran parte dei benefici del miglioramento del contrasto senza introdurre modifiche eccessive all'immagine originale.

Guided Filtering e Guided_LOW hanno mostrato una buona capacità di riduzione del rumore, mentre SRAD e A-NLM hanno prodotto immagini con perdita di dettagli. SWT ha generato immagini visivamente pulite ma con una moderata attenuazione delle strutture più sottili.

Le valutazioni effettuate in questa fase hanno quindi fornito le basi per gli esperimenti successivi, dedicati alla valutazione dei modelli di segmentazione basati su architetture Transformer e foundation models.

6. Metodi

6.1 MMSegmentation

La fase iniziale degli esperimenti di segmentazione automatica è stata condotta utilizzando MMSegmentation [28], [29], un framework open-source sviluppato all'interno dell'ecosistema OpenMMLab e basato su PyTorch. MMSegmentation non rappresenta un singolo modello di rete neurale, ma piuttosto un ambiente di lavoro che consente di costruire, addestrare, validare e testare modelli di segmentazione semantica in modo strutturato e riproducibile. In altre parole, il framework fornisce l'infrastruttura necessaria per gestire l'intera pipeline sperimentale: caricamento del dataset, applicazione delle trasformazioni sulle immagini, definizione dell'architettura, addestramento del modello, salvataggio dei pesi migliori e valutazione finale delle prestazioni.

Nel contesto della segmentazione semantica, l'obiettivo del modello è assegnare a ciascun pixel dell'immagine una determinata classe. Nel caso più generale, ogni pixel può appartenere a una delle diverse classi anatomiche oppure allo sfondo. Questo principio è stato applicato alla segmentazione di immagini ecografiche muscolari, con lo scopo di ottenere una maschera binaria o multi-classe capace di delimitare automaticamente la regione muscolare di interesse. La segmentazione automatica assume un ruolo particolarmente rilevante nell'analisi quantitativa dell'ecografia muscolare, poiché la definizione della regione di interesse rappresenta il punto di partenza per qualunque misurazione successiva, come l'analisi dell'eointensità, della texture o di altri parametri quantitativi.

Uno degli aspetti più importanti di MMSegmentation è la sua struttura modulare [28]. Secondo la pipeline, mostrata in Figura 6.1, il framework separa le diverse componenti del modello, come il backbone, il modulo di aggregazione delle feature e la segmentation head, permettendo di combinare tra loro architetture differenti senza dover riscrivere da zero l'intero codice di addestramento. Questa organizzazione è particolarmente utile in ambito sperimentale, perché consente di modificare singoli elementi della rete, mantenendo invariata la struttura generale della pipeline.

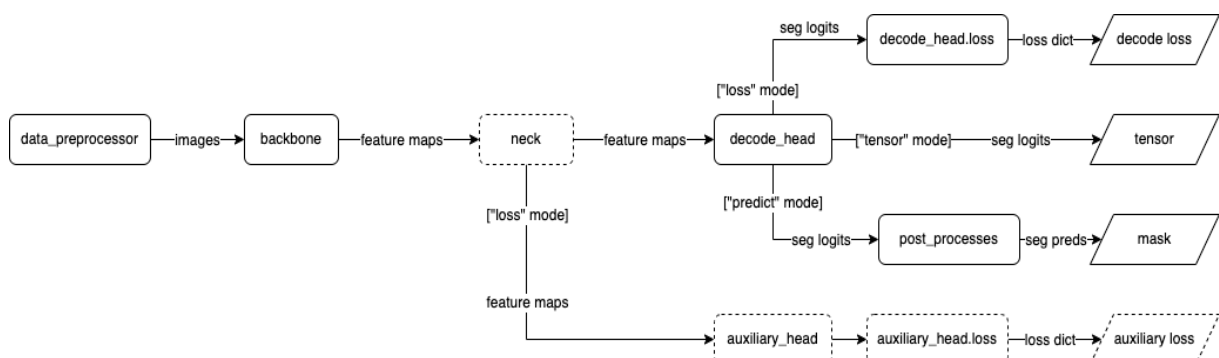


Figura 6.1: Schema generale del flusso dati in un modello di segmentazione con MMSegmentation. Adottato dalla documentazione ufficiale di MMSegmentation, [OpenMMLab,2020] [28].

È possibile ricondurre le principali componenti del framework alle scelte operative adottate. Questa distinzione è utile perché MMSegmentation non coincide con una singola rete neurale, ma rappresenta l'ambiente in cui vengono definiti e collegati tra loro il dataset, le trasformazioni sui dati, l'architettura del modello, la funzione di perdita, l'ottimizzatore e le metriche di valutazione.

La Tabella 6.1 riassume quindi il ruolo dei principali blocchi della pipeline e la loro implementazione nel caso specifico della segmentazione ecografica muscolare.

Tabella 6.1: Principali componenti della pipeline MMSegmentation utilizzata, con i casi specifici utilizzati nel progetto

Componente	Funzione generale	Utilizzo effettuato
Dataset	Gestisce il caricamento delle immagini e delle relative annotazioni manuali.	Immagini ecografiche muscolari e maschere di ground truth.
Pipeline dati	Applica le trasformazioni necessarie prima dell'ingresso nella rete, come ridimensionamento, normalizzazione e augmentations.	Immagini normalizzate nell'intervallo $[-1, 1]$, data augmentation in tempo reale, rotazioni casuali fino a $\pm 15^\circ$, flip orizzontali e verticali, scaling tra 0.9 e 1.1, modifiche di brightness $\pm 10\%$, batch size 2 e risoluzione 512×512 .
Backbone	Estrae rappresentazioni profonde dell'immagine, trasformando i pixel in feature maps informative.	Swin Transformer
Decoder/Segmentation Head	Integra le feature estratte e produce la predizione della maschera di segmentazione.	UPerNet + K-Net
Loss function	Quantifica la differenza tra la maschera predetta e la maschera manuale.	Confronto supervisionato con la ground truth
Ottimizzatore	Aggiorna i pesi della rete durante l'addestramento per ridurre progressivamente l'errore.	AdamW
Metriche di valutazione	Misurano quantitativamente la qualità della segmentazione ottenuta.	Dice coefficient, IoU

MMSegmentation è stato quindi utilizzato come piattaforma di riferimento per valutare l'effetto delle tecniche di preprocessing. Le immagini ecografiche e le relative maschere manuali sono state organizzate secondo la struttura richiesta dal framework, suddividendo i dati in insiemi di training, validation e test. Durante l'addestramento, la pipeline basata su Swin Transformer, UPerNet e K-Net apprende progressivamente a riconoscere le caratteristiche visive associate al muscolo da segmentare, confrontando la maschera predetta con la maschera manuale fornita come ground truth. Al termine della fase di training, le prestazioni vengono valutate sul test set mediante metriche quantitative, come il Dice coefficient e l'Intersection over Union, e mediante un'analisi qualitativa delle maschere prodotte.

6.1.1 Architettura Utilizzata

L'architettura adottata deriva direttamente dal lavoro sviluppato presso il Radboud University Medical Center [17] e combina tre componenti principali:

- Swin Transformer come backbone;
- UPerNet come modulo di aggregazione delle caratteristiche;
- K-Net come segmentation head.

Questa configurazione è stata selezionata nello studio originale per la sua capacità di gestire efficacemente immagini caratterizzate da elevata variabilità anatomica e dalla presenza di strutture di dimensioni differenti.

L'integrazione di questi tre moduli consente di sfruttare simultaneamente le capacità di rappresentazione gerarchica offerte dai Transformer, la fusione multi-scala delle informazioni e il raffinamento iterativo delle maschere di segmentazione.

6.1.2 Swin Transformer

Il backbone rappresenta il componente responsabile dell'estrazione delle caratteristiche dall'immagine di ingresso. Nel presente studio è stato utilizzato Swin Transformer [30], una delle architetture Transformer più influenti nell'ambito della computer vision moderna.

Nel presente lavoro, lo Swin Transformer è stato utilizzato come backbone dell'architettura di segmentazione. Il suo ruolo non è quello di produrre direttamente la maschera finale, ma di estrarre dall'immagine ecografica una rappresentazione gerarchica delle informazioni visive. Rispetto al Vision Transformer tradizionale (ViT), che tende a lavorare su una rappresentazione a singola scala, lo Swin Transformer introduce una struttura gerarchica più simile a quella delle reti convoluzionali, producendo feature maps a risoluzioni progressivamente ridotte (riportato in Figura 6.2). Questo aspetto è particolarmente utile nei compiti di segmentazione, nei quali è necessario combinare informazioni locali, come bordi e texture, con informazioni più globali sul contesto anatomico.

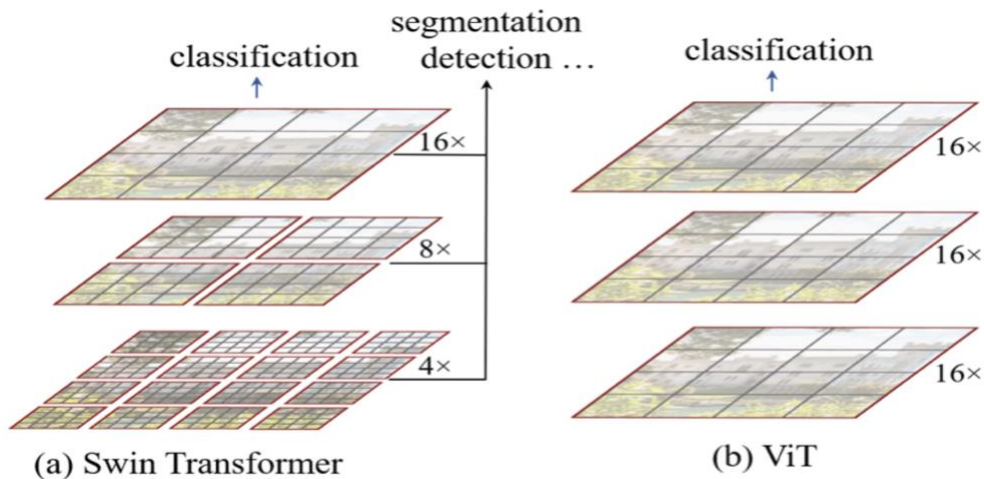


Figura 6.2: La figura mostra schematicamente la differenza tra Swin Transformer e Vision Transformer classico (ViT), evidenziando la natura gerarchica del primo e la sua maggiore adattabilità a compiti di segmentazione semantica, Figura tratta da Liu et al. [30].

Un altro elemento caratteristico dello Swin Transformer è l'utilizzo di finestre locali di attenzione, dette windows, che vengono spostate tra livelli successivi della rete. In questo modo, il modello riduce il costo computazionale rispetto alla self-attention globale e, allo stesso tempo, permette lo scambio di informazioni tra regioni vicine dell'immagine.

Nel contesto della segmentazione ecografica, questa impostazione risulta utile perché le immagini muscolari presentano strutture con dimensioni e caratteristiche visive differenti. La possibilità di estrarre feature a più scale permette al modello di considerare sia dettagli locali, come variazioni di ecogenicità e margini muscolari, sia informazioni più ampie relative alla posizione del muscolo rispetto ai tessuti circostanti.

6.1.3 UPerNet

Le caratteristiche estratte dal backbone vengono successivamente elaborate da UPerNet (Unified Perceptual Parsing Network) [31], un modulo progettato per aggregare informazioni provenienti da differenti scale spaziali.

La segmentazione di immagini mediche richiede infatti la capacità di identificare contemporaneamente dettagli locali e contesto anatomico globale. L'analisi esclusiva di una singola scala potrebbe infatti portare alla perdita di informazioni rilevanti per il corretto riconoscimento delle strutture anatomiche.

UPerNet affronta questo problema attraverso una strategia di fusione multi-scala basata su Feature Pyramid Networks (FPN) e Pyramid Pooling Modules (PPM). Questa combinazione consente di integrare informazioni provenienti dai diversi livelli gerarchici del backbone, generando rappresentazioni più ricche e informative.

Nel caso specifico delle immagini ecografiche muscolari, la disponibilità di informazioni multi-scala è particolarmente importante. I muscoli facciali occupano una porzione relativamente ridotta dell'immagine e presentare dettagli anatomici sottili che rischierebbero di essere persi durante le operazioni di elaborazione. La fusione tra dettaglio locale e contesto anatomico

permette al modello di utilizzare sia le informazioni fini, legate alla texture e ai margini, sia le informazioni più globali, relative alla posizione del muscolo nell'immagine ecografica.

UPerNet è stato quindi considerato come il modulo responsabile dell'aggregazione delle feature estratte dallo Swin Transformer. Le rappresentazioni ottenute vengono successivamente fornite a K-Net, che produce la predizione finale della maschera di segmentazione.

6.1.4 K-Net

L'ultimo componente dell'architettura è rappresentato da K-Net (Kernel Update Network), utilizzato come Segmentation Head per la generazione della maschera finale. Nel flusso complessivo del modello, K-Net riceve in ingresso le feature elaborate dallo Swin Transformer e aggregate da UPerNet, e le utilizza per produrre la predizione della regione muscolare [32].

Nel flusso complessivo del modello, K-Net riceve in ingresso le feature elaborate dallo Swin Transformer e aggregate da UPerNet, e le utilizza per produrre la predizione della regione muscolare. L'idea alla base di K-Net è diversa rispetto a molte architetture tradizionali di segmentazione, nelle quali ogni pixel viene classificato direttamente in una determinata classe. In questo caso, invece, i kernel apprendibili possono essere interpretati come rappresentazioni interne associate alle regioni da segmentare, che vengono progressivamente aggiornate in funzione delle feature dell'immagine [32].

Questo approccio permette di correggere errori iniziali e migliorare progressivamente la qualità della segmentazione finale. Il raffinamento iterativo risulta particolarmente utile nel caso di strutture anatomiche caratterizzate da contorni complessi o poco definiti, come nei casi delle immagini ultrasonografiche sotto esame.

L'obiettivo è ottenere una delimitazione più precisa delle regioni anatomiche e ridurre la presenza di errori localizzati lungo i bordi della struttura segmentata.

L'impiego di architetture basate su kernel condizionali ha trovato applicazione anche in altri contesti biomedici, ad esempio nella segmentazione di cellule in immagini citologiche, confermando l'interesse di questo tipo di approccio per problemi in cui le strutture da identificare presentano forma e aspetto variabili [33].

6.1.5 Organizzazione degli esperimenti

Gli esperimenti sono stati condotti seguendo una strategia di cross-validation a cinque fold. Tale approccio è stato scelto per garantire una valutazione robusta e statisticamente affidabile delle prestazioni del modello.

Il task valutato è stato formulato come segmentazione binaria della regione muscolare di interesse. Le immagini erano già associate al relativo codice muscolare e l'obiettivo del modello non era quindi classificare il tipo di muscolo, ma produrre una maschera binaria capace di distinguere il muscolo dallo sfondo e dai tessuti circostanti. Le metriche di classificazione riportate dallo studio di riferimento sono state utilizzate soltanto per motivare la scelta dei

muscoli facciali più critici, mentre la valutazione sperimentale di questa tesi si è basata su Dice Similarity Coefficient, Intersection over Union e segmentation failure.

Per ciascun fold, il dataset è stato suddiviso in insiemi distinti di training, validation e test. Questa organizzazione ha consentito di utilizzare efficacemente l'intero dataset disponibile mantenendo separati i dati impiegati per l'addestramento da quelli utilizzati per la valutazione finale.

Particolare attenzione è stata dedicata ai muscoli facciali identificati come maggiormente problematici nello studio originale. Le immagini corrispondenti ai muscoli Depressor Anguli Oris, Mentalis, Orbicularis Oris e Zygomaticus sono state quindi estratte e analizzate separatamente al fine di valutare con maggiore precisione l'impatto delle tecniche di preprocessing.

Per ogni configurazione sperimentale sono state generate le relative maschere di segmentazione e successivamente confrontate con le annotazioni manuali disponibili nel dataset.

6.2 Foundation Models per la segmentazione di Immagini Mediche

Negli ultimi anni, il campo dell'intelligenza artificiale ha attraversato una fase di profonda evoluzione grazie alla diffusione dei cosiddetti Foundation Models. Questi modelli rappresentano un cambiamento importante rispetto agli approcci tradizionali di Deep Learning, poiché non vengono progettati per risolvere un unico compito specifico, ma per apprendere rappresentazioni generali che possano essere successivamente adattate a problemi differenti [34].

Questo paradigma si è affermato inizialmente nel campo del Natural Language Processing, con modelli come BERT e GPT, e si è poi progressivamente esteso anche alla computer vision [35], [36]. In questo ambito, l'obiettivo è sviluppare modelli capaci di riconoscere e interpretare strutture visive molto diverse tra loro, riducendo la necessità di addestrare da zero una nuova architettura per ogni singolo problema.

Nel settore medico, tale approccio risulta particolarmente interessante. L'annotazione delle immagini cliniche richiede infatti competenze specialistiche, tempi lunghi e un'elevata accuratezza. Per questo motivo, disporre di modelli già pre-addestrati e adattabili a diversi scenari può rappresentare un vantaggio significativo, soprattutto quando i dataset disponibili sono limitati o caratterizzati da elevata variabilità [37].

Nel presente capitolo vengono quindi introdotti i Foundation Models per la segmentazione di immagini mediche, con particolare attenzione al Segment Anything Model e alla sua estensione in ambito medico, MedSAM. Quest'ultimo verrà poi analizzato nel contesto specifico della segmentazione ecografica dei muscoli facciali [38].

6.2.1 Segmentazione e Foundation Models

La segmentazione automatica costituisce uno dei compiti più importanti nell'ambito della computer vision applicata alla medicina. L'obiettivo consiste nell'identificare e delimitare

automaticamente regioni anatomiche, organi, tessuti o lesioni all'interno delle immagini diagnostiche.

Per molti anni, le architetture dominanti in questo settore sono state basate su reti neurali convoluzionali, tra cui U-Net e le sue numerose varianti. Sebbene tali modelli abbiano raggiunto risultati eccellenti in numerose applicazioni, essi presentano alcune limitazioni legate alla capacità di generalizzazione e alla necessità di dataset accuratamente annotati.

L'avvento dei Foundation Models ha aperto nuove prospettive anche nel campo della segmentazione. L'idea fondamentale consiste nello sviluppare sistemi in grado di segmentare oggetti mai osservati durante l'addestramento mantenendo prestazioni elevate. Questo approccio risulta particolarmente interessante per l'imaging medico, dove la variabilità anatomica e patologica rende difficile coprire tutte le possibili condizioni cliniche mediante dataset tradizionali.

6.2.2 Segment Anything Model (SAM)

Uno dei più importanti sviluppi recenti nel campo della segmentazione automatica è rappresentato dal Segment Anything Model (SAM), introdotto da Meta AI nel 2023 [38].

SAM è stato progettato con l'obiettivo di creare un sistema di segmentazione general-purpose capace di identificare automaticamente oggetti appartenenti a categorie molto diverse tra loro. Per raggiungere questo risultato, il modello è stato addestrato utilizzando il dataset SA-1B, una delle più grandi raccolte di dati mai realizzate per applicazioni di segmentazione.

Il dataset SA-1B contiene oltre un miliardo di maschere di segmentazione associate a milioni di immagini provenienti da differenti contesti applicativi. Questa enorme quantità di dati ha consentito al modello di apprendere rappresentazioni estremamente generali delle strutture visive presenti nelle immagini.

L'architettura di SAM è composta da tre moduli principali, schematizzati in Figura 6.3:

- Un Image Encoder
- Un Prompt Encoder
- Un Mask Decoder

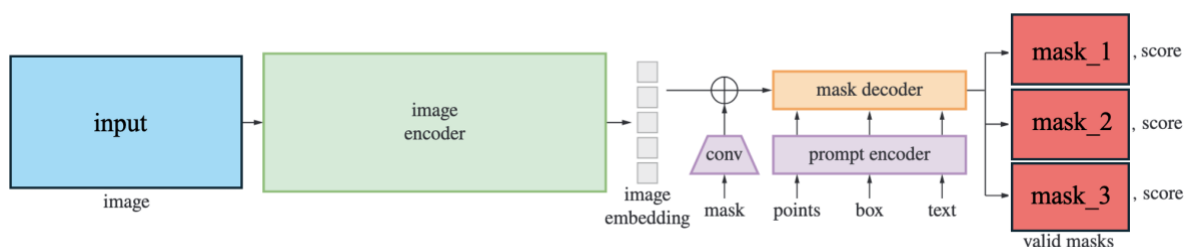


Figura 6.3: Schema generale del Segment Anything Model. Il modello riceve in ingresso un'immagine e uno o più prompt, come punti o bounding box, e genera una maschera di segmentazione attraverso la combinazione di Image Encoder, Prompt Encoder e Mask Decoder. Immagine adattata da J. Ma et al. [16]

L'Image Encoder è responsabile dell'estrazione delle caratteristiche visive dell'immagine. Attraverso questa fase il modello genera una rappresentazione compatta ma altamente informativa del contenuto dell'immagine.

Il Prompt Encoder, invece, codifica le informazioni fornite dall'utente. Queste possono essere costituite da punti positivi o negativi, bounding box o maschere precedenti. In questo modo, il modello riceve una guida esplicita su quale regione dell'immagine debba essere segmentata. Questa caratteristica rappresenta una delle innovazioni principali di SAM, poiché consente un'interazione flessibile tra utente e modello, rendendo il processo di segmentazione più adattabile rispetto ai metodi completamente automatici tradizionali.

Infine, il Mask Decoder combina le informazioni estratte dall'immagine con quelle provenienti dal prompt e genera la maschera finale. Il decoder è progettato per essere rapido e leggero, in modo da permettere una risposta quasi immediata dopo l'inserimento del prompt. In caso di ambiguità, il modello può produrre più maschere candidate, lasciando all'utente o al sistema la possibilità di selezionare quella più appropriata.

Tuttavia, è importante sottolineare che SAM è stato originariamente sviluppato e addestrato principalmente su immagini naturali, cioè fotografie appartenenti al mondo reale. Questa caratteristica, come discusso nel paragrafo successivo, rappresenta un limite quando il modello viene applicato direttamente all'imaging medico.

6.2.3 Limiti di SAM nell'Imaging Medico

Nonostante i risultati molto rilevanti ottenuti da SAM nel campo delle immagini naturali, la sua applicazione diretta all'imaging medico non è priva di criticità. Le immagini mediche, infatti, presentano caratteristiche profondamente diverse rispetto alle fotografie utilizzate nella fase di pre-addestramento del modello. Una fotografia naturale contiene in genere oggetti con bordi relativamente riconoscibili, colori e texture simili a quelle osservate nel mondo reale. Al contrario, immagini come ecografie, risonanze magnetiche o tomografie computerizzate rappresentano strutture anatomiche interne, spesso con livelli di contrasto ridotti, rumore elevato e confini non sempre chiaramente distinguibili.

Questa differenza tra il dominio di addestramento e il dominio di applicazione viene comunemente indicata con il termine *domain shift*. Nel caso di SAM, il *domain shift* è particolarmente rilevante perché il modello è stato addestrato su immagini naturali, mentre le immagini mediche seguono logiche visive e statistiche completamente differenti. Di conseguenza, anche se SAM mantiene una certa capacità di generalizzazione, le sue prestazioni possono diminuire quando viene utilizzato senza adattamento su dati biomedicali [16].

Il problema diventa ancora più evidente nella segmentazione di strutture anatomiche piccole, poco contrastate o caratterizzate da margini sfumati. In molte applicazioni mediche, infatti, il confine tra la regione di interesse e i tessuti circostanti non è netto. Questo aspetto rende difficile anche la segmentazione manuale da parte di un operatore esperto e può portare il modello a produrre maschere incomplete, eccessivamente estese o non perfettamente aderenti alla struttura anatomica.

Nel caso specifico dell'ecografia, queste difficoltà risultano ulteriormente amplificate. Le immagini ecografiche sono caratterizzate dalla presenza di rumore *speckle*, da una qualità fortemente dipendente dall'operatore e da un contrasto spesso limitato tra strutture adiacenti.

Inoltre, la stessa struttura anatomica può apparire in modo diverso a seconda dell'orientamento della sonda, della profondità di acquisizione, delle impostazioni del dispositivo e delle caratteristiche del paziente. Questi fattori rendono complesso l'utilizzo diretto di un modello generalista non specificamente adattato al dominio ecografico.

Un ulteriore limite riguarda la necessità di fornire prompt accurati. SAM è pensato come modello interattivo: la qualità della segmentazione dipende quindi anche dalla qualità del prompt utilizzato. Nel contesto clinico, tuttavia, un sistema di segmentazione ideale dovrebbe ridurre il più possibile l'intervento manuale, soprattutto quando devono essere analizzate grandi quantità di immagini. Per questo motivo, diversi lavori recenti hanno cercato di adattare SAM al dominio medico e, in particolare, all'ecografia, proponendo strategie di auto-prompting o modifiche architetturali mirate [39].

Queste limitazioni non riducono l'importanza di SAM, ma evidenziano la necessità di adattarlo al contesto biomedicale. In altre parole, SAM rappresenta un punto di partenza molto potente, ma non sempre sufficiente quando viene applicato direttamente a immagini cliniche. Da questa esigenza nasce lo sviluppo di modelli specificamente progettati per il dominio medico, tra cui MedSAM.

6.2.4 MedSAM

Per superare le limitazioni osservate nell'applicazione diretta di SAM all'imaging medico, è stato sviluppato MedSAM (Medical Segment Anything Model).

MedSAM rappresenta una versione specializzata del framework originale progettata specificamente per l'analisi di immagini mediche. L'obiettivo principale era preservare la capacità di SAM di utilizzare prompt flessibili, adattandola alle peculiarità dei dati biomedicali.

Per adattare SAM al dominio medico, MedSAM è stato sviluppato attraverso una fase di addestramento su una vasta raccolta di immagini diagnostiche (pipeline schematizzata in Figura 6.4). Il dataset utilizzato comprende oltre un milione e mezzo di coppie immagine-maschera e include diverse modalità di imaging, tra cui tomografia computerizzata, risonanza magnetica, ecografia, endoscopia e altre tecniche di acquisizione clinica [16]. Questa varietà ha permesso al modello di apprendere rappresentazioni più vicine alle caratteristiche delle immagini mediche rispetto a quelle apprese dal SAM originale.

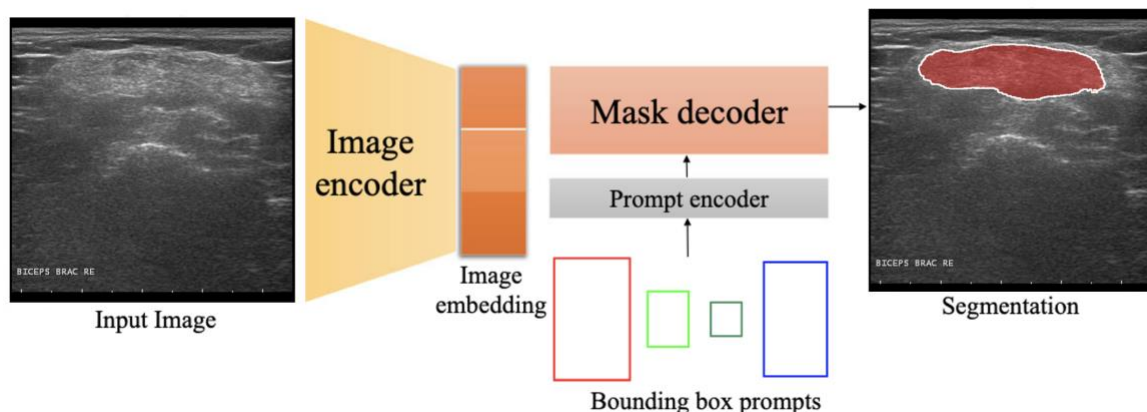


Figura 6.4: Esempio di applicazione di MedSAM alla segmentazione ecografica dei muscoli scheletrici (00083_001_00_1). La bounding box viene utilizzata come prompt per guidare il modello verso la regione anatomica di interesse; la maschera ottenuta viene successivamente confrontata con l'annotazione manuale. Immagine adattata da Ma et al. [16].

Attraverso questo processo di adattamento, il modello ha potuto apprendere caratteristiche specifiche del dominio medico mantenendo contemporaneamente la flessibilità dell'architettura originale.

Uno degli aspetti più interessanti di MedSAM è la sua capacità di affrontare compiti di segmentazione anche in scenari caratterizzati da una disponibilità limitata di dati annotati. Questa caratteristica risulta particolarmente vantaggiosa nel caso delle immagini ecografiche muscolari, dove la creazione di annotazioni manuali richiede competenze altamente specializzate e tempi considerevoli.

MedSAM assume particolare interesse perché permette di valutare un approccio diverso rispetto a quello seguito con MMSegmentation. Mentre i modelli implementati tramite MMSegmentation sono stati addestrati in modo specifico sul dataset di muscoli facciali, MedSAM viene utilizzato come modello fondazionale già pre-addestrato sul dominio medico. Il confronto tra questi due approcci consente quindi di analizzare non solo le prestazioni finali, ma anche il diverso grado di generalizzazione e robustezza rispetto alle immagini ecografiche considerate.

6.2.5 Motivazioni della Scelta di MedSAM

La scelta di introdurre MedSAM nella seconda fase sperimentale del presente lavoro nasce dai risultati ottenuti con l'approccio basato su MMSegmentation. Sebbene i modelli addestrati con MMSegmentation abbiano permesso di costruire una pipeline completa per la segmentazione dei muscoli facciali, le prestazioni ottenute hanno evidenziato alcune difficoltà ancora presenti, soprattutto in relazione alla complessità delle immagini ecografiche e alla limitata definizione dei confini muscolari.

La scelta di MedSAM è stata motivata da diversi fattori. In primo luogo, il modello aveva già dimostrato prestazioni particolarmente promettenti in numerose applicazioni di imaging medico. In secondo luogo, la sua natura di foundation model suggeriva una possibile maggiore stabilità rispetto a variazioni moderate dell'immagine e una buona capacità di adattamento a compiti di segmentazione guidata da prompt.

Un ulteriore elemento che ha motivato la scelta di MedSAM riguarda la sua possibile robustezza rispetto alle variazioni introdotte dal preprocessing. Nel presente lavoro, infatti, sono state considerate immagini originali e immagini sottoposte a 2 diverse tecniche di miglioramento del contrasto, come CLAHE e CLAHE_LOW. L'utilizzo di un modello fondazionale permette di valutare se una rappresentazione appresa su larga scala nel dominio medico sia meno sensibile alle variazioni di preprocessing rispetto a un modello addestrato in modo più specifico sul dataset disponibile.

La scelta di MedSAM è inoltre coerente con l'evoluzione recente della letteratura, che mostra un crescente interesse verso modelli SAM-based adattati all'imaging medico e, più in particolare, all'ecografia. Lin et al. [39] evidenziano come la segmentazione ecografica possa beneficiare di adattamenti specifici di SAM, orientati a migliorare la rappresentazione delle informazioni locali e a ridurre la dipendenza da prompt manuali. Questo conferma che l'ecografia rappresenta un dominio particolarmente sfidante, ma anche molto promettente per l'applicazione di modelli fondazionali.

Per queste ragioni, MedSAM è stato selezionato come principale candidato per la seconda fase sperimentale del presente lavoro. I capitoli successivi descrivono la metodologia adottata per la sua implementazione e analizzano in dettaglio i risultati ottenuti nel confronto con il framework MMSegmentation.

6.3 Metriche di Valutazione

Le prestazioni del modello sono state valutate mediante metriche standard utilizzate nella letteratura scientifica dedicata alla segmentazione semantica.

Le principali metriche considerate sono state:

- Dice Similarity Coefficient (DSC);
- Intersection over Union (IoU);
- Segmentation Failure

DSC e IoU sono metriche di sovrapposizione ampiamente utilizzate nella valutazione della segmentazione di immagini mediche, poiché permettono di quantificare in modo diretto il grado di accordo spaziale tra la regione segmentata automaticamente e quella annotata manualmente [40], [41].

Indicando con P_i l'insieme dei pixel appartenenti alla maschera predetta per l'immagine i e con G_i l'insieme dei pixel appartenenti alla corrispondente maschera manuale, il confronto tra le due maschere può essere espresso attraverso i veri positivi, i falsi positivi e i falsi negativi, come riportato nell'Equazione (6.1).

$$TP_i = |P_i \cap G_i|, FP_i = |P_i \setminus G_i|, FN_i = |G_i \setminus P_i| \quad (6.1)$$

In cui TP_i rappresenta il numero di pixel correttamente classificati come muscolo, FP_i indica i pixel classificati erroneamente come muscolo dal modello e NP_i rappresenta i pixel appartenenti al muscolo nella ground truth ma non riconosciuti dalla segmentazione automatica.

Il Dice Similarity Coefficient misura il grado di sovrapposizione tra la maschera predetta e la maschera di riferimento. Questa metrica è particolarmente diffusa nella segmentazione medica perché fornisce una misura intuitiva della concordanza tra due regioni segmentate. Il Dice coefficient è stato calcolato secondo l'Equazione (6.2).

$$DSC_i = \frac{2|P_i \cap G_i|}{|P_i| + |G_i|} \quad (6.2)$$

Utilizzando la formulazione in termini di veri positivi, falsi positivi e falsi negativi, la stessa metrica può essere scritta come indicato nell'Equazione (6.3).

$$DSC_i = \frac{2TP_i}{2TP_i + FP_i + FN_i} \quad (6.3)$$

Il valore del Dice coefficient varia tra 0 e 1. Un valore pari a 0 indica assenza di sovrapposizione tra predizione e ground truth, mentre un valore pari a 1 indica una segmentazione perfettamente sovrapponibile alla maschera manuale.

L'Intersection over Union, indicata anche come IoU o Jaccard Index, misura invece il rapporto tra l'intersezione e l'unione della maschera predetta e della maschera di riferimento. Questa metrica è molto utilizzata nella computer vision e nella segmentazione semantica, in quanto valuta quanto l'area predetta coincida con l'area reale dell'oggetto da segmentare [40], [42]. La IoU è stata calcolata secondo l'Equazione (6.4).

$$IoU_i = \frac{|P_i \cap G_i|}{|P_i \cup G_i|} \quad (6.4)$$

In termini di veri positivi, falsi positivi e falsi negativi, la IoU può essere espressa come riportato nell'Equazione (6.5).

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i} \quad (6.5)$$

Anche la IoU assume valori compresi tra 0 e 1. Valori elevati indicano una buona sovrapposizione tra la maschera predetta e quella manuale, mentre valori bassi indicano una segmentazione meno accurata. Rispetto al Dice coefficient, la IoU tende a essere più severa, poiché il denominatore considera l'intera area appartenente ad almeno una delle due maschere.

Dice coefficient e IoU sono metriche strettamente correlate. La relazione tra le due può essere espressa come riportato nell'Equazione (6.6).

$$IoU_i = \frac{DSC_i}{2 - DSC_i} \quad (6.6)$$

Questa relazione evidenzia come, a parità di segmentazione, il valore del Dice Coefficient risulti generalmente più alto rispetto alla IoU. Per questo motivo, riportare entrambe le metriche consente una valutazione più completa delle prestazioni del modello.

Oltre alle metriche continue di sovrapposizione, è stata considerata anche la Segmentation Failure, definita come la presenza di una segmentazione particolarmente scarsa e redatta come nell'Equazione 6.7. Nel lavoro di riferimento [7], i fallimenti di segmentazione sono stati definiti utilizzando una soglia basata sulla IoU. Nel presente lavoro, invece, per rendere omogeneo il confronto tra le diverse configurazioni sperimentali, la segmentation failure è stata definita come il caso in cui il Dice Similarity Coefficient risulta inferiore a 0,5.

$$SF_i = \begin{cases} 1, & \text{per } IoU_i < 0.5 \\ 0, & \text{per } IoU_i > 0.5 \end{cases} \quad (6.7)$$

La percentuale complessiva di fallimenti sul dataset è stata quindi calcolata come riportato nell'Equazione (6.8), dove N rappresenta il numero totale di immagini valutate.

$$SF(\%) = \frac{1}{N} \sum_{i=1}^N SF_i \cdot 100 \quad (6.8)$$

Questa metrica non descrive il grado medio di sovrapposizione, ma permette di quantificare la frequenza dei casi critici, cioè delle immagini in cui il modello non riesce a produrre una segmentazione sufficientemente affidabile.

Per confrontare in modo sintetico le diverse condizioni sperimentali, i valori di Dice e IoU sono stati mediati sull'intero insieme di immagini valutate. La media del Dice coefficient è stata calcolata secondo l'Equazione (6.9).

$$\overline{DSC} = \frac{1}{N} \sum_{i=1}^N DSC_i \quad (6.9)$$

Analogamente, la media della IoU è stata calcolata secondo l'Equazione (6.10).

$$\overline{IoU} = \frac{1}{N} \sum_{i=1}^N IoU_i \quad (6.10)$$

L'utilizzo combinato di DSC, IoU e Segmentation Failure consente quindi di valutare le prestazioni del modello da prospettive complementari. Il Dice coefficient fornisce una misura intuitiva della sovrapposizione globale, la IoU offre una valutazione più severa dell'accordo tra predizione e ground truth, mentre la Segmentation Failure permette di evidenziare la frequenza dei casi peggiori.

Questa impostazione risulta particolarmente utile comprendere quanto le diverse tecniche di preprocessing influenzino la robustezza del modello nei casi più complessi.

7. Risultati sperimentali sui muscoli facciali

7.1 *MMSegmentation*

In questo capitolo vengono riportati i risultati ottenuti mediante il framework *MMSegmentation* sui quattro muscoli facciali considerati nel presente lavoro: Depressor Anguli Oris, Mentalis, Orbicularis Oris e Zygomaticus. L'obiettivo dell'analisi è valutare in modo comparativo l'effetto di alcune strategie di preprocessing sulle maschere di segmentazione prodotte dal modello.

Sono state considerate quattro condizioni sperimentali: immagini originali, immagini preprocessate mediante CLAHE, immagini preprocessate mediante CLAHE_LOW e immagini preprocessate mediante GUIDED_LOW. Per ciascuna condizione, le maschere predette sono state confrontate con le corrispondenti annotazioni manuali, utilizzate come ground truth.

Le prestazioni sono state valutate mediante Dice Similarity Coefficient, Intersection over Union e percentuale di segmentation failure. In questa analisi, la segmentation failure è stata definita come il caso in cui il Dice Similarity Coefficient risulta inferiore a 0,5. I risultati vengono presentati principalmente attraverso grafici a barre, in modo da evidenziare le differenze relative tra le diverse condizioni sperimentali.

7.1.1 Confronto quantitativo globale

L'analisi quantitativa condotta sull'insieme dei quattro muscoli facciali evidenzia prestazioni complessivamente limitate del framework *MMSegmentation* in questo specifico distretto anatomico.

Considerando le 6654 immagini analizzate, le immagini originali hanno ottenuto un Dice medio pari a $0,504 \pm 0,289$ e una IoU media pari a $0,383 \pm 0,244$, con una percentuale di segmentation failure pari al 38,28%.

L'applicazione delle tecniche di preprocessing non ha determinato un miglioramento globale netto delle prestazioni. CLAHE ha ottenuto un Dice medio pari a $0,492 \pm 0,288$ e una IoU media pari a $0,373 \pm 0,243$, mentre CLAHE_LOW ha raggiunto un Dice medio pari a $0,500 \pm 0,291$ e una IoU media pari a $0,381 \pm 0,246$. GUIDED_LOW ha mostrato valori simili a CLAHE, con Dice medio pari a $0,492 \pm 0,288$ e IoU media pari a $0,372 \pm 0,242$.

Nel complesso, le differenze tra le configurazioni risultano contenute. Le immagini originali mantengono il valore medio più elevato di Dice e la percentuale più bassa di segmentation failure, mentre CLAHE_LOW mostra valori molto vicini a quelli ottenuti sulle immagini originali. Questo risultato suggerisce che, nel caso di *MMSegmentation*, il preprocessing non produce un vantaggio globale sistematico sui quattro muscoli facciali.

Il confronto globale tra le quattro condizioni sperimentali è riportato in Figura 7.1 e in Figura 7.2. Come si può osservare, i valori medi di Dice risultano molto vicini tra loro, senza un miglioramento netto associato all'applicazione del preprocessing. Le immagini originali mantengono il valore medio leggermente più elevato, mentre CLAHE_LOW mostra prestazioni molto simili.

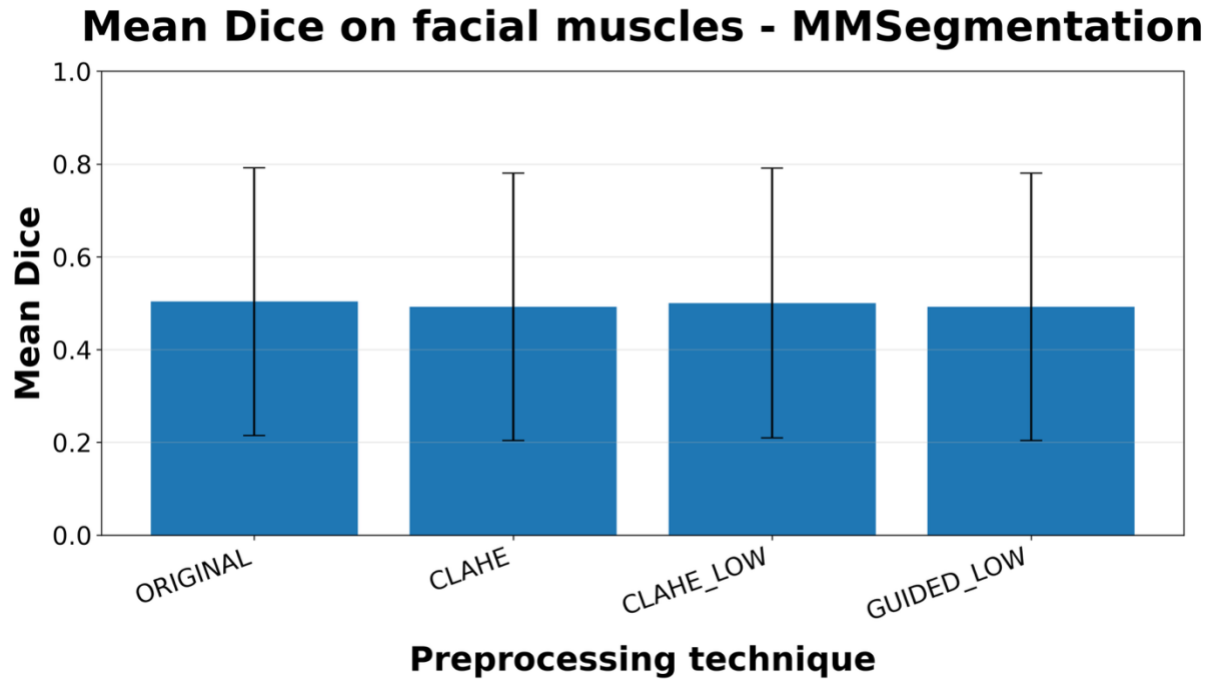


Figura 7.1: Dice medio ottenuto mediante MMSegmentation sui quattro muscoli facciali nelle quattro condizioni sperimentali considerate: ORIGINAL, CLAHE, CLAHE_LOW e GUIDED_LOW. Le barre rappresentano il valore medio, mentre le linee verticali indicano la deviazione standard.

Un andamento analogo si osserva considerando la percentuale di segmentation failure, riportata in Figura 7.2. Anche in questo caso le differenze tra le tecniche sono contenute, con valori compresi tra circa il 38% e il 40%. Questo risultato conferma che il preprocessing non riduce in modo sostanziale la frequenza dei casi critici.

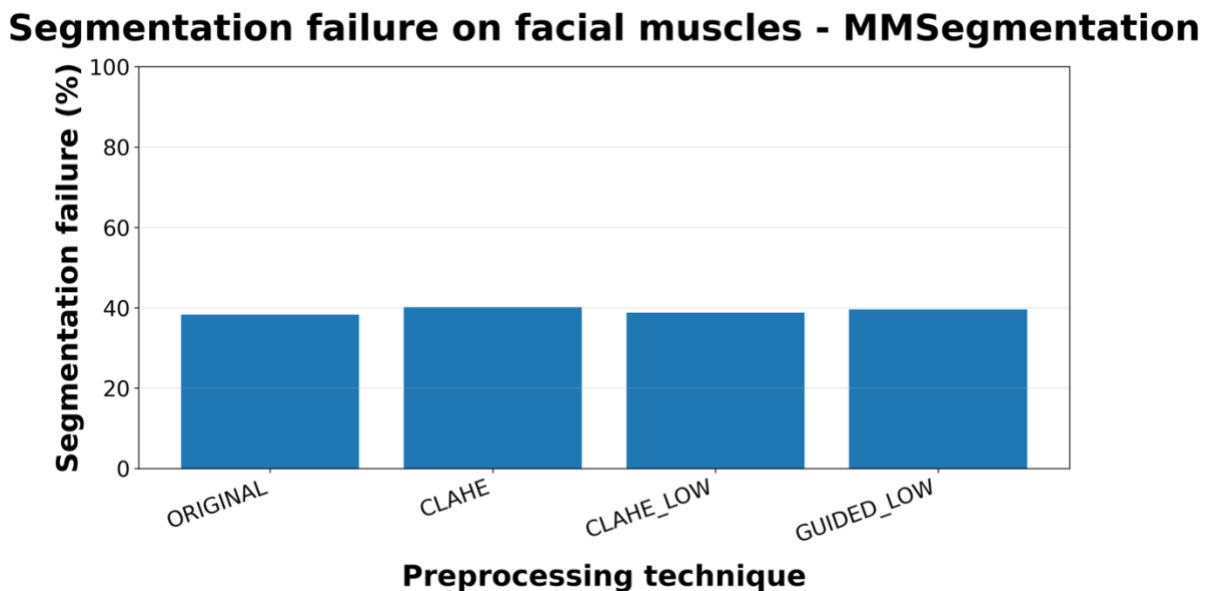


Figura 7.2: Percentuale di segmentation failure ottenuta mediante MMSegmentation sui quattro muscoli facciali. La failure è stata definita come Dice Similarity Coefficient inferiore a 0,5.

7.1.2 Confronto per singolo muscolo

L'analisi per singolo muscolo mostra un comportamento più articolato rispetto al confronto globale. Il preprocessing non produce infatti lo stesso effetto su tutti i distretti anatomici considerati.

Per il Depressor Anguli Oris, la configurazione CLAHE_LOW ha ottenuto il Dice medio più elevato, pari a $0,571 \pm 0,309$, rispetto a $0,564 \pm 0,312$ delle immagini originali. Anche la percentuale di segmentation failure risulta leggermente inferiore con CLAHE_LOW, pari al 27,53%, rispetto al 28,53% delle immagini originali.

Nel caso del Mentalis, le differenze tra le tecniche sono molto ridotte. CLAHE_LOW ha ottenuto il valore medio di Dice più alto, pari a $0,556 \pm 0,189$, mentre le immagini originali hanno raggiunto un valore pari a $0,556 \pm 0,182$. Anche in questo caso, quindi, il preprocessing produce solo variazioni numeriche molto contenute.

L'Orbicularis Oris è risultato il muscolo più critico tra quelli analizzati. Le immagini originali hanno ottenuto un Dice medio pari a $0,400 \pm 0,313$, mentre CLAHE, CLAHE_LOW e GUIDED_LOW hanno mostrato valori rispettivamente pari a $0,384 \pm 0,317$, $0,383 \pm 0,320$ e $0,396 \pm 0,312$. La percentuale di segmentation failure rimane superiore al 54% in tutte le configurazioni, confermando la particolare difficoltà di segmentazione di questo muscolo.

Anche lo Zygomaticus ha mostrato prestazioni limitate. Il Dice medio più elevato è stato ottenuto sulle immagini originali, con un valore pari a $0,497 \pm 0,288$. CLAHE_LOW ha raggiunto un valore leggermente inferiore, pari a $0,493 \pm 0,284$, mentre CLAHE e GUIDED_LOW hanno ottenuto rispettivamente $0,486 \pm 0,280$ e $0,488 \pm 0,281$.

Per comprendere meglio l'effetto delle diverse tecniche, il confronto è stato successivamente analizzato separatamente per ciascun muscolo facciale. I risultati sono riportati in Figura 7.3. Questo confronto evidenzia come l'effetto del preprocessing non sia uniforme tra i diversi muscoli.

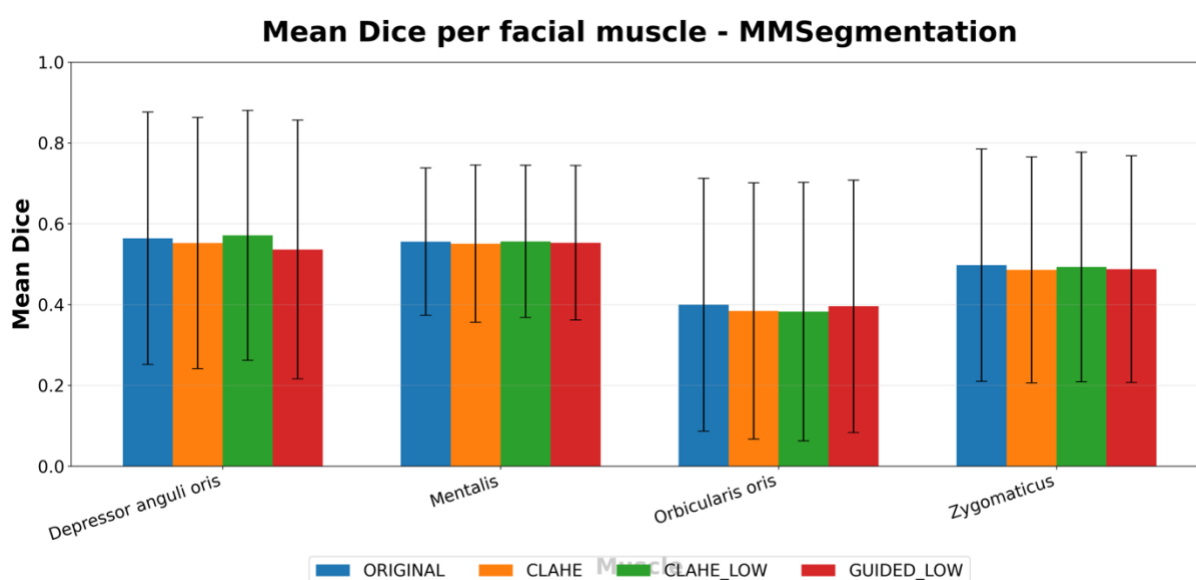


Figura 7.3: Dice medio per ciascuno dei quattro muscoli facciali analizzati mediante MMsegmentation. Le barre rappresentano le quattro condizioni sperimentali considerate. L'ampiezza delle deviazioni standard evidenzia l'elevata variabilità delle prestazioni, particolarmente marcata per Orbicularis Oris e Zygomaticus.

Questi risultati indicano che l'effetto del preprocessing dipende dalle caratteristiche anatomiche e visive del singolo muscolo. In particolare, CLAHE_LOW sembra fornire un lieve vantaggio per Depressor Anguli Oris e Mentalis, mentre non migliora le prestazioni per Orbicularis Oris e Zygomaticus. Il preprocessing, quindi, può modificare il comportamento del modello, ma non appare sufficiente a risolvere in modo stabile le difficoltà associate alla segmentazione dei muscoli facciali.

7.1.3 Limiti del solo preprocessing

L'analisi quantitativa descritta nei paragrafi precedenti mostra che le tecniche di preprocessing considerate modificano il comportamento del modello, ma non producono un miglioramento globale netto e uniforme delle prestazioni. Per questo motivo, oltre ai valori medi di Dice, IoU e segmentation failure, è utile osservare direttamente alcune maschere predette dal modello.

A questo scopo, sono stati selezionati due esempi qualitativi relativi all'Orbicularis oris, uno dei muscoli risultati più critici nell'analisi quantitativa. Il primo esempio, riportato in Figura 7.4, mostra un caso favorevole, caratterizzato da una buona sovrapposizione tra la maschera predetta e la ground truth.

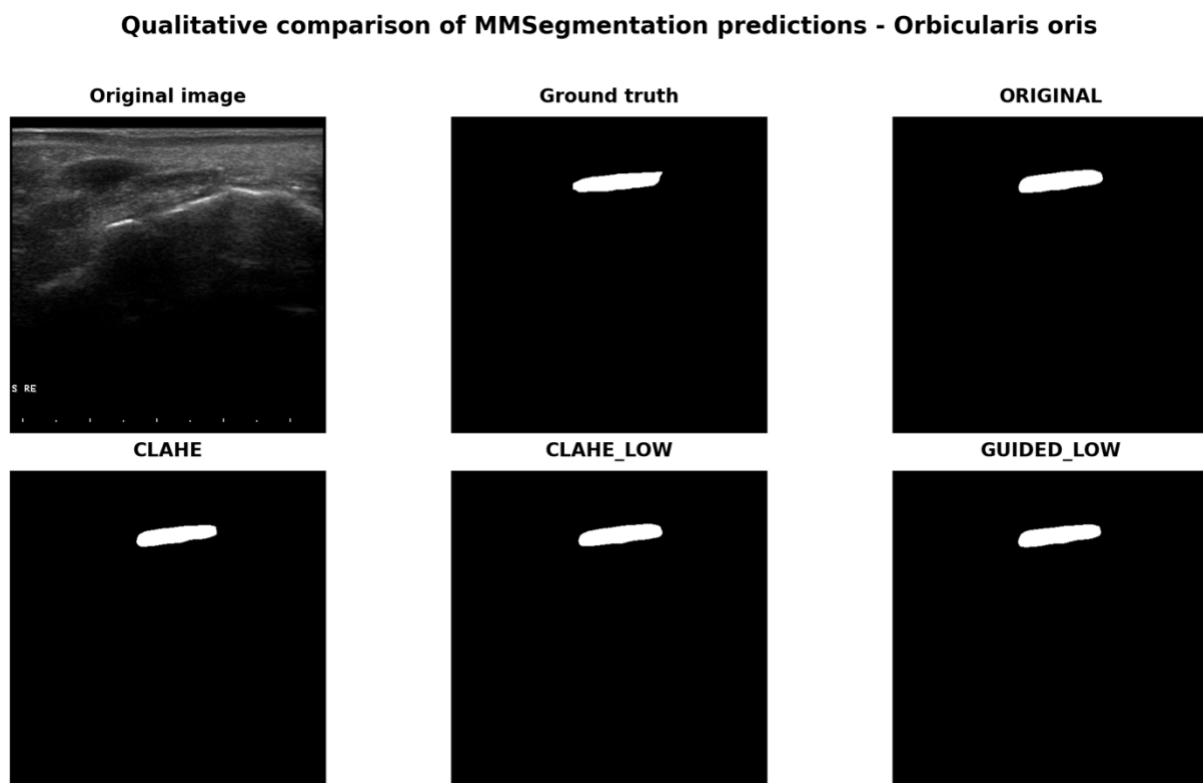


Figura 7.4: Confronto qualitativo tra immagine originale (01815_013_00_3), ground truth e maschere predette mediante MMSegmentation per un caso di Orbicularis oris caratterizzato da buona sovrapposizione tra predizione e annotazione manuale. La figura mostra come, nei casi più favorevoli, il modello riesca a produrre segmentazioni coerenti con la ground truth, pur in presenza di lievi differenze tra le diverse condizioni di preprocessing.

Come si osserva in Figura 7.4, nei casi più favorevoli il modello è in grado di individuare correttamente la regione muscolare e di produrre maschere compatibili con l'annotazione manuale. Le differenze tra le diverse condizioni di preprocessing risultano contenute e

riguardano principalmente piccole variazioni nella forma o nell'estensione della maschera predetta.

Tuttavia, questo comportamento non è costante su tutto il dataset. Un secondo esempio qualitativo è riportato in Figura 7.5 e mostra un caso più critico, in cui le predizioni risultano meno coerenti con la ground truth.

Qualitative comparison of MMSegmentation predictions - Orbicularis oris

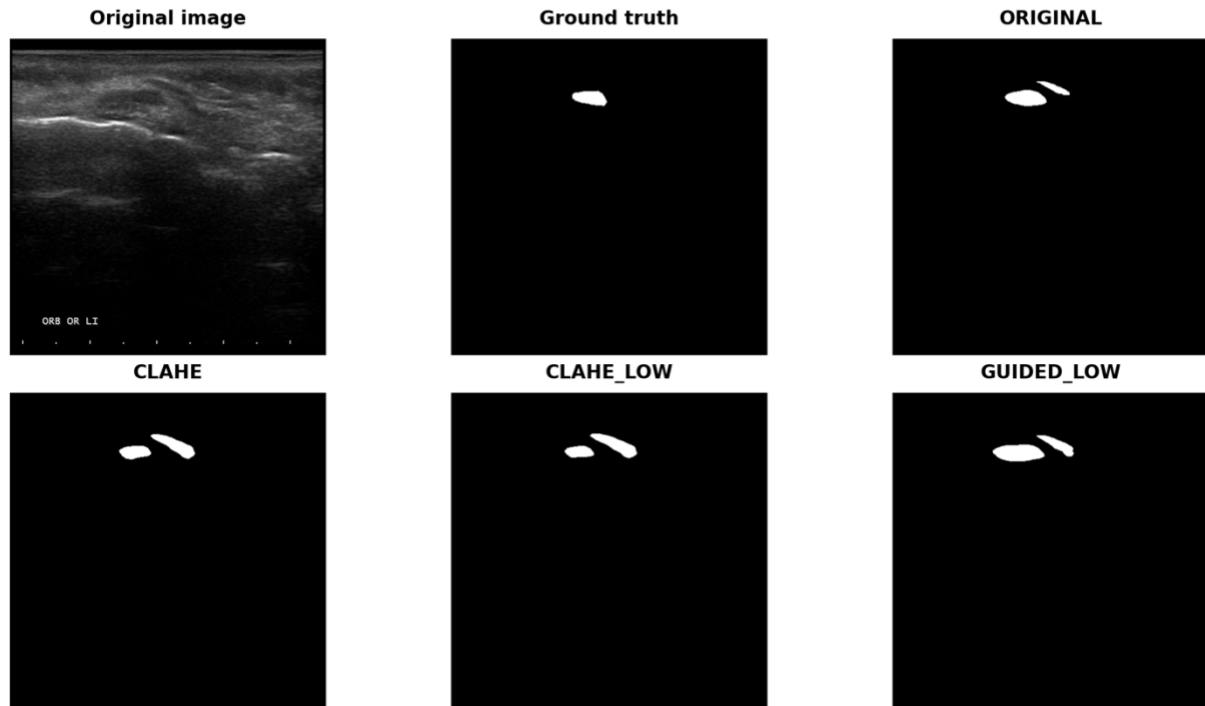


Figura 7.5: Confronto qualitativo tra immagine originale (02240_013_01_1), ground truth e maschere predette mediante MMSegmentation per un caso critico di Orbicularis oris. In questo esempio, le diverse configurazioni mostrano una sovrapposizione meno accurata rispetto alla ground truth, evidenziando i limiti del solo preprocessing nel migliorare in modo consistente la segmentazione dei muscoli facciali.

La Figura 7.5 evidenzia come, nei casi più complessi, il preprocessing non sia sufficiente a correggere gli errori del modello. Le maschere predette possono presentare una localizzazione non ottimale, una forma differente rispetto alla ground truth o una parziale sovrastima della regione muscolare. Questo aspetto è particolarmente rilevante nei muscoli facciali, dove anche piccoli errori spaziali possono determinare una riduzione marcata delle metriche di sovrapposizione.

Il confronto qualitativo conferma quanto osservato nei risultati quantitativi: il preprocessing può modificare localmente la segmentazione prodotta da MMSegmentation, ma non garantisce un miglioramento sistematico. Le difficoltà principali sembrano quindi dipendere non solo dalla qualità dell'immagine, ma anche dalla complessità anatomica dei muscoli facciali e dalla capacità del modello di riconoscere strutture piccole, poco contrastate e con margini non sempre ben definiti.

7.1.4 Considerazioni

I risultati ottenuti con MMSegmentation confermano la complessità della segmentazione automatica dei muscoli facciali in immagini ecografiche. Le prestazioni medie risultano

inferiori rispetto a quelle osservate in altri distretti muscolari più estesi e con margini anatomici più riconoscibili.

L'applicazione delle tecniche di preprocessing non ha prodotto un miglioramento globale netto. CLAHE_LOW ha mostrato un comportamento leggermente più favorevole in alcuni muscoli, ma le differenze rispetto alle immagini originali sono contenute e non uniformi. Per questo motivo, i risultati devono essere interpretati come differenze numeriche e non come evidenza di un miglioramento statisticamente significativo.

Nel complesso, questa fase sperimentale suggerisce che il solo preprocessing non sia sufficiente a superare le difficoltà associate alla segmentazione dei muscoli facciali. Queste osservazioni hanno motivato la successiva valutazione di MedSAM, con l'obiettivo di analizzare il comportamento di un foundation model medico in un contesto caratterizzato da strutture anatomiche piccole, basso contrasto e confini poco definiti.

7.2 MedSAM

Dopo aver completato gli esperimenti basati sul framework MMSegmentation, è stato valutato MedSAM come modello alternativo per la segmentazione di immagini ecografiche muscolari. L'obiettivo di questa fase non era addestrare un nuovo modello sul dataset considerato, ma analizzare il comportamento di un foundation model medico pre-addestrato in un contesto caratterizzato da immagini ecografiche, strutture anatomiche di piccole dimensioni e confini muscolari non sempre ben definiti.

MedSAM è stato utilizzato a partire dal checkpoint pre-addestrato `medsam_vit_b.pth`, senza effettuare fine-tuning sul dataset di ecografia muscolare analizzato. Il modello è stato quindi impiegato in modalità di inferenza, mantenendo invariati i pesi appresi durante l'addestramento originale sul dominio medico. Questa scelta ha permesso di valutare la capacità del modello di trasferire le rappresentazioni apprese a un dominio specifico, come quello dell'ecografia muscolare.

È importante precisare che, nel presente lavoro, MedSAM è stato utilizzato in modalità prompt-guided. In particolare, per ciascuna immagine è stata fornita al modello una bounding box associata alla regione muscolare di interesse. La bounding box è stata generata automaticamente a partire dalla maschera manuale di riferimento, calcolando il rettangolo minimo contenente tutti i pixel appartenenti alla ground truth. La stessa procedura è stata applicata alle immagini ORIGINAL, CLAHE e CLAHE_LOW, in modo da mantenere confrontabili le tre condizioni di preprocessing.

Di conseguenza, l'esperimento con MedSAM non deve essere interpretato come una pipeline di segmentazione completamente automatica end-to-end. Il modello non localizza autonomamente il muscolo all'interno dell'immagine, ma produce la maschera a partire da una regione di interesse già indicata tramite prompt. I risultati ottenuti valutano quindi la capacità di MedSAM di delineare correttamente il muscolo quando la sua localizzazione è già disponibile.

Gli esperimenti sono stati condotti su tre versioni del dataset: immagini originali, immagini preprocessate mediante CLAHE e immagini preprocessate mediante CLAHE_LOW. Le immagini sono state mantenute organizzate secondo i cinque fold sperimentali già definiti. Poiché MedSAM non è stato addestrato sui fold, tale suddivisione non rappresenta una cross-

validation di training del modello, ma una valutazione sulle porzioni di test dei cinque fold, successivamente aggregate per ottenere le metriche complessive.

Le prestazioni sono state valutate mediante Dice Similarity Coefficient, Intersection over Union e segmentation failure. In coerenza con l'analisi condotta per MMSegmentation, la segmentation failure è stata definita come il caso in cui il Dice Similarity Coefficient risulta inferiore a 0,5. I valori sono stati calcolati per ciascuna immagine e successivamente aggregati sia sui quattro muscoli facciali principali, sia sull'intero dataset dei 16 muscoli.

7.2.1 Risultati sui quattro muscoli facciali principali

Considerando esclusivamente i quattro muscoli facciali principali analizzati, MedSAM ha ottenuto prestazioni superiori rispetto a quelle osservate con MMSegmentation (risultati visibili nella Tabella 7.1). Sulle immagini originali, il Dice medio è risultato pari a $0,741 \pm 0,124$, con una IoU media pari a $0,602 \pm 0,143$ e una segmentation failure del 5,91%. Le immagini preprocessate mediante CLAHE e CLAHE_LOW hanno mostrato valori molto simili, rispettivamente pari a $0,733 \pm 0,125$ e $0,737 \pm 0,124$ in termini di Dice medio.

Tabella 7.1: Prestazioni complessive di MedSAM sui quattro muscoli facciali principali nelle tre condizioni analizzate. I valori sono riportati come media $\pm$ deviazione standard. La segmentation failure è stata definita come Dice Similarity Coefficient inferiore a 0,5.

Preprocessing	Support	DSC	IoU	Segmentation failure (%)
ORIGINAL	6654	$0,741 \pm 0,124$	$0,602 \pm 0,143$	5,91
CLAHE	6654	$0,733 \pm 0,125$	$0,593 \pm 0,143$	6,18
CLAHE_LOW	6654	$0,737 \pm 0,124$	$0,597 \pm 0,142$	6,01

L'analisi aggregata mostra che MedSAM ottiene prestazioni superiori rispetto a quelle osservate con MMSegmentation sui quattro muscoli facciali principali. Sulle immagini originali, il modello raggiunge un DSC medio pari a $0,741 \pm 0,124$ e una IoU media pari a $0,602 \pm 0,143$, con una segmentation failure del 5,91%. Le immagini preprocessate mediante CLAHE e CLAHE_LOW mostrano valori molto simili, rispettivamente pari a $0,733 \pm 0,125$ e $0,737 \pm 0,124$ in termini di DSC medio.

Nel complesso, il preprocessing non determina un miglioramento netto delle prestazioni di MedSAM sui quattro muscoli facciali principali. Le immagini originali mantengono il valore medio di DSC più elevato, mentre CLAHE_LOW mostra prestazioni leggermente superiori rispetto a CLAHE standard. Le differenze tra le tre condizioni rimangono tuttavia contenute, suggerendo una limitata sensibilità di MedSAM alle modifiche moderate introdotte dal preprocessing.

Per comprendere meglio il comportamento del modello nei diversi distretti facciali, l'analisi aggregata è stata successivamente dettagliata per singolo muscolo. Questa distinzione è importante perché la media complessiva può mascherare differenze rilevanti tra strutture anatomiche con dimensioni, morfologia e visibilità ecografica differenti. I risultati sono riportati nella Tabella 7.2.

Tabella 7.2: Prestazioni di MedSAM sui quattro muscoli facciali principali, suddivise per muscolo e per tecnica di preprocessing. Il Mentalis mostra le prestazioni più critiche, con valori di DSC e IoU inferiori e una percentuale di segmentation failure nettamente più elevata rispetto agli altri muscoli.

Muscle	Preprocessing	Support	DSC	IoU	Segmentation failure (%)
Depressor Anguli Oris	ORIGINAL	1500	$0,792 \pm 0,092$	$0,664 \pm 0,116$	1,13
Depressor Anguli Oris	CLAHE	1500	$0,779 \pm 0,096$	$0,648 \pm 0,119$	1,47
Depressor Anguli Oris	CLAHE_LOW	1500	$0,786 \pm 0,091$	$0,656 \pm 0,115$	0,93
Mentalis	ORIGINAL	1244	$0,609 \pm 0,148$	$0,454 \pm 0,153$	25,80
Mentalis	CLAHE	1244	$0,604 \pm 0,142$	$0,447 \pm 0,146$	25,08
Mentalis	CLAHE_LOW	1244	$0,608 \pm 0,144$	$0,452 \pm 0,149$	24,76
Orbicularis Oris	ORIGINAL	1341	$0,765 \pm 0,102$	$0,630 \pm 0,123$	1,42
Orbicularis Oris	CLAHE	1341	$0,751 \pm 0,106$	$0,612 \pm 0,128$	2,39
Orbicularis Oris	CLAHE_LOW	1341	$0,759 \pm 0,103$	$0,621 \pm 0,124$	2,16
Zygomaticus	ORIGINAL	2569	$0,763 \pm 0,089$	$0,624 \pm 0,109$	1,40
Zygomaticus	CLAHE	2569	$0,760 \pm 0,095$	$0,622 \pm 0,114$	1,75
Zygomaticus	CLAHE_LOW	2569	$0,760 \pm 0,095$	$0,621 \pm 0,113$	1,91

L'analisi per singolo muscolo evidenzia un comportamento non uniforme. Il Depressor Anguli Oris, l'Orbicularis Oris e lo Zygomaticus presentano valori di DSC compresi approssimativamente tra 0,76 e 0,79, con percentuali di segmentation failure generalmente basse. In questi casi, MedSAM riesce a produrre segmentazioni complessivamente coerenti con le maschere manuali di riferimento, pur mantenendo piccole variazioni tra le diverse condizioni di preprocessing.

Il Mentalis rappresenta invece il caso più critico. Sulle immagini originali, questo muscolo ottiene un DSC medio pari a $0,609 \pm 0,148$ e una IoU media pari a $0,454 \pm 0,153$, con una segmentation failure pari al 25,80%. Valori simili sono osservati anche dopo CLAHE e CLAHE_LOW, confermando che il preprocessing non è sufficiente a risolvere le difficoltà associate a questo distretto anatomico. Tale comportamento può essere attribuito alle ridotte dimensioni del muscolo, alla scarsa definizione dei margini e alla presenza di tessuti circostanti con caratteristiche ecografiche simili.

Questi risultati indicano che MedSAM, quando guidato da bounding box, migliora sensibilmente la qualità della segmentazione rispetto alla pipeline basata su MMSegmentation, ma non elimina completamente le criticità associate ai piccoli muscoli facciali. Inoltre, il beneficio del preprocessing rimane limitato: CLAHE_LOW mostra un comportamento

leggermente più equilibrato rispetto a CLAHE standard, ma le immagini originali mantengono le prestazioni medie più elevate nell'analisi aggregata.

Dopo l'analisi focalizzata sui quattro muscoli facciali principali, MedSAM è stato valutato anche sull'intero dataset dei 16 muscoli. Questa analisi ha lo scopo di fornire una visione più generale del comportamento del modello su distretti anatomici differenti, includendo sia muscoli facciali sia muscoli degli arti e del tronco. I risultati globali ottenuti sull'intero dataset non devono però essere confusi con le prestazioni specifiche sui quattro muscoli facciali principali, che rappresentano il focus primario del presente lavoro.

7.2.2 Risultati estesi sull'intero dataset - Original

La prima valutazione quantitativa è stata condotta considerando le immagini ecografiche originali, cioè non sottoposte ad alcuna tecnica di preprocessing. Questa analisi rappresenta il punto di partenza per comprendere il comportamento di MedSAM sul dataset nella sua forma iniziale e consente di stabilire un riferimento rispetto al quale confrontare successivamente i risultati ottenuti con CLAHE e CLAHE_LOW.

La Tabella 7.3 riporta le prestazioni ottenute da MedSAM sulle immagini originali. Per ciascun muscolo sono indicati il numero di immagini analizzate, il valore medio del Dice coefficient con relativa deviazione standard, il valore medio dell'Intersection over Union con deviazione standard e la percentuale di segmentation failure.

Tabella 7.3: Risultati quantitativi ottenuti da MedSAM sulle immagini originali

Muscle type	Region	DSC	IoU	Segmentation failure (%)
Biceps brachii	ARM	$0,842 \pm 0,051$	$0,731 \pm 0,071$	0,29
Deltoideus	ARM	$0,832 \pm 0,051$	$0,716 \pm 0,071$	0,00
Depressor anguli_oris, DAO	FACE	$0,792 \pm 0,092$	$0,664 \pm 0,116$	1,13
Digastricus	FACE	$0,855 \pm 0,061$	$0,751 \pm 0,088$	0,07
Gastrocnemius_medial_head	LEG	$0,847 \pm 0,039$	$0,737 \pm 0,057$	0,00
Geniohyoideus	FACE	$0,875 \pm 0,041$	$0,780 \pm 0,063$	0,00
Masseter	FACE	$0,849 \pm 0,047$	$0,740 \pm 0,069$	0,00
Mentalis	FACE	$0,609 \pm 0,148$	$0,454 \pm 0,153$	25,80
Orbicularis_oris, OO	FACE	$0,765 \pm 0,102$	$0,630 \pm 0,123$	1,42
Rectus_abdominis	CHEST	$0,791 \pm 0,077$	$0,661 \pm 0,100$	0,46
Rectus_femoris	LEG	$0,807 \pm 0,066$	$0,681 \pm 0,088$	0,29
Temporalis	FACE	$0,796 \pm 0,088$	$0,669 \pm 0,112$	1,02
Tibialis_anterior	LEG	$0,844 \pm 0,063$	$0,735 \pm 0,090$	0,00
Trapezius	ARM	$0,803 \pm 0,077$	$0,678 \pm 0,101$	0,39
Vastus_lateralis	LEG	$0,786 \pm 0,071$	$0,652 \pm 0,089$	0,75
Zygomaticus	FACE	$0,763 \pm 0,089$	$0,624 \pm 0,109$	1,40

Dall'analisi dei risultati emerge che MedSAM raggiunge prestazioni complessivamente buone sulle immagini originali, con un Dice medio globale pari a $0,802 \pm 0,095$ e un valore medio di IoU pari a $0,679 \pm 0,118$. La percentuale complessiva di segmentation failure risulta pari

all'1,80%, indicando che nella maggior parte dei casi il modello è in grado di produrre una segmentazione coerente con la maschera manuale di riferimento.

I risultati migliori sono stati osservati per alcuni muscoli facciali profondi o relativamente ben delimitabili nell'immagine ecografica, come il Geniohyoideus, il Digastricus e il Masseter. In particolare, il Geniohyoideus ha raggiunto il valore medio di Dice più elevato, pari a $0,875 \pm 0,041$, seguito dal Digastricus, con Dice pari a $0,855 \pm 0,061$, e dal Masseter, con Dice pari a $0,849 \pm 0,047$. Questi risultati suggeriscono che, quando il muscolo presenta margini sufficientemente riconoscibili e una regione di interesse ben rappresentata nell'immagine, MedSAM è in grado di adattarsi in maniera efficace anche senza una fase di addestramento specifica sul singolo distretto anatomico.

Prestazioni elevate sono state osservate anche in alcuni muscoli degli arti inferiori e superiori, come il Gastrocnemius medial head, il Tibialis anterior e il Biceps brachii. Questi muscoli presentano generalmente una regione segmentata più ampia rispetto ai piccoli muscoli facciali e, di conseguenza, risultano meno penalizzati da piccoli errori locali di segmentazione. In questi casi, anche una lieve imprecisione del contorno predetto ha un impatto minore sul valore complessivo delle metriche di sovrapposizione.

Il risultato più critico è stato invece osservato per il Mentalis, che ha ottenuto un Dice medio pari a $0,609 \pm 0,148$ e una IoU media pari a $0,454 \pm 0,153$. Inoltre, questo muscolo presenta una percentuale di segmentation failure pari al 25,80%, nettamente superiore rispetto agli altri distretti anatomici analizzati.

È importante sottolineare che, nei muscoli di piccole dimensioni, anche errori apparentemente limitati nella localizzazione della maschera possono determinare una riduzione significativa del Dice e dell'IoU (come nel caso del muscolo Mentalis). Al contrario, nei muscoli più estesi, la stessa entità di errore spaziale può avere un impatto percentualmente inferiore.

La Figura 7.6 riporta un esempio qualitativo di segmentazione ottenuta su immagine originale, mostrando l'immagine ecografica di input, la maschera manuale di riferimento, la predizione generata da MedSAM e la sovrapposizione tra le due maschere.



Figura 7.6: La figura mostra, da sinistra a destra, l'immagine ecografica originale (01175_011_01_4), la maschera manuale di riferimento, la maschera predetta da MedSAM e la sovrapposizione tra ground truth (verde) e predizione (rossa).

7.2.3 Risultati estesi sull'intero dataset – CLAHE

La seconda valutazione quantitativa è stata condotta considerando le immagini preprocessate mediante CLAHE. L'obiettivo di questa analisi era verificare se l'aumento del contrasto locale potesse modificare le prestazioni di MedSAM rispetto alle immagini originali. La Tabella 7.4 riporta i risultati ottenuti sulle immagini preprocessate con CLAHE, suddivisi per singolo muscolo.

Tabella 7.4: Risultati quantitativi ottenuti da MedSAM dopo preprocessing CLAHE.

Muscle type	Region	DSC	IoU	Segmentation failure (%)
Biceps brachii	ARM	$0,842 \pm 0,051$	$0,730 \pm 0,072$	0,17
Deltoides	ARM	$0,832 \pm 0,051$	$0,715 \pm 0,071$	0,00
Depressor anguli_oris, DAO	FACE	$0,779 \pm 0,096$	$0,648 \pm 0,119$	1,47
Digastricus	FACE	$0,861 \pm 0,058$	$0,760 \pm 0,085$	0,00
Gastrocnemius_medial_head	LEG	$0,848 \pm 0,039$	$0,739 \pm 0,057$	0,00
Geniohyoideus	FACE	$0,880 \pm 0,041$	$0,787 \pm 0,063$	0,00
Masseter	FACE	$0,856 \pm 0,045$	$0,751 \pm 0,067$	0,07
Mentalis	FACE	$0,604 \pm 0,142$	$0,447 \pm 0,146$	25,08
Orbicularis_oris, OO	FACE	$0,751 \pm 0,106$	$0,612 \pm 0,128$	2,39
Rectus_abdominis	CHEST	$0,800 \pm 0,069$	$0,672 \pm 0,092$	0,06
Rectus_femoris	LEG	$0,816 \pm 0,060$	$0,694 \pm 0,081$	0,23
Temporalis	FACE	$0,793 \pm 0,086$	$0,664 \pm 0,108$	1,09
Tibialis_anterior	LEG	$0,842 \pm 0,064$	$0,731 \pm 0,090$	0,00
Trapezius	ARM	$0,812 \pm 0,071$	$0,690 \pm 0,095$	0,20
Vastus_lateralis	LEG	$0,802 \pm 0,065$	$0,673 \pm 0,085$	0,35
Zygomaticus	FACE	$0,760 \pm 0,095$	$0,622 \pm 0,114$	1,75

Nel complesso, l'applicazione di CLAHE ha prodotto risultati molto simili a quelli ottenuti sulle immagini originali. Il Dice medio globale è risultato pari a $0,804 \pm 0,095$, mentre l'IoU media è risultata pari a $0,681 \pm 0,119$. La percentuale complessiva di segmentation failure è rimasta pari all'1.80%. Rispetto alle immagini originali, dove il Dice medio era pari a $0,802 \pm 0,095$ e l'IoU media pari a $0,679 \pm 0,118$, si osserva quindi un lieve incremento delle metriche medie, ma senza un cambiamento marcato delle prestazioni complessive.

I muscoli che hanno mostrato le prestazioni migliori dopo l'applicazione di CLAHE sono stati il Geniohyoideus, il Digastricus e il Masseter. In particolare, il Geniohyoideus ha raggiunto un Dice medio pari a $0,880 \pm 0,041$. Anche il Digastricus ha ottenuto un valore elevato, pari a $0,861 \pm 0,058$, così come il Masseter, con Dice pari a $0,856 \pm 0,045$. Questi risultati suggeriscono che l'aumento del contrasto locale possa contribuire a rendere più evidenti alcune strutture muscolari, soprattutto quando i margini anatomici sono già sufficientemente riconoscibili nell'immagine di partenza.

Anche alcuni muscoli appartenenti agli arti e al tronco hanno mantenuto prestazioni elevate. Il Gastrocnemius medial head, il Biceps brachii, il Tibialis anterior e il Deltoides hanno mostrato valori di Dice superiori o prossimi a 0,83. Questo conferma che MedSAM tende a comportarsi in maniera stabile nei muscoli caratterizzati da una regione di interesse più ampia e da una morfologia più regolare rispetto ai piccoli muscoli facciali.

Tuttavia, il miglioramento introdotto da CLAHE non è uniforme su tutti i muscoli. Alcuni distretti facciali di dimensioni ridotte, come il Depressor anguli oris, l'Orbicularis oris e lo Zygomaticus, non mostrano un incremento evidente rispetto alle immagini originali. In questi casi, l'aumento del contrasto locale potrebbe non essere sufficiente a compensare la complessità anatomica del distretto oppure potrebbe enfatizzare anche componenti non informative dell'immagine.

Il caso più critico rimane il Mentalis. Dopo preprocessing CLAHE, questo muscolo ha ottenuto un Dice medio pari a $0,604 \pm 0,142$ e un IoU media pari a $0,447 \pm 0,146$. La percentuale di segmentation failure è risultata pari al 25,08%, valore nettamente superiore rispetto a quello degli altri muscoli analizzati. Sebbene tale valore sia leggermente inferiore rispetto alla failure osservata sulle immagini originali, pari al 25,80%, il Mentalis continua a rappresentare il principale limite della segmentazione automatica con MedSAM.

Questo risultato suggerisce che il preprocessing CLAHE, pur migliorando localmente il contrasto dell'immagine, non sia sufficiente a risolvere le difficoltà legate alla segmentazione dei muscoli facciali più piccoli. La ridotta dimensione anatomica del Mentalis, la vicinanza a tessuti con caratteristiche ecografiche simili e la possibile scarsa definizione dei margini rendono la segmentazione particolarmente sensibile anche a piccoli errori di localizzazione della maschera.

Nel complesso, i risultati ottenuti con CLAHE confermano la buona stabilità di MedSAM. Il preprocessing determina un lieve miglioramento dei valori medi di Dice e IoU, ma non produce una riduzione evidente della failure complessiva. Questo comportamento suggerisce che MedSAM sia già in grado di estrarre informazioni utili dalle immagini originali e che il miglioramento del contrasto locale non modifichi in modo sostanziale la qualità della segmentazione. La Figura 7.7 riporta un esempio qualitativo, sempre del Masseter (che oltre ad essere un muscolo facciale ha buoni valori di DSC e IoU) utile a visualizzare l'effetto del preprocessing CLAHE e il risultato della segmentazione ottenuta dal modello.

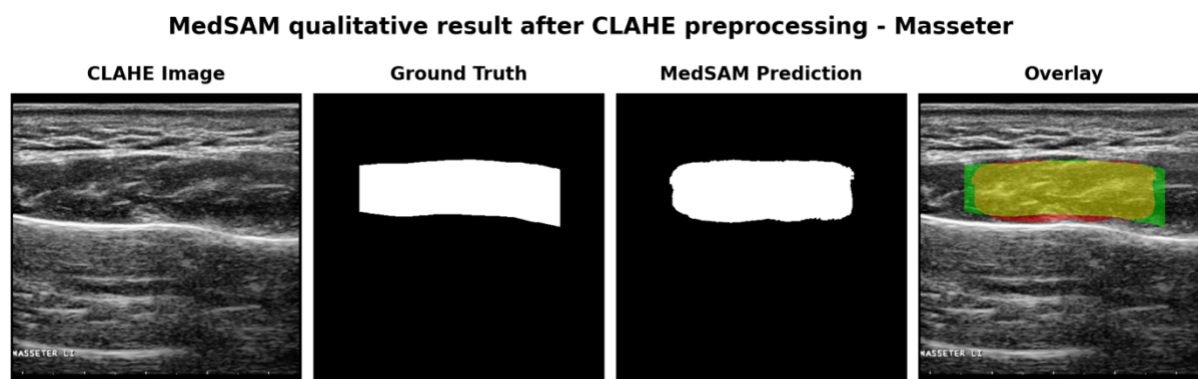


Figura 7.7: La figura mostra, da sinistra a destra, l'immagine ecografica preprocessata mediante CLAHE (01856_011_01_2), la maschera manuale di riferimento, la maschera predetta da MedSAM e l'overlay tra ground truth (verde) e predizione (rossa).

7.2.4 Risultati estesi sull'intero dataset – CLAHE_LOW

La terza analisi quantitativa è stata condotta considerando le immagini preprocessate mediante CLAHE_LOW, cioè una versione meno aggressiva del Contrast Limited Adaptive Histogram

Equalization. L'obiettivo di questa configurazione è quello di migliorare il contrasto locale dell'immagine in maniera più moderata rispetto al CLAHE standard, cercando di preservare maggiormente la naturale distribuzione dei livelli di grigio e di limitare l'enfaticizzazione di rumore o artefatti.

La Tabella 7.5 riporta i risultati ottenuti da MedSAM sulle immagini preprocessate con CLAHE_LOW. Anche in questo caso, i risultati dei cinque fold sono stati aggregati e successivamente raggruppati per singolo muscolo. Per ogni muscolo sono riportati il numero di immagini considerate, il valore medio del Dice coefficient con relativa deviazione standard, il valore medio dell'Intersection over Union con deviazione standard e la percentuale di segmentation failure.

Tabella 7.5: Risultati quantitativi ottenuti da MedSAM dopo preprocessing CLAHE_LOW.

Muscle type	Region	DSC	IoU	Segmentation failure (%)
Biceps brachii	ARM	$0,842 \pm 0,054$	$0,730 \pm 0,074$	0,23
Deltoides	ARM	$0,834 \pm 0,047$	$0,718 \pm 0,066$	0,00
Depressor anguli_oris, DAO	FACE	$0,786 \pm 0,091$	$0,656 \pm 0,115$	0,93
Digastricus	FACE	$0,860 \pm 0,058$	$0,758 \pm 0,085$	0,00
Gastrocnemius_medial_head	LEG	$0,848 \pm 0,039$	$0,738 \pm 0,058$	0,00
Geniohyoideus	FACE	$0,877 \pm 0,041$	$0,784 \pm 0,063$	0,00
Masseter	FACE	$0,855 \pm 0,045$	$0,749 \pm 0,067$	0,00
Mentalis	FACE	$0,608 \pm 0,144$	$0,452 \pm 0,149$	24,76
Orbicularis_oris, OO	FACE	$0,759 \pm 0,103$	$0,621 \pm 0,124$	2,16
Rectus_abdominis	CHEST	$0,803 \pm 0,069$	$0,676 \pm 0,092$	0,06
Rectus_femoris	LEG	$0,817 \pm 0,063$	$0,695 \pm 0,085$	0,35
Temporalis	FACE	$0,792 \pm 0,087$	$0,663 \pm 0,110$	0,95
Tibialis_anterior	LEG	$0,841 \pm 0,065$	$0,731 \pm 0,093$	0,00
Trapezius	ARM	$0,809 \pm 0,075$	$0,685 \pm 0,099$	0,20
Vastus_lateralis	LEG	$0,802 \pm 0,066$	$0,674 \pm 0,085$	0,52
Zygomaticus	FACE	$0,760 \pm 0,095$	$0,621 \pm 0,113$	1,91

Nel complesso, i risultati ottenuti con CLAHE_LOW sono molto simili a quelli osservati sia nelle immagini originali sia nelle immagini preprocessate con CLAHE. Il Dice medio globale è risultato pari a $0,804 \pm 0,095$, mentre l'IoU media è risultata pari a $0,682 \pm 0,119$. La percentuale complessiva di segmentation failure è risultata pari all'1,77%, rappresentando il valore più basso tra le tre condizioni analizzate. Sebbene le differenze siano contenute, questi dati suggeriscono che CLAHE_LOW costituisca la configurazione complessivamente più equilibrata.

Anche in questa condizione, i muscoli che hanno mostrato le prestazioni migliori sono stati il Geniohyoideus, il Digastricus e il Masseter. In particolare, il Geniohyoideus ha ottenuto un Dice medio pari a $0,877 \pm 0,041$, il Digastricus un Dice pari a $0,860 \pm 0,058$ e il Masseter un Dice pari a $0,855 \pm 0,045$. Questi valori confermano che MedSAM mantiene prestazioni elevate nei muscoli facciali che presentano una struttura più chiaramente identificabile e una delimitazione anatomica relativamente ben visibile all'ecografia.

Prestazioni elevate sono state osservate anche in muscoli di altri distretti corporei, come il Gastrocnemius medial head, il Biceps brachii, il Tibialis anterior e il Deltoides. In questi casi,

la maggiore estensione della regione segmentata e la presenza di contorni generalmente più regolari rendono il compito di segmentazione meno sensibile a piccoli errori locali. Questo comportamento era già stato osservato nelle condizioni precedenti e viene confermato anche dopo l'applicazione del preprocessing CLAHE_LOW.

Dal confronto con i risultati ottenuti mediante CLAHE standard emerge che CLAHE_LOW fornisce prestazioni sostanzialmente equivalenti in termini di Dice medio globale, ma leggermente superiori in termini di IoU media e di percentuale di failure. Questo suggerisce che una modifica più moderata del contrasto locale possa risultare più adatta al comportamento di MedSAM, probabilmente perché consente di migliorare la leggibilità di alcune strutture senza alterare in modo eccessivo l'aspetto originario dell'immagine ecografica.

Anche nel caso di CLAHE_LOW, tuttavia, il muscolo più problematico rimane il Mentalis. Per questo distretto anatomico il Dice medio è risultato pari a $0,608 \pm 0,144$, con un'IoU media pari a $0,452 \pm 0,149$ e una percentuale di segmentation failure pari al 24,76%. Sebbene anche in questo caso si osservi un lieve miglioramento rispetto alle immagini originali e a quelle preprocessate con CLAHE standard, il Mentalis continua a rappresentare il principale limite della segmentazione automatica ottenuta con MedSAM.

La criticità di questo muscolo è probabilmente legata, ancora una volta, alla sua ridotta dimensione anatomica, alla complessità del contesto ecografico circostante e alla maggiore sensibilità delle metriche di sovrapposizione a piccoli errori di segmentazione. Pur in presenza di un leggero miglioramento quantitativo, CLAHE_LOW non elimina completamente le difficoltà associate alla segmentazione del Mentalis.

Nel complesso, i risultati ottenuti con CLAHE_LOW mostrano che questa tecnica di preprocessing garantisce una buona stabilità e un lieve vantaggio rispetto alle altre due condizioni analizzate. Il miglioramento non è netto, ma appare coerente soprattutto in termini di IoU media e di riduzione della segmentation failure. La Figura 7.8 riporta un esempio qualitativo utile a visualizzare il risultato della segmentazione ottenuta da MedSAM su immagine preprocessata con CLAHE_LOW.

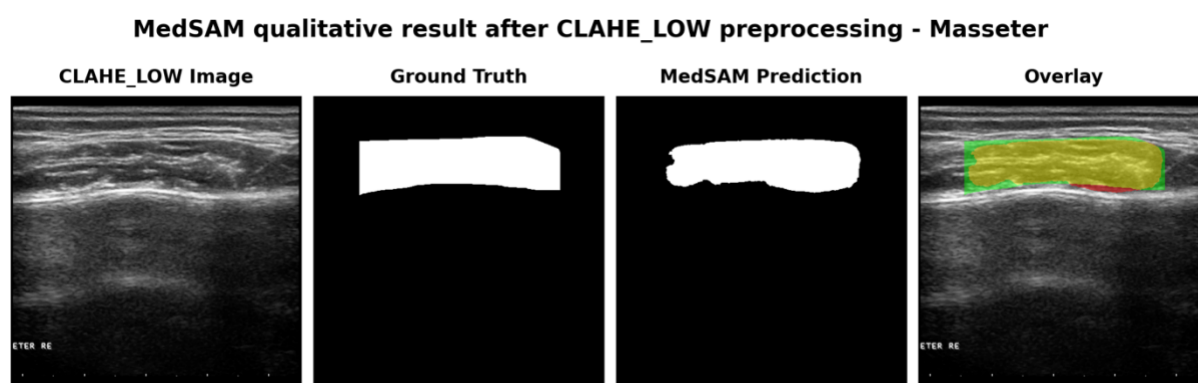


Figura 7.8: La figura mostra, da sinistra a destra, l'immagine ecografica preprocessata mediante CLAHE_LOW (01815_011_00_3), la maschera manuale di riferimento, la maschera predetta da MedSAM e la sovrapposizione tra ground truth (verde) e predizione (rossa).

7.2.5 Confronto collettivo tra le tecniche

Dopo aver analizzato separatamente i risultati ottenuti sulle immagini originali, sulle immagini preprocessate con CLAHE e sulle immagini preprocessate con CLAHE_LOW, è stato eseguito un confronto complessivo tra le tre condizioni. L'obiettivo di questa analisi è valutare se

l'applicazione del preprocessing abbia determinato un miglioramento effettivo delle prestazioni di MedSAM e individuare quale configurazione risulti complessivamente più favorevole.

La Tabella 7.6 riporta il confronto globale tra le tre tecniche, considerando tutti i muscoli analizzati. Per ciascuna configurazione sono riportati il Dice medio con deviazione standard, l'IoU media con deviazione standard e la percentuale complessiva di segmentation failure.

Tabella 7.6: Confronto complessivo delle prestazioni di MedSAM nelle tre condizioni analizzate.

Preprocessing	Support	DSC	IoU	Segmentation failure (%)
ORIGINAL	24756	$0,802 \pm 0,095$	$0,679 \pm 0,118$	1,80
CLAHE	24756	$0,804 \pm 0,095$	$0,681 \pm 0,119$	1,80
CLAHE_LOW	24756	$0,804 \pm 0,095$	$0,682 \pm 0,119$	1,77

Dal confronto globale emerge che le tre tecniche producono risultati molto simili. Le immagini originali hanno ottenuto un Dice medio pari a $0,802 \pm 0,095$ e una IoU media pari a $0,679 \pm 0,118$, con una percentuale di failure pari all'1,80%. L'applicazione di CLAHE ha prodotto un lieve incremento del Dice medio, pari a $0,804 \pm 0,095$, e dell'IoU media, pari a $0,681 \pm 0,119$, mantenendo invariata la percentuale di failure complessiva, pari all'1,80%.

La configurazione CLAHE_LOW ha ottenuto valori molto simili, con Dice medio pari a $0,804 \pm 0,095$ e IoU media pari a $0,682 \pm 0,119$, ma con una percentuale di failure leggermente inferiore, pari all'1,77%.

Sebbene le differenze quantitative siano contenute, CLAHE_LOW può essere considerata la configurazione complessivamente più equilibrata. Questa tecnica ottiene infatti il valore medio di IoU più elevato e la percentuale di segmentation failure più bassa tra le tre condizioni analizzate.

Tuttavia, il vantaggio rispetto alle immagini originali e al CLAHE standard è limitato, suggerendo che MedSAM sia già relativamente robusto alle caratteristiche dell'immagine di input e non dipenda in modo marcato dal preprocessing.

Per rendere il confronto più coerente con l'analisi condotta anche con MMSegmentation, è stato inoltre realizzato un confronto grafico limitato ai quattro muscoli facciali principali considerati in questo studio: Depressor anguli oris, Mentalis, Orbicularis oris e Zygomaticus. La Figura 7.9 riporta il Dice medio con deviazione standard ottenuto dalle tre tecniche per ciascuno di questi muscoli.

MedSAM performance across preprocessing techniques

Dice comparison on facial muscles

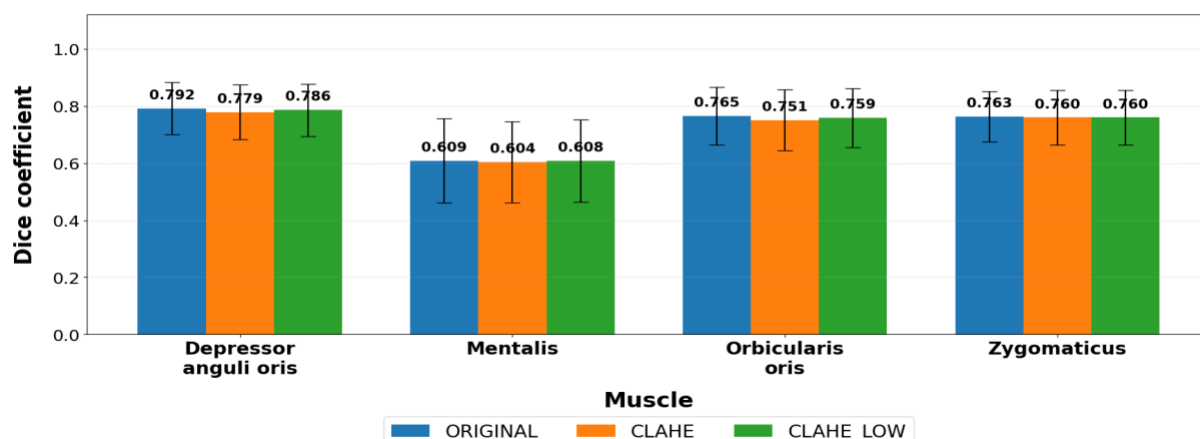


Figura 7.9: Confronto complessivo delle prestazioni di MedSAM nelle tre condizioni analizzate.

Per affiancare al confronto quantitativo anche una valutazione qualitativa coerente con il focus del presente lavoro, è stato riportato un confronto visivo delle maschere ottenute da MedSAM su due casi appartenenti ai muscoli facciali. La Figura 7.10 mostra, per ciascun caso selezionato, l'immagine ecografica, la maschera manuale di riferimento e gli overlay tra ground truth e maschera predetta nelle tre condizioni considerate: ORIGINAL, CLAHE e CLAHE_LOW. Questa rappresentazione permette di osservare sia un caso con buona sovrapposizione tra predizione e ground truth, sia un caso più critico relativo al Mentalis, muscolo che ha mostrato le prestazioni più limitate nell'analisi quantitativa.

Qualitative comparison of MedSAM segmentations on facial muscles

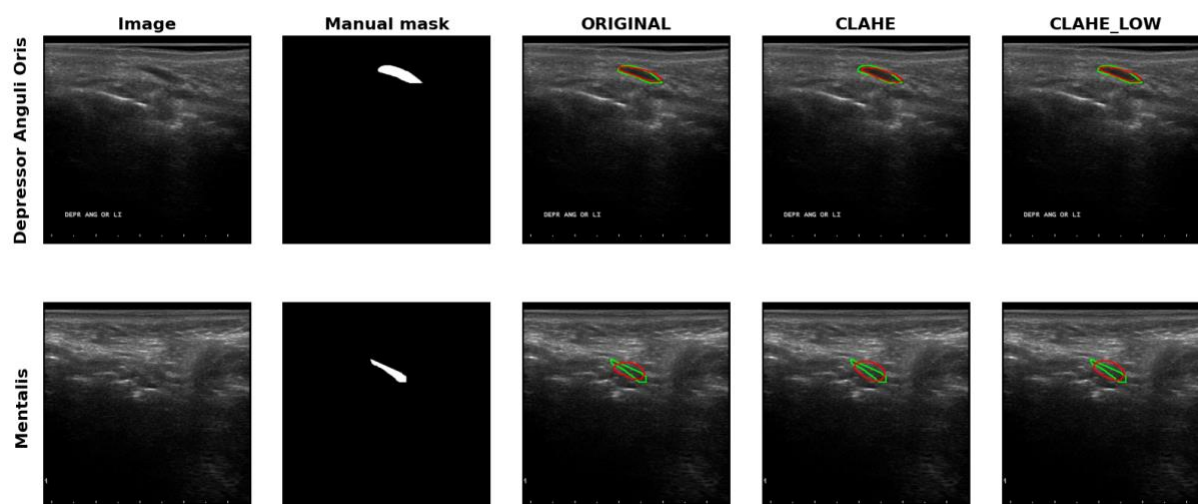


Figura 7.10: Confronto qualitativo delle segmentazioni ottenute da MedSAM su due casi appartenenti ai muscoli facciali (Depressor Anguli Oris, 02495_003_01_2, e Mentalis, 00323_012_01_3). Per ciascun caso sono riportati l'immagine ecografica, la maschera manuale di riferimento e gli overlay tra ground truth e maschera predetta nelle condizioni ORIGINAL, CLAHE e CLAHE_LOW. I contorni verdi indicano la ground truth, mentre i contorni rossi indicano la predizione di MedSAM.

Nel complesso, il confronto quantitativo e qualitativo conferma che le tre condizioni producono risultati molto simili. CLAHE_LOW mostra una lieve differenza numerica favorevole in termini di segmentation failure complessiva, ma il vantaggio rispetto alle immagini originali e

a CLAHE rimane limitato. Questo risultato suggerisce che MedSAM, nella configurazione valutata, sia relativamente stabile rispetto alle modifiche moderate introdotte dal preprocessing.

7.3 Confronto tra MMSegmentation e MedSAM

Uno degli obiettivi del presente lavoro è stato confrontare due approcci differenti alla segmentazione automatica delle immagini ecografiche muscolari: da un lato una pipeline basata su MMSegmentation, con architettura Swin Transformer, UPerNet e K-Net, dall'altro MedSAM, un foundation model sviluppato per la segmentazione di immagini mediche.

Il confronto tra i due approcci è utile perché permette di analizzare due strategie metodologicamente diverse. La pipeline implementata in MMSegmentation rappresenta un approccio supervisionato, basato su Swin Transformer, UPerNet e K-Net, addestrato sul dataset di interesse e applicato direttamente alle immagini ecografiche. MedSAM, invece, è stato utilizzato come modello pre-addestrato sul dominio medico e guidato mediante bounding-box prompt. Questa differenza è fondamentale per interpretare correttamente i risultati: i due sistemi non svolgono esattamente lo stesso compito e, di conseguenza, il confronto non deve essere inteso come una valutazione completamente equivalente tra due pipeline end-to-end.

Nel caso della pipeline implementata in MMSegmentation, l'immagine ecografica viene fornita in ingresso e la maschera di segmentazione viene prodotta direttamente in output. Nel caso di MedSAM, invece, la regione di interesse viene indicata attraverso una bounding box, che fornisce al modello un'informazione preliminare sulla localizzazione del muscolo da segmentare. Per questo motivo, i risultati ottenuti da MedSAM devono essere interpretati come una valutazione della capacità del modello di delineare il muscolo quando la localizzazione della struttura è già fornita, e non come dimostrazione di una segmentazione completamente automatica.

7.3.1 Differenze metodologiche tra i due approcci

La prima differenza tra MMSegmentation e MedSAM riguarda la modalità di apprendimento. La pipeline basata su MMSegmentation è stata sviluppata attraverso un addestramento supervisionato sul dataset disponibile. In questo contesto, le prestazioni del modello dipendono dalla qualità e dalla quantità delle immagini annotate, dalla suddivisione del dataset, dalla procedura di training, dagli iperparametri e dall'efficacia dell'architettura scelta.

Questo approccio consente di ottenere un modello specializzato per il problema considerato, ma richiede una fase di addestramento dedicata e una quantità sufficiente di dati annotati. Nel presente lavoro, MMSegmentation è stato utilizzato soprattutto per valutare l'effetto delle tecniche di preprocessing sui quattro muscoli facciali principali, mantenendo costante la pipeline di segmentazione e modificando l'aspetto delle immagini in ingresso.

MedSAM segue invece una logica diversa. Il modello non è stato addestrato specificamente sul dataset ecografico utilizzato in questa tesi, ma è stato impiegato a partire da un checkpoint pre-addestrato su immagini mediche. In questo caso, l'obiettivo non è stato ottimizzare i pesi del modello sul dataset di muscoli facciali, ma valutare il comportamento di un foundation model medico in un contesto caratterizzato da immagini ecografiche, strutture anatomiche piccole e confini non sempre ben definiti.

Un'altra differenza rilevante riguarda il tipo di informazione fornita al modello. MMSegmentation opera a partire dall'intera immagine e produce direttamente la maschera predetta. MedSAM, invece, richiede un prompt, nel presente lavoro rappresentato da una bounding box associata alla regione anatomica di interesse. Questo rende il compito di MedSAM meno ambiguo, poiché il modello riceve già un'indicazione spaziale della struttura da segmentare.

Di conseguenza, il confronto tra i due approcci deve essere interpretato con cautela. MMSegmentation valuta una pipeline di segmentazione automatica più completa, mentre MedSAM valuta la capacità di un modello pre-addestrato di produrre una maschera accurata quando la localizzazione del muscolo è nota tramite prompt.

7.3.2 Confronto delle prestazioni

I risultati ottenuti con MMSegmentation sui quattro muscoli facciali hanno evidenziato prestazioni complessivamente limitate. Considerando le immagini originali, il Dice medio globale è risultato pari a $0,504 \pm 0,289$, con una IoU media pari a $0,383 \pm 0,244$ e una percentuale di segmentation failure pari al 38,28%. Le configurazioni preprocessate hanno prodotto valori molto simili, senza un miglioramento globale netto rispetto alle immagini originali.

In particolare, CLAHE_LOW ha mostrato valori vicini a quelli ottenuti sulle immagini originali, con un Dice medio pari a $0,500 \pm 0,291$ e una IoU media pari a $0,381 \pm 0,246$. Tuttavia, le differenze tra le configurazioni sono risultate contenute e non uniformi tra i diversi muscoli. Questo risultato suggerisce che, nel caso di MMSegmentation, il preprocessing può modificare localmente il comportamento del modello, ma non è sufficiente a risolvere in modo stabile le difficoltà associate alla segmentazione dei muscoli facciali.

MedSAM ha mostrato prestazioni quantitative più elevate rispetto a MMSegmentation anche considerando lo stesso sottoinsieme dei quattro muscoli facciali principali. Sulle immagini originali, MedSAM ha ottenuto un DSC medio pari a $0,741 \pm 0,124$ e una IoU media pari a $0,602 \pm 0,143$, con una segmentation failure pari al 5,91%. Nelle condizioni CLAHE e CLAHE_LOW, i valori di DSC medio sono risultati rispettivamente pari a $0,733 \pm 0,125$ e $0,737 \pm 0,124$, confermando una maggiore stabilità rispetto alla pipeline basata su MMSegmentation.

Il confronto con MMSegmentation evidenzia quindi una netta riduzione dei casi critici, passando da una segmentation failure pari al 38,28% sulle immagini originali con MMSegmentation al 5,91% con MedSAM. Tuttavia, questo risultato deve essere interpretato alla luce della diversa impostazione metodologica dei due approcci. MMSegmentation produce direttamente la maschera a partire dall'immagine ecografica, mentre MedSAM riceve una bounding box derivata dalla maschera manuale di riferimento. Pertanto, il confronto dimostra il potenziale di MedSAM nella fase di delineazione della maschera quando la regione di interesse è già localizzata, ma non una superiorità diretta come pipeline completamente automatica end-to-end.

Considerando invece l'intero dataset dei 16 muscoli, MedSAM ha ottenuto valori medi globali di DSC superiori a 0,80 in tutte le condizioni analizzate. Questo risultato conferma la buona stabilità del modello nel dominio considerato, ma deve essere interpretato come analisi

complementare, poiché include anche muscoli più estesi e generalmente più semplici da segmentare rispetto ai piccoli muscoli facciali.

7.3.3 Sensibilità al preprocessing

Un elemento interessante emerso dal confronto riguarda la diversa sensibilità dei due approcci alle strategie di preprocessing.

Nel caso di MMSegmentation, l'applicazione delle tecniche di preprocessing ha prodotto variazioni delle prestazioni, ma senza un miglioramento globale uniforme. Le immagini originali hanno mantenuto il valore medio di Dice leggermente più elevato nel confronto globale sui quattro muscoli facciali, mentre CLAHE_LOW ha mostrato un comportamento leggermente più favorevole in alcuni muscoli, come Depressor Anguli Oris e Mentalis. Al contrario, per Orbicularis Oris e Zygomaticus le immagini originali hanno ottenuto valori medi superiori o molto simili rispetto alle immagini preprocessate.

Questo andamento suggerisce che l'effetto del preprocessing dipenda dal singolo muscolo e dalle sue caratteristiche anatomiche ed ecografiche. Tecniche come CLAHE e CLAHE_LOW possono rendere più evidenti alcune strutture locali, ma possono anche modificare la texture o enfatizzare componenti non informative dell'immagine. Nel caso dei piccoli muscoli facciali, tali variazioni possono avere effetti non sempre prevedibili sulle maschere prodotte dal modello.

MedSAM ha invece mostrato una minore variazione delle prestazioni tra immagini originali, CLAHE e CLAHE_LOW. Le differenze quantitative tra le tre configurazioni sono risultate limitate, indicando una maggiore stabilità rispetto alle modifiche moderate dell'aspetto dell'immagine. Questo comportamento può essere interpretato come una forma di robustezza intra-dominio rispetto al preprocessing, cioè come una capacità del modello di mantenere prestazioni simili all'interno dello stesso dataset e dello stesso contesto di acquisizione.

È importante sottolineare che questa stabilità non equivale a una validazione esterna. I dati analizzati provengono dallo stesso dominio sperimentale e non permettono di concludere che il modello sia automaticamente trasferibile ad altri centri, dispositivi ecografici o protocolli di acquisizione. Per questo motivo, è più corretto parlare di stabilità rispetto alle condizioni di preprocessing valutate, piuttosto che di vera generalizzazione clinica.

7.3.4 Limiti del confronto diretto

Il confronto tra MMSegmentation e MedSAM presenta alcuni limiti metodologici che devono essere considerati nell'interpretazione dei risultati.

Il primo limite riguarda la diversa natura delle due pipeline. MMSegmentation è stato utilizzato come sistema di segmentazione automatica a partire dall'intera immagine, mentre MedSAM è stato valutato mediante bounding-box prompt. Questo significa che MedSAM riceve un'informazione preliminare sulla posizione del muscolo, riducendo l'ambiguità del compito rispetto a una segmentazione completamente automatica.

Un secondo limite riguarda l'insieme di immagini valutate. Gli esperimenti con MMSegmentation sono stati focalizzati sui quattro muscoli facciali principali, mentre

MedSAM è stato valutato anche sull'intero dataset dei muscoli disponibili. Di conseguenza, i valori medi globali di MedSAM possono essere influenzati dalla presenza di muscoli più grandi e più facilmente segmentabili, che tendono a ottenere prestazioni superiori rispetto ai piccoli muscoli facciali. Per una valutazione pienamente omogenea, il confronto dovrebbe essere condotto sugli stessi muscoli, sulle stesse immagini e con la stessa modalità di aggregazione delle metriche.

Un ulteriore aspetto riguarda la definizione della segmentation failure. Nel presente lavoro è stata adottata una definizione unica basata sul Dice Similarity Coefficient inferiore a 0,5, in modo da rendere più coerente il confronto tra le diverse fasi sperimentali. Tuttavia, la letteratura e lo studio di riferimento possono utilizzare soglie differenti, ad esempio basate sulla IoU. Questo aspetto deve essere considerato quando si confrontano direttamente i risultati con lavori precedenti.

Infine, il confronto non include una valutazione automatica della fase di generazione del prompt per MedSAM. Dal punto di vista pratico, un sistema completamente automatico dovrebbe essere in grado non solo di segmentare il muscolo a partire da una bounding box, ma anche di generare tale bounding box senza intervento manuale o senza utilizzare direttamente informazioni derivate dalla ground truth. Questo rappresenta un passaggio fondamentale per trasformare MedSAM da strumento di segmentazione guidata a componente di una pipeline end-to-end.

7.3.5 Considerazioni finali

Il confronto tra MMSegmentation e MedSAM mostra che i due approcci offrono informazioni complementari sul problema della segmentazione ecografica muscolare.

MMSegmentation rappresenta una pipeline supervisionata e automatica, utile per valutare l'effetto del preprocessing su modelli addestrati direttamente sul dataset. I risultati ottenuti sui quattro muscoli facciali indicano però che il solo preprocessing non è sufficiente a superare in modo stabile le difficoltà associate a strutture anatomiche piccole, poco contrastate e con margini poco definiti.

MedSAM, invece, ha mostrato prestazioni complessivamente elevate e una ridotta sensibilità alle tre condizioni di preprocessing considerate. Questo suggerisce che i foundation models medici possano rappresentare strumenti utili per la segmentazione di immagini ecografiche muscolari, soprattutto quando la regione di interesse viene fornita mediante prompt. Tuttavia, il risultato deve essere interpretato come una valutazione di segmentazione guidata da bounding box e non come dimostrazione di una pipeline completamente automatica.

Nel complesso, i dati raccolti suggeriscono che MedSAM possa costituire una direzione interessante per futuri sviluppi, ma anche che siano necessari ulteriori passaggi metodologici prima di una possibile applicazione clinica. In particolare, sarà necessario valutare la generazione automatica dei prompt, confrontare il modello con baseline addestrate in modo controllato sugli stessi dati, e validare le prestazioni su dataset esterni acquisiti con dispositivi e protocolli differenti.

8. Discussione

I risultati ottenuti in questo lavoro permettono di analizzare il problema della segmentazione automatica dei muscoli facciali in ecografia da due prospettive complementari: da un lato il ruolo del preprocessing, dall'altro l'impiego di un foundation model medico guidato da prompt. I capitoli precedenti hanno mostrato in dettaglio le prestazioni quantitative e qualitative dei diversi approcci. In questa sezione, tali risultati vengono discussi in termini di significato metodologico, limiti interpretativi e possibili implicazioni per sviluppi futuri.

8.1 Interpretazione complessiva dei risultati

Un primo elemento emerso dall'analisi è la conferma della complessità della segmentazione dei muscoli facciali. Questo risultato era atteso, considerando le caratteristiche anatomiche ed ecografiche di tali strutture, ma viene ulteriormente confermato dalle prestazioni ottenute nella fase sperimentale. I muscoli facciali occupano una porzione ridotta dell'immagine, presentano spesso margini poco definiti e possono essere circondati da tessuti con caratteristiche ecografiche simili. Di conseguenza, anche piccoli errori nella localizzazione o nella definizione del contorno possono determinare una riduzione marcata delle metriche di sovrapposizione.

Questo aspetto è particolarmente importante perché mostra che la difficoltà del problema non dipende esclusivamente dalla scelta del modello o dalla qualità visiva dell'immagine. La segmentazione dei muscoli facciali è influenzata da una combinazione di fattori: dimensione ridotta della struttura, basso contrasto, variabilità anatomica, rumore ecografico e possibili alterazioni patologiche associate alla FSHD. Per questo motivo, un singolo intervento sulla pipeline, come il solo miglioramento del contrasto, difficilmente può risolvere in modo completo il problema.

8.2 Ruolo e limiti del preprocessing

Le tecniche di preprocessing considerate hanno modificato in modo evidente l'aspetto delle immagini ecografiche, soprattutto in termini di contrasto locale, uniformità della texture e visibilità dei margini. Tuttavia, i risultati sperimentali mostrano che il miglioramento visivo dell'immagine non corrisponde necessariamente a un miglioramento stabile della segmentazione automatica.

Questo punto è uno degli aspetti più rilevanti dello studio. In ambito ecografico è intuitivo pensare che un'immagine più contrastata o più pulita possa facilitare il lavoro del modello. Tuttavia, i risultati ottenuti suggeriscono che il rapporto tra preprocessing e segmentazione sia più complesso. Una tecnica può rendere più evidenti alcune strutture anatomiche, ma allo stesso tempo può enfatizzare il rumore, alterare la texture o modificare caratteristiche che il modello aveva imparato a utilizzare durante l'addestramento.

CLAHE_LOW rappresenta un caso interessante perché introduce una modifica più moderata rispetto alla CLAHE standard. Questo approccio appare più coerente con la necessità di migliorare la leggibilità dell'immagine senza alterarne eccessivamente l'aspetto originario. Tuttavia, anche in questo caso il beneficio osservato rimane limitato e dipendente dal muscolo considerato. Il preprocessing deve quindi essere interpretato come un possibile supporto alla segmentazione, non come una soluzione autonoma al problema.

Nel complesso, questa parte del lavoro evidenzia che l'ottimizzazione dell'immagine può contribuire alla pipeline, ma non è sufficiente a superare le criticità principali della segmentazione dei piccoli muscoli facciali. Tali criticità richiedono anche modelli più robusti, strategie di localizzazione più efficaci e una valutazione più approfondita dell'impatto anatomico degli errori di segmentazione.

8.3 Interpretazione dei risultati ottenuti con MedSAM

L'introduzione di MedSAM ha permesso di valutare un approccio diverso rispetto alla segmentazione supervisionata tradizionale. I risultati ottenuti indicano che il modello è in grado di produrre maschere complessivamente stabili nelle condizioni sperimentali analizzate e che le variazioni introdotte dal preprocessing hanno un impatto limitato sulle prestazioni globali.

Questo comportamento suggerisce che MedSAM riesca a sfruttare rappresentazioni apprese su un ampio dominio medico, mantenendo una buona capacità di delineare la struttura anatomica quando questa viene indicata attraverso un prompt. Tuttavia, questo punto deve essere interpretato con cautela. Nel presente lavoro, infatti, MedSAM è stato valutato come modello di segmentazione guidata da bounding box. La localizzazione della regione di interesse è quindi fornita al modello e non viene risolta autonomamente.

Questa distinzione è fondamentale. Le prestazioni ottenute da MedSAM mostrano il potenziale del modello nella fase di delineazione della maschera, ma non dimostrano ancora la realizzazione di una pipeline completamente automatica. Per arrivare a questo obiettivo sarebbe necessario integrare una fase preliminare di localizzazione automatica del muscolo, capace di generare il prompt senza intervento manuale o senza informazioni derivate dalla ground truth.

Pertanto, il contributo principale di MedSAM in questo lavoro non va interpretato come una sostituzione diretta e definitiva dei modelli supervisionati, ma come l'indicazione che i foundation models medici possono rappresentare una componente promettente all'interno di pipeline più complesse. In particolare, potrebbero risultare utili se combinati con sistemi automatici di detection o con strategie di auto-prompting specifiche per l'ecografia muscolare.

8.4 Limiti dello studio

Il primo limite dello studio riguarda la non perfetta equivalenza tra le pipeline confrontate. MMSegmentation produce direttamente una maschera a partire dall'immagine, mentre MedSAM utilizza una bounding box come informazione di guida. Questo rende il confronto utile dal punto di vista metodologico, ma non completamente sovrapponibile dal punto di vista operativo.

Un secondo limite riguarda l'assenza di una validazione esterna. Gli esperimenti sono stati condotti su dati appartenenti allo stesso contesto di acquisizione e non permettono quindi di trarre conclusioni definitive sulla trasferibilità dei risultati ad altri centri, scanner o protocolli ecografici. Per questo motivo, i risultati devono essere interpretati come indicativi della stabilità dei modelli nel dominio analizzato, non come prova di generalizzazione clinica.

Un ulteriore limite riguarda il fatto che l'analisi si è concentrata principalmente sulle metriche di sovrapposizione tra maschere. Dice e IoU sono strumenti fondamentali per valutare la qualità

geometrica della segmentazione, ma non descrivono direttamente l'impatto degli errori sulle analisi successive. In futuro sarà quindi importante valutare come le differenze tra maschere manuali e automatiche influenzino parametri quantitativi quali area, ecointensità, texture o classificazione del grado di compromissione muscolare.

Infine, il lavoro non affronta ancora il problema dell'integrazione clinica della pipeline. Per un utilizzo reale, un sistema automatico dovrebbe essere robusto rispetto alla variabilità di acquisizione, facilmente applicabile a nuovi dati e in grado di fornire risultati interpretabili e riproducibili. Questi aspetti richiedono ulteriori validazioni e rappresentano un passaggio necessario prima di qualsiasi possibile applicazione clinica.

8.5 Sintesi della discussione

Nel complesso, il lavoro mostra che la segmentazione automatica dei muscoli facciali in ecografia rimane un problema complesso e non completamente risolto. Il preprocessing può migliorare alcuni aspetti dell'immagine e influenzare localmente le prestazioni, ma non garantisce un beneficio uniforme. MedSAM mostra risultati promettenti nella segmentazione guidata da prompt, ma richiede ulteriori sviluppi per essere integrato in una pipeline completamente automatica.

Questi risultati suggeriscono che la soluzione più realistica non sia affidata a un singolo elemento della pipeline, ma alla combinazione di più componenti: immagini di qualità adeguata, modelli di segmentazione robusti, sistemi automatici di localizzazione e validazione su dati esterni. In questa prospettiva, il presente lavoro contribuisce a chiarire sia le potenzialità sia i limiti delle strategie analizzate, fornendo una base metodologica per futuri sviluppi nell'analisi automatica dell'ecografia muscolare nella FSHD.

9. Conclusioni e sviluppi futuri

9.1 Conclusioni

Il presente lavoro di tesi ha avuto l'obiettivo di analizzare possibili strategie per migliorare la segmentazione automatica dei muscoli facciali in immagini ecografiche muscolari, con particolare riferimento al contesto della distrofia muscolare facio-scapolo-omeroale. La scelta di concentrarsi sui muscoli facciali è stata motivata dalla loro rilevanza clinica nella FSHD e dalle difficoltà già evidenziate nei precedenti studi di segmentazione automatica, in cui tali strutture risultavano tra le più complesse da identificare correttamente.

Il primo aspetto analizzato ha riguardato l'impatto delle tecniche di preprocessing sulle prestazioni di segmentazione ottenute mediante MMSegmentation. I risultati hanno mostrato che il preprocessing modifica in modo evidente l'aspetto delle immagini ecografiche, ma non determina necessariamente un miglioramento globale e uniforme delle prestazioni. In particolare, le immagini originali hanno mantenuto il valore medio globale di Dice leggermente più elevato, mentre CLAHE_LOW ha prodotto risultati molto vicini e lievemente favorevoli in alcuni muscoli specifici. Questo indica che il miglioramento visivo dell'immagine non è automaticamente associato a un miglioramento della segmentazione automatica.

Nel complesso, il preprocessing deve quindi essere interpretato come uno strumento di supporto, utile per modificare alcune caratteristiche dell'immagine e potenzialmente migliorare la leggibilità di specifici distretti anatomici, ma non sufficiente da solo a risolvere le difficoltà associate alla segmentazione dei piccoli muscoli facciali. Le prestazioni limitate ottenute in alcuni casi confermano che il problema dipende da una combinazione di fattori anatomici, ecografici e modellistici, tra cui le ridotte dimensioni delle strutture, il basso contrasto, la variabilità morfologica e la scarsa definizione dei margini.

La seconda parte del lavoro ha valutato l'utilizzo di MedSAM, un foundation model sviluppato per la segmentazione di immagini mediche. I risultati ottenuti hanno mostrato prestazioni complessivamente elevate e stabili nelle tre condizioni analizzate: immagini originali, CLAHE e CLAHE_LOW. Le differenze limitate tra le configurazioni indicano che MedSAM risulta meno sensibile alle variazioni introdotte dal preprocessing rispetto alla pipeline basata su MMSegmentation.

Tuttavia, l'interpretazione di questi risultati richiede cautela. MedSAM è stato utilizzato mediante bounding-box prompt e quindi non rappresenta, nella configurazione valutata, una pipeline completamente automatica. Il modello non localizza autonomamente il muscolo all'interno dell'immagine, ma produce la maschera a partire da una regione di interesse già indicata. Di conseguenza, i risultati ottenuti dimostrano soprattutto la capacità di MedSAM di delineare la struttura anatomica quando questa è già localizzata, più che la capacità di sostituire integralmente un sistema automatico end-to-end.

Il confronto tra MMSegmentation e MedSAM ha quindi evidenziato due approcci differenti e complementari. MMSegmentation rappresenta una pipeline automatica supervisionata, ma risulta fortemente penalizzata dalla complessità dei muscoli facciali. MedSAM, invece, mostra un comportamento più stabile e prestazioni più elevate, ma richiede un'informazione preliminare sotto forma di prompt. Per questo motivo, il contributo principale del lavoro non

consiste nell'individuare un metodo definitivamente superiore, ma nel chiarire i vantaggi e i limiti di due strategie diverse applicate a un problema complesso di segmentazione ecografica.

In conclusione, i risultati ottenuti suggeriscono che la segmentazione automatica dei muscoli facciali in ecografia rimane una sfida aperta. Il solo preprocessing non è sufficiente a garantire un miglioramento stabile, mentre i foundation models medici rappresentano una direzione promettente, soprattutto se integrati in pipeline capaci di generare automaticamente prompt affidabili. Ulteriori studi saranno necessari per validare questi approcci su dati esterni, ridurre la dipendenza da annotazioni manuali e valutare l'impatto delle segmentazioni automatiche su analisi quantitative clinicamente rilevanti.

9.2 Sviluppi futuri

A partire dai risultati ottenuti, è possibile individuare diverse direzioni future per proseguire e ampliare il lavoro svolto.

Una prima direzione riguarda l'evoluzione degli AI foundation models applicati all'imaging medico. Negli ultimi anni, modelli derivati da SAM e adattati al dominio biomedicale hanno mostrato un potenziale crescente in diversi compiti di segmentazione. In questo contesto, sarà importante confrontare MedSAM con altre architetture sviluppate specificamente per immagini mediche ed ecografiche, come SAMUS, LiteMedSAM o ulteriori modelli SAM-based. Tale confronto dovrebbe essere condotto in modo controllato, utilizzando gli stessi dataset, le stesse metriche e le stesse condizioni sperimentali, così da valutare in modo più affidabile quale approccio sia più adatto alla segmentazione ecografica muscolare.

Un secondo sviluppo riguarda il fine-tuning o l'adattamento leggero dei foundation models al dominio specifico dell'ecografia muscolare. Nel presente lavoro MedSAM è stato utilizzato come modello pre-addestrato, senza una fase di addestramento dedicata sul dataset considerato. Tecniche di adattamento domain-specific, eventualmente basate su strategie parameter-efficient, potrebbero permettere di combinare la conoscenza generale appresa dal modello su ampie raccolte di immagini mediche con una maggiore specializzazione verso le caratteristiche dell'ecografia muscolare e dei muscoli facciali.

Un aspetto particolarmente importante riguarda la generazione automatica del prompt. Nella configurazione attuale, MedSAM richiede una bounding box che identifichi la regione da segmentare. Per trasformare questo approccio in una pipeline realmente automatica, sarà necessario integrare un modulo preliminare di localizzazione del muscolo, capace di generare in modo autonomo il prompt richiesto dal modello. Questa fase potrebbe essere realizzata mediante modelli di detection, reti leggere di localizzazione oppure strategie di auto-prompting sviluppate specificamente per immagini ecografiche.

Un'ulteriore direzione riguarda la validazione esterna. Gli esperimenti condotti nel presente lavoro sono stati svolti all'interno dello stesso dominio sperimentale. Per valutare realmente la robustezza dei modelli sarà necessario testarli su immagini acquisite presso centri differenti, con dispositivi ecografici diversi, operatori diversi e protocolli di acquisizione non identici. Questo passaggio è fondamentale per distinguere una stabilità osservata all'interno del dataset da una reale capacità di trasferimento a scenari clinici differenti.

Dal punto di vista metodologico, sarà inoltre utile confrontare i foundation models con baseline più tradizionali ma solide, come U-Net, nnU-Net o altre architetture supervisionate

specificamente ottimizzate per la segmentazione medica. Un confronto di questo tipo permetterebbe di comprendere meglio se il vantaggio dei foundation models derivi dalla loro maggiore capacità rappresentativa, dalla modalità prompt-based o da altri aspetti della pipeline sperimentale.

Un altro sviluppo rilevante riguarda l'utilizzo delle maschere prodotte per analisi quantitative successive. La segmentazione rappresenta infatti solo il primo passaggio di una pipeline più ampia. Una volta ottenuta una maschera affidabile, diventa possibile estrarre parametri come area, spessore, ecointensità media, caratteristiche di texture e descrittori radiomici. Questi parametri potrebbero essere utilizzati per studiare la progressione della FSHD, confrontare soggetti sani e patologici o supportare sistemi automatici di classificazione del grado di compromissione muscolare.

In prospettiva più ampia, se questi sistemi raggiungessero una robustezza adeguata, potrebbero essere integrati direttamente nei software ecografici per fornire un supporto alla valutazione in tempo reale. In ambito neuromuscolare, questo potrebbe facilitare il monitoraggio longitudinale dei pazienti e ridurre il tempo necessario per l'analisi delle immagini. In altri contesti muscolo-scheletrici, come la medicina dello sport, tecnologie simili potrebbero supportare la valutazione ecografica di lesioni muscolari, ad esempio negli infortuni degli atleti o dei calciatori. In questi scenari, un sistema automatico potrebbe aiutare a delimitare rapidamente la regione di interesse, misurare l'estensione della lesione e monitorarne l'evoluzione nel tempo.

È importante sottolineare che tali applicazioni non dovrebbero essere intese come sostituzione del medico o dell'operatore esperto, ma come strumenti di supporto alla valutazione clinica. L'integrazione dell'intelligenza artificiale nell'ecografia richiederà infatti validazioni approfondite, attenzione alla sicurezza, interpretabilità dei risultati e verifica della riproducibilità in condizioni reali di utilizzo.

In conclusione, il futuro della segmentazione ecografica muscolare sembra orientato verso pipeline integrate, capaci di combinare modelli di localizzazione, foundation models medici, validazione multicentrica e analisi quantitativa automatica. Il presente lavoro si inserisce in questa direzione, mostrando sia le potenzialità sia i limiti degli approcci analizzati e fornendo una base sperimentale per sviluppi successivi nel campo dell'ecografia muscolare assistita dall'intelligenza artificiale.

Bibliografia

- [1] P. Casey *et al.*, «The current use of ultrasound to measure skeletal muscle and its ability to predict clinical outcomes: a systematic review», *J. Cachexia Sarcopenia Muscle*, vol. 13, fasc. 5, pp. 2298–2309, ott. 2022, doi: 10.1002/jcsm.13041.
- [2] K. Mul, S. Lassche, N. C. Voermans, G. W. Padberg, C. G. Horlings, e B. G. Van Engelen, «What's in a name? The clinical features of facioscapulohumeral muscular dystrophy», *Pract. Neurol.*, vol. 16, fasc. 3, pp. 201–207, giu. 2016, doi: 10.1136/practneurol-2015-001353.
- [3] M. K. Preston e L. H. Wang, «Facioscapulohumeral Muscular Dystrophy».
- [4] N. Rahmani, M. A. Mohseni-Bandpei, R. Vameghi, M. Salavati, e I. Abdollahi, «Application of Ultrasonography in the Assessment of Skeletal Muscles in Children with and without Neuromuscular Disorders: A Systematic Review», *Ultrasound Med. Biol.*, vol. 41, fasc. 9, pp. 2275–2283, set. 2015, doi: 10.1016/j.ultrasmedbio.2015.04.027.
- [5] O. Ronneberger, P. Fischer, e T. Brox, «U-Net: Convolutional Networks for Biomedical Image Segmentation», in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, e A. F. Frangi, A c. di, in *Lecture Notes in Computer Science*, vol. 9351. , Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
- [6] H. H. Lee *et al.*, «Foundation Models for Biomedical Image Segmentation: A Survey».
- [7] F. Marzola, N. Van Alfen, J. Doorduyn, e K. M. Meiburger, «Machine learning-driven Heckmatt grading in facioscapulohumeral muscular dystrophy: A novel pathway for musculoskeletal ultrasound analysis», *Clin. Neurophysiol.*, vol. 172, pp. 61–69, apr. 2025, doi: 10.1016/j.clinph.2025.01.016.
- [8] S. C. C. Vincenten *et al.*, «The other face of facioscapulohumeral muscular dystrophy: Exploring orofacial weakness using muscle ultrasound», *Muscle Nerve*, vol. 70, fasc. 5, pp. 1062–1071, nov. 2024, doi: 10.1002/mus.28254.
- [9] M. S. Tihaya *et al.*, «Facioscapulohumeral muscular dystrophy: the road to targeted therapies», *Nat. Rev. Neurol.*, vol. 19, fasc. 2, pp. 91–108, feb. 2023, doi: 10.1038/s41582-022-00762-2.
- [10] T. Schätzl, L. Kaiser, e H.-P. Deigner, «Facioscapulohumeral muscular dystrophy: genetics, gene activation and downstream signalling with regard to recent therapeutic approaches: an update», *Orphanet J. Rare Dis.*, vol. 16, fasc. 1, p. 129, mar. 2021, doi: 10.1186/s13023-021-01760-1.
- [11] K. Mul, C. G. C. Horlings, S. C. C. Vincenten, N. C. Voermans, B. G. M. Van Engelen, e N. Van Alfen, «Quantitative muscle MRI and ultrasound for facioscapulohumeral muscular dystrophy: complementary imaging biomarkers», *J. Neurol.*, vol. 265, fasc. 11, pp. 2646–2655, nov. 2018, doi: 10.1007/s00415-018-9037-y.
- [12] L. Fionda, «Comparison of quantitative muscle ultrasound and whole-body muscle MRI in facioscapulohumeral muscular dystrophy type 1 patients».
- [13] J. Wijntjes e N. Van Alfen, «Muscle ultrasound: Present state and future opportunities», *Muscle Nerve*, vol. 63, fasc. 4, pp. 455–466, apr. 2021, doi: 10.1002/mus.27081.
- [14] J. Wijntjes, C. Saris, J. Doorduyn, N. Van Alfen, B. Van Engelen, e K. Mul, «Improving Heckmatt muscle ultrasound grading scale through Rasch analysis», *Neuromuscul. Disord.*, vol. 42, pp. 14–21, set. 2024, doi: 10.1016/j.nmd.2024.07.001.
- [15] P. Ardianto *et al.*, «A Review of the Challenges in Deep Learning for Skeletal and Smooth Muscle Ultrasound Images», *Appl. Sci.*, vol. 11, fasc. 9, p. 4021, apr. 2021, doi: 10.3390/app11094021.
- [16] J. Ma, Y. He, F. Li, L. Han, C. You, e B. Wang, «Segment anything in medical images», *Nat. Commun.*, vol. 15, fasc. 1, p. 654, gen. 2024, doi: 10.1038/s41467-024-44824-z.

- [17] F. Marzola, N. Van Alfen, J. Doorduyn, e K. M. Meiburger, «Machine learning-driven Heckmatt grading in facioscapulohumeral muscular dystrophy: A novel pathway for musculoskeletal ultrasound analysis», *Clin. Neurophysiol.*, vol. 172, pp. 61–69, apr. 2025, doi: 10.1016/j.clinph.2025.01.016.
- [18] D. Müller, I. Soto-Rey, e F. Kramer, «Towards a guideline for evaluation metrics in medical image segmentation», *BMC Res. Notes*, vol. 15, fasc. 1, p. 210, dic. 2022, doi: 10.1186/s13104-022-06096-y.
- [19] F. Marzola, N. Van Alfen, J. Doorduyn, e K. M. Meiburger, «Deep learning segmentation of transverse musculoskeletal ultrasound images for neuromuscular disease assessment», *Comput. Biol. Med.*, vol. 135, p. 104623, ago. 2021, doi: 10.1016/j.compbiomed.2021.104623.
- [20] C. Caresio, M. Salvi, F. Molinari, K. M. Meiburger, e M. A. Minetto, «Fully Automated Muscle Ultrasound Analysis (MUSA): Robust and Accurate Muscle Thickness Measurement», *Ultrasound Med. Biol.*, vol. 43, fasc. 1, pp. 195–205, gen. 2017, doi: 10.1016/j.ultrasmedbio.2016.08.032.
- [21] L. J. Brattain, B. A. Telfer, M. Dhyani, J. R. Grajo, e A. E. Samir, «Machine learning for medical ultrasound: status, methods, and future opportunities», *Abdom. Radiol.*, vol. 43, fasc. 4, pp. 786–799, apr. 2018, doi: 10.1007/s00261-018-1517-0.
- [22] M. Moinuddin, S. Khan, A. U. Alsaggaf, M. J. Abdulaal, U. M. Al-Saggaf, e J. C. Ye, «Medical ultrasound image speckle reduction and resolution enhancement using texture compensated multi-resolution convolution neural network», *Front. Physiol.*, vol. 13, p. 961571, nov. 2022, doi: 10.3389/fphys.2022.961571.
- [23] M. Triantafyllou *et al.*, «The Effect of Ultrasound Image Pre-Processing on Radiomics Feature Quality: A Study on Shoulder Ultrasound», *J. Imaging Inform. Med.*, vol. 38, fasc. 6, pp. 4184–4195, feb. 2025, doi: 10.1007/s10278-025-01421-w.
- [24] K. He, J. Sun, e X. Tang, «Guided Image Filtering», *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, fasc. 6, pp. 1397–1409, giu. 2013, doi: 10.1109/TPAMI.2012.213.
- [25] H. Choi e J. Jeong, «Despeckling Algorithm for Removing Speckle Noise from Ultrasound Images», *Symmetry*, vol. 12, fasc. 6, p. 938, giu. 2020, doi: 10.3390/sym12060938.
- [26] P. Coupe, P. Hellier, C. Kervrann, e C. Barillot, «Nonlocal means-based speckle filtering for ultrasound images», *IEEE Trans. Image Process.*, vol. 18, fasc. 10, pp. 2221–2229, ott. 2009, doi: 10.1109/TIP.2009.2024064.
- [27] P. N. Devi e R. Asokan, «An improved adaptive wavelet shrinkage for ultrasound despeckling», *Sadhana*, vol. 39, fasc. 4, pp. 971–988, ago. 2014, doi: 10.1007/s12046-014-0254-5.
- [28] OpenMMLab, «Models — MMSegmentation 1.2.2 documentation». [Online]. Disponibile su: https://mmssegmentation.readthedocs.io/en/latest/advanced_guides/models.html
- [29] OpenMMLab, «Overview — MMSegmentation 1.2.2 documentation». [Online]. Disponibile su: <https://mmssegmentation.readthedocs.io/en/latest/overview.html>
- [30] Z. Liu *et al.*, «Swin Transformer: Hierarchical Vision Transformer using Shifted Windows», 17 agosto 2021, *arXiv*: arXiv:2103.14030. doi: 10.48550/arXiv.2103.14030.
- [31] T. Xiao, Y. Liu, B. Zhou, Y. Jiang, e J. Sun, «Unified Perceptual Parsing for Scene Understanding», in *Computer Vision – ECCV 2018*, vol. 11209, V. Ferrari, M. Hebert, C. Sminchisescu, e Y. Weiss, A c. di, in Lecture Notes in Computer Science, vol. 11209. , Cham: Springer International Publishing, 2018, pp. 432–448. doi: 10.1007/978-3-030-01228-1_26.
- [32] W. Zhang, J. Pang, K. Chen, e C. C. Loy, «K-Net: Towards Unified Image Segmentation».

- [33] M. Salvi *et al.*, «CYTO-KNET : An instance segmentation approach for multiple myeloma plasma cells using conditional kernels», *Int. J. Imaging Syst. Technol.*, vol. 34, fasc. 1, p. e22984, gen. 2024, doi: 10.1002/ima.22984.
- [34] R. Bommasani *et al.*, «On the Opportunities and Risks of Foundation Models», 12 luglio 2022, *arXiv*: arXiv:2108.07258. doi: 10.48550/arXiv.2108.07258.
- [35] J. Devlin, M.-W. Chang, K. Lee, e K. Toutanova, «BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding», 24 maggio 2019, *arXiv*: arXiv:1810.04805. doi: 10.48550/arXiv.1810.04805.
- [36] T. B. Brown *et al.*, «Language Models are Few-Shot Learners», 22 luglio 2020, *arXiv*: arXiv:2005.14165. doi: 10.48550/arXiv.2005.14165.
- [37] M. Moor *et al.*, «Foundation models for generalist medical artificial intelligence», *Nature*, vol. 616, fasc. 7956, pp. 259–265, apr. 2023, doi: 10.1038/s41586-023-05881-4.
- [38] A. Kirillov *et al.*, «Segment Anything», 5 aprile 2023, *arXiv*: arXiv:2304.02643. doi: 10.48550/arXiv.2304.02643.
- [39] X. Lin, Y. Xiang, L. Yu, e Z. Yan, «Beyond Adapting SAM: Towards End-to-End Ultrasound Image Segmentation via Auto Prompting», 8 luglio 2024, *arXiv*: arXiv:2309.06824. doi: 10.48550/arXiv.2309.06824.
- [40] A. A. Taha e A. Hanbury, «Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool», *BMC Med. Imaging*, vol. 15, fasc. 1, p. 29, dic. 2015, doi: 10.1186/s12880-015-0068-x.
- [41] K. H. Zou *et al.*, «Statistical validation of image segmentation quality based on a spatial overlap index1», *Acad. Radiol.*, vol. 11, fasc. 2, pp. 178–189, feb. 2004, doi: 10.1016/S1076-6332(03)00671-8.
- [42] D. Müller, I. Soto-Rey, e F. Kramer, «Towards a guideline for evaluation metrics in medical image segmentation», *BMC Res. Notes*, vol. 15, fasc. 1, p. 210, dic. 2022, doi: 10.1186/s13104-022-06096-y.

Ringraziamenti

Inizio col ringraziare i miei relatori, per il loro impegno, tempo e supporto dedicatomi senza eguali. Ringrazio anche per la pazienza e aiuto fornitomi durante questo mio ultimo passo verso la laurea

Ringrazio i miei genitori e mia sorella senza i quali non sarei qui e tutta questa mia esperienza non sarebbe stata possibile, siete stati il mio porto sicuro quando ne avevo bisogno e le ali di riserva quando mi serviva rispaccare il volo.

Ringrazio tutti e 4 i miei nonni che anche se non tutti fisicamente sento sempre vicino e con i quali c'è un'affinità inspiegabile.

Ringrazio i miei cugini (Tori, Josh ed Enri) e cugine (Carli, Otti, Cri; Joe, Ali, Giu, Vitto, Ade e Ceci), più fratelli e sorelle sarebbe più corretto dire, che riempiono il mio mondo da quando ho memoria.

Ai miei zii e zie che riescono sempre a mettermi di buon umore anche quando fuori sembra buio.

Ai miei amici più cari, chi c'è da sempre e che ci sarà sempre (Fu), chi mi ha preso per man e accompagnato in questo percorso accademico (Penni), e chi mi ha fatto ricredere che anche con meno tempo si possano creare dei rapporti che sembrano di una vita (Davi).

Ringrazio la FAM (Fi, Ali e Marti) per avermi accolto a Torino e avermi fatto sentire meno lo stacco con Palermo, grazie al “gruppo bomba” (di cui citerò Matti, Tzu, Robi e Palli come rappresentanza di tutti i personaggi che ne hanno fatto parte).

Grazie anche a tutte le persone di Palermo che mi hanno accompagnato durante la vita e che reputo importanti: Ale, Ludo, Simo, Eri, Peppe, Gaia, Manfri, Renna, Giu, Ele, Anto, Chia, Michi, Ciccio, Raffa, GSG, Fede, Vale e tutti compagni, squadre e colleghi del mio percorso.

Tutti voi mi avete fatto arrivare fino a qui, e non potrei esserne più grato. Ricordatevi che “L'Amore Sopravvive Sempre”.