Politecnico di Torino

Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Biomedica

A.a. 2021/2022

Sessione di Laurea Luglio 2022

Metodi di machine learning per il riconoscimento di metaplasie gastrointestinali

a partire da immagini gastroscopiche

Relatori:

Prof. Monica Visintin

Prof. Guido Pagana

Candidato:

Marcella Toma

Abstract

Quando cellule differenziate si evolvono in modo diverso da quello fissato dall'ontogenesi si parla di metaplasia. Il persistere dello stimolo che ha indotto la modificazione può indurre una trasformazione neoplastica. Oggetto di questa tesi è la metaplasia gastrointestinale e la progettazione di un metodo di machine learning che sia di sostegno nella relativa diagnosi a partire da immagini gastroscopiche. Le lesioni metaplastiche gastrointestinali presentano una struttura ghiandolare di aspetto tubulo-villoso, pertanto, si propone un riconoscimento automatico di tali pattern al fine di segmentare le zone a rischio. Il database a disposizione si compone di 102 acquisizioni gastroscopiche RGB associate agli esiti istologici relativi alla presenza di metaplasie. Alle immagini patologiche sono affiancate le maschere delle zone metaplastiche, segmentate manualmente da un esperto dell'*Azienda Ospedaliera Ordine Mauriziano di Torino*. Le immagini sono state acquisite attraverso gastroscopio Pentax MEDICAL in modalità i-SCAN 3 e sono rappresentative di diverse zone anatomiche dello stomaco.

Le immagini vengono processate allo scopo di isolarne le porzioni di interesse. Si è scelto di effettuare un *pre-processing* che prevede l'equalizzazione dell'istogramma delle luminosità e l'applicazione di un filtro gaussiano per la riduzione del rumore, per poi valutarne l'efficacia. Una volta convertite in scala di grigi, le immagini vengono divise in ROI (*Regions of Interest*) 25x25 con un overlap bidirezionale di 10 pixel. Per ogni ROI, si estraggono diverse tipologie di *feature*, valutate a partire da *Gray-Level Co-Occurrence Matrix* (GLCM), istogramma delle luminosità e *Gray-Level Run-Length Matrix* (GLRLM), per un totale di 20 variabili. Si utilizzano tre diversi *Construction Set* (CS) che differiscono nella rappresentazione della classe sana. Il CS1 prevede che questa sia

rappresentata unicamente da ROI sane estratte da immagini sane. Il CS2 introduce le ROI sane da immagini patologiche come classe a parte, passando ad un problema di classificazione multiclasse con due classi sane ed una patologica. Il CS3 suddivide le ROI sane da immagini patologiche in due cluster i quali vengono assegnati a due classi differenti, per un totale di tre classi sane ed una patologica. Tale clustering si effettua con l'ipotesi di poter distinguere elementi riconducibili a ROI effettivamente sane da misclassificazioni della segmentazione manuale. Ognuna delle tre configurazioni viene combinata con l'assenza o la presenza di *pre-processing* sulle immagini al fine di costruire sei classificatori. La dimensionalità dei CS viene ridotta attraverso l'impiego della *Principal Component Analysis* (PCA). Per tutti i classificatori si utilizza un modello ottimizzabile di Random Forest, allenato in concomitanza ad un tuning degli iperparametri mediante un processo di ottimizzazione bayesiana valutato in cross-validation.

Dei due classificatori più prestanti nella classificazione pixel per pixel si comparano le performance con e senza l'impiego della PCA. Il classificatore che presenta le prestazioni migliori è il Random Forest costruito a partire dal CS1 estratto da immagini sottoposte a *pre-processing* e senza l'utilizzo della PCA, con accuratezza bilanciata dell'80,67% ed F1-score del 57,18%. L'applicazione di operatori morfologici sulle maschere di segmentazione automatica permette di ottenere un'accuratezza dell'81,78% ed F1-score del 62,61%.

Indice

Indice delle figure 8
Indice delle tabelle 12
Acronimi 14
Capitolo 1. Introduzione e stato dell'arte 17
1.1. Che cos'è una metaplasia 18
1.2. La metaplasia intestinale 21
1.2.1. L'esofago di Barrett 21
1.2.2. La metaplasia gastrointestinale 23
1.3. Intelligenza Artificiale e Metaplasie Gastrointestinali 33
Capitolo 2. Materiali e metodi 36
2.1. Dataset 36
2.2. Pre-elaborazione 38
2.2.1. Isolamento della porzione di interesse 38
2.2.2. Isolamento della sola componente tissutale 40
2.2.3. Equalizzazione dell'istogramma delle luminosità 47
2.2.4. Filtraggio con filtro gaussiano 48
2.3. Estrazione delle feature 50
2.3.1. Feature dell'istogramma delle luminosità 51
2.3.2. Feature della GLCM 53
2.3.3. Feature della GLRLM 56
2.4. Definizione dei Construction Set 60
2.4.1. Construction Set 1 60
2.4.2. Construction Set 2 61
2.4.3. Construction Set 3 62
2.4.4. Definizione di Training Set e Test Set 63
2.5. Costruzione delle feature 68
2.6. Allenamento dei classificatori 70

Capitolo 3. Risultati e post-elaborazione 76
3.1. Classificazione delle ROI 77
3.1.1. Prestazioni di RF1 77
3.1.2. Prestazioni di RF2 79
3.1.3. Prestazioni di RF3 82
3.1.4. Prestazioni di RF4 85
3.1.5. Prestazioni di RF5 87
3.1.6. Prestazioni di RF6 90
3.1.7. Prestazioni a confronto 93
3.2. Classificazione dei pixel 94
3.3. Effetto della Principal Component Analysis 95
3.4. Estrazione delle maschere di segmentazione automatica 99
Capitolo 4. Conclusioni 103
4.1. Possibili miglioramenti e sviluppi futuri 105
Bibliografia 108

Indice delle figure

Figura 1. Acquisizioni endoscopiche di metaplasie intestinali: due focolai di mucosa gastrica tubulo-villosa (a sinistra) e presenza di LBC (a destra) [1]. 17

Figura 2. Epitelio squamoso stratificato e metaplasia intestinale a confronto nell'esofago di Barrett [3]. 21

Figura 3. Esofago colonnare [5]. 22

Figura 4. Cascata di Correa: step del processo di degenerazione tissutale che porta allo sviluppo di un adenocarcinoma di tipo intestinale [51]. 26

Figura 5. Immagini acquisite con tecnologia NBI. A) mucosa gastrica normale; B) metaplasia gastrointestinale [14]. 27

Figura 6. Immagine esemplificativa acquisita attraverso (a) HDWLE (i-SCAN disabilitato); (b) i-SCAN 1; (c) i-SCAN 2; (d) i-SCAN 3 [15]. 31

Figura 7. Zone anatomiche dello stomaco comprese nel dataset delle acquisizioni [53]. 36

Figura 8. Estrazione di una maschera binaria a partire dalla segmentazione manuale fornita dal medico. In bianco si raffigura la zona rappresentante la MGI. 38

Figura 9. Esempio di ritaglio automatico di un'immagine secondo gli step a, b e c descritti nella sezione 2.2.1. 40

Figura 10. Rimozione della porzione inferiore di un'immagine contente scritte sovrapposte. 1) Maschera binaria; 2) eliminazione dell'oggetto ad area maggiore; 3) localizzazione del livello per il ritaglio ed aggiunta di un margine; 4) immagine RGB finale. 42

Figura 11. Selezione del numero ottimale di cluster tramite la minimizzazione della WSS. 43

Figura 12. Immagine in grayscale (a sinistra) con relativa suddivisione in cluster (a destra). 43

Figura 13. Esempio di maschera rappresentativo delle zone sovraesposte da rimuovere. 44

Figura 14. Esempio di eliminazione dei riflessi di luce. 44

Figura 15. Esempio di maschera delle zone a bassa intensità da rimuovere: prima (a sinistra) e dopo (a destra) l'applicazione di operatori morfologici. 45

Figura 16. Esempio di applicazione della maschera totale per la rimozione delle zone di non interesse. 46

Figura 17. Esempio di un processo di equalizzazione: istogrammi ed immagini. 48

Figura 18. Esempio di filtraggio gaussiano: istogrammi ed immagini. 50

Figura 19. Confronto tra ROI di diverse dimensioni. 51

Figura 20. Esempio di istogramma delle luminosità. 51

Figura 21. GLCM (0°) di una ROI 4x4 con dinamica [0, 3]. 54

Figura 22. GLRLM (0°) di una ROI 4x4 con dinamica [1, 3]. 56

Figura 23. Esempio di immagine patologica nel CS1: la parte patologica inclusa nella linea verde viene utilizzata nel dataset delle ROI patologiche, la parte esterna non viene usata. 60

Figura 24. Esempio di immagine patologica nel CS2: la parte patologica inclusa nella linea verde viene utilizzata nel dataset delle ROI patologiche, la parte esterna viene inclusa in una seconda classe sana con label “2”. 61

Figura 25. Esempio di immagine patologica nel CS3: la parte patologica inclusa nella linea verde viene utilizzata nel dataset delle ROI patologiche, le ROI della parte esterna vengono smistate attraverso K-Means in due classi di ROI sane con label “2” e “3”. 62

Figura 26. Scelta numero di cluster per la classe patologica (a sinistra) e sana di label “0” (a destra) nelle configurazioni RF1, RF2 ed RF3. 64

Figura 27. Scelta numero di cluster classe patologica (a sinistra) e classe sana di label "0" (a destra) nelle configurazioni RF4, RF5, RF6. 65

Figura 28. Scelta numero di cluster per la classe sana di label "2" nella configurazione RF2. 65

Figura 29. Scelta numero di cluster per la classe sana di label "2" e "3" nella configurazione RF3. 66

Figura 30. Esempio di impostazione del numero di elementi da estrarre da ogni cluster per la costruzione del TRS (N) e del TS (M) a partire dal cluster di minore

numerosità. I cerchi rappresentano i cluster e la relativa estensione ne indica la numerosità. 67

Figura 31. Schema esplicativo di un processo di cross-validazione 5-fold. 73

Figura 32. Fasi della valutazione delle performance per la selezione del modello migliore da post-elaborare 76

Figura 33. Andamento dell'errore di classificazione della configurazione RF1 durante l'ottimizzazione degli iperparametri 78

Figura 34. Matrici di confusione del modello RF1. 79

Figura 35. Andamento dell'errore di classificazione della configurazione RF2 durante l'ottimizzazione degli iperparametri 80

Figura 36. Matrici di confusione del modello RF2. 81

Figura 37. Matrici di confusione del modello RF2 dopo l'aggregazione dei risultati. 81

Figura 38. Andamento dell'errore di classificazione della configurazione RF3 durante l'ottimizzazione degli iperparametri 82

Figura 39. Matrici di confusione del modello RF3. 83

Figura 40. Matrici di confusione del modello RF3 dopo l'aggregazione dei risultati tramite il metodo 1 84

Figura 41. Matrici di confusione del modello RF3 dopo l'aggregazione dei risultati tramite il metodo 2. 85

Figura 42. Andamento dell'errore di classificazione della configurazione RF4 durante l'ottimizzazione degli iperparametri 86

Figura 43. Matrici di confusione del modello RF4. 87

Figura 44. Andamento dell'errore di classificazione della configurazione RF5 durante l'ottimizzazione degli iperparametri 88

Figura 45. Matrici di confusione del modello RF5. 89

Figura 46. Matrici di confusione del modello RF5 dopo l'aggregazione dei risultati. 89

Figura 47. Andamento dell'errore di classificazione della configurazione RF6 durante l'ottimizzazione degli iperparametri. 90

Figura 48. Matrici di confusione del modello RF6. 91

Figura 49. Matrici confusionali del modello RF6 dopo l'aggregazione dei risultati tramite il metodo 1 (Paragrafo 3.1.3). 92

Figura 50. Matrici confusionali del modello RF6 dopo l'aggregazione dei risultati tramite il metodo 2 (Paragrafo 3.1.3). 92

Figura 51. Andamento dell'errore di classificazione del modello RF1 senza PCA durante l'ottimizzazione degli iperparametri 96

Figura 52. Matrici di confusione del modello RF1 senza PCA. 96

Figura 53. Andamento dell'errore di classificazione del modello RF4 senza PCA durante l'ottimizzazione degli iperparametri 97

Figura 54. Matrici di confusione del modello RF4 senza PCA. 97

Figura 55. Immagini patologiche: segmentazione manuale ed automatica a confronto 99

Figura 56. Immagini sane: segmentazione automatica 100

Figura 57. Immagini patologiche: confronto maschere automatiche originali e post-elaborate (Mod.). 101

Figura 58. Immagini sane: confronto maschere automatiche originali e post-elaborate (Mod.). 102

Indice delle tabelle

Tabella 1. Performance di Multilayer Perceptron Neural Network (MLPNN), Support Vector Machine (SVM), Random Forest e Random Fern a confronto. 35

Tabella 2. Composizione del dataset. 37

Tabella 3. Soglie imposte nell'isolamento della porzione di interesse. 39

Tabella 4. Feature estratte dall'istogramma delle luminosità. 53

Tabella 5. Feature estratte dalla GLCM. 56

Tabella 6. Feature estratte dalla GLRLM. 59

Tabella 7. Costituzione del CS1. 61

Tabella 8. Costituzione del CS2. 61

Tabella 9. Costituzione del CS3. 62

Tabella 10. Descrizione delle sei configurazioni RF. 63

Tabella 11. Numero di cluster rappresentativi delle singole classi nelle configurazioni RF 66

Tabella 12. Risultati della PCA per tutte le configurazioni RF. 69

Tabella 13. Spazi di ricerca degli iperparametri. 70

Tabella 14. Iperparametri selezionati per ogni configurazione RF. 73

Tabella 15. Metriche per la valutazione ed il confronto delle prestazioni dei classificatori. 77

Tabella 16. Costituzione di Training e Test Set nella configurazione RF1. 79

Tabella 17. Costituzione di Training e Test Set nella configurazione RF2. 80

Tabella 18. Costituzione di Training e Test Set nella configurazione RF3. 83

Tabella 19. Costituzione di Training e Test Set nella configurazione RF4. 87

Tabella 20. Costituzione di Training e Test Set nella configurazione RF5. 88

Tabella 21. Costituzione di Training e Test Set nella configurazione RF6. 91

Tabella 22. Prestazioni complessive dei modelli RF nella classificazione binaria S-P delle ROI. 93

Tabella 23. Confronto delle prestazioni dei modelli RF nella classificazione binaria pixel per pixel. 95

Tabella 24. Iperparametri selezionati per i modelli RF1 ed RF4 senza PCA. 98

Tabella 25. Valutazione dell'effetto della PCA sulle prestazioni delle configurazioni RF1 ed RF4 nella classificazione binaria pixel per pixel. 98

Tabella 26. Confronto performance del modello RF4 senza PCA prima e dopo la post-elaborazione (Mod.) delle maschere automatiche. 101

Tabella 27. Configurazione e prestazioni del modello finale. 104

Acronimi

A.O. Azienda ospedaliera

CAD Computer-Aided Diagnosis

CE Contrast Enhancement (aumento del contrasto)

CG Carcinoma gastrico

CS Construction Set

CULBP Color uniform local binary pattern

D-CNN Deep convolutional neural network

DL Deep learning

GGE Giunzione gastroesofagea

GLCM Gray-level co-occurrence matrix
(matrice delle co-occorrenze dei livelli di grigio)

GLN Gray-level non-uniformity

GLRLM Gray-level run-length matrix
(matrice delle estensioni delle sequenze di livelli di grigio)

GSC Giunzione squamoso-colonnare

HDWLE High definition white light endoscopy
(endoscopia a luce bianca ad alta definizione)

HGRE High gray-level run emphasis

IA Intelligenza artificiale

LBC Light blue crest

LGRE Low gray-level run emphasis

LRE Long run emphasis

LRHGE Long rum high gray-level emphasis

LRLGE Long run low gray-level emphasis

MGI Metaplasia gastrointestinale

MI Metaplasia intestinale

ML Machine Learning

MRGE Malattia da reflusso gastroesofageo

NIB Narrow band imaging (diagnostica per immagini a banda stretta)

NPV Negative predictive value

PC Principal Component (componente principale)

PCA Principal Component Analysis

PPV Positive predictive value

RF Random Forest

RLN Run-length non-uniformity

ROI Regione di interesse

RP Run percentage

SE Surface enhancement (evidenziazione della superficie)

SRE Short run emphasis

SRHGE Short run high gray-level emphasis

SRLGE Short run low gray-level emphasis

SVM Support vector machine

TE Tone enhancement (miglioramento delle tonalità)

TRS Training Set

TS Test Set

WLE White light endoscopy (endoscopia a luce bianca)

WSS Within cluster sum of squares (somma dei quadrati all'interno dei cluster)

Capitolo 1.
Introduzione e stato dell'arte

Questo studio si pone l'obiettivo di applicare metodi di *machine learning* (ML) su immagini gastroscopiche per il riconoscimento di *metaplasie gastrointestinali* (MGI) attraverso la ricerca di eventuali alterazioni morfologiche, caratteristiche del tessuto patologico. L'individuazione di abnormità tissutali riconducibili a metaplasia è, di fatti, di notevole interesse, poiché la presenza di MGI rappresenta un fattore di rischio non trascurabile di trasformazioni neoplastiche, irreversibili. Pertanto, un'identificazione accurata di alterazioni metaplastiche è alla base di un corretto iter di prevenzione tumorale. Nella procedura standard utilizzata per la diagnosi, il medico osserva la struttura ghiandolare della mucosa gastrica alla ricerca di eventuali anomalie caratterizzanti. Una sospetta MGI viene segnalata in presenza di una struttura allungata, nello specifico di tipo tubulo-villoso. Meno usuale è invece il riconoscimento di MGI a partire dal riscontro di una colorazione alterata dell'epitelio. Questa anomalia, nota come *Light Blue Crest* (LBC), si presenta sulle creste della superficie epiteliale ed è rappresentata da una sottile linea blu-chiaro, caratteristica meno comune della MGI oltre che di difficile rilevamento [1].

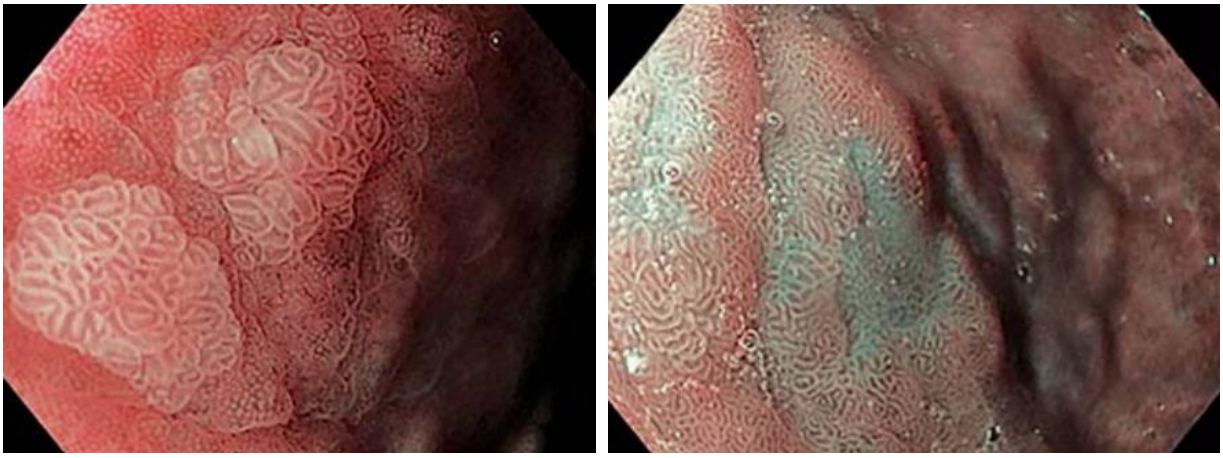

Figura 1. *Acquisizioni endoscopiche di metaplasie intestinali: due focolai di mucosa gastrica tubulo-villosa (a sinistra) e presenza di LBC (a destra) [1].*

Il lavoro condotto in questa tesi ha come fine ultimo quello di trasferire l'intero iter di riconoscimento della MGI in un algoritmo di ML, allo scopo di fornire al medico uno strumento di supporto nell'individuazione di eventuali porzioni di tessuto metaplastico. A tal proposito, il criterio di identificazione su cui si basa l'algoritmo è quello indicato come più affidabile dagli esperti del settore: l'analisi della struttura ghiandolare. Nell'algoritmo progettato, questa analisi si traduce nell'impiego di *feature* di texture con l'ipotesi che siano discriminanti per la classificazione delle diverse *regioni di interesse* (ROI) contenute nell'immagine. Questo studio si è concentrato su immagini acquisite tramite endoscopi PENTAX in modalità i-SCAN, una tecnologia di potenziamento dell'immagine che fornisce un tracciamento accentuato della texture della superficie mucosa e dei vasi sanguigni. La scelta di utilizzare tecniche di ML per l'implementazione si deve alla ridotta disponibilità di osservazioni per la costruzione del classificatore, ostacolo importante per un eventuale impiego di algoritmi di *deep learning* (DL). Inoltre, nonostante il DL sia di gran lunga la tecnica più indicata per l'analisi di immagini, d'altra parte, il ML garantisce un maggior controllo del processo di classificazione, con conseguente miglior interpretabilità dei risultati. In questo primo capitolo si intende fornire una visione d'insieme sulla patologia in analisi, assieme ad una panoramica dello stato dell'arte relativo alle tecniche di *intelligenza artificiale* (IA) impiegate nella costruzione di strumenti di supporto alla diagnosi.

1.1. Che cos'è una metaplasia

La metaplasia è la trasformazione di un tessuto adulto e completamente differenziato in un tessuto di un altro tipo. Tale fenomeno può avvenire per modificazione diretta delle cellule o, più frequentemente, viene indotto da un precedente sdifferenziamento a cui fa seguito il passaggio verso una diversa

condizione differenziata. La modificazione metaplastica è potenzialmente reversibile, pertanto, al cessare dello stimolo che ne ha provocato l'insorgenza, vi è la possibilità di recuperare funzione e morfologia della porzione di tessuto coinvolta. Tale caratteristica si deve al fatto che la variazione in questione si ha nell'espressione genica (fenotipica), mentre il genoma cellulare (genotipo) si conserva senza alcuna modificazione. Per questo motivo, ripristinate le condizioni ottimali originarie, i geni repressi durante la trasformazione metaplastica hanno la possibilità di tornare nuovamente ad esprimersi.

Ad esempio, nel caso di metaplasia gastrointestinale (MGI) (Vedi 1.2.2), da uno studio di Walker M. M. [2] si evince che, a lungo termine e con un follow-up di almeno cinque anni, esiste evidenza epidemiologica che la trasformazione metaplastica può essere reversibile. Per il raggiungimento di tale obiettivo, può tuttavia essere necessario l'impiego di una combinazione di agenti antiossidanti in concomitanza all'eradicazione dell'Helicobacter Pylori, principale fattore scatenante della MGI. D'altra parte, però, il persistere dello stimolo che porta alla formazione di metaplasie può indurre trasformazioni neoplastiche del tessuto metaplastico, raggiungendo quindi una condizione patologica irreversibile. Per questo motivo, nonostante una modificazione metaplastica preveda il passaggio da un tipo di epitelio sano ad un altro, deve essere considerata come una condizione patologica, poiché riconducibile a fattore di rischio importante e quindi a stadio pre-tumorale.

Tra le varie possibili cause che inducono l'insorgenza di una metaplasia si può individuare un denominatore comune: il tentativo di adattamento delle cellule in risposta ad un cambiamento nel loro ambiente vitale verso condizioni meno favorevoli, a seguito di stimoli ormonali, infiammatori, chimici o fisici.

Un esempio molto diffuso è quello della metaplasia del fumatore, che, come suggerisce la denominazione stessa, insorge a causa dell'effetto termico ed irritativo del fumo di sigaretta. La trasformazione metaplasica, in questo caso, prevede la sostituzione dell'epitelio cilindrico ciliato di trachea e bronchi con un epitelio pavimentoso pluristratificato. Tale trasformazione determina una maggior resistenza alle irritazioni indotte dal fumo, ma, allo stesso tempo, disabilita meccanismi protettivi caratteristici dell'epitelio cilindrico che usualmente agiscono contro infezioni e sostanze tossiche provenienti dall'ambiente esterno [2]. Di notevole diffusione è anche la *metaplasia intestinale* (MI), descritta nel dettaglio nel seguente sottocapitolo, poiché oggetto di studio di questo lavoro.

1.2. La metaplasia intestinale

La metaplasia intestinale è la sostituzione di una o più porzioni di tessuto con epitelio dai connotati riconducibili alla mucosa dell'intestino. Il termine "intestinale" non fa quindi riferimento all'organo bersaglio ma alla tipologia di tessuto che viene sviluppato con l'incombenza della patologia. Questo fenomeno ha luogo, tipicamente, negli organi interni cavi rivestiti di tessuto epiteliale di carattere ghiandolare. Maggiormente soggetti a questa modificazione sono l'esofago e lo stomaco. Nel primo caso si parla di esofago di Barrett, nel secondo di MGI [3].

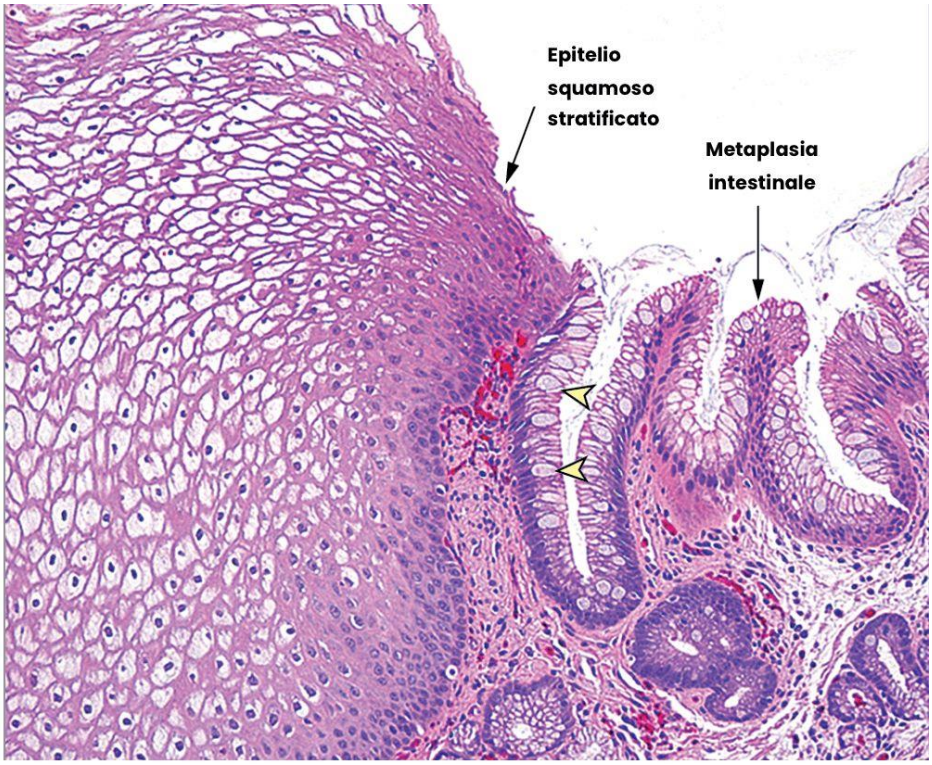

Figura 2. *Epitelio squamoso stratificato e metaplasia intestinale a confronto nell'esofago di Barrett [3].*

1.2.1. L'esofago di Barrett

L'esofago di Barrett è una condizione precancerosa che si manifesta, più soventemente, nella parte inferiore dell'esofago. Il nome della patologia deriva dal suo scopritore, Norman Rupert Barrett, un chirurgo di origini australiane che

ne pubblicò un report sul British Journal of Surgery nel 1950 [4]. Si tratta di un'alterazione che coinvolge il tessuto di rivestimento interno dell'esofago (mucosa squamosa), il quale viene sostituito da epitelio colonnare, simile a quello delle pareti duodenali.

I fattori di rischio maggiormente evidenziati dalla comunità scientifica per lo sviluppo di metaplasia intestinale dell'esofago sono: la *malattia da reflusso gastroesofageo* (MRGE), il fumo di sigaretta ed uno scorretto regime alimentare.

La diagnosi dell'esofago di Barrett prevede due step: una preliminare endoscopia che evidenzi la presenza di un esofago colonnare seguita da un'eventuale biopsia esofagea che confermi la presenta di metaplasia intestinale. La mucosa colonnare si distingue da quella squamosa esofagea poiché ha un caratteristico colore rosa-salmone e, soprattutto, presenta una texture più grossolana. Individuata tale modificazione, il medico quantifica l'estensione dell'esofago colonnare per determinare la presenza di esofago di Barrett a segmento lungo (>3 cm) o a segmento corto (<3 cm) [5].

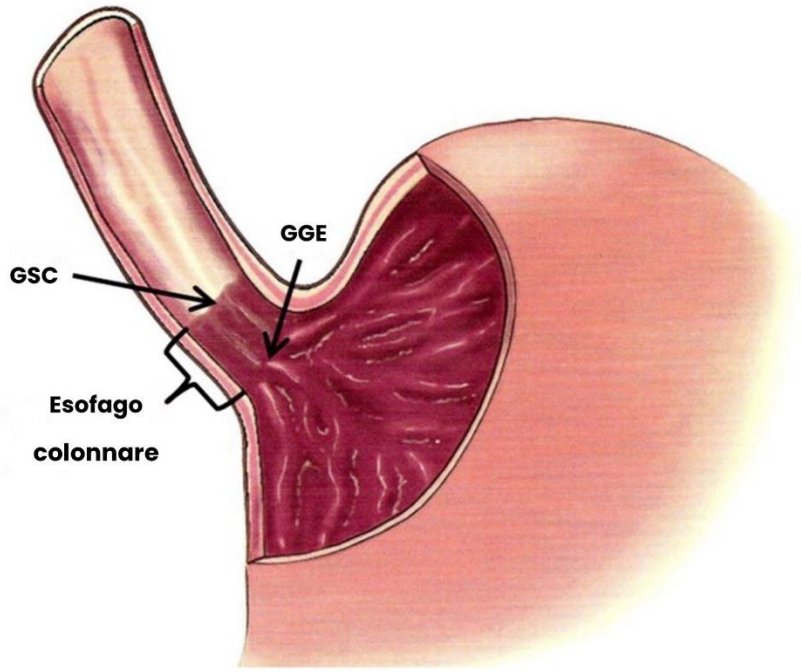

Figura 3. *Esofago colonnare [5].*

L'esofago colonnare è definito come la porzione di esofago delimitata dalla *giunzione gastroesofagea* (GGE) e quella *squamoso-colonnare* (GSC), come mostrato nella **Figura 3**.

1.2.2. La metaplasia gastrointestinale

La MGI è definita come la sostituzione della mucosa gastrica con dell'epitelio dalle caratteristiche tipiche dell'intestino. Questa modificazione è solitamente dovuta ad un'irritazione persistente della mucosa gastrica. Tra i numerosi fattori di rischio quelli più accreditati sono: le infezioni da Helicobacter Pylori, il fumo di sigaretta, la gastrite atrofica autoimmune, la predisposizione genetico-ereditaria al tumore allo stomaco e cattive abitudini alimentari. In particolare, la gastrite atrofica indotta da Helicobacter Pylori sembra essere il fattore scatenante più comune [6].

Una MGI può più facilmente degenerare in neoplasia rispetto al tessuto originario. Infatti, dal punto di vista quantitativo, si stima un rischio dieci volte maggiore di incorrere in un tumore allo stomaco in presenza di metaplasie. In particolare, di notevole diffusione a livello mondiale vi è l'adenocarcinoma gastrico che è caratterizzato da un elevato tasso di mortalità, pertanto, risulta importante monitorarne i fattori di rischio [7].

Analizzare le caratteristiche della MGI rilevata può, in questo senso, risultare di notevole utilità nella valutazione dei rischi associati ad una sua eventuale evoluzione patologica. Attraverso l'impiego di specifiche tecniche istologiche ed istochimiche, la MGI può essere classificata in diversi sottotipi. Tale procedura può avere alla base differenti sistemi di classificazione tra i quali il più utilizzato è quello di Jass & Filipe [8]. Questo metodo si basa sul tipo di cellule dell'intestino tenue che compaiono nella modificazione metaplastica: cellule assorbenti,

cellule caliciformi, cellule colonnari e/o cellule di Paneth [9]. Nello specifico, si distinguono tre tipi di MGI:

- *Tipo I* (completa): si rilevano cellule assorbenti, cellule di Paneth e cellule caliciformi che secernono sialomucina; viene replicato in modo esatto il fenotipo dell'intestino tenue.
- *Tipo II* (incompleta): sono presenti cellule colonnari e cellule caliciformi che secernono principalmente sialomucina; viene replicato il fenotipo dell'intestino crasso ed è introdotta una lieve distorsione ghiandolare.
- *Tipo III* (incompleta): si rilevano cellule caliciformi e cellule colonnari che secernono principalmente sulfomucina; viene replicato il fenotipo dell'intestino crasso, è presente una distorsione ghiandolare più severa.

Tale classificazione concorre alla previsione di eventuali risvolti e complicazioni. MGI di tipo I sono associate ad un minor rischio di sviluppare un tumore allo stomaco rispetto a MGI di tipo III [9]. Infatti, dalla letteratura si evince che, analizzando dati retrospettivi, si può notare come pazienti affetti da MGI di tipo III siano caratterizzati da un rischio di carcinoma gastrico quattro volte maggiore rispetto a quelli affetti da MGI di tipo I.

Molto frequentemente, i campioni di biopsia gastrica con MGI contengono focolai sia di metaplasia completa che incompleta (metaplasia mista) [10]. Nella valutazione del rischio di insorgenza di carcinoma gastrico, oltre al tipo di metaplasia, si considera anche la sua dimensione, pertanto, è di notevole interesse poterne quantificare l'estensione in fase diagnostica [11].

L'evoluzione del tessuto gastrico che porta allo sviluppo di formazioni tumorali è detta *cascata di Correa* [12]. L'ipotesi introdotta da tale concetto prevede l'insorgenza di una serie di modificazioni della mucosa gastrica verso alterazioni

sempre più severe fino a raggiungere una trasformazione di tipo neoplastico (**Figura 4**).

La cascata di Correa rappresenta un processo di carcinogenesi che consiste in una degenerazione progressiva del tessuto attraverso diversi stadi [12]:

- *Gastrite non-atrofica*: infiammazione cronica della mucosa gastrica caratterizzata da presenza di infiltrato infiammatorio e assenza di perdita di ghiandole gastriche appropriate.
- *Gastrite atrofica*: processo infiammatorio cronico della mucosa gastrica caratterizzato da presenza di infiltrato infiammatorio associato ad una perdita della componente ghiandolare, rimpiazzata da tessuto fibroso. Si distingue in:
    - gastrite atrofica autoimmune (o gastrite atrofica di tipo A): insorge a causa di un malfunzionamento del sistema immunitario;
    - gastrite atrofica multifocale (o gastrite atrofica di tipo B): si deve, tipicamente, a flogosi gastriche croniche indotte, ad esempio, da Helicobacter Pylori [13].
- *Metaplasia gastrointestinale di tipo I, II o III.*
- *Displasia*: è una modificazione caratterizzata da un fenotipo neoplastico sia in termini morfologici che di organizzazione architetturale. L'epitelio displastico mostra nuclei grandi, ipercromatici ed affollati, con perdita di polarità rispetto alla membrana basale.
- *Adenocarcinoma:* equivalente maligno di adenoma, tumore epiteliale caratterizzato da un aspetto tubulare o tubulovilloso insorto per proliferazione displastica. Viene diagnosticato come tale quando la modificazione si estende alla lamina propria o attraverso la muscularis mucosae.

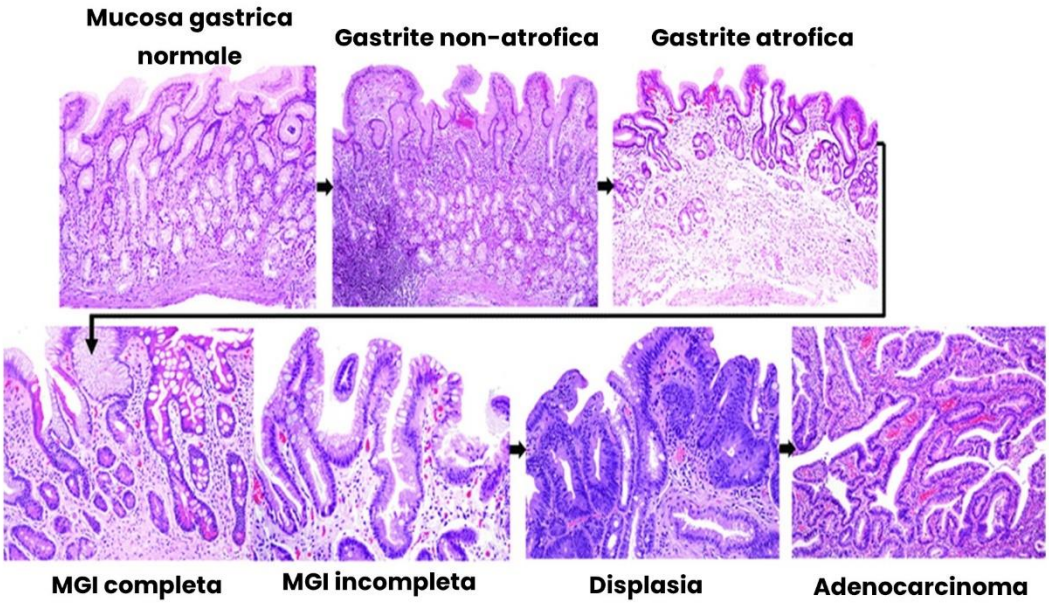

Figura 4. *Cascata di Correa: step del processo di degenerazione tissutale che porta allo sviluppo di un adenocarcinoma di tipo intestinale [51].*

1.2.2.1. Diagnosi endoscopica di una MGI

Il metodo diagnostico di base per il rilevamento di MGI è quello endoscopico. La convenzionale endoscopia a luce bianca (*White Light Endoscopy*, WLE) non fornisce un accuratezza tale da poter distinguere tra le lesioni gastriche pre-tumorali. Per questo motivo, differenti modalità di acquisizione sono state proposte per ottenere strumenti più sensibili nel rilevamento di MGI e *carcinoma gastrico* (CG) allo stadio iniziale.

Tra le tecnologie endoscopiche più diffuse vi è l'imaging a banda stretta (*Narrow Band Imaging*, NBI) che prevede il filtraggio della luce bianca secondo alcune specifiche lunghezze d'onda, al fine di massimizzare l'assorbimento da parte dell'emoglobina e di limitare la penetrazione della luce oltre la superficie mucosa. In particolare, la luce blu, a causa della sua ridotta lunghezza d'onda, non penetra in profondità e per questo fornisce una migliore resa della struttura fine presente in superficie, anche senza l'impiego di coloranti. Le immagini

ottenute attraverso questo metodo di acquisizione permettono di distinguere la mucosa gastrica normale, dal pattern tondeggiante e regolare, da quella affetta da metaplasia intestinale, caratterizzata da pattern tubulovilloso ed irregolare. Inoltre, una mucosa senza alterazioni presenta fossette gastriche spaziate in modo omogeneo, mentre nella MGI è evidente una densità vascolare molto variabile (**Figura 5**).

In alcuni casi, in presenza di tessuto metaplastico, è possibile individuare le LBC, tuttavia, oltre al fatto che tale caratteristica non è sempre presente, essa può essere rilevata solo sotto lunghezze d'onda di 400-430 nm, risultando di difficile rilevamento.

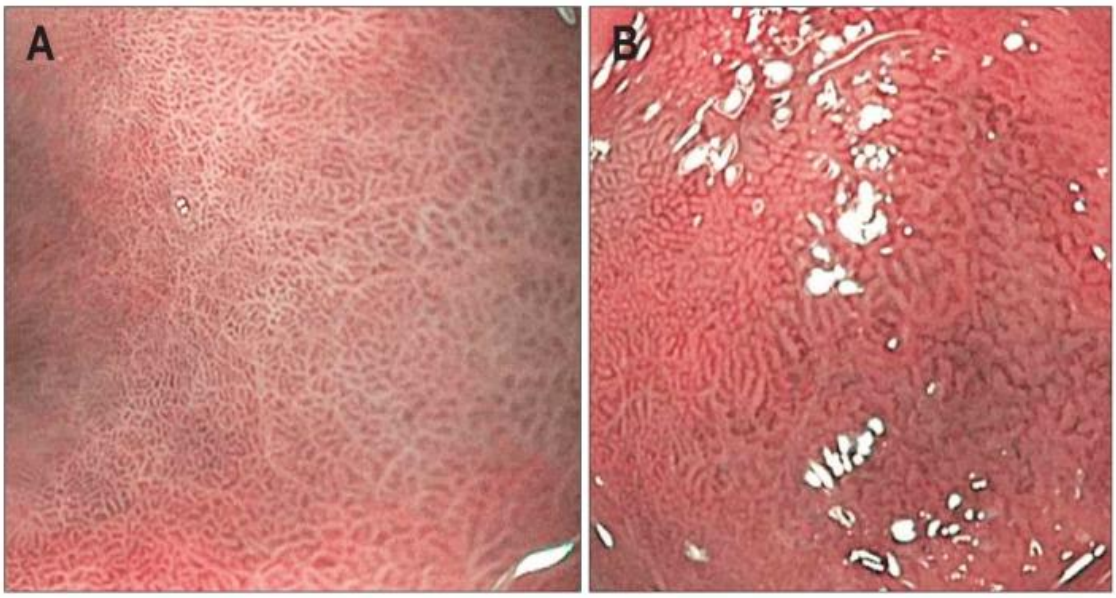

Figura 5. *Immagini acquisite con tecnologia NBI. A) mucosa gastrica normale; B) metaplasia gastrointestinale [14].*

Un'altra tecnica diagnostica di largo impiego è la cromoendoscopia, metodo che prevede la colorazione della mucosa gastrica attraverso l'uso di coloranti selettivamente assorbibili da mucose non-acide ed assorbenti, come quella presente nelle cellule intestinali. In questo modo, tale tecnica consente di delineare l'estensione di piccole irregolarità superficiali ed anomalie istologiche

dello stomaco. Tra i coloranti più utilizzati vi è il blu di metilene e l'indigotina. Se, da un lato, la cromoendoscopia può essere molto d'aiuto nella detezione di lesioni precancerose, questa tecnica allunga le tempistiche della pratica endoscopica, aumenta il carico di lavoro del personale e può intaccare la tolleranza del paziente alla procedura [14].

Per ovviare a tale inconveniente nella routine clinica, diversi approcci alternativi, innovativi ed altamente tecnologici sono stati proposti dai produttori di endoscopi di ultima generazione. Tra gli esempi più noti vi è l'Olympus NBI, il Fujinon FICE (Flexile spectral Imaging Color Enhancement) ed il PENTAX i-SCAN. Tali tecniche permettono di sostituire l'applicazione diretta di coloranti sulla mucosa gastrica con metodi di intensificazione del colore che agiscono sulle immagini in fase di acquisizione. In questo modo, la texture della mucosa e i vasi sanguigni vengono messi in risalto.

1.2.2.2. La funzione PENTAX i-SCAN di PENTAX Medical

La tecnologia i-SCAN di PENTAX è una tecnica di potenziamento di immagini che esegue modifiche in tempo reale pixel per pixel su frame acquisiti tramite HDWLE (High Definition WLE). Il dispositivo di acquisizione è provvisto di tre principali operazioni che vengono eseguite in diverse combinazioni per ottenere obiettivi differenti nel processo di potenziamento [15]. Tali funzioni sono:

- Aumento del contrasto (*Contrast Enhancement*, CE).

  Algoritmo che si applica a ciascun pixel del segnale video. Il singolo pixel in esame, pixel target, viene confrontato con quelli circostanti. Se il pixel target ha luminosità inferiore a quelli vicini almeno lungo una delle tre direzioni, è considerato "localmente scuro" ed il suo tono di blu viene leggermente incrementato. Se, invece, il pixel target ha luminosità maggiore o uguale a quella dei pixel circostanti, rimane invariato. Il

risultato che ne deriva è un aumento nel contrasto ed un'evidenziazione delle aree superficiali caratterizzate da micro-irregolarità.

- Evidenziazione della superficie (*Surface Enhancement*, SE).

  Algoritmo che si applica a ciascun pixel del segnale video. Il pixel target rappresenta il centro di una matrice di 25 pixel (5x5). Si calcola la "variabile di potenziamento" come la differenza tra la luminanza del pixel target e la media delle luminanze dei pixel circostanti appartenenti alla matrice 5x5. Quando il valore ottenuto è positivo, ovvero la luminanza del pixel target risulta maggiore di quella media circostante, il pixel target è reso ancora più chiaro dall'algoritmo. D'altra parte, quando la variabile di potenziamento è negativa, ovvero la luminanza del pixel target risulta minore di quella media circostante, il pixel target è considerato "pixel di bordo" e viene reso più scuro dall'algoritmo. Il risultato che ne deriva è una messa in risalto di pattern dettagliati e micro-irregolarità delle strutture superficiali della mucosa, attraverso un puntuale aumento del contrasto tra i pixel target ed il loro intorno.

- Miglioramento delle tonalità (*Tone Enhancement*, TE).

  Funzione che prevede l'applicazione di sei filtri digitali su ciascun frame del segnale video. Ogni filtro rappresenta una distinta curva di tono per ciascun canale R, G e B. Tali curve specificano una relazione di input-output che prevede una rimappatura di tutti i valori dei pixel per ogni canale di colore. L'obiettivo dei filtri utilizzati è quello di aumentare il contrasto nei canali G e B riducendo la luminosità del canale R. Questo allo scopo di mettere in risalto i vasi sanguigni, poiché l'emoglobina presenta un'elevata capacità di assorbimento di componenti di luce blu

e verde mentre mostra caratteristiche di basso assorbimento per la componente di luce rossa. PENTAX fornisce due modalità di TE predefinite che prevedono una notevole soppressione della luminosità del canale R: TEc e TEg.

Diverse possibili combinazioni degli algoritmi descritti portano alla definizione di tre diverse modalità di utilizzo di PENTAX i-SCAN, con lo scopo di evidenziare caratteristiche differenti del tessuto su cui si indaga:

- i-SCAN 1: comprende operazioni di CE e SE, mettendo in risalto i vasi superficiali e la texture della mucosa. Fornisce all'utente un'immagine luminosa come nella convenzionale HDWLE, pur aumentandone la nitidezza. Evidenzia aree depresse o sopraelevate della mucosa e i vasi sanguigni, più nitidi, mantengono la propria naturale colorazione rossa.

- i-SCAN 2: comprende stessi livelli di SE e CE di i-SCAN 1 ed introduce un'operazione di TE in modalità c (TEc). L'algoritmo di TEc aumenta il contrasto tra la mucosa e i vasi sanguigni, incrementandone la visibilità rispetto ad i-SCAN 1. Tale procedura può comportare una riduzione della luminosità di porzioni dell'immagine già caratterizzate da basse intensità.

- i-SCAN 3: comprende stessi livelli di SE e CE di i-SCAN 1 ed introduce un'operazione di TE in modalità g (TEg). L'algoritmo di TEg impiega curve di tono differenti rispetto a quello di TEc presente in i-SCAN 2. La modalità i-SCAN 3 introduce una maggior visibilità dei vasi sanguigni anche in regioni lontane scarsamente illuminate attraverso un bilanciamento delle alterazioni introdotte sulle tonalità, contro un eccessivo oscuramento delle aree a bassa luminosità.

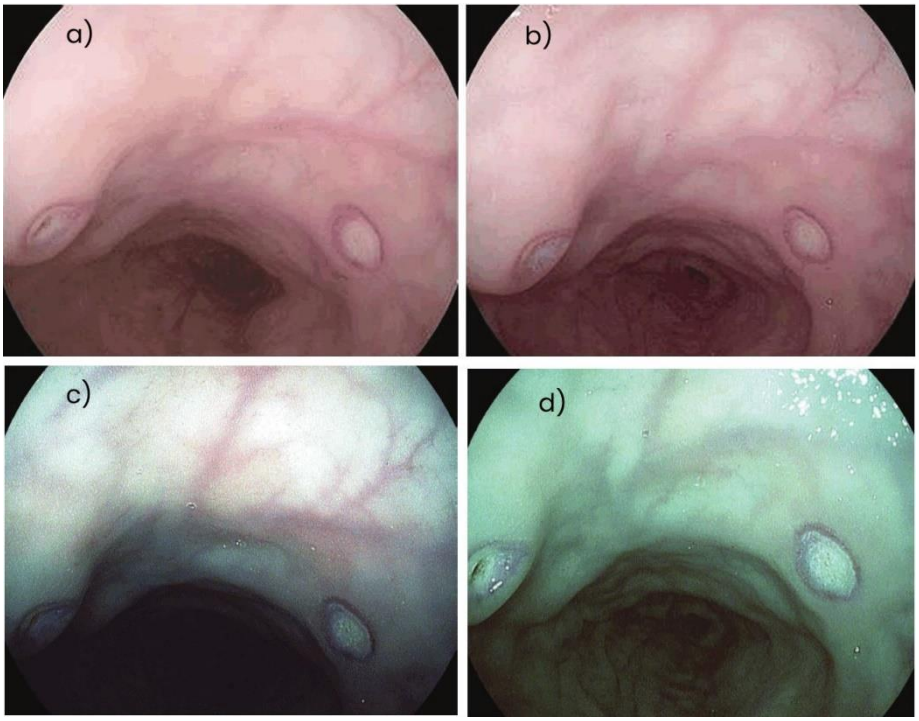

Figura 6. *Immagine esemplificativa acquisita attraverso (a) HDWLE (i-SCAN disabilitato); (b) i-SCAN 1; (c) i-SCAN 2; (d) i-SCAN 3 [15].*

La resa delle diverse modalità è fornita nella **Figura 6**.

1.2.2.3. Confronto tra prestazioni di tecniche endoscopiche

Al fine di valutare le prestazioni delle singole tecniche diagnostiche che si basano su acquisizioni endoscopiche, è necessario individuare un gold standard che faccia da riferimento. Tipicamente, questo coincide con il risultato istologico fornito da un prelievo bioptico. Di fatti, il completo iter di diagnosi della MGI prevede un check delle ipotesi avanzate in fase di osservazione delle immagini endoscopiche attraverso una biopsia che permette, inoltre, di assegnare un punteggio istologico, indicatore del rischio di progressione verso CG. In questo modo, è possibile confrontare le performance dei metodi di diagnosi endoscopica precedentemente descritti.

Uno studio olandese su 47 pazienti affetti da MGI ha comparato i risultati ottenuti tramite l'impiego dell'endoscopia a luce bianca (*White Light Endoscopy*, WLE)

con quelli ricavati da endoscopie a banda stretta (*Narrow Band Imaging*, NBI). Sono stati calcolati sensibilità, specificità, valore predittivo positivo (*positive predictive value*, PPV) e valore predittivo negativo (*negative predictive value*, NPV) trovando valori rispettivamente di 71%, 58%, 65% e 65% per l'NBI rispetto ai 51%, 67% 62% e 55% per la WLE [16].

Secondo uno studio prospettico multicentrico condotto da Pimentel-Nunes et al., l'NBI incrementa la sensibilità nella diagnosi della MGI in modo significativo rispetto alla WLE (87% vs 53%; P<0,001), con un generale aumento dell'accuratezza diagnostica (94% vs 83%; P<0,001) [1].

D'altra parte, al momento, le prestazioni della tecnologia i-SCAN, essendo questa una tecnica relativamente nuova nel campo della diagnosi di patologie del tratto gastrointestinale, sono attestate da una limitata disponibilità di letteratura clinica. Tuttavia, è presente un crescente numero di risultati che indicano che i-SCAN occupa una posizione di rilievo fra le tecnologie endoscopiche avanzate. Nello specifico, secondo lo studio *randomizzato* condotto da Hong et al., i-SCAN 2 migliora la previsione istologica delle lesioni colon-rettali iperplastiche ed adenomatose rispetto all'HDWLE [17].

Inoltre, uno studio di Lee et al. [18] ha presentato un confronto tra i-SCAN e NBI nella previsione della patologia del polipo diminutivo del colon. Si è notato che, in modo statisticamente significativo, entrambe le tecnologie presentano sensibilità e accuratezza maggiori rispetto alla HDWE. Non si rilevano, però, differenze significative tra NBI e i-SCAN (sensibilità: 88,8% vs 94,6%; specificità: 86,8% vs 86,4% e accuratezza: 87,8% vs 90,7%; P<0,05) [18].

Un altro recente lavoro condotto da Cammarota et al. ha studiato la rilevanza del ruolo di i-SCAN nella valutazione dei modelli villosi duodenali [19].

1.3. Intelligenza Artificiale e Metaplasie Gastrointestinali

Negli anni, una serie di studi hanno introdotto e valutato metodi di intelligenza artificiale su immagini endoscopiche del tratto gastrointestinale. L'obiettivo di fondo è quello di fornire al medico uno strumento di supporto alla diagnosi di patologie e/o lesioni pretumorali. Il carcinoma gastrico trova maggior incidenza a livello mondiale nell'Asia dell'Est (58,1%) dove, per questo motivo, si prevede una procedura di *screening* endoscopico del CG. D'altra parte, nei paesi occidentali, si riscontra un'incidenza di gran lunga inferiore rispetto a quella dei paesi orientali, sebbene non sia in alcun modo trascurabile (14,7% in Europa e 7,8% nell'America Centrale e Latina) [20]. Per questo motivo, nella società occidentale, nonostante l'elevato tasso di mortalità dovuta al carcinoma gastrico, non è prevista una procedura preventiva di *screening*. Infatti, a differenza di quanto riscontrato in alcuni paesi dell'Asia orientale, non è ancora stato provato che la diagnosi endoscopica preventiva di questa tipologia di cancro sia efficace quanto sostenibile. Questo si deve principalmente al fatto che la correlazione tra l'endoscopia convenzionale e l'istologia viene considerata inadeguata [1]. Allo stesso tempo, mentre nell'Asia Orientale non ci sono particolari limitazioni sul numero e sulle condizioni contestuali per l'esecuzione di un prelievo bioptico, nei Paesi occidentali queste sono molto più presenti [20]. Per questo motivo, a livello scientifico è stato introdotto l'obiettivo di fornire agli esperti endoscopisti dei sistemi *CAD* (Computer-Aided Diagnosis) che siano di supporto all'individuazione di condizioni e/o lesioni pre-tumorali, allo scopo di poter raggiungere accuratezze diagnostiche comparabili con quelle dei referti istopatologici. Le tecnologie di IA impiegate si distinguono tra quelle che

prevedono estrazione e costruzione manuale delle *feature* (ML) e quelle di *deep learning* (DL).

Diversi sistemi di detezione automatici sono stati sviluppati impiegando algoritmi di ML, poiché tali metodi non necessitano di una disponibilità eccessiva di dati per la fase di costruzione. Nello studio condotto da Van Der Sommen et al [21], ad esempio, si utilizza un classificatore Support Vector Machine (SVM) per il riconoscimento di aree riconducibili a cancro esofageo allo stadio iniziale. L'algoritmo basa il suo apprendimento su dati etichettati che comprendono *feature* di texture e di colore relative ad immagini endoscopiche. Le caratteristiche di texture sono estratte a partire dall'applicazione di filtri di Gabor 2D lungo diverse orientazioni. Le prestazioni raggiunte sono descritte da un valore di recall pari al 95% ed una precision del 75% [21].

Analogamente, la pubblicazione di Kanesaka et al [22]. introduce l'impiego di due classificatori SVM per la costruzione di un sistema computer-aided per la diagnosi precoce del cancro gastrico. Un classificatore lavora a livello immagine, l'altro a livello ROI ed entra in gioco solo nel caso in cui l'immagine venga identificata come patologica dal primo algoritmo SVM, fornendo informazioni sulla presunta estensione della lesione individuata. Preliminarmente, le immagini sono state sottoposte ad una fase di *pre-processing* con l'applicazione di un filtro gaussiano per la riduzione del rumore e l'equalizzazione dell'istogramma delle luminosità. In questo caso, sono state utilizzate solo *feature* descrittive della texture, calcolate a partire da una *Gray-Level Co-Occurrence Matrix* (GLCM). L'accuratezza diagnostica dichiarata è del 96,3%, con recall pari a 96,7% e precision del 96,3% [22].

Un altro interessante studio è quello condotto da Zhou et al [23]. nel quale vengono messe a confronto le performance di diversi classificatori che hanno

come compito quello di classificare immagini acquisite tramite capsula endoscopica. L'apprendimento è basato sull'impiego di *feature* di texture e di colore, estratte a partire da un innovativo algoritmo di Color Uniform Local Binary Pattern (CULBP). Le performance raggiunte dai diversi classificatori sono raccolti nella **Tabella 1**. L'utilizzo di Random Forest e Random Fern porta a delle prestazioni significativamente migliorate [23].

	MLPNN	*SVM*	*Random Forest*	*Random Fern*
Sensibilità (%)	90	93,3	98	98,7
Specificità (%)	86,3	89,5	98	98,5
Accuratezza (%)	87,27	90,5	98	98,6

Tabella 1. *Performance di Multilayer Perceptron Neural Network (MLPNN), Support Vector Machine (SVM), Random Forest e Random Fern a confronto.*

D'altra parte, tra i metodi di DL più utilizzati per la diagnosi di lesioni a carico del tratto gastrointestinale vi sono le deep convolutional neural network (D-CNN).

Ad esempio, uno studio di Xu et al. [24] ne fa uso per lo sviluppo di un sistema di supporto alla diagnosi di condizioni pre-cancerose a livello gastrico su frame acquisiti tramite endoscopia con potenziamento di immagine. L'accuratezza diagnostica ottenuta su di dati multicentrici è dell'85,9% [24].

Analogamente, la ricerca di Yan et al. [25] ha portato allo sviluppo di un sistema intelligente di diagnosi di MGI attraverso l'impiego di una CNN. Questo studio mette a confronto gli indicatori delle performance dello strumento sviluppato con quelle dei medici esperti che si occupano della diagnosi. Secondo tale pubblicazione, passando dalle diagnosi avanzate dai medici a quelle fornite dal sistema implementato, la sensibilità passa dall'86,5% al 91,9%, la specificità dall'81,4% all'86% e l'accuratezza dall'83,8% all'88,8% [25].

Capitolo 2.
Materiali e metodi

2.1. Dataset

Il *dataset* utilizzato in questo studio è composto da diverse immagini endoscopiche, fornite dal reparto di gastroenterologia *dall'Azienda Ospedaliera Ordine Mauriziano di Torino* (*A.O. Ordine Mauriziano di Torino*), una struttura dotata di aree di alta specializzazione. Le acquisizioni risalgono prevalentemente alla primavera del 2021 e sono rappresentative di differenti aree anatomiche dello stomaco, comprendendo prevalentemente la zona dell'altro, del corpo e quella pilorica (Vedi **Figura 7**). Sono state acquisite attraverso l'impiego di endoscopi PENTAX Medical in modalità i-SCAN 3. Pertanto, i frame di HDWLE sono stati filtrati in real-time mediante la combinazione di sei filtri digitali, allo scopo di mettere in risalto la texture mucosale ed i vasi sanguigni.

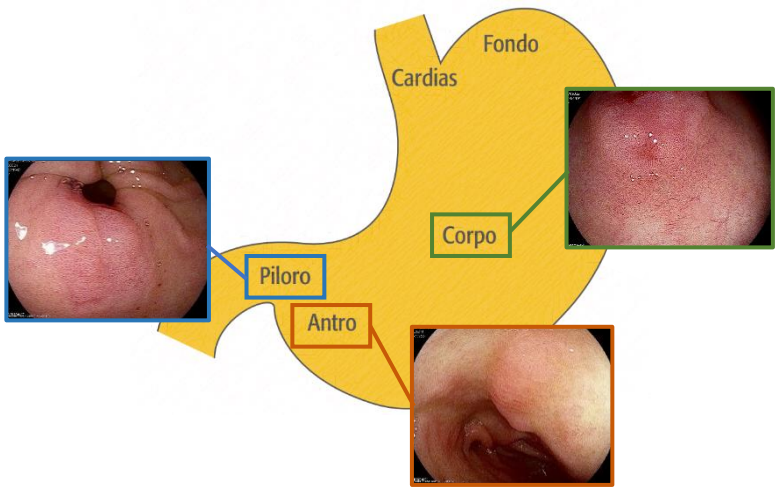

Figura 7. *Zone anatomiche dello stomaco comprese nel dataset delle acquisizioni [53].*

Nello specifico, il *dataset* comprende 102 immagini suddivise in questo modo:

- 57 immagini sane;
- 45 immagini contenenti lesioni metaplastiche (MGI).

I soggetti inclusi nello studio sono 24 ed ognuno di essi è rappresentato da acquisizioni di diverse zone anatomiche dello stomaco. Per questo motivo, non è stato possibile considerare una divisione tra pazienti sani e patologici, poiché uno stesso paziente può presentare assenza di patologia in alcune aree dello stomaco e, allo stesso tempo, essere affetto da MGI in una diversa vista endoscopica.

	Immagini sane	*Immagini patologiche*
Numero di immagini	57	45
Sesso maschile	43,86 %	37,78 %
Età media	60±9	66±10

Tabella 2. *Composizione del dataset.*

Come evidenziato nella **Tabella 2**, prevalgono acquisizioni su pazienti di sesso femminile. Inoltre, i dati riportati si riferiscono alle immagini effettivamente utilizzate per la costruzione e la validazione del classificatore, al netto di acquisizioni non utilizzabili. Sono state escluse, infatti, immagini che comprendessero parti di endoscopio o nettamente sfocate.

Le acquisizioni sono state etichettate come sane o patologiche sulla base di un referto istologico. Alle immagini contenenti MGI sono state affiancante delle segmentazioni manuali della presunta area patologica, effettuate da un esperto dell'*A.O. Ordine Mauriziano* al fine di caratterizzarne l'estensione. Da queste segmentazioni fornite, sono state ricavate delle maschere binarie mediante l'impiego di *Matlab*.

Figura 8. *Estrazione di una maschera binaria a partire dalla segmentazione manuale fornita dal medico. In bianco si raffigura la zona rappresentante la MGI.*

La maschera in **Figura 8** verrà utilizzata per fornire la corretta etichetta (*label*) ad ogni porzione delle immagini analizzate.

2.2. Pre-elaborazione

Allo scopo di estrarre le *feature* per la costruzione del classificatore, è stato necessario isolare la porzione dell'immagine contenente il tessuto da analizzare. Tale procedura è stata implementata mediante l'impiego di *Matlab* secondo gli step seguentemente descritti ed è pensata per il formato fornito in output dal tool utilizzato dall'*A.O. Ordine Mauriziano*.

2.2.1. Isolamento della porzione di interesse

Le acquisizioni grezze erano comprensive di doppia vista endoscopica, come evidenziato nella prima immagine della **Figura 9**. Per questo motivo, è stato implementato un primo passaggio di ritaglio automatico delle immagini.

Tale procedura prevede:

a. **Rimozione della banda laterale destra con seconda vista.**

   Si mantiene la sola porzione di immagine che rappresenta i 2/3 della sua larghezza, a partire dalla prima colonna di sinistra (Vedi **Figura 9**a).

b. **Estrazione di una maschera binaria dell'oggetto di interesse.**

   Attraverso un'operazione di sogliatura sulle intensità (**Tabella 3**, prima riga), si ricava una maschera che rappresenta in bianco le zone da

rimuovere, ovvero le bande laterali scure, in nero le aree di interesse (Vedi **Figura 9**b).

c. Individuazione della zona di interesse.

Nella maschera binaria ricavata al punto b sono rappresentati in bianco i pixel che presentano un'intensità inferiore al 5% della massima, i quali sono contenuti in quantità più importanti nelle zone di non interesse dell'immagine. Per questo motivo, a partire dalla maschera binaria, si valuta il numero di pixel bianchi (da rimuovere) presenti per ogni riga e per ogni colonna. Si impone una soglia di numerosità (**Tabella 3**, seconda riga) oltre la quale la riga e/o colonna risulta da rimuovere (Vedi **Figura 9**c).

	Tipo di soglia	*Soglia*	*% rispetto a*
Step b	Intensità	<5 %	Intensità massima
Step c	# pixel bianchi	per colonne: >80 %	# pixel/colonna
		per righe: >50 %	# pixel/riga

Tabella 3. *Soglie imposte nell'isolamento della porzione di interesse.*

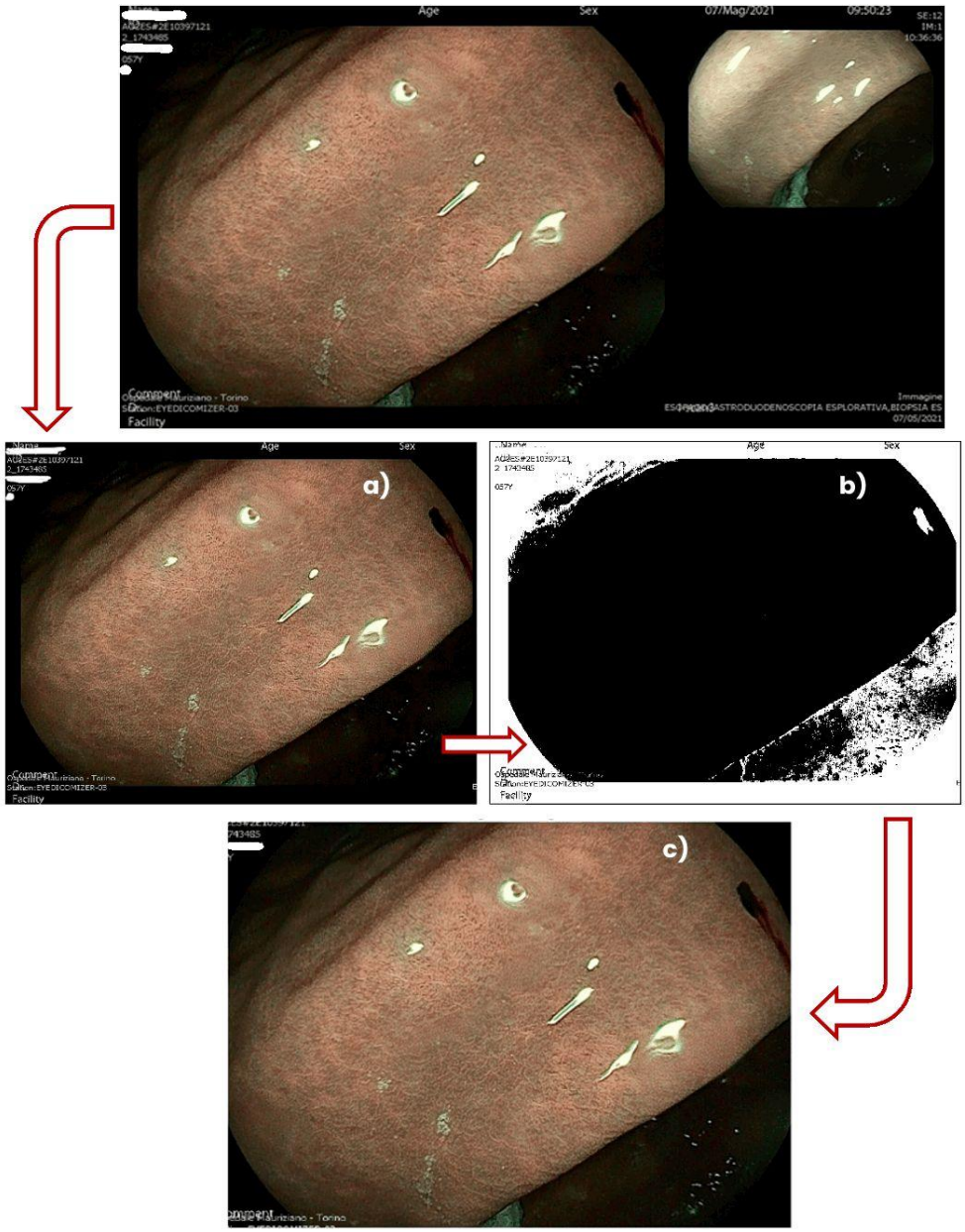

Figura 9. *Esempio di ritaglio automatico di un'immagine secondo gli step a, b e c descritti nella sezione 2.2.1.*

2.2.2. Isolamento della sola componente tissutale

Le immagini ottenute dallo step precedente sono ancora comprensive di elementi di non interesse e/o disturbo quali: scritte relative a metadati clinici e riflessi di luce.

Per ovviare a questa problematica, un altro script interviene sull'immagine al fine di estrarne la sola componente tissutale. Se ne elencano di seguito gli step implementativi.

a. **Rimozione della porzione inferiore dell'immagine**.

   Un algoritmo di riconoscimento dell'estensione della scritta sovrapposta determina un indice riga all'altezza del quale viene applicato un ulteriore ritaglio dell'immagine. Lo script si articola nei seguenti passaggi:

   - binarizzazione dell'immagine attraverso una soglia globale di intensità fornita dal metodo di Otsu (Vedi **Figura 10**.1). Il metodo di Otsu prevede una segmentazione binaria dell'immagine secondo l'impostazione di una soglia colore selezionata in modo automatico dall'algoritmo. Nello specifico, si analizza la varianza intra-classe delle due classi di pixel ottenute al variare della soglia. La soglia prescelta è quella che minimizza tale varianza [26] [27].
   - rimozione dell'oggetto ad area maggiore appartenente all'immagine binaria ottenuta, sicuramente non rappresentativo delle scritte da rimuovere (Vedi **Figura 10**.2);
   - localizzazione delle scritte nella porzione inferiore dell'immagine attraverso un'analisi della distribuzione dei pixel bianchi nelle prime due colonne dell'immagine. Nello specifico, si seleziona la sequenza di pixel chiari che rappresenterà la porzione da rimuovere (Vedi **Figura 10**.3).

   Considerato un margine di 15 pixel verso l'alto, si valuta l'indice riga a livello del quale viene applicato il nuovo ritaglio (Vedi **Figura 10**.3-4).

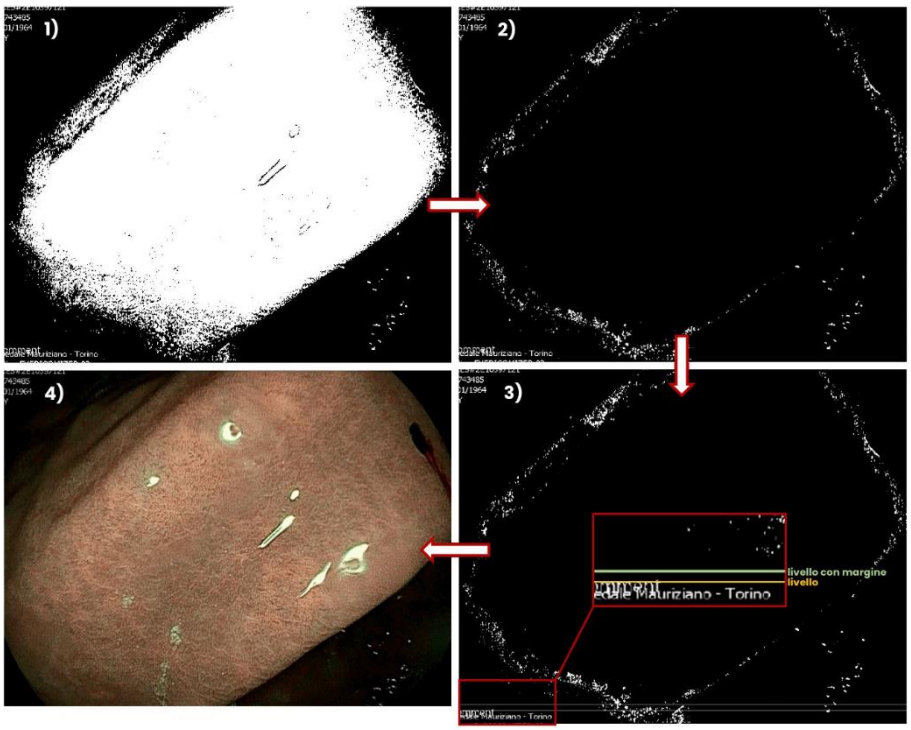

Figura 10. *Rimozione della porzione inferiore di un'immagine contente scritte sovrapposte. 1) Maschera binaria; 2) eliminazione dell'oggetto ad area maggiore; 3) localizzazione del livello per il ritaglio ed aggiunta di un margine; 4) immagine RGB finale.*

b. Rimozione dei riflessi di luce.

Si utilizza l'algoritmo *K-Means* per immagini attraverso la funzione built-in di *Matlab*. Questo allo scopo di effettuare un clustering delle intensità e rimuovere il cluster ad intensità media più elevata, con l'ipotesi che questo rappresenti le aree sovraesposte e soggette a riflessi di luce. Il numero di cluster più adatto è stato selezionato attraverso il metodo della *Within Cluster Sum of Squares* (WSS). Tale metodo prevede il calcolo di un indicatore di compattezza come la somma dei quadrati delle distanze tra i punti e i corrispettivi centroidi, per ogni cluster.

Nello specifico si ha [28]:

$$WSS = \sum_{i=1}^{N_c} \sum_{x \in C_i} d(x, \bar{x}_{C_i})^2$$

$$con\ C_i = cluster; N_c = \#\ cluster;\ \bar{x}_{C_i} = centroide\ del\ cluster.$$

Pertanto, con lo scopo di ottenere cluster più compatti, è importante minimizzare la WSS.

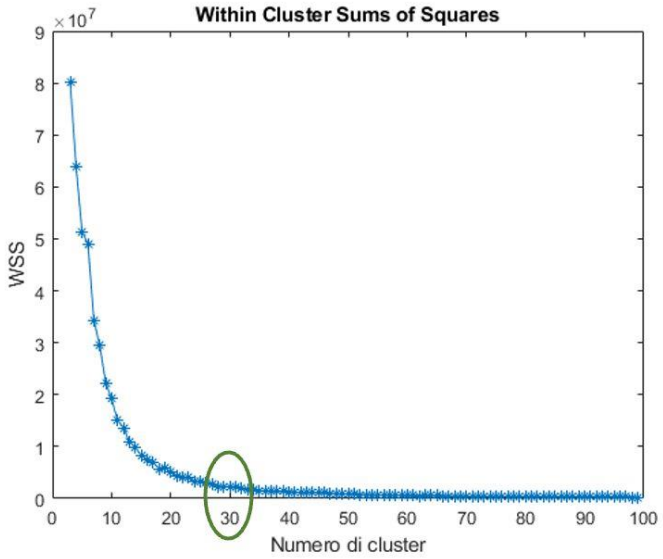

Figura 11. *Selezione del numero ottimale di cluster tramite la minimizzazione della WSS.*

Come evidenziato nella **Figura 11** si sceglie un numero di cluster pari a 30 poiché tale valore identifica l'inizio di una fase di plateau della curva.

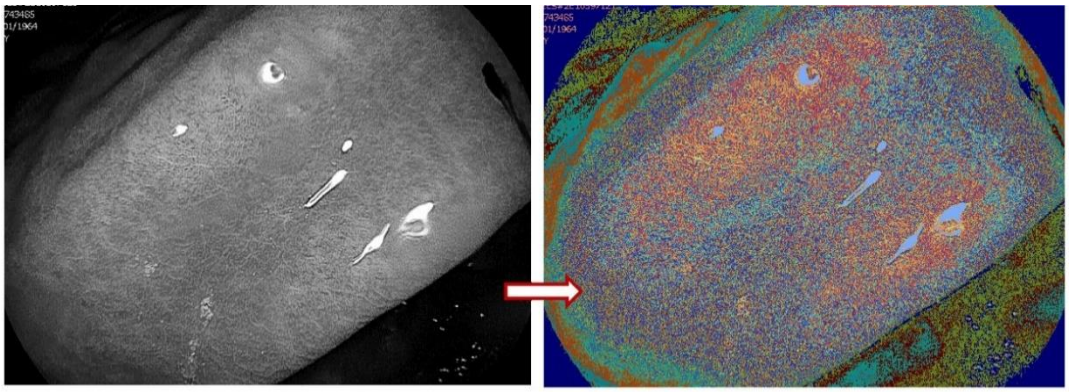

Figura 12. *Immagine in grayscale (a sinistra) con relativa suddivisione in cluster (a destra).*

La **Figura 12** rappresenta il risultato del clustering effettuato tramite algoritmo *K-Means* sull'immagine in scala di grigi.

Figura 13. *Esempio di maschera rappresentativo delle zone sovraesposte da rimuovere.*

Allo stesso tempo, si ricava una maschera binaria rappresentativa delle zone da rimuovere (**Figura 13**).

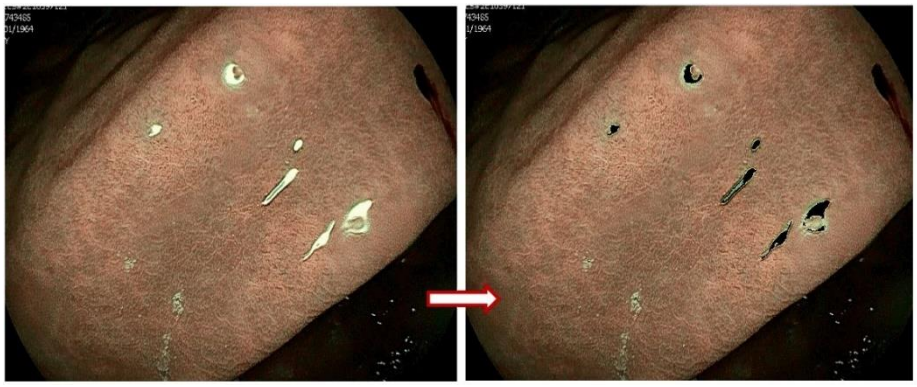

Figura 14. *Esempio di eliminazione dei riflessi di luce.*

Un esempio del risultato dato dalla rimozione delle zone a più elevata intensità media è fornito nella **Figura 14**.

c. Eliminazione di porzioni in secondo piano e/o non a fuoco.

Tipicamente, si nota che le zone di non interesse sono caratterizzate da tonalità inferiori. Per questo motivo, è stata applicata una soglia percentuale sulle intensità, al fine di individuare tali porzioni

dell'immagine. Il risultato è stato poi perfezionato tramite l'impiego di operatori morfologici.

In particolare, a partire dall'immagine ricavata dallo step a), si effettua la conversione in grayscale e si ricava una binarizzazione dell'immagine secondo una soglia pari al 18 % della massima intensità (valore selezionato a seguito di un tuning empirico del parametro). La maschera ottenuta evidenzia in bianco le zone a bassa intensità, da rimuovere. Al fine di rendere tale maschera più omogenea, si adoperano operazioni morfologiche di chiusura e di rimozione di pixel bianchi isolati.

Praticando l'operazione di unione tra le due maschere ricavate ai punti b e c, si ricava un'unica rappresentazione binaria che delimita tutte porzioni di non interesse. Applicati nuovamente operatori morfologici di chiusura con lo scopo di rendere la rappresentazione più omogenea, si procede alla rimozione di tali aree dall'immagine. Nella **Figura 16** è schematizzato un esempio di isolamento della sola porzione tissutale. In particolare, la terza immagine rappresentata

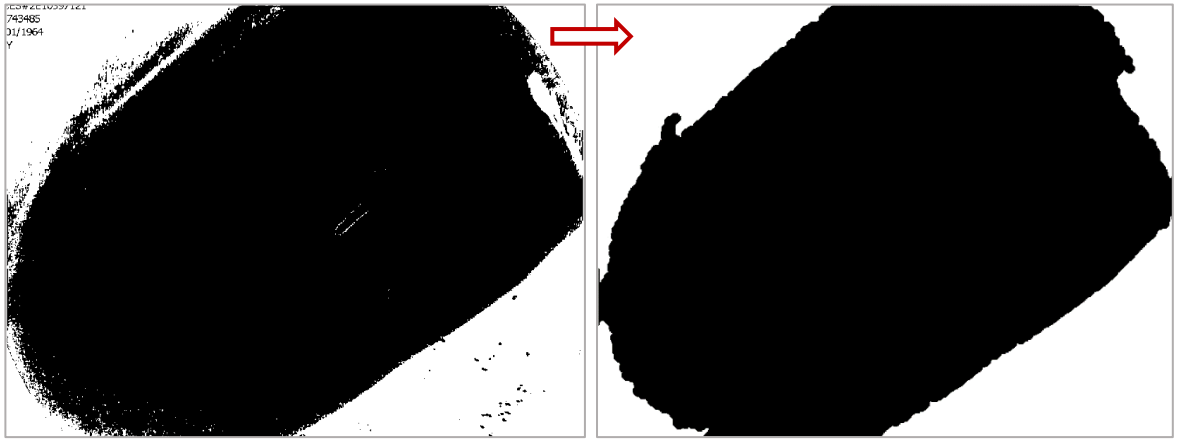

Figura 15. *Esempio di maschera delle zone a bassa intensità da rimuovere: prima (a sinistra) e dopo (a destra) l'applicazione di operatori morfologici.*

coincide con la versione definitiva dell'acquisizione la quale sarà processata ed analizzata per la costruzione del classificatore.

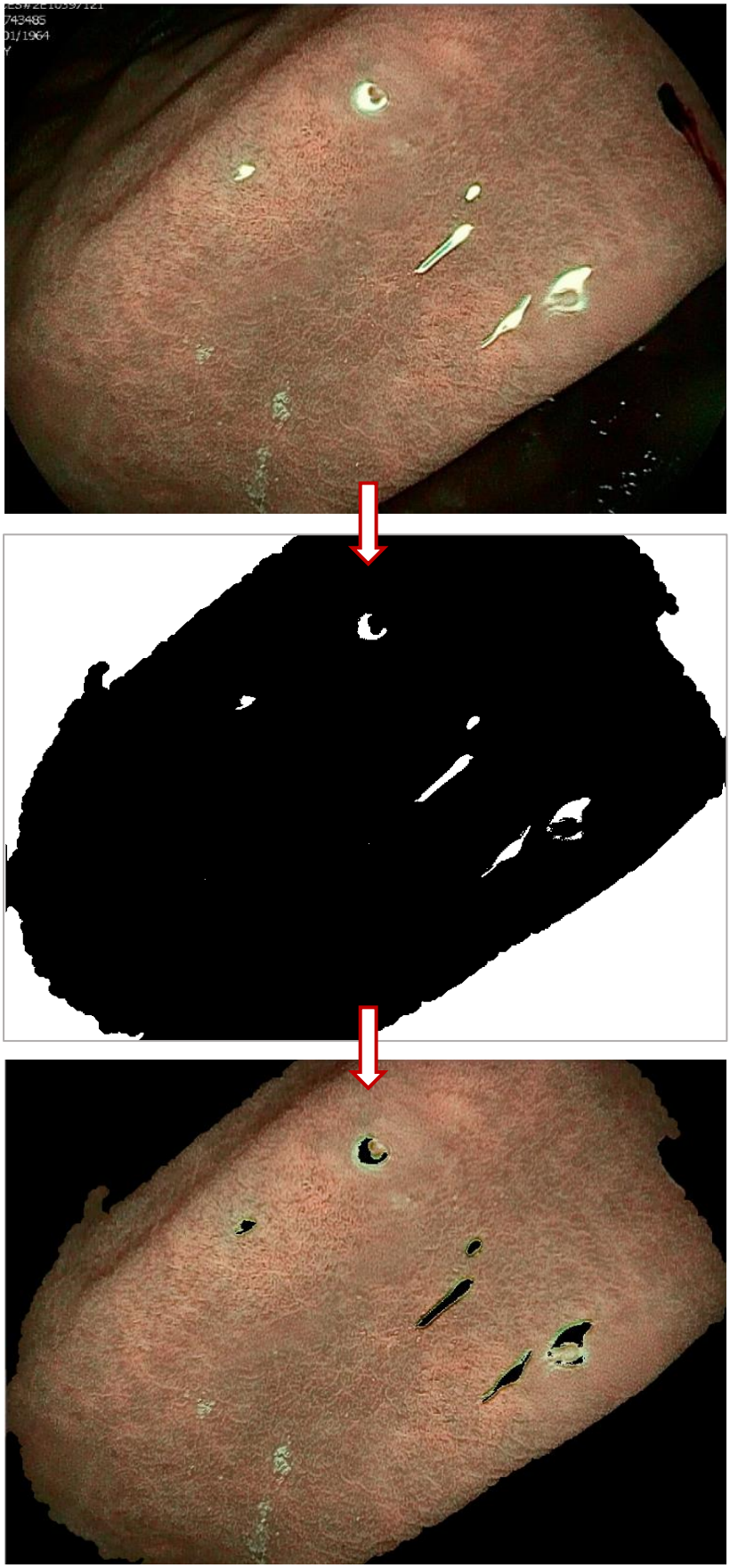

Figura 16. *Esempio di applicazione della maschera totale per la rimozione delle zone di non interesse.*

2.2.3. Equalizzazione dell'istogramma delle luminosità

Le immagini rappresentative della sola porzione tissutale di interesse vengono sottoposte ad equalizzazione dell'istogramma delle luminosità. La tecnica in questione consiste nel riassegnare i valori di intensità dei pixel per far sì che la distribuzione dell'intensità diventi uniforme. In questo modo si ottiene un incremento del contrasto complessivo dell'immagine, con maggiore effetto sulle aree a basso contrasto locale.

Il metodo di equalizzazione adottato nel corrente studio si basa sulla trasformazione integrale di probabilità che prevede il passaggio da una variabile casuale K avente una qualsiasi distribuzione continua ad un'altra variabile casuale S caratterizzata da distribuzione uniforme. In particolare, se la variabile casuale continua R ha funzione cumulata $T_K(k)$ monotona crescente allora la variabile casuale S data dalla cumulata stessa avrà una distribuzione $f_S(s)$ uniforme, pertanto la trasformazione in questione coinciderebbe con la funzione cumulata [29].

$$s = T_K(k) = \int_0^k f_K(w)\, dw$$

dove $f_K(k)$ è la funzione densità di probabilità di K.

Nello specifico caso, le tonalità di grigio possono assumere valori discreti nell'intervallo [0, 1]. Per questo motivo, il processo di equalizzazione viene approssimato e si traduce in una procedura di minimizzazione dello scarto con la distribuzione target (uniforme). La distribuzione discreta delle luminosità viene rappresentata attraverso un istogramma e l'obiettivo è quello di applicare una trasformazione alle intensità dei pixel per fare in modo che la nuova distribuzione si avvicini il più possibile ad un istogramma piatto. Nello specifico, si utilizza la

funzione integrata di *Matlab histeq* che, a partire dall'immagine di partenza, ne restituisce una versione equalizzata ottenuta attraverso la trasformazione T che minimizza la seguente espressione:

$$|c_1(T(k)) - c_0(k)|$$

con c_0 istogramma cumulativo dell'immagine originaria contenente valori di luminosità k e c_1 somma cumulativa dell'istogramma piatto desiderato [30].

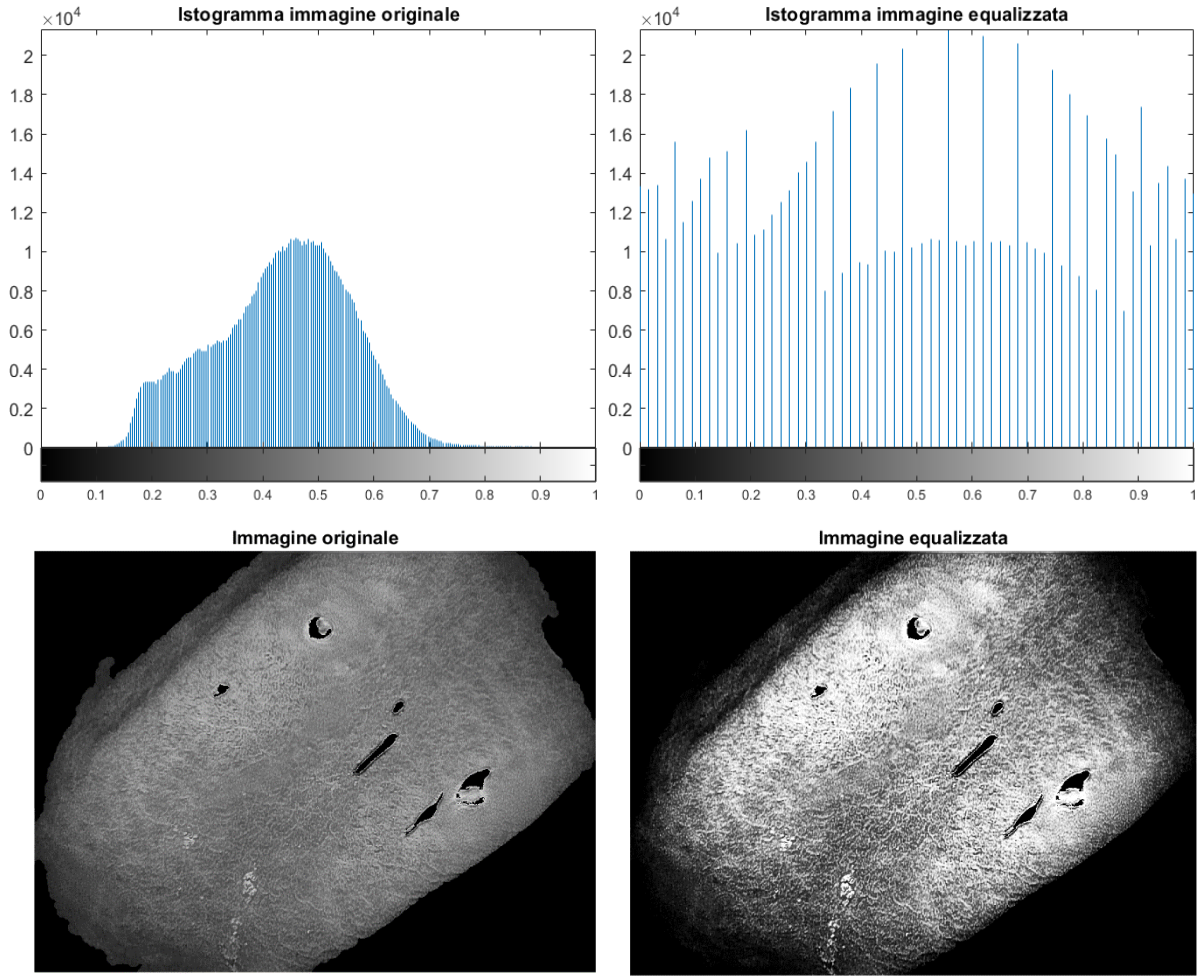

Figura 17. *Esempio di un processo di equalizzazione: istogrammi ed immagini.*

2.2.4. Filtraggio con filtro gaussiano

Il filtro gaussiano è un tipo di filtro passabasso che si applica nei primi step della catena di *processing* al fine di ridurre il rumore sovrapposto all'immagine, con lo

scopo di scongiurare cattivi funzionamenti nelle fasi successive dell'elaborazione. Si definisce un *kernel* (risposta all'impulso, h) con l'obiettivo di introdurre un'operazione di media che va ad agire localmente nelle diverse aree dell'immagine attraverso un processo di convoluzione bidimensionale. Tale kernel, nel caso del filtraggio gaussiano, è definito da tre parametri: le due dimensioni K ed L e l'apertura σ. Al fine di ottenere un kernel $h[k,l]$ la cui somma degli elementi sia unitaria, si scrive:

$$h_g[k,l] = e^{\frac{-(l^2+k^2)}{2\sigma^2}}$$

$$h[k,l] = \frac{h_g[k,l]}{\sum_K \sum_L h_g[k,l]}$$

con $k \in \left[-\frac{K-1}{2}, \frac{K-1}{2}\right]$ e $l \in \left[-\frac{L-1}{2}, \frac{L-1}{2}\right]$. Nel corrente studio, si utilizza la funzione integrata di *Matlab imgaussfilt* che richiede come unico input la deviazione standard σ, dalla quale si ricava la dimensione K imposta uguale ad L come segue:

$$K = L = 2\ ceil(2\sigma) + 1$$

dove $ceil$ opera un arrotondamento all'intero successivo. Osservando empiricamente l'effetto delle diverse aperture sul filtraggio delle immagini, si è scelto un valore di σ pari a 1 che corrisponde a $K = L = 5$ [31].

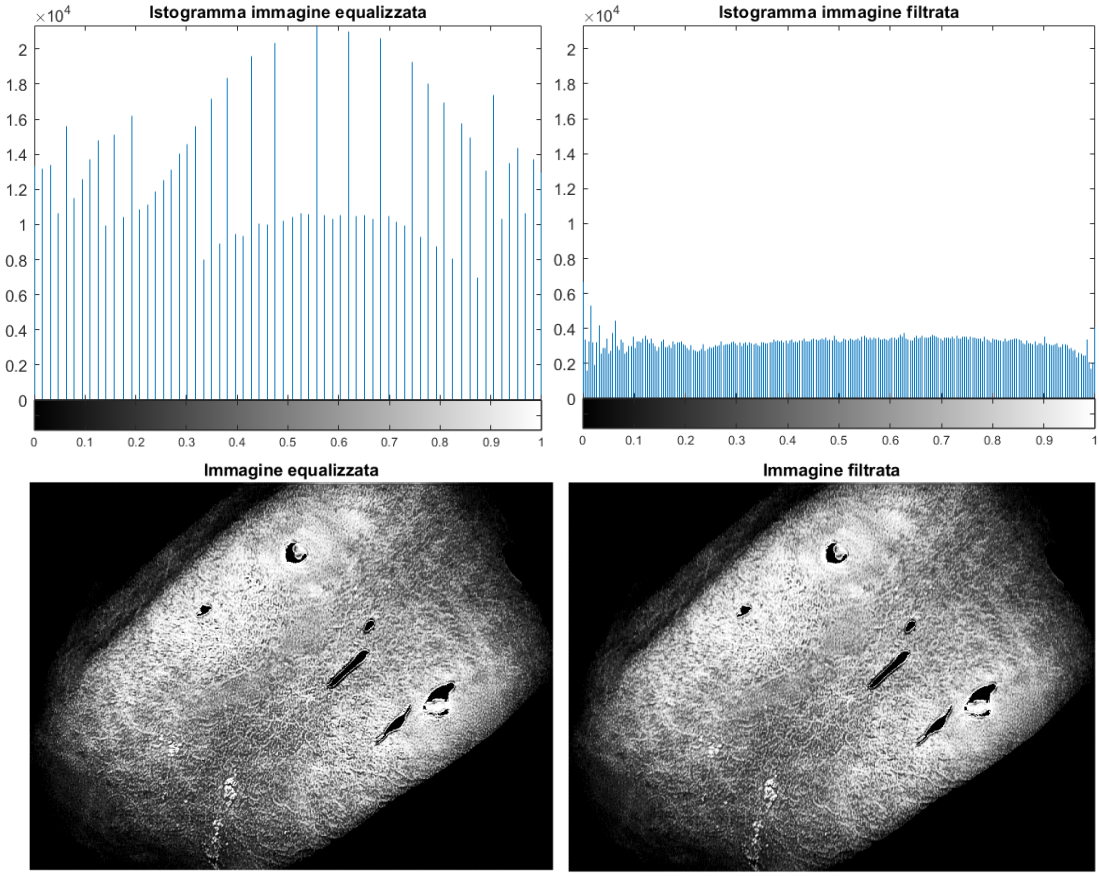

Figura 18. *Esempio di filtraggio gaussiano: istogrammi ed immagini.*

2.3. Estrazione delle feature

Ogni immagine è stata suddivisa in ROI (*Region of Interest*) quadrate 25x25 con un overlap di 10 pixel lungo la direzione orizzontale e lungo quella verticale. La scelta delle dimensioni della singola ROI è stata fatta attraverso un'analisi diretta della texture rappresentabile dalle diverse opzioni. Nella **Figura 19** si può notare come una ROI troppo piccola (5x5) non sia sufficientemente rappresentativa e, allo stesso tempo, dimensioni troppo elevate (50x50) risultano in una resa ridondante del pattern presente.

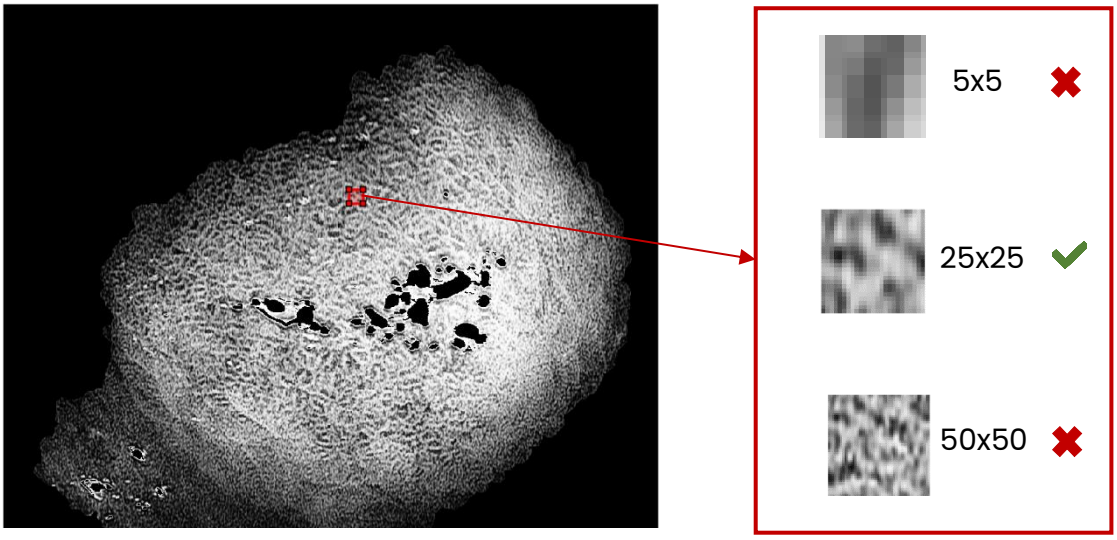

Figura 19. *Confronto tra ROI di diverse dimensioni.*

Per la caratterizzazione delle singole ROI, questo studio propone l'estrazione di 20 *feature*, distinguibili in 3 tipologie.

2.3.1. Feature dell'istogramma delle luminosità

L'istogramma delle luminosità rappresenta il numero di occorrenze, in pixel, per ogni livello di grigio di cui è composta l'immagine in formato gray-scale. In questo modo, si associa all'immagine una visione probabilistica in cui il valore del pixel diventa una variabile casuale la cui distribuzione di probabilità è data dall'istogramma delle luminosità.

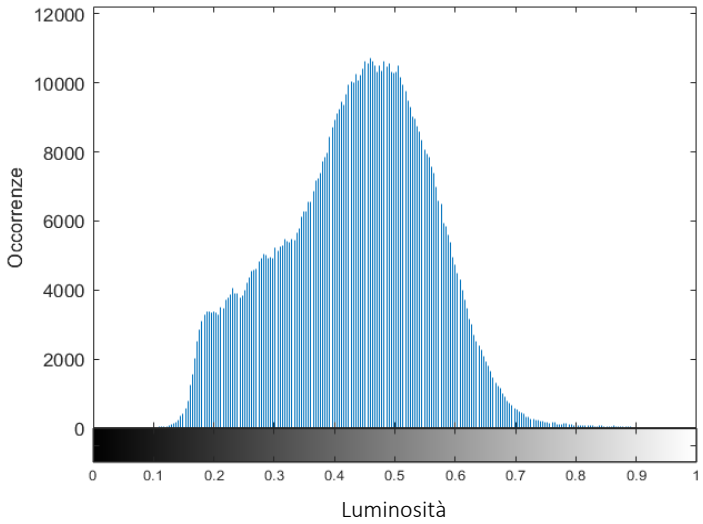

Figura 20. *Esempio di istogramma delle luminosità.*

Pertanto, è di notevole interesse estrarre dei parametri descrittivi di tale distribuzione. In questo studio, sono stati valutati i seguenti indicatori (**Tabella 4**):

- **Media**

  Misura dell'intensità complessiva della ROI. Si tratta di una media campionaria dei valori di intensità che i pixel assumono della regione di interesse.

- **Varianza**

  Misura di variabilità dei toni di grigio all'interno della ROI. Nello specifico, la varianza indica la media del quadrato dello scarto tra i valori dei singoli pixel rispetto ed il valore medio.

- **Skewness**

  Indice di asimmetria dei valori di grigio della ROI intorno alla media campionaria. Quando S_k assume valori negativi, i toni di grigio si diffondono maggiormente verso toni più bassi della media, viceversa, se S_k assume valori positivi, questi si diffondono per lo più verso toni più alti della media. Se la distribuzione delle intensità è simmetrica rispetto al valore medio, allora la skewness sarà pari a 0 [32].

- **Kurtosi**

  Misura della tendenza della distribuzione delle intensità ad assumere valori interpretabili come outlier. Maggiore è K_t, più alta sarà la probabilità di incorrere in toni anomali, lontani dalla media [33].

- **Entropia**

  Misura statistica del livello di casualità con cui può essere caratterizzata la texture della ROI. L'entropia è, quindi, un indicatore che rappresenta quanto l'alternarsi di determinati livelli di grigio all'interno della ROI sia dovuto al caso o meno [34].

Si consideri N il numero totale di pixel nella ROI, $I(x_i, y_i)$ il valore del pixel i-esimo e p_j la frequenza normalizzata dell'istogramma in corrispondenza del j-esimo livello di grigio. La **Tabella 4** riporta le formule implementate per l'estrazione delle *feature* sopra descritte.

Feature	***Formula***	***Descrizione***
Media	$m = \frac{1}{N}\sum_{i=1}^{N} I(x_i, y_i)$	*Intensità* media della ROI.
Varianza	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}[I(x_i, y_i) - m]^2$	Misura di *variabilità* dei livelli grigio all'interno della ROI.
Skewness	$S_k = \frac{1}{N}\frac{\sum_{i=1}^{N}[I(x_i, y_i) - m]^3}{\sigma^3}$	Indice di *asimmetria* della distribuzione di toni di grigio nella ROI.
Kurtosi	$K_t = \frac{1}{N}\frac{\sum_{i=1}^{N}[I(x_i, y_i) - m]^4}{\sigma^4}$	Misura della tendenza della distribuzione di toni di grigio della ROI a presentare *outlier*.
Entropia	$E = -\sum_{j=1}^{N\ bins} p_j\ log_2(p_j)$	Misura della *casualità* che caratterizza la texture della ROI.

Tabella 4. *Feature estratte dall'istogramma delle luminosità.*

2.3.2. Feature della GLCM

La *Gray-Level Co-Occurrence Matrix* (GLCM) [35] è una matrice quadrata NxN, dove N rappresenta il numero di livelli di grigio di cui è composta l'immagine. Si considerino i e j le coordinate della GLCM. Questa è definita per una direzione specifica e tiene conto di quante volte il tono di grigio i-esimo si trova vicino a quello j-esimo lungo la direzione selezionata, risultando in una matrice

simmetrica. Le *feature* estraibili da tale matrice rappresentano degli importanti descrittori di texture dell'immagine analizzata. Questo si può anche evincere dalla definizione stessa della GLCM che prevede una descrizione numerica dell'alternarsi dei toni di grigio.

Una volta ottenuta la GLCM, questa viene riassunta attraverso alcune sue *feature*.

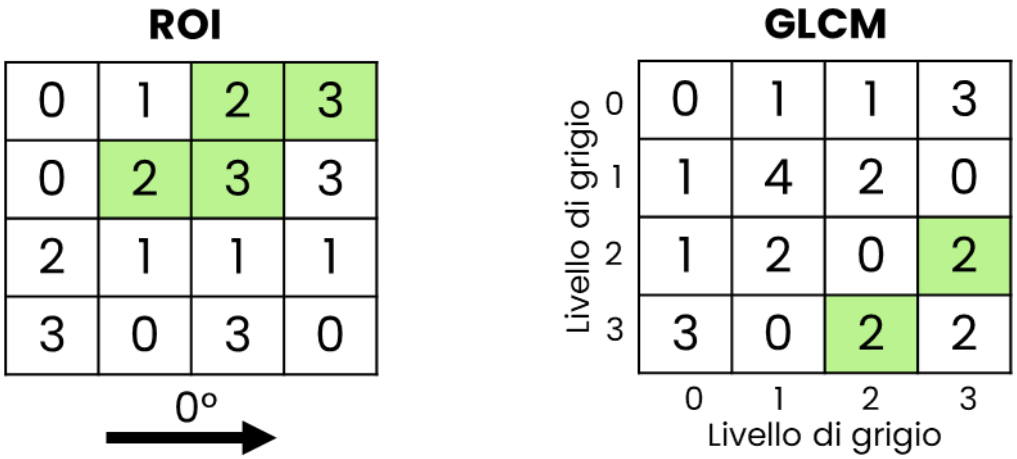

Figura 21. *GLCM (0°) di una ROI 4x4 con dinamica [0, 3].*

Le *feature* utilizzate in questo studio sono (**Tabella 5**):

- ***Contrasto***

  Misura del contrasto locale all'interno della ROI, ovvero tra il singolo pixel del il suo vicinato. Il contrasto indica quindi l'ammontare di variazioni locali all'interno della regione di interesse. Se la ROI è costituita da toni di grigio costanti, il contrasto sarà pari a 0.

- ***Correlazione***

  Misura quanto ogni valore di grigio assunto dal singolo pixel nella ROI sia correlato con quelli assunti dal suo vicinato. Indica l'eventuale dipendenza lineare tra pixel aventi diversi livelli di grigio e, quindi, se ci sono relazioni tra pattern di tonalità diverse. La correlazione è pari a 1 o -1 nel caso di ROI perfettamente correlate, positivamente o negativamente.

- ***Omogeneità*** (o Inverse Difference Moment)

  Misura la vicinanza della distribuzione degli elementi della GLCM alla diagonale principale. Può variare nell'intervallo [0, 1] e, maggiore è questo indicatore, maggiore sarà l'omogeneità dell'area. In particolare, un valore di omogeneità pari a 1 si riferisce ad una GLCM diagonale. Si dimostra che tale indicatore è inversamente proporzionale al contrasto.
- ***Energia*** (o Uniformità o Angular Second Moment)

  Misura l'uniformità della texture della singola ROI. In caso di periodicità all'interno dell'immagine, assume un valore più alto. Può variare nell'intervallo [0, 1] ed è uguale ad 1 se la ROI è costituita da toni di grigio costanti.

Considerando $p(i,j)$ i valori assunti dalla GLCM in corrispondenza dei toni di grigio i e j e μ il valor medio, si riportano nella **Tabella 5** le formule utilizzate per la quantificazione delle *feature* sopra descritte [35] [36].

Feature	***Formula***	***Descrizione***
Contrasto	$c = \sum_{i,j} \lvert i-j \rvert^2 \, p(i,j)$	*Misura delle variazioni locali nella matrice GLCM.*
Correlazione	$co = \sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j}$	Indice delle dipendenze lineari tra pixel con diverso livello di grigio.
Omogeneità	$Om = \sum_{i,j} \frac{p(i,j)}{1+\lvert i-j \rvert}$	Misura l'omogeneità locale della ROI.

Energia	$En = \sum_{i,j} p(i,j)^2$	Indice di eventuali periodicità all'interno della ROI.

Tabella 5. *Feature estratte dalla GLCM.*

2.3.3. Feature della GLRLM

La *Gray-Level Run-Length Matrix* (GLRLM) [37] è una matrice N_gxN_r, dove N_g rappresenta il numero di livelli di grigio di cui è composta l'immagine. La dimensione N_r, invece, è definita a seconda della direzione scelta per il calcolo della GLRLM e rappresenta la massima estensione dell'immagine lungo tale direzione. Infatti, considerate i e j le coordinate della GLRLM, tale matrice indica quante volte il tono di grigio j-esimo compare in una "run" di $n = i$ pixel. È importante ricordare che per "run" si intende una sequenza consecutiva di pixel costanti nella direzione prescelta.

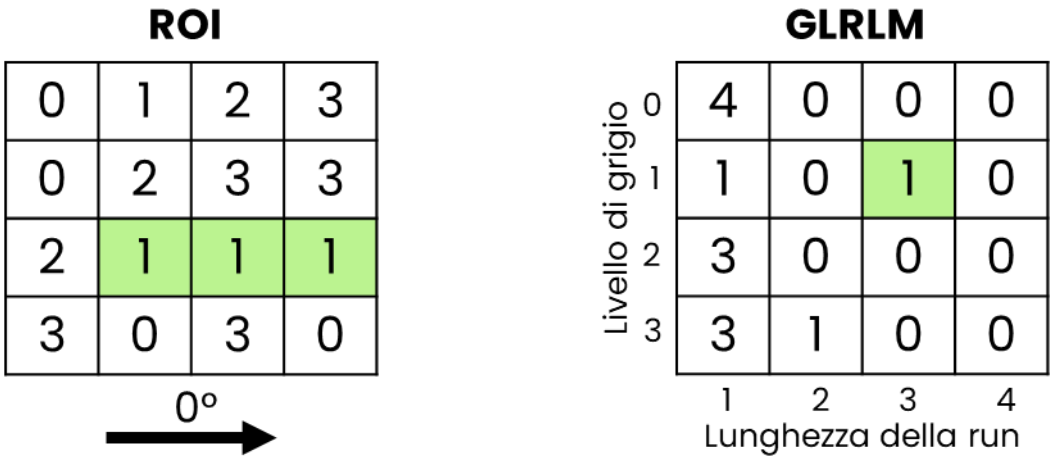

Figura 22. *GLRLM (0°) di una ROI 4x4 con dinamica [1, 3].*

Analogamente alla GLCM, anche la GLRLM fornisce una serie di *feature* molto utili al fine di descrivere le caratteristiche di texture delle immagini. In questo lavoro, sono state utilizzate le *feature* descritte di seguito:

- ***Short Run Emphasis*** (SRE)

  Misura la prevalenza di run corte all'interno della ROI. Pattern costituiti da piccole sequenze di toni di grigio che si intervallano fra loro sono caratterizzati da un valore di SRE più elevato.
- ***Long Run Emphasis*** (LRE)

  Valuta la prevalenza di run lunghe all'interno della ROI. La presenza di texture caratterizzate da lunghe sequenze di valori di grigio è associata ad un LRE maggiore.
- ***Gray-Level Non-uniformity*** (GLN)

  Indica quanto i valori di grigio nelle run sono uniformi. Essendo GLN una misura di non uniformità, ROI dalle tonalità uniformi corrisponderanno a GLN minori.
- ***Run-Length Non-uniformity*** (RLN)

  Indica quanto le lunghezze delle run sono uniformi. Trattandosi anche in questo caso di una misura di non uniformità, ROI dalle sequenze uniformi corrisponderanno a RLN minori.
- ***Run Percentage*** (RP)

  Indica l'omogeneità delle run presenti della regione analizzata. Relaziona il numero di run complessive al numero totale di pixel di cui è costituita la ROI.
- ***Low Gray-Level Run Emphasis*** (LGRE)

  Valuta la prevalenza di run contenenti basse tonalità di grigio all'interno della ROI. Pattern caratterizzati da abbondanza di sequenze di toni di grigio bassi avranno LGRE maggiore.

- ***High Gray-Level Run Emphasis*** (HGRE)

  Misura la prevalenza di run contenenti toni di grigio elevati all'interno della ROI. Profili di texture caratterizzati da sequenze di luminosità elevate avranno HGRE maggiore.

- ***Short-Run Low Gray-Level Emphasis*** (SRLGE)

  Valuta la prevalenza di run corte contenenti bassi livelli di grigio nella regione di interesse. Quando basse tonalità di grigio si susseguono prevalentemente in piccole sequenze di pixel, si ottiene un valore di SRLGE maggiore.

- ***Short-Run High Gray-Level Emphasis*** (SRHGE)

  Quantifica la prevalenza di run corte contenenti alti livelli di grigio all'interno della ROI. In presenza di numerosi pixel ad elevata luminosità che si alternano in sequenze ristrette di pixel, l'indicatore SRHGE assume valori più alti.

- ***Long-Run Low Gray-Level Emphasis*** (LRLGE)

  Valuta la prevalenza di sequenze lunghe costituite da bassi livelli di grigio. Se le basse luminosità sono numerose e si organizzano in run relativamente lunghe, l'LRLGE avrà valori più elevati.

- ***Long-Run High Gray-Level Emphasis*** (LRHGE)

  Misura la prevalenza di run lunghe costituite da elevati toni di grigio. Qualora siano presenti prevalentemente sequenze estese di pixel a luminosità elevata, si otterranno valori maggiori di LRHGE.

Considerando P_{ij} il valore assunto dalla GLRLM in corrispondenza del livello di grigio i e lunghezza di run j ed N il numero totale di pixel appartenenti alla ROI in analisi, si riportano nella **Tabella 6** le formule implementative delle *feature* sopra descritte [37].

Feature	*Formula*	*Descrizione*
SRE	$SRE = \frac{\sum_{i \in N_g} \sum_{j \in N_r} \frac{P_{ij}}{j^2}}{\sum_{i \in N_g} \sum_{j \in N_r} P_{ij}}$	*Prevalenza di run corte.*
LRE	$LRE = \frac{\sum_{i \in N_g} \sum_{j \in N_r} j^2 P_{ij}}{\sum_{i \in N_g} \sum_{j \in N_r} P_{ij}}$	Prevalenza di run lunghe.
GLN	$GLN = \frac{\sum_{i \in N_g} (\sum_{j \in N_r} P_{ij})^2}{\sum_{i \in N_g} \sum_{j \in N_r} P_{ij}}$	Indice di uniformità dei valori di grigio delle run.
RLN	$RLN = \frac{\sum_{j \in N_r} (\sum_{i \in N_g} P_{ij})^2}{\sum_{i \in N_g} \sum_{j \in N_r} P_{ij}}$	Indice di uniformità delle lunghezze delle run.
RP	$RP = \sum_{i \in N_g} \sum_{j \in N_r} \frac{P_{ij}}{N}$	Indice di omogeneità delle run.
LGRE	$LGRE = \frac{\sum_{i \in N_g} \sum_{j \in N_r} \frac{P_{ij}}{i^2}}{\sum_{i \in N_g} \sum_{j \in N_r} P_{ij}}$	Prevalenza di run a bassi livelli di grigio.
HGRE	$HGRE = \frac{\sum_{i \in N_g} \sum_{j \in N_r} i^2 P_{ij}}{\sum_{i \in N_g} \sum_{j \in N_r} P_{ij}}$	Prevalenza di run ad alti livelli di grigio.
SRLGE	$SRLGE = \frac{\sum_{i \in N_g} \sum_{j \in N_r} \frac{P_{ij}}{i^2 j^2}}{\sum_{i \in N_g} \sum_{j \in N_r} P_{ij}}$	Prevalenza di run ad alti livelli di grigio.
SRHGE	$SRHGE = \frac{\sum_{i \in N_g} \sum_{j \in N_r} \frac{i^2 P_{ij}}{j^2}}{\sum_{i \in N_g} \sum_{j \in N_r} P_{ij}}$	Prevalenza di run corte a bassi livelli di grigio.
LRLGE	$LRLGE = \frac{\sum_{i \in N_g} \sum_{j \in N_r} \frac{j^2 P_{ij}}{i^2}}{\sum_{i \in N_g} \sum_{j \in N_r} P_{ij}}$	Prevalenza di run lunghe a bassi livelli di grigio.
LRHGE	$LRHGE = \frac{\sum_{i \in N_g} \sum_{j \in N_r} i^2 j^2 P_{ij}}{\sum_{i \in N_g} \sum_{j \in N_r} P_{ij}}$	Prevalenza di run lunghe ad alti livelli di grigio.

Tabella 6. *Feature estratte dalla GLRLM.*

2.4. Definizione dei Construction Set

Le osservazioni contenute nel *dataset* coincidono con le singole ROI caratterizzate dalle 20 *feature* precedentemente descritte. A partire da tale insieme di dati, sono stati costruiti tre diversi tipi di *Construction Set* (CS) che differiscono nella rappresentazione della classe sana.

2.4.1. Construction Set 1

Il *Construction Set 1* (CS1) prevede l'assegnazione di due sole tipologie di etichette: sano o patologico (metaplastico). In questo caso, le ROI si considerano sane esclusivamente nel caso in cui siano estratte a partire da immagini raffiguranti tessuto completamente sano, con assenza di MGI. Sono state quindi escluse le osservazioni sane provenienti da immagini contenenti lesioni metaplastiche (**Figura 23**). Questa scelta mira a limitare l'introduzione nel CS di elementi che possano derivare da errori di segmentazione da parte del medico.

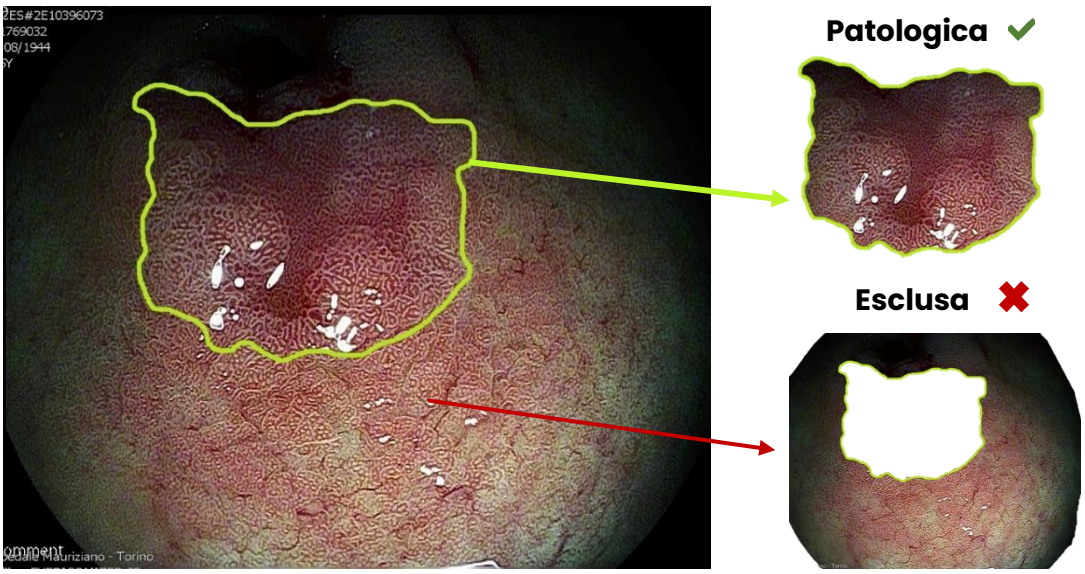

Figura 23. *Esempio di immagine patologica nel CS1: la parte patologica inclusa nella linea verde viene utilizzata nel dataset delle ROI patologiche, la parte esterna non viene usata.*

	# classi	Label assegnata	Immagini di provenienza
ROI patologiche	1	"**1**"	patologiche
ROI sane	1	"**0**"	sane

Tabella 7. *Costituzione del CS1.*

2.4.2. Construction Set 2

Il *Construction Set 2* (CS2) prevede una divisione del *dataset* in tre classi totali. Nello specifico, si considerano due diverse tipologie di ROI sane: quelle estratte da immagini sane e quelle provenienti dall'intorno di lesioni metaplastiche nelle immagini patologiche. Lo scopo di questa suddivisione è quello di poter utilizzare le osservazioni sane da immagini patologiche pur mantenendo una discriminazioni di tali ROI da quelle certamente sane.

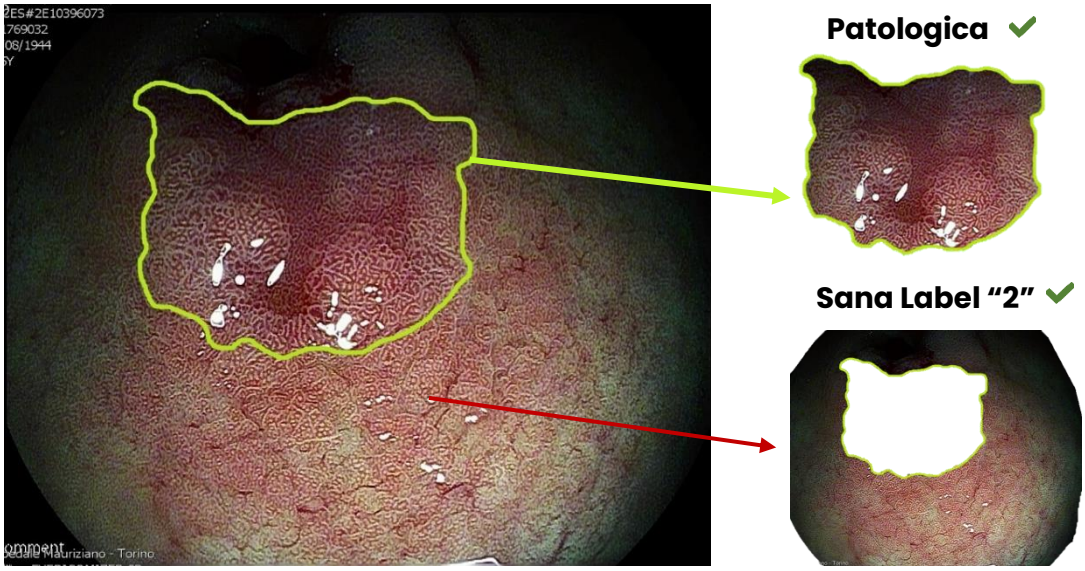

Figura 24. *Esempio di immagine patologica nel CS2: la parte patologica inclusa nella linea verde viene utilizzata nel dataset delle ROI patologiche, la parte esterna viene inclusa in una seconda classe sana con label "2".*

	# classi	Label assegnata	Immagini di provenienza
ROI patologiche	1	"**1**"	patologiche
ROI sane	2	"**0**"	sane
		"**2**"	patologiche

Tabella 8. *Costituzione del CS2.*

2.4.3. Construction Set 3

Il Construction Set 3 (CS3) prevede un'ulteriore suddivisione della classe sana. Infatti, le ROI sane estratte da immagini patologiche sono sottoposte ad un processo di clustering mediante *K-Means* sulle 20 *feature* estratte con k imposto pari a 2. Tale procedura si impiega con l'ipotesi di poter distinguere ROI effettivamente sane da quelle etichettate erroneamente come sane nella segmentazione manuale effettuata dall'esperto.

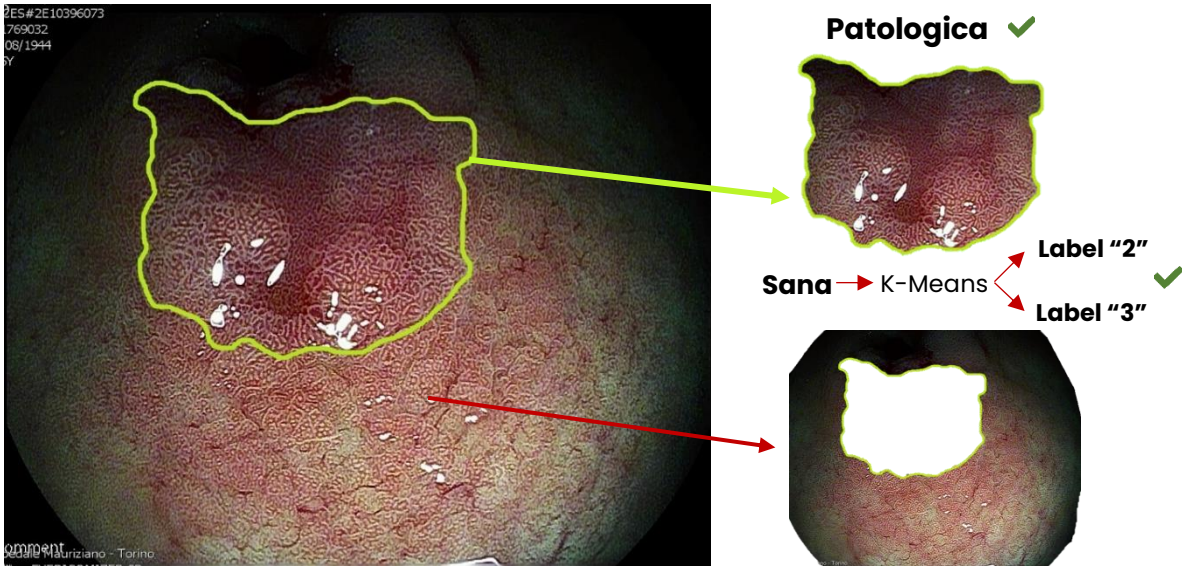

Figura 25. *Esempio di immagine patologica nel CS3: la parte patologica inclusa nella linea verde viene utilizzata nel dataset delle ROI patologiche, le ROI della parte esterna vengono smistate attraverso K-Means in due classi di ROI sane con label "2" e "3".*

	# classi	*Label assegnata*	*Immagini di provenienza*
ROI patologiche	1	**"1"**	patologiche
ROI sane	3	**"0"**	sane
		"2"	patologiche
		"3"	patologiche

Tabella 9. *Costituzione del CS3.*

2.4.4. Definizione di Training Set e Test Set

Le tre tipologie di Construction Set si utilizzano su due diversi tipi di dati: quelli estratti da immagini sottoposte agli step di equalizzazione e filtraggio (Vedi 2.2.3 e 2.2.4) e da immagini su cui queste due fasi di *pre-processing* sono state saltate. In questo modo, si definiscono sei configurazioni per la costruzione di sei classificatori Random Forest (RF). Nella **Tabella 10** se ne elencano denominazioni e caratteristiche.

Denominazione	***Equalizzazione** e **filtraggio***	*Tipologia di **Construction Set***
RF1	**No**	CS**1**
RF2	**No**	CS**2**
RF3	**No**	CS**3**
RF4	**Si**	CS**1**
RF5	**Si**	CS**2**
RF6	**Si**	CS**3**

***Tabella 10**. Descrizione delle sei configurazioni RF.*

A partire da queste configurazioni, gli elementi sono stati divisi in Training e Test Set bilanciati, ovvero in cui ogni classe è rappresentata dallo stesso numero di osservazioni. Il metodo utilizzato si propone di incrementare la rappresentatività dei due set attraverso un'estrazione casuale degli elementi delle singole classi precedentemente sottoposte a processo di clustering. Nello specifico, ogni classe viene divisa in cluster attraverso il metodo del *K-Means*, scegliendo il numero k di cluster ottimale per ogni classe attraverso la minimizzazione della variabilità intra-cluster ($varIntra$) e la massimizzazione della distanza inter-cluster ($distInter$). Si è utilizzata la funzione *kmeans* di *Matlab* impostando un numero di iterazioni pari a 1000 e di ripetizioni pari a 10 per ogni k, definito nell'intervallo $[2, 20]$. Per ognuna delle 10 ripetizioni sono stati calcolati i seguenti indicatori:

$$varIntra = \max\left(\frac{1}{N_i}\sum_{ROI \in cl_i} dist(ROI, c_i)\right)$$

$$distInter = min\big(dist(c_i, c_j)\big)$$

La varianza intra-cluster rappresenta la media delle distanze euclidee dei singoli elementi (ROI) dal centroide (c_j) all'interno del j-esimo cluster ed è, quindi, indice della compattezza dei cluster. Se ne prende il valore massimo per un approccio più conservativo. La distanza inter-cluster, invece, è indicativa della bontà di discriminazione del clustering e rappresenta la minima distanza tra coppie di cluster in termini di distanza fra centroidi. Si vuole minimizzare la varianza intra-cluster e massimizzare la distanza inter-cluster. Per una maggior stabilità dei risultati, i due indicatori vengono mediati sulle dieci ripetizioni per ogni valore di k. Per quanto riguarda le ROI patologiche e quelle sane di *label* "0", il clustering viene eseguito sui dati da immagini equalizzate e filtrate (configurazioni RF 1-3) e su quelli da immagini non pre-processate (configurazioni RF 4-6), con scelta del numero di cluster schematizzata nella **Figura 26** e nella **Figura 27**.

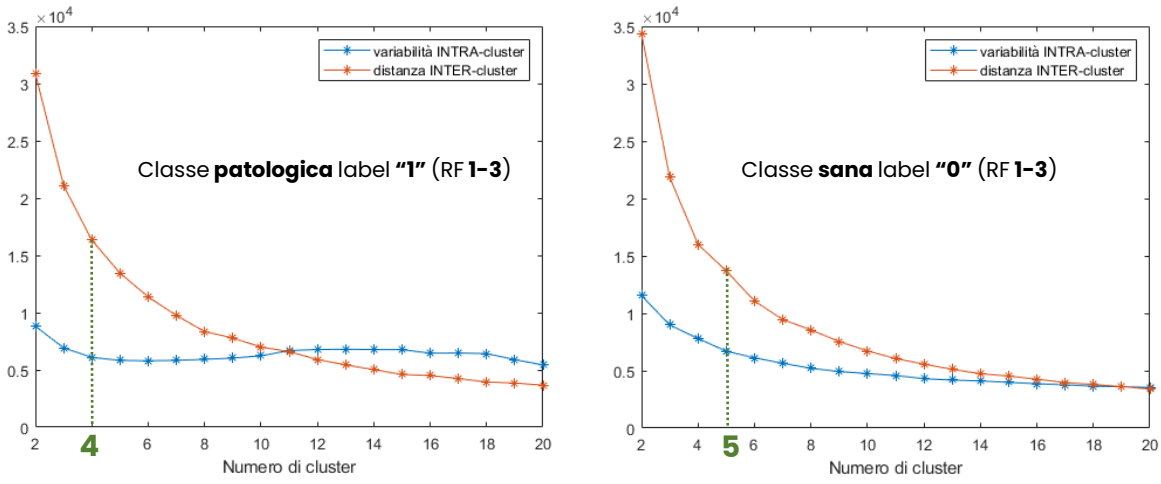

Figura 26. *Scelta numero di cluster per la classe patologica (a sinistra) e sana di label "0" (a destra) nelle configurazioni RF1, RF2 ed RF3.*

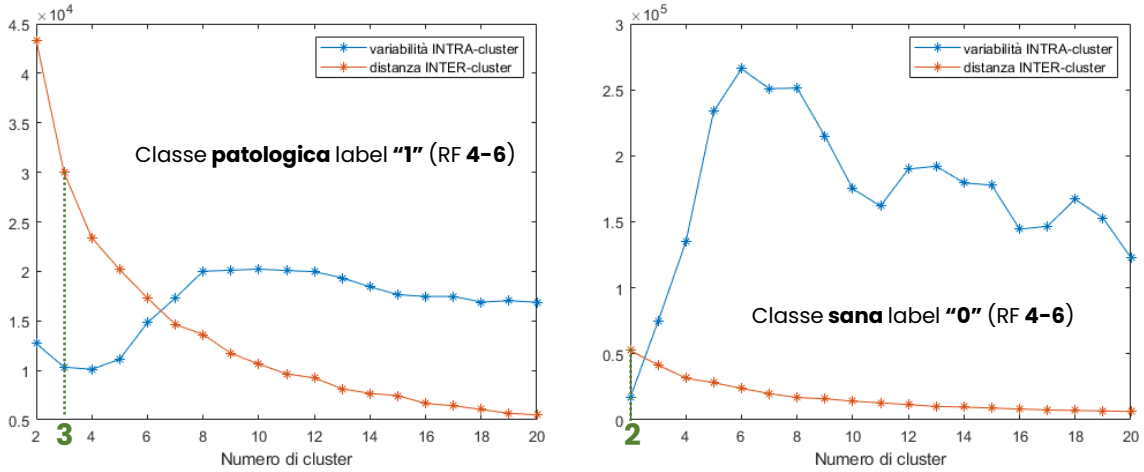

Figura 27. *Scelta numero di cluster classe patologica (a sinistra) e classe sana di label "0" (a destra) nelle configurazioni RF4, RF5, RF6.*

Le ROI sane derivanti da immagini patologiche assumono *label* diversa a seconda della tipologia di CS selezionata, pertanto diverse procedure di clustering sono state effettuate per caratterizzare le configurazioni RF2 ed RF3 (Vedi **Figura 28** e **Figura 29**).

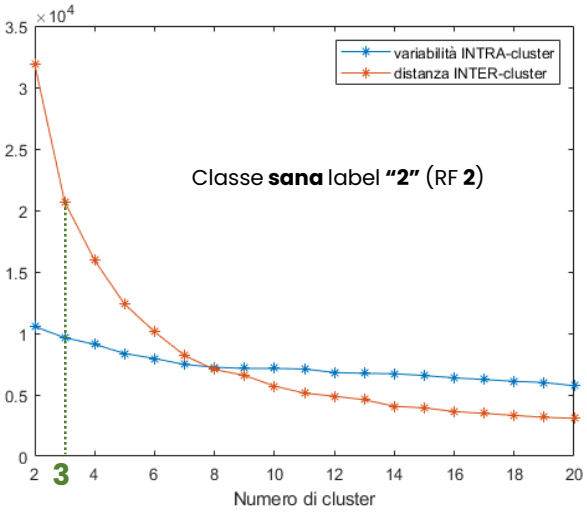

Figura 28. *Scelta numero di cluster per la classe sana di label "2" nella configurazione RF2.*

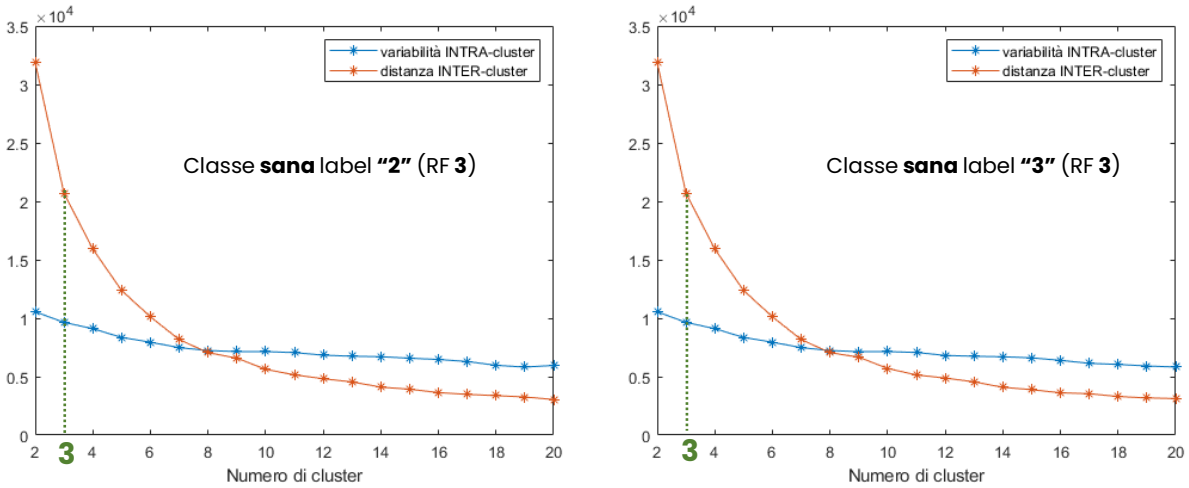

Figura 29. *Scelta numero di cluster per la classe sana di label "2" e "3" nella configurazione RF3.*

La stessa questione dovrebbe essere estesa alle configurazioni RF5 ed RF6 in cui si aggiunge solo la fase di *pre-processing* sulle immagini. Tuttavia, si è notato che in questi due casi il clustering sulle ROI sane da immagini patologiche portava alla definizione di gruppi a numerosità troppo bassa. La **Tabella 11** riporta il numero di cluster selezionati per ogni classe delle singole configurazioni RF.

	Numero di cluster			
	Classe **patologica**	Classe **sana "0"**	Classe **sana "2"**	Classe **sana "3"**
RF1	4	5	-	-
RF2			3	-
RF3			3	3
RF4	3	2	-	-
RF5			1	-
RF6			1	1

Tabella 11*. Numero di cluster rappresentativi delle singole classi nelle configurazioni RF.*

Per ogni configurazione, al fine di costruire il Training Set (TRS), si estrae in modo casuale un numero fissato ed uguale di elementi a partire da ogni cluster. Tale valore si fa corrispondere al 70% della minima numerosità presente tra i cluster della configurazione. Il restante 30% corrisponde al numero di osservazioni estratte a partire dagli elementi residui nei cluster per la costruzione del Test Set (TS) (Vedi **Figura 30**).

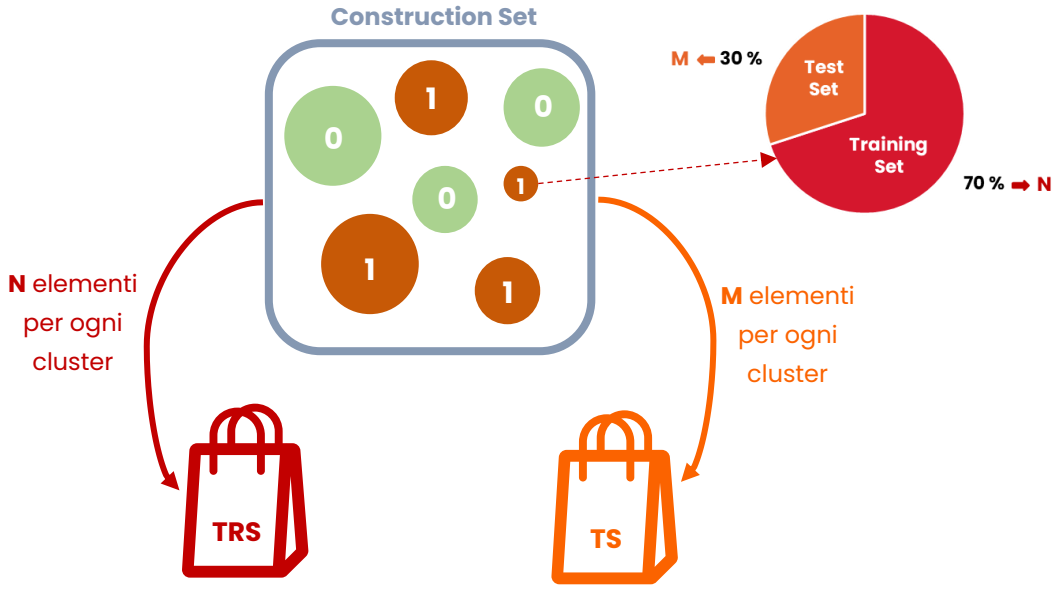

Figura 30. *Esempio di impostazione del numero di elementi da estrarre da ogni cluster per la costruzione del TRS (N) e del TS (M) a partire dal cluster di minore numerosità. I cerchi rappresentano i cluster e la relativa estensione ne indica la numerosità.*

Le *feature* sono state poi normalizzate mediante *"Min-Max Scaling"* rispetto alle osservazioni contenute nel Training Set. Considerando x l'insieme dei valori originali di una *feature*, la *feature* normalizzata x' si ottiene come segue:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Pertanto, minimo e massimo di ogni *feature* relativi al Training Set vengono salvati come parametri di normalizzazione, al fine di essere riutilizzati su Test Set e nuovi input.

2.5. Costruzione delle feature

Una volta ottenuti i diversi set di dati, si utilizza un metodo per ridurne la dimensionalità. In questo caso, la procedura utilizzata è stata di costruzione delle *feature* (*feature construction, FC*) che, a differenza di tecniche di selezione delle *feature* (*feature selection, FS)*, genera un set di *feature* completamente nuove a partire da quelle originali. A tale scopo e con l'intento di rimuovere la dipendenza mutua tra le variabili considerate, è stata utilizzata la *Principal Component Analysis* (PCA). La PCA è una tecnica impiegata nel campo della statistica multivariata al fine di semplificare i dati, sviluppata nel 1901 da Karl Pearson. Il fine ultimo è quello di generare un nuovo set di variabili, tra di loro non correlate, chiamate componenti principali (*principal components*, PC). Ogni PC è una combinazione lineare delle variabili originali ed è ortogonale alle altre, pertanto, non vi sono informazioni ridondanti. In questo modo, viene costituita una base ortogonale per la rappresentazione dello spazio delle osservazioni. L'obiettivo della PCA è spiegare il massimo quantitativo di varianza complessiva con il minor numero di componenti principali. Si tratta, quindi, di una trasformazione lineare delle variabili in uno spazio dimensionalmente inferiore che contiene il massimo quantitativo di informazione sulle stesse. Nello specifico, le PC corrispondono agli autovettori che definiscono il nuovo spazio ai quali vengono associati degli autovalori che ne indicano l'apporto singolo di varianza spiegata [38]. Nel corrente studio, la PCA è stata implementata nel contesto della *Classification Learner App* di *Matlab*. Per selezionare il numero ottimale di PC è stata imposta una soglia di minima varianza spiegata pari al 95%. La **Tabella 12** riporta il numero di componenti principali selezionate per ogni configurazione, la varianza spiegata da ognuna di esse e la varianza complessiva rappresentata. Si nota che in tutti i casi questa è maggiore del 95%.

	# di PC	***Varianza** spiegata per **componente***	***Varianza** spiegata **complessiva***
RF1	4	66,5% - 21,8% - 6,2% - 2,3%	96,8%
RF2	4	71,6% - 17,4% - 5,9% - 2,2%	97,1%
RF3	3	70,1% - 17,6% - 7,8%	95,5%
RF4	4	71,6% - 16,7% - 5,6% - 3,9%	97.8%
RF5	3	76,2% - 14,9% - 4,2%	95,3%
RF6	3	75,8% - 15,2% - 4,2%	95,2%

Tabella 12. *Risultati della PCA per tutte le configurazioni RF.*

2.6. Allenamento dei classificatori

L'allenamento dei diversi modelli è stato svolto nell'ambiente *Classification Learner App* di *Matlab*. In questo studio, i classificatori utilizzati sono le foreste di alberi decisionali (*Random Forest*, RF). Le RF sono metodi *ensemble* costituite da semplici alberi decisionali allenati su differenti sottoinsiemi del Training Set. A differenza degli alberi decisionali, la RF offre una migliore gestione del problema *dell'overfitting* sui dati di *training* e, per questo motivo, consente di ottenere delle prestazioni migliori nella classificazione. Solitamente, maggiore è il numero di alberi utilizzati, migliori sono le prestazioni della RF, a discapito di un elevato costo computazionale. Così come nell'impiego degli alberi decisionali semplici, ci sono diversi iperparametri che influenzano il comportamento delle RF. Ad esempio, l'iperparametro più importante per questo tipo di classificatore è senza dubbio il numero di alberi di cui è costituita la foresta. La lista completa degli iperparametri con i relativi intervalli di ricerca è riportata nella **Tabella 13** [39].

Iperparametro	***Spazio** di **ricerca***
***Algoritmo** Ensemble*	AdaBoost, RUSBoost, LogitBoost, GentleBoost and Bag
*Massimo numero di **partizioni***	[1, max(2, n-1)] n = numero di osservazioni
*Numero di **alberi***	[10-500]
Learning rate	[0,001-1]
*Numero di **predittori** da campionare*	[1, max(2, p)] p = numero delle variabili predittive

Tabella 13. *Spazi di ricerca degli iperparametri.*

Per selezionare i valori ottimali da assegnare, dato l'elevato costo computazionale nel testare diversi alberi simultaneamente, si è deciso di usare il metodo dell'ottimizzazione bayesiana al posto del metodo più diffuso della ricerca a griglia ("*grid search*"). Al contrario della *grid search*, il metodo utilizzato

consente di campionare solo un certo numero di valori per ogni iperparametro, estratti dallo spazio di ricerca definito nella **Tabella 13**. Questo metodo, pertanto, riduce il tempo computazionale, ma esegue una ricerca meno esaustiva per tutte le possibili combinazioni di iperparametri.

L'algoritmo di ottimizzazione bayesiana, implementato nella funzione *bayesopt* di *Matlab* [40], tenta di minimizzare una funzione obiettivo scalare $f(x)$, con x iperparametro appartenente ad un dominio limitato. La funzione può essere deterministica o stocastica e x può essere un insieme di valori reali continui, interi o categorici. Nel ricercare il punto ottimale nello spazio degli iperparametri, si ottimizza una *funzione di acquisizione* $a(x)$ con la quale si sceglie il punto successivo da esplorare fino al raggiungimento di una condizione di terminazione che, nel corrente studio coincide con un numero di iterazioni pari a 30. La funzione di acquisizione utilizzata in questo studio va sotto la dicitura "*probability of improvement*". Questa funzione valuta la "bontà" di un punto x basandosi sulla funzione di distribuzione di probabilità a posteriori Q. Tale distribuzione viene aggiornata ad ogni punto esplorato poiché viene aggiornato il modello gaussiano con cui si modella la funzione obiettivo la quale cambia valore punto per punto.

In primo luogo si valuta x_{best} come la posizione della media posteriore più bassa $(\mu_Q(x_{best}))$ Successivamente, si calcola la probabilità PI che un nuovo punto x porti ad un valore migliorativo della funzione obiettivo, con un margine m.

$$PI(x, Q) = P_Q(f(x) < \mu_{Q(x_{best})} - m)$$

Secondo la definizione standard di *Matlab*, m coincide con la deviazione standard del rumore stimata [41]. Pertanto, un punto nello spazio degli iperparametri x viene selezionato come il successivo da esplorare se massimizza la probabilità di incremento nelle performance, ovvero nel caso in cui,

selezionando il punto x, si potrebbe riscontrare una minimizzazione significativa della funzione obiettivo $f(x)$. In questo contesto, un generico punto x rappresenta una combinazione di iperparametri e la funzione obiettivo è la funzione di perdita, ovvero quella che restituisce l'errore di classificazione calcolato come segue [42]:

$$Err = \sum_{j=1}^{N} w_j I\{\hat{y}_j \neq y_j\}$$

dove w_j è il peso della j-esima osservazione, $\hat{y}_j$ e y_j sono classe predetta e classe effettiva della j-esima osservazione e I è la funzione indicatrice definita in questo modo:

$$I\{A\} = \begin{cases} 1, & se\ A\ vero \\ 0, & se\ A\ falso \end{cases}$$

Il software computa i pesi delle osservazioni in modo da ottenere una somma sul totale delle osservazioni pari a 1:

$$\sum_{j=1}^{N} w_j = 1$$

Questa procedura di ottimizzazione è stata eseguita secondo un processo di *cross-validazione* K*-fold*. Con questo metodo, il Training set viene diviso in K sottogruppi di uguale numerosità. In questo modo, $K-1$ sottogruppi vengono utilizzati per allenare il modello, mentre il sottogruppo rimanente rappresenta il Validation Set (VS). La funzione principale del VS è valutare le prestazioni di una specifica combinazione degli iperparametri associati al modello allenato sui $K-1$ *fold* rimanenti. Questa procedura viene ripetuta K volte in modo che tutti i sottogruppi abbiano avuto il ruolo di Validation Set e abbiano fatto parte del Training Set per almeno una volta. In questo modo, le prestazioni di ogni

combinazione di iperparametri esplorata sono valutate su K diversi sottogruppi del Training Set per poi stimarne l'andamento medio.

Diversi studi suggeriscono che la varianza della predizione è minore nel caso in cui si selezioni un numero di *fold* $K = 5$ o $K = 10$ [43] [44]. Nel lavoro corrente, è stato utilizzato un numero di sottoinsiemi K pari a 5.

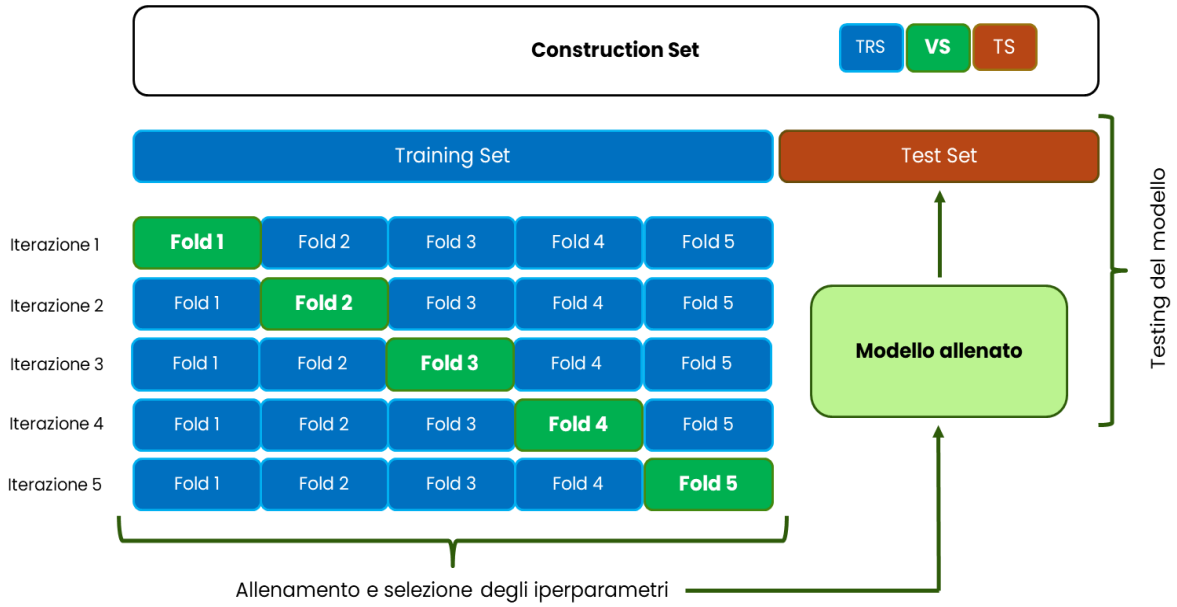

Figura 31. *Schema esplicativo di un processo di cross-validazione 5-fold.*

In questo modo, si ottengono le combinazioni ottimali di iperparametri per ogni configurazione (Vedi **Tabella 14**).

	Algoritmo	*# max* ***partizioni***	***# alberi***	***Learning rate***	***# predittori***
RF1	AdaBoost	676	491	0,017656	-
RF2	Bag	553	443	-	4
RF3	AdaBoost	149	40	0,95986	-
RF4	LogitBoost	73	107	0,083627	-
RF5	Bag	403	316	0,47589	2
RF6	AdaBoost	85	497	0,64845	-

Tabella 14. *Iperparametri selezionati per ogni configurazione RF.*

La foresta di alberi può essere costruita secondo il modello "*bagging*" o "*boosting*". L'algoritmo *Bag* prevede che il Training Set sia suddiviso in sottogruppi con possibilità di ripetizione degli elementi. Ogni sottoinsieme è dato in pasto a tutti gli alberi decisionali che compongono la foresta; questa fase viene detta di "*bootstrapping*". Per ogni nodo decisionale all'interno dei singoli alberi, si seleziona un sottoinsieme di *feature* per poi suddividere il singolo nodo usando le *feature* che forniscono il maggior guadagno in termini di informazione. Una volta che tutti gli alberi nella foresta hanno predetto la classe secondo le *feature* nel sottoinsieme di training, la predizione di ogni albero viene aggregata per assegnare la label della classe usando la tecnica del *majority voting* (fase denominata "*aggregating*"). Pertanto, possiamo notare che la predizione dei diversi alberi viene eseguita in parallelo ed ognuno di essi ha uguale voce in capitolo nella classificazione finale.

Gli algoritmi *AdaBoost* e *LogitBoost*, invece, sono due versioni del metodo del *boosting*. Il *boosting* lavora sequenzialmente applicando un algoritmo di classificazione a diverse versioni ripesate del Training set, per poi considerare il majority voting pesato della sequenza di classificatori. L'algoritmo *AdaBoost* prevede l'impiego di alberi decisionali composti da un solo nodo ed ognuno di essi può avere più o meno voce in capitolo nella classificazione. Inoltre, ogni albero prende in considerazione gli errori compiuti dall'albero precedente, pertanto l'ordine in cui questi vengono interrogati è importante.

Considerata la funzione di perdita quella funzione che valuta la distanza tra la predizione fornita da quella aspettata, si possono fare le seguenti considerazioni. La funzione di perdita di *AdaBoost* cambia esponenzialmente con l'errore di classificazione, pertanto, l'algoritmo risulta vulnerabile durante la gestione di dati affetti da rumore. Per ovviare a tale problema, talvolta, si preferisce utilizzare

l'algoritmo *LogitBoost*, introdotto da Friedman et al. [45], il quale presenta una funzione di perdita che varia linearmente rispetto all'errore di classificazione, risultando in una ridotta sensibilità al rumore e ad eventuali valori anomali. Con questo algoritmo, Friedman et al. introducono una visione statistica del boosting che viene visto come un'approssimazione di un modello additivo su scala logistica [45] [46].

Ognuno degli algoritmi *ensemble* esplorati è stato valutato in concomitanza con diverse combinazione dei relativi iperparametri. Nello specifico, il numero massimo di partizioni sta ad indicare un limite superiore sulla numerosità di punti di diramazione. Questo perché la presenza di un numero eccessivo di diramazioni può accentuare il fenomeno dell'*overfitting*, mentre alberi meno ramificati risultano più robusti e di più facile interpretazione.

Il numero di alberi si seleziona nel tentativo di cercare un compromesso tra accuratezza e tempo computazionale. Un maggior numero di alberi porta, tipicamente, ad un'accuratezza maggiore ma, allo stesso tempo, comporta un costo computazionale più elevato. Al fine di ottenere un buon potere predittivo, solitamente si seleziona un numero di alberi pari a qualche centinaio.

A seconda che il metodo ensemble esplorato sia di tipo *boosting* (*AdaBoost* e *LogitBoost*) o *bagging* (*Bag*), si ottimizzano due diversi iperparametri: il *learning rate* nel primo caso e il numero di predittori da campionare nel secondo caso.

Per numero di predittori da campionare si intende la numerosità delle *feature* da selezionare in modo casuale per ogni diramazione degli alberi decisionali. Il learning rate, invece, può essere selezionato in un range che va da $0{,}001$ a 1. Tanto minore di 1 è il learning rate, tante più iterazioni di apprendimento sono necessarie e questo comporta, spesso, un incremento nell'accuratezza. Una valore usuale per questo iperparametro è di circa $0{,}1$ [47].

Capitolo 3.
Risultati e post-elaborazione

Una volta allenati i sei modelli RF, se ne valutano le prestazioni su Training e Test Set. Successivamente, si passa da una classificazione a livello ROI ad una a livello pixel tramite la tecnica del *majority voting* (MV) e si ricalcolano le performance sull'intero Construction Set. Sulla base di queste, sono stati selezionati i due modelli RF migliori e se ne sono confrontate le prestazioni con e senza l'utilizzo della Principal Component Analysis (PCA). Scelto il classificatore migliore, sono state effettuate delle procedure di post-elaborazione per ottimizzarne le prestazioni nella classificazione delle intere immagini.

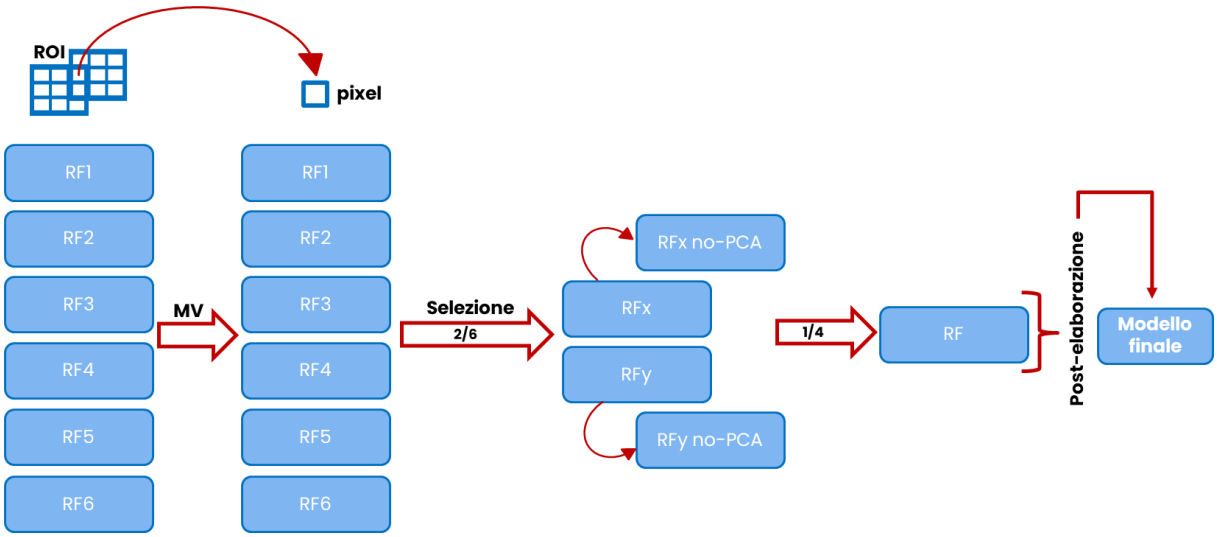

Figura 32. *Fasi della valutazione delle performance per la selezione del modello migliore da post-elaborare.*

Nel corrente capitolo si riportano nel dettaglio i passaggi riassunti sopra e rappresentati nella **Figura 32**.

Per la quantificazione delle prestazioni sono state utilizzate diverse metriche. Se ne riportano formule implementative e descrizioni nella **Tabella 15**.

	Formula	Note	Altre denominazioni
Accuratezza	$\frac{VP + VN}{P + N}$	Frazione di predizioni corrette sul totale delle osservazioni.	-
Precisione	$\frac{VP}{VP + FP}$	Frazione di elementi correttamente predetti positivi sul totale dei predetti positivi.	*Positive predictive value* (PPV)
Recall	$\frac{VP}{P}$	Frazione di elementi *positivi* riconosciuti come tali.	Sensibilità o *True Positive Rate* (TPR)
Specificità	$\frac{VN}{N}$	Frazione di elementi *negativi* riconosciuti come tali.	*True Negative Rate* (TNR)
Accuratezza bilanciata	$\frac{Recall + Specificità}{2}$	Accuratezza non influenzata da eventuale sbilanciamento nelle numerosità delle classi [48].	-
F1-score	$\frac{2\ Precision \cdot Recall}{Precision + Recall}$	Media armonica di *Precision* e *Recall,* misura di accuratezza con maggior rilievo ai VP [49].	-

Tabella 15. *Metriche per la valutazione ed il confronto delle prestazioni dei classificatori.*

3.1. Classificazione delle ROI

In questo sottocapitolo si riportano le prestazioni dei sei classificatori RF nel classificare le singole ROI contenute nel Training Set e nel Test Set.

3.1.1. Prestazioni di RF1

Il modello RF1 allenato sul Training Set (estratto dal CS1) le cui osservazioni rappresentano le singole ROI presenta un valore di accuratezza del $\mathbf{74,84}\%$. Il modello corrisponde a quello determinato dalla combinazione di iperparametri

che minimizza l'estremo superiore dell'intervallo di confidenza al 99% dell'errore di classificazione [39] [50] che, in questo caso, risulta pari a $0{,}2516$ (Vedi **Figura 33**).

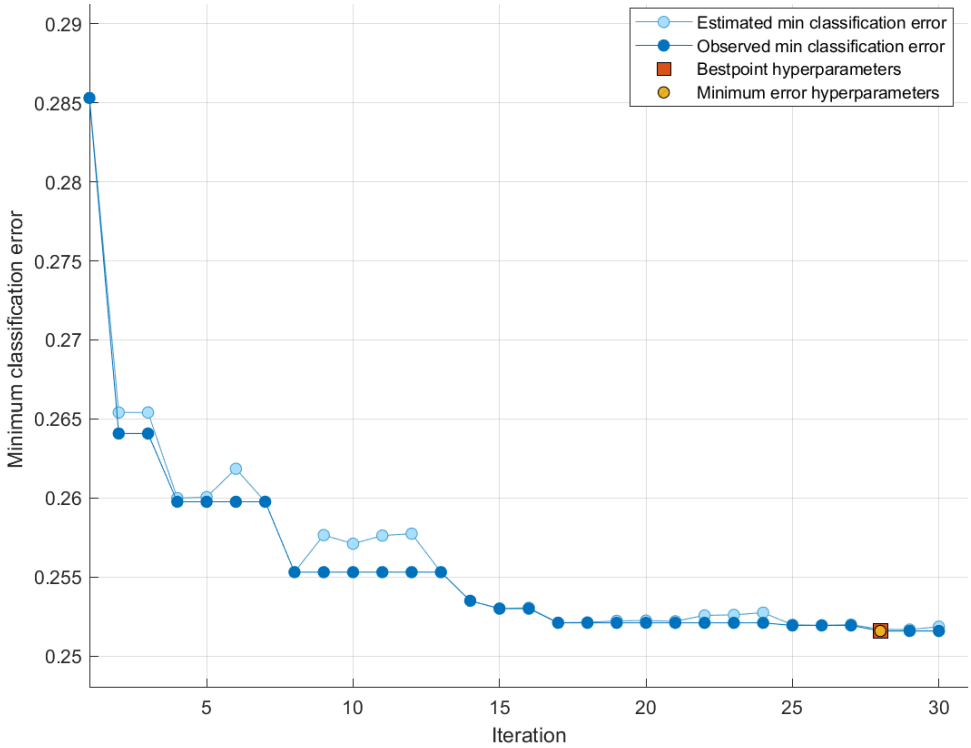

Figura 33. *Andamento dell'errore di classificazione della configurazione RF1 durante l'ottimizzazione degli iperparametri.*

Il classificatore viene successivamente valutato sul Test Set al fine di verificare che non sia affetto dal fenomeno di *overfitting* (sovra-adattamento sul Training Set). L'accuratezza ottenuta è del $\mathbf{74{,}50\%}$. Si riportano i risultati della classificazione sotto forma di matrici di confusione che rappresentano il numero di osservazioni per ogni combinazione di classe effettiva e classe predetta (Vedi **Figura 34**). Per un'interpretazione più efficace dei dati rappresentati si elencano nella **Tabella 16** le numerosità delle singole classi all'interno di Training e Test Set.

	Numerosità		
	ROI **patologiche**	ROI **sane "0"**	ROI **totali**
Training Set	30105	30104	60209
Test Set	14581	14581	29162

Tabella 16. *Costituzione di Training e Test Set nella configurazione RF1.*

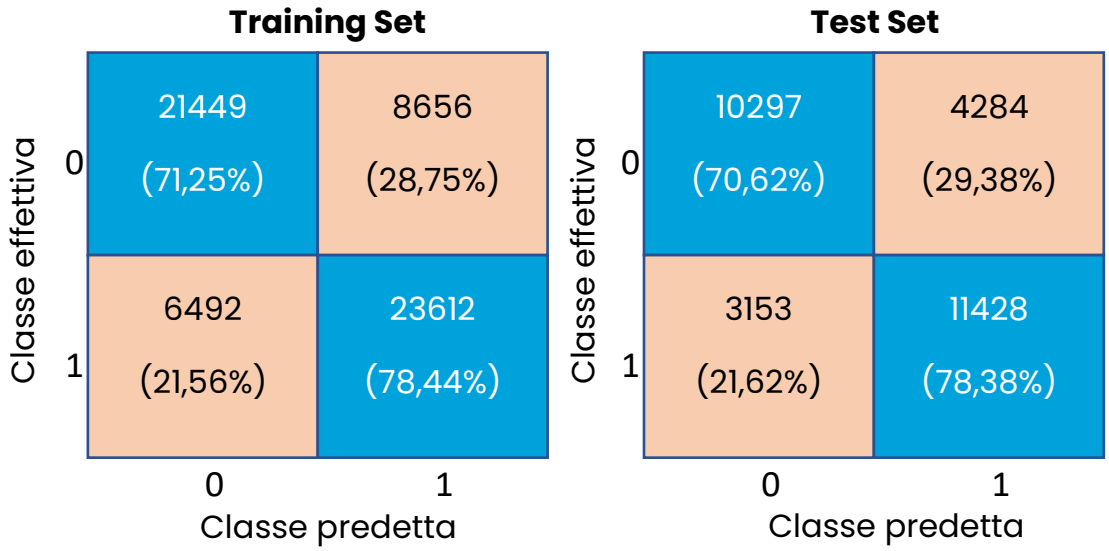

Figura 34. *Matrici di confusione del modello RF1.*

3.1.2. Prestazioni di RF2

Il modello RF2 allenato sul Training Set (estratto dal CS2) le cui osservazioni rappresentano le singole ROI presenta un valore di accuratezza del $\mathbf{53,17\%}$. Questo modello corrisponde a quello che nell'ottimizzazione degli iperparametri presenta un errore di classificazione pari a $0{,}4668$ (Vedi **Figura 35**).

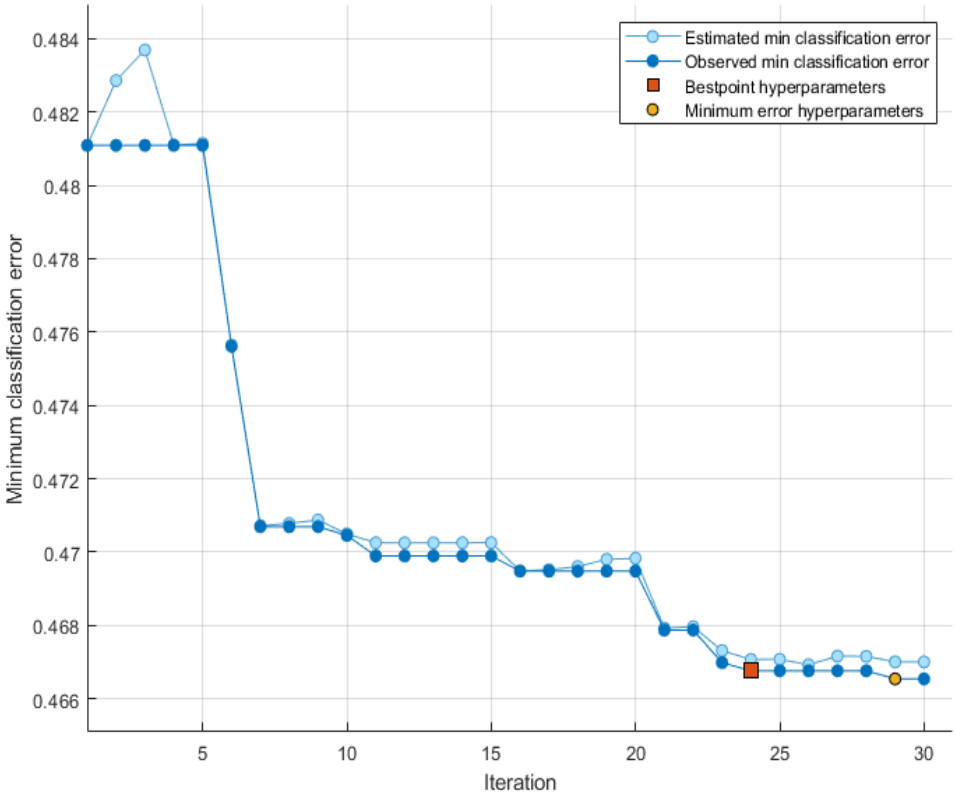

Figura 35. *Andamento dell'errore di classificazione della configurazione RF2 durante l'ottimizzazione degli iperparametri.*

Il classificatore viene successivamente valutato sul Test Set e si ottiene un'accuratezza del $\mathbf{53,60\%}$. Si riportano le matrici di confusione con i risultati della classificazione (Vedi **Figura 36**). Per un'interpretazione più efficace dei dati rappresentati si elencano nella **Tabella 17** le numerosità delle singole classi all'interno di Training e Test Set.

	Numerosità			
	ROI **patologiche**	ROI **sane "0"**	ROI **sane "2"**	ROI **totali**
Training Set	22578	22578	22578	67734
Test Set	22107	22107	22107	66321

Tabella 17. *Costituzione di Training e Test Set nella configurazione RF2.*

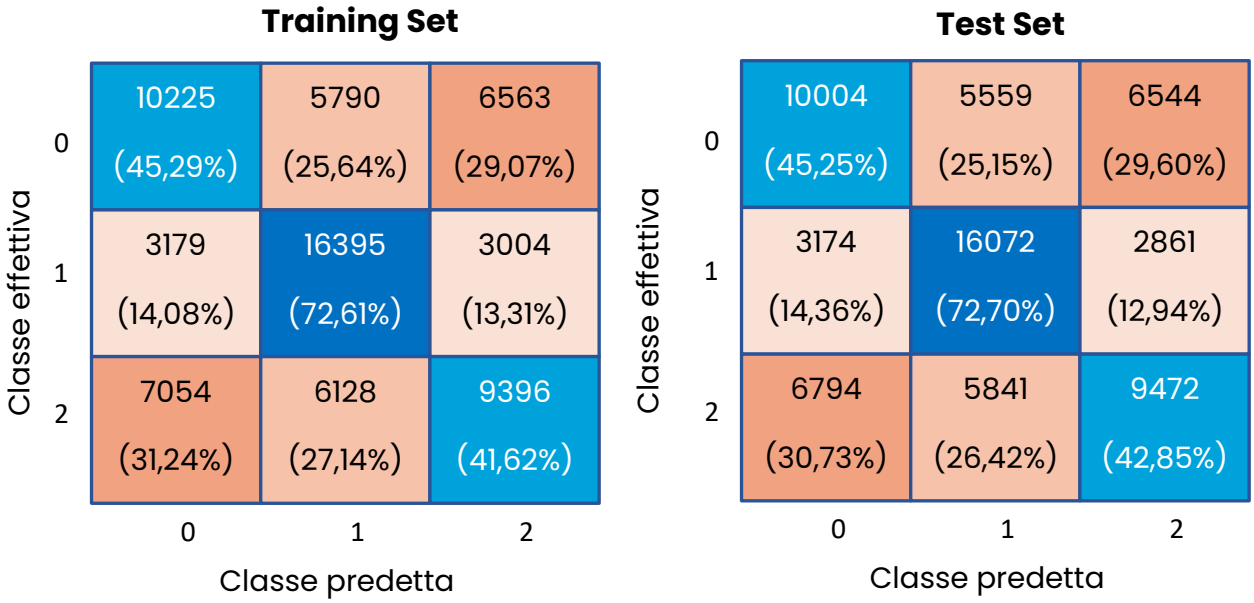

Figura 36. *Matrici di confusione del modello RF2.*

Allo scopo di confrontare le prestazioni dei diversi modelli, i risultati dei classificatori multiclasse, tra cui il corrente modello RF2, vengono aggregati per essere ricondotti ad un problema di classificazione binaria (ROI sane vs ROI patologiche, S vs P). Nello specifico, le ROI con etichetta "2" vengono aggregate a quelle di etichetta "0" al fine di rappresentare la classe sana "S", le ROI con label

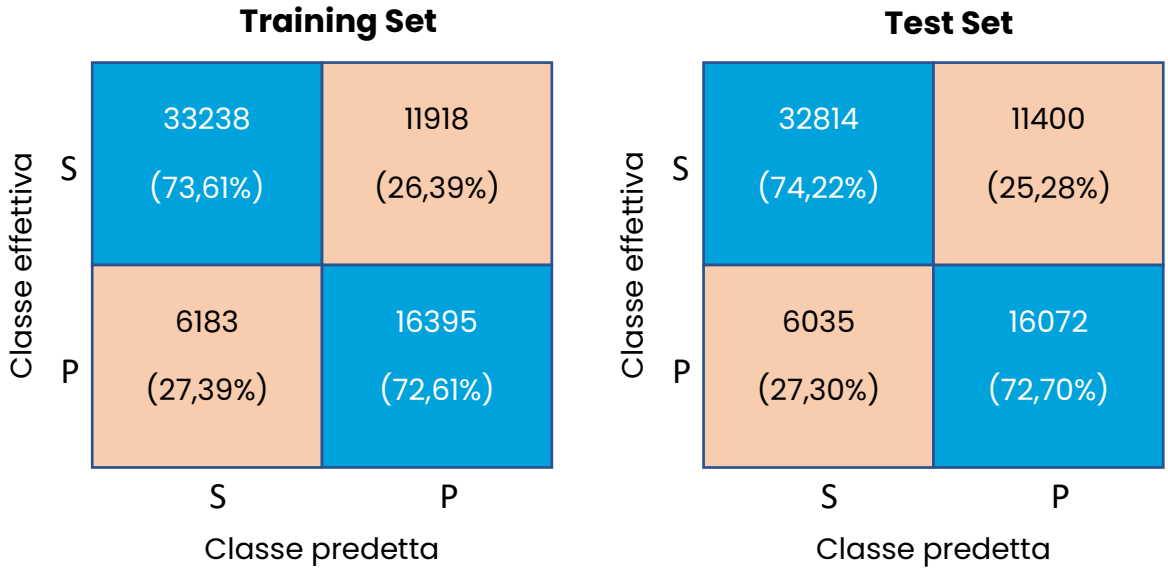

Figura 37. *Matrici di confusione del modello RF2 dopo l'aggregazione dei risultati.*

"1" coincidono, invece, con la classe sana "P".

Le nuove prestazioni vengono riassunte nelle matrici di confusione rappresentate nella **Figura 37**. In questo modo, l'accuratezza nella classificazione binaria diventa pari al $\mathbf{73,28\%}$ sul Training Set e uguale al $\mathbf{73,71\%}$ sul Test Set.

3.1.3. Prestazioni di RF3

Il modello RF3 allenato sul Training Set estratto dal CS3 le cui osservazioni rappresentano le singole ROI presenta un valore di accuratezza del $\mathbf{39,58\%}$. Questo modello corrisponde a quello che nell'ottimizzazione degli iperparametri presenta un errore di classificazione pari a $0{,}6042$ (Vedi **Figura 38**).

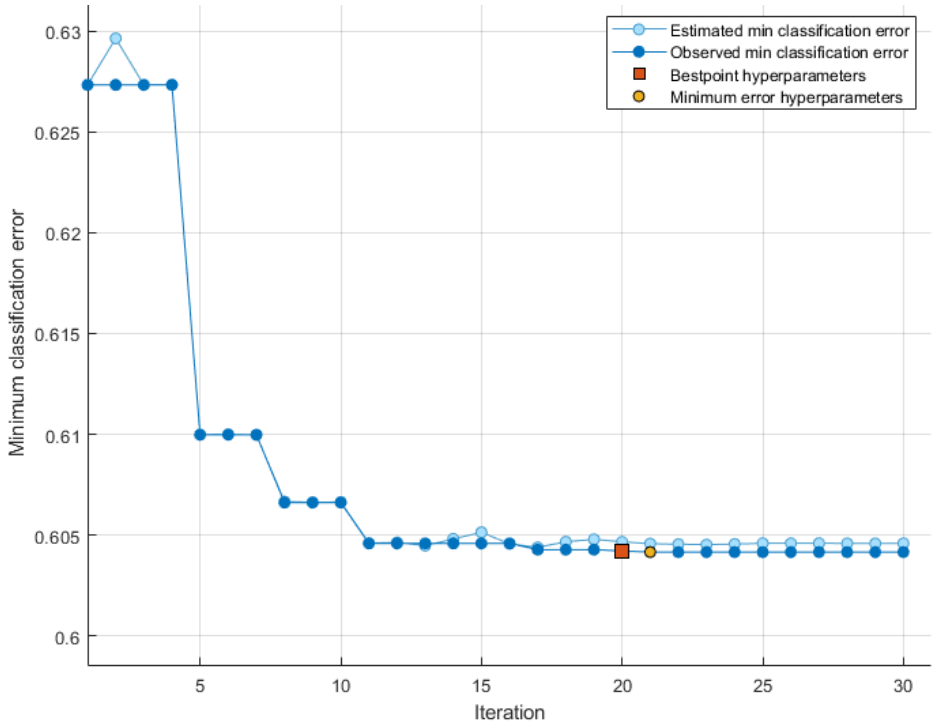

Figura 38. *Andamento dell'errore di classificazione della configurazione RF3 durante l'ottimizzazione degli iperparametri.*

Il classificatore viene successivamente valutato sul Test Set e si ottiene un'accuratezza del $\mathbf{39,53\%}$. Si riportano le matrici di confusione con i risultati della classificazione (Vedi **Figura 39**). Per un'interpretazione più efficace dei dati rappresentati si elencano nella **Tabella 18** le numerosità delle singole classi all'interno di Training e Test Set.

	Numerosità				
	ROI **patologiche**	ROI **sane** **"0"**	ROI **sane** **"2"**	ROI **sane** **"3"**	ROI **totali**
Training Set	22578	22578	22579	22579	90314
Test Set	22107	22107	22107	22107	88428

Tabella 18. *Costituzione di Training e Test Set nella configurazione RF3.*

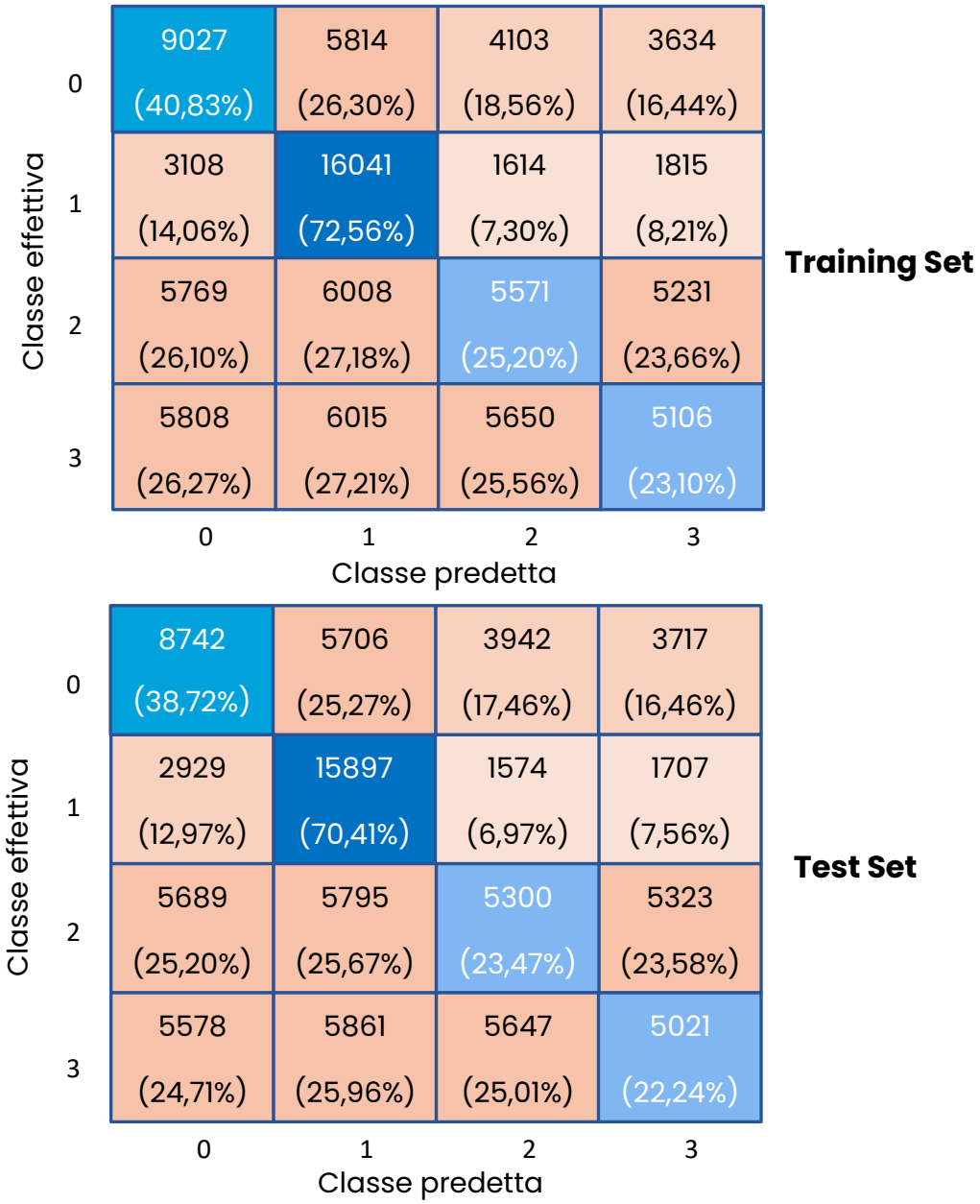

Figura 39. *Matrici di confusione del modello RF3.*

Anche in questo caso, essendo RF3 un modello di classificazione multiclasse, i risultati sono stati aggregati per essere ricondotti ad un problema di classificazione binaria. Per i modelli allenati sui dati provenienti dal CS3 che prevedeva la rappresentazione di 3 sottogruppi rappresentativi della classe sana (label "0", "2" e "3"), sono stati ideati due metodi di aggregazione dei risultati.

Il primo metodo è coerente con l'ipotesi fatta nella definizione del CS3 in cui è stata fatta una divisione in due cluster delle ROI sane estratte da immagini patologiche con l'idea di poter discriminare più o meno grossolanamente le ROI sane effettive da quelle misclassificate dal medico in fase di segmentazione manuale. Uno dei due cluster (classi "2" e "3") dovrebbe, quindi, poter essere affiliato alla classe patologica, mentre il restante alla classe sana.

Osservando le matrici di confusione nella **Figura 39**, si nota che non è presente una tendenza netta di una delle due classi ad esser confusa con etichetta sana o patologica. L'unica prevalenza che si può notare è sulla base del numero di ROI patologiche misclassificate come classe "3" rispetto a quelle misclassificate

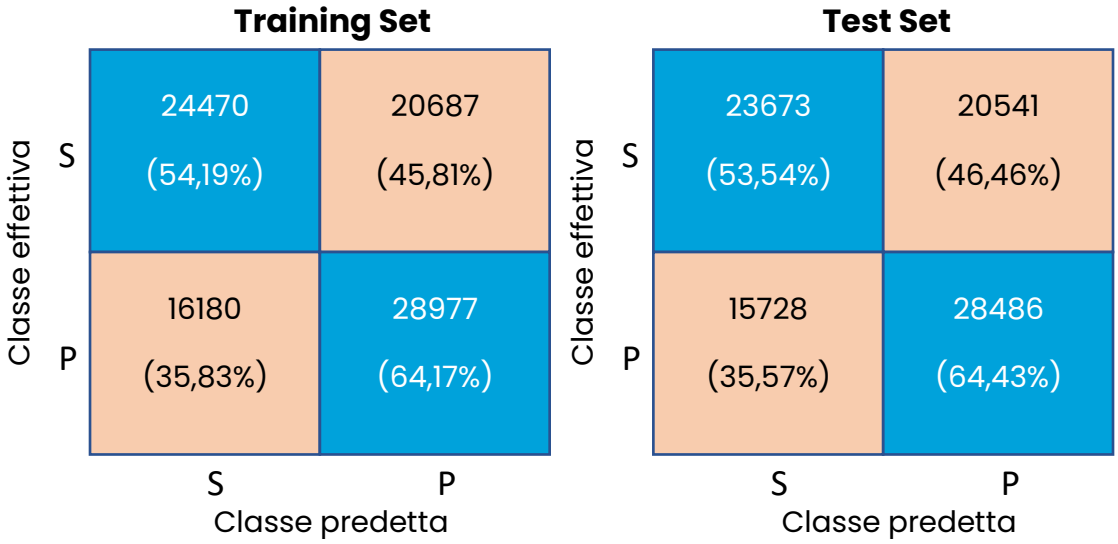

Figura 40. *Matrici di confusione del modello RF3 dopo l'aggregazione dei risultati tramite il metodo 1.*

come classe “2”. Si effettua quindi il tentativo di considerare la classe “3” come patologica “P” assieme alle ROI “1”. La classe “2” viene, invece, associata alle ROI “0” nella rappresentazione della classe sana “S”. I risultati sottoposti a tale aggregazione si riportano nelle matrici di confusione in **Figura 40**. Il valore di accuratezza sul Training Set è pari al $\mathbf{59,18\%}$, sul Test Set al $\mathbf{58,98\%}$.

Il secondo metodo di aggregazione prevede che entrambe le classi “2” e “3”, derivando da porzioni di immagine dichiarate sane in fase di segmentazione manuale, rappresentino assieme alla classe “0” la classe sana “S”.

In questo modo si ottengono delle prestazioni migliori rispetto a quelle ricavate con il primo metodo di aggregazione (Vedi **Figura 41**), con valori di accuratezza

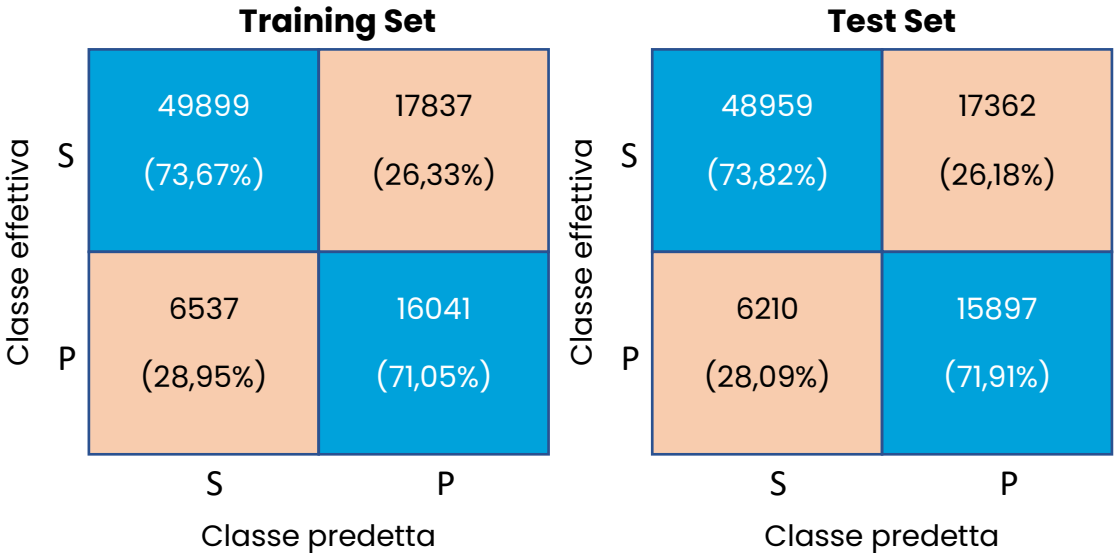

Figura 41. *Matrici di confusione del modello RF3 dopo l'aggregazione dei risultati tramite il metodo 2.*

pari al $\mathbf{73,01\%}$ sul Training Set ed al $\mathbf{73,34\%}$ sul Test Set.

3.1.4. Prestazioni di RF4

Il modello RF4 è stato allenato sul Training Set estratto a partire da un CS di configurazione CS1 e, in questo caso, ricavato a partire da immagini sottoposte

a *pre-processing* (equalizzazione e filtraggio gaussiano). Le osservazioni rappresentano le singole ROI e l'accuratezza ottenuta nella classificazione vale $\mathbf{74,90\%}$. Questo modello corrisponde a quello che nell'ottimizzazione degli iperparametri un errore di classificazione pari a $0{,}2510$ (Vedi **Figura 42**).

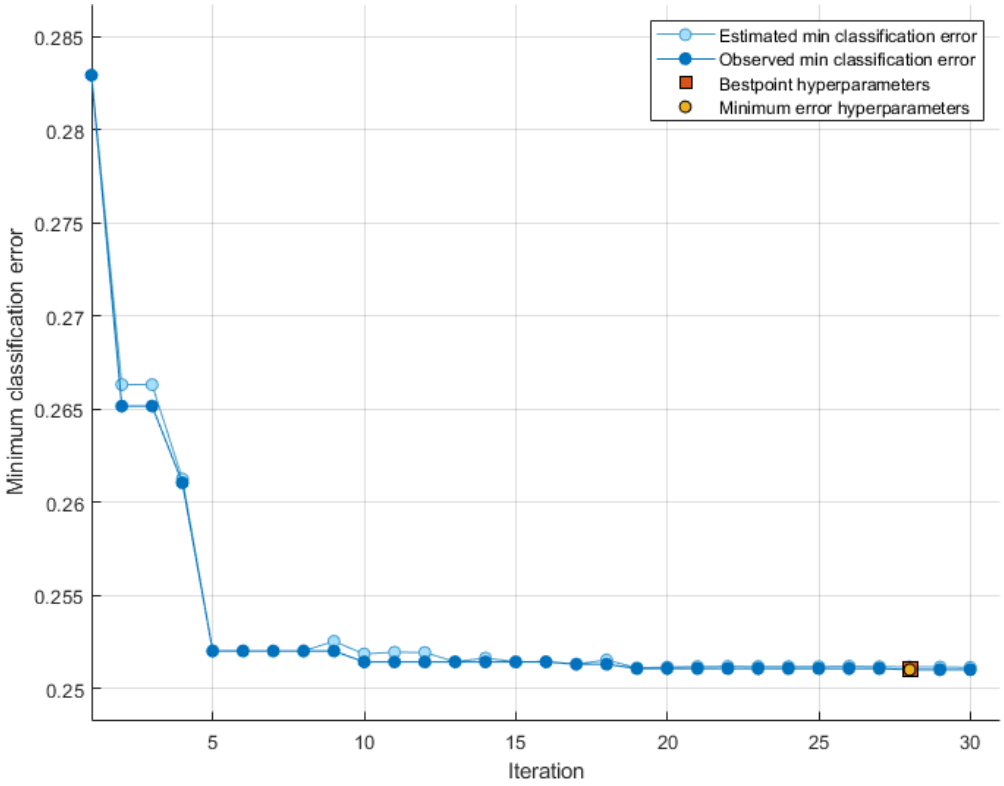

Figura 42. *Andamento dell'errore di classificazione della configurazione RF4 durante l'ottimizzazione degli iperparametri.*

Il classificatore viene successivamente valutato sul Test Set ottenendo un'accuratezza del $\mathbf{74,86\%}$. I risultati della classificazione su entrambi i set si riportano sotto forma di matrici di confusione nella **Figura 43**. Per un'interpretazione più efficace dei dati rappresentati si elencano nella **Tabella 19** le numerosità delle singole classi all'interno di Training e Test Set.

	Numerosità		
	ROI **patologiche**	ROI **sane "0"**	ROI **totali**
Training Set	23923	23923	47846
Test Set	20762	20762	41524

Tabella 19. *Costituzione di Training e Test Set nella configurazione RF4.*

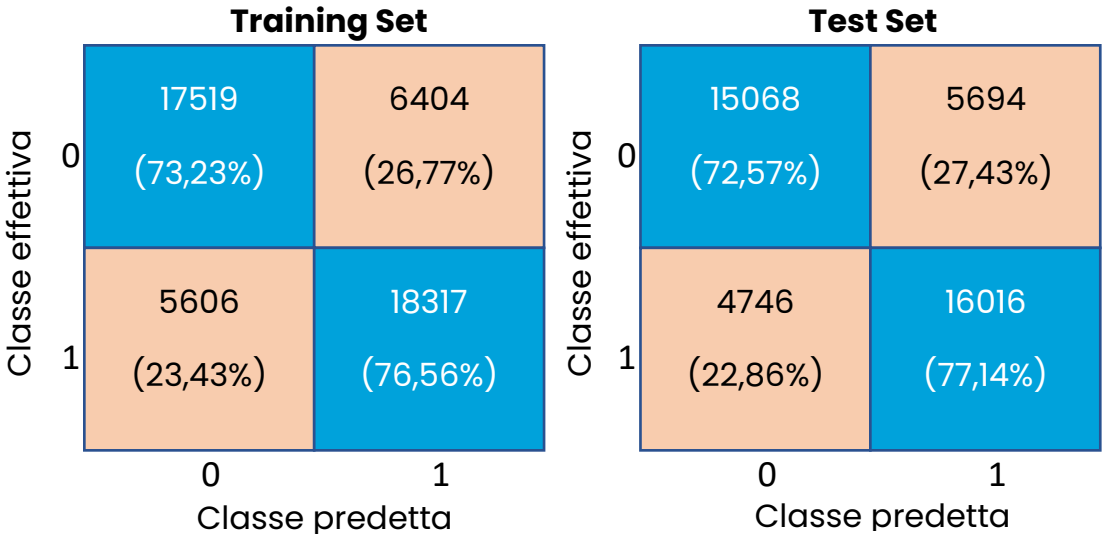

Figura 43. *Matrici di confusione del modello RF4.*

3.1.5. Prestazioni di RF5

Il modello RF5 è stato allenato sul Training Set estratto a partire da un CS di configurazione CS2 ricavato a partire da immagini sottoposte a *pre-processing*. Le osservazioni rappresentano le singole ROI e l'accuratezza ottenuta nella classificazione vale $\mathbf{52,32\%}$. Questo modello corrisponde a quello che nell'ottimizzazione degli iperparametri presenta un errore di classificazione pari a $0{,}4759$ (Vedi **Figura 44**).

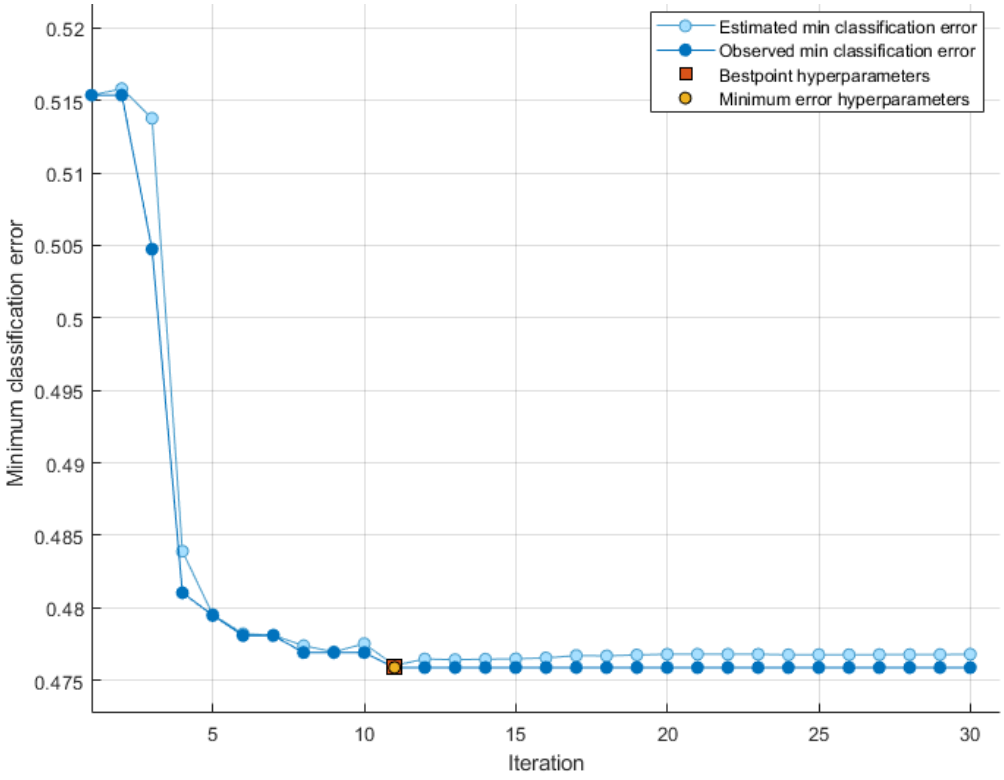

Figura 44. *Andamento dell'errore di classificazione della configurazione RF5 durante l'ottimizzazione degli iperparametri.*

Il classificatore viene successivamente valutato sul Test Set e si ottiene un'accuratezza del $52,57\%$. Si riportano le matrici di confusione con i risultati della classificazione (Vedi **Figura 45**). Per un'interpretazione più efficace dei dati rappresentati si elencano nella **Tabella 20** le numerosità delle singole classi all'interno di Training e Test Set.

	Numerosità			
	ROI **patologiche**	ROI **sane "0"**	ROI **sane "2"**	ROI **totali**
Training Set	11960	11960	11960	35880
Test Set	32725	32725	32725	98175

Tabella 20. *Costituzione di Training e Test Set nella configurazione RF5.*

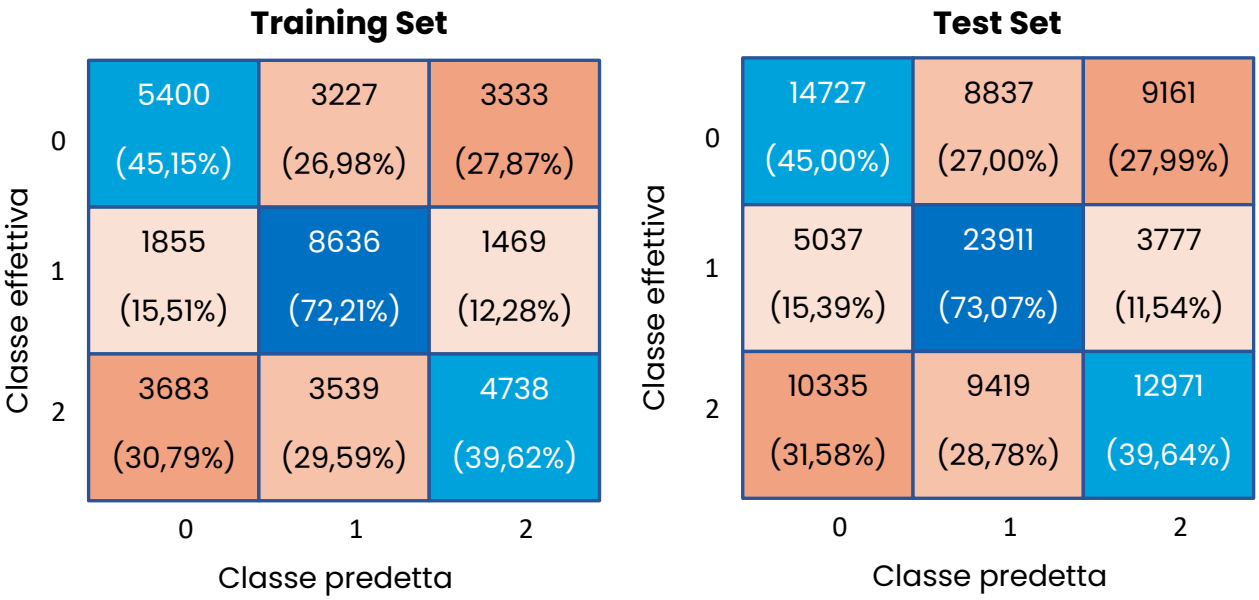

Figura 45. *Matrici di confusione del modello RF5.*

Essendo il corrente modello RF4 un classificatore multiclasse, i risultati ottenuti vengono aggregati per essere ricondotti al problema di classificazione binaria (S-P). Nello specifico, le ROI con etichetta "2" vengono aggregate a quelle di etichetta "0" al fine di rappresentare la classe sana "S", le ROI con label "1" coincidono, invece, con la classe sana "P".

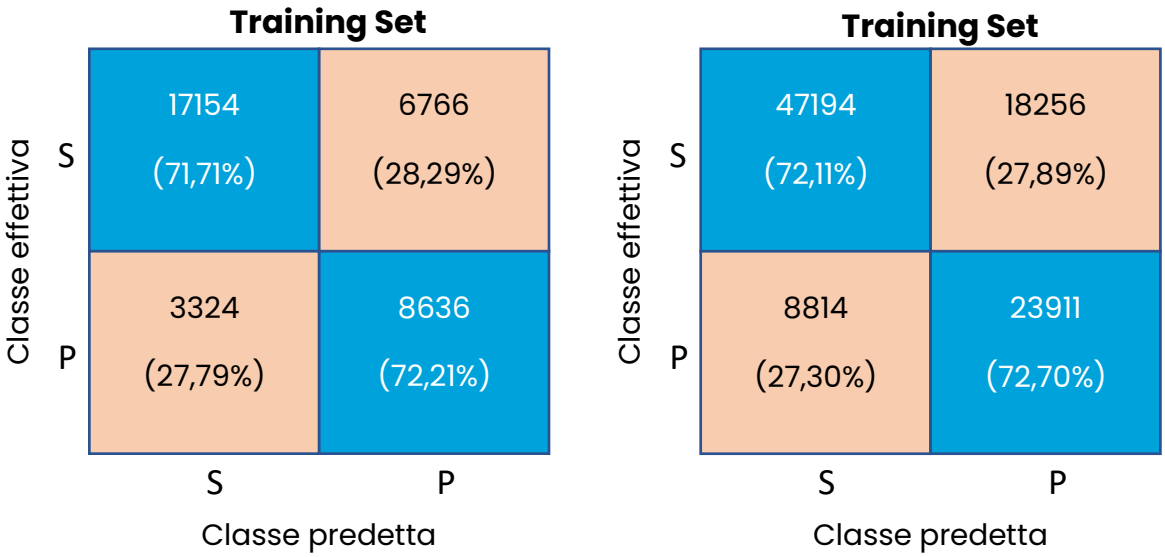

Figura 46. *Matrici di confusione del modello RF5 dopo l'aggregazione dei risultati.*

Le nuove prestazioni vengono riassunte nelle matrici di confusione rappresentate nella **Figura 46**. In questo modo, l'accuratezza nella classificazione binaria diventa pari al $\mathbf{71,88\%}$ sul Training Set e uguale al $\mathbf{72,43\%}$ sul Test Set.

3.1.6. Prestazioni di RF6

Il modello RF5 è stato allenato sul Training Set estratto a partire da un CS di configurazione CS3 ricavato a partire da immagini sottoposte a *pre-processing*. Le osservazioni rappresentano le singole ROI e l'accuratezza ottenuta nella classificazione vale $\mathbf{38,93\%}$. Questo modello corrisponde a quello che nell'ottimizzazione degli iperparametri presenta un errore di classificazione pari a $0{,}6110$ (Vedi **Figura 47**).

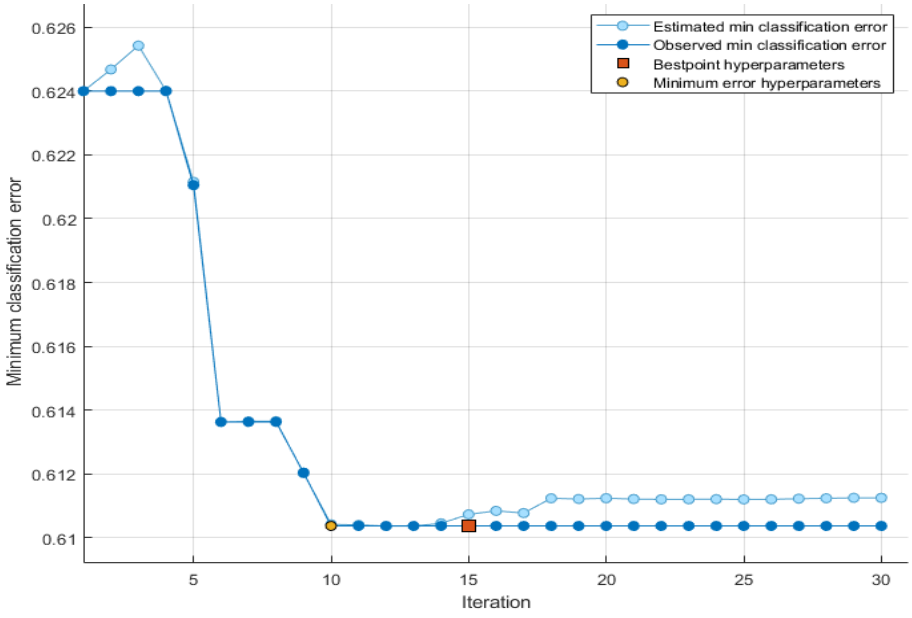

Figura 47. *Andamento dell'errore di classificazione della configurazione RF6 durante l'ottimizzazione degli iperparametri.*

Il classificatore viene successivamente valutato sul Test Set e si ottiene un'accuratezza del $\mathbf{38,59\%}$. Si riportano le matrici di confusione con i risultati della classificazione (Vedi **Figura 48**). Per un'interpretazione più efficace dei dati rappresentati si elencano nella **Tabella 21** le numerosità delle singole classi all'interno di Training e Test Set.

	Numerosità				
	ROI **patologiche**	ROI **sane** **"0"**	ROI **sane** **"2"**	ROI **sane** **"3"**	ROI **totali**
Training Set	11960	11960	11960	11960	47840
Test Set	32725	32725	32725	32725	130900

Tabella 21. *Costituzione di Training e Test Set nella configurazione RF6.*

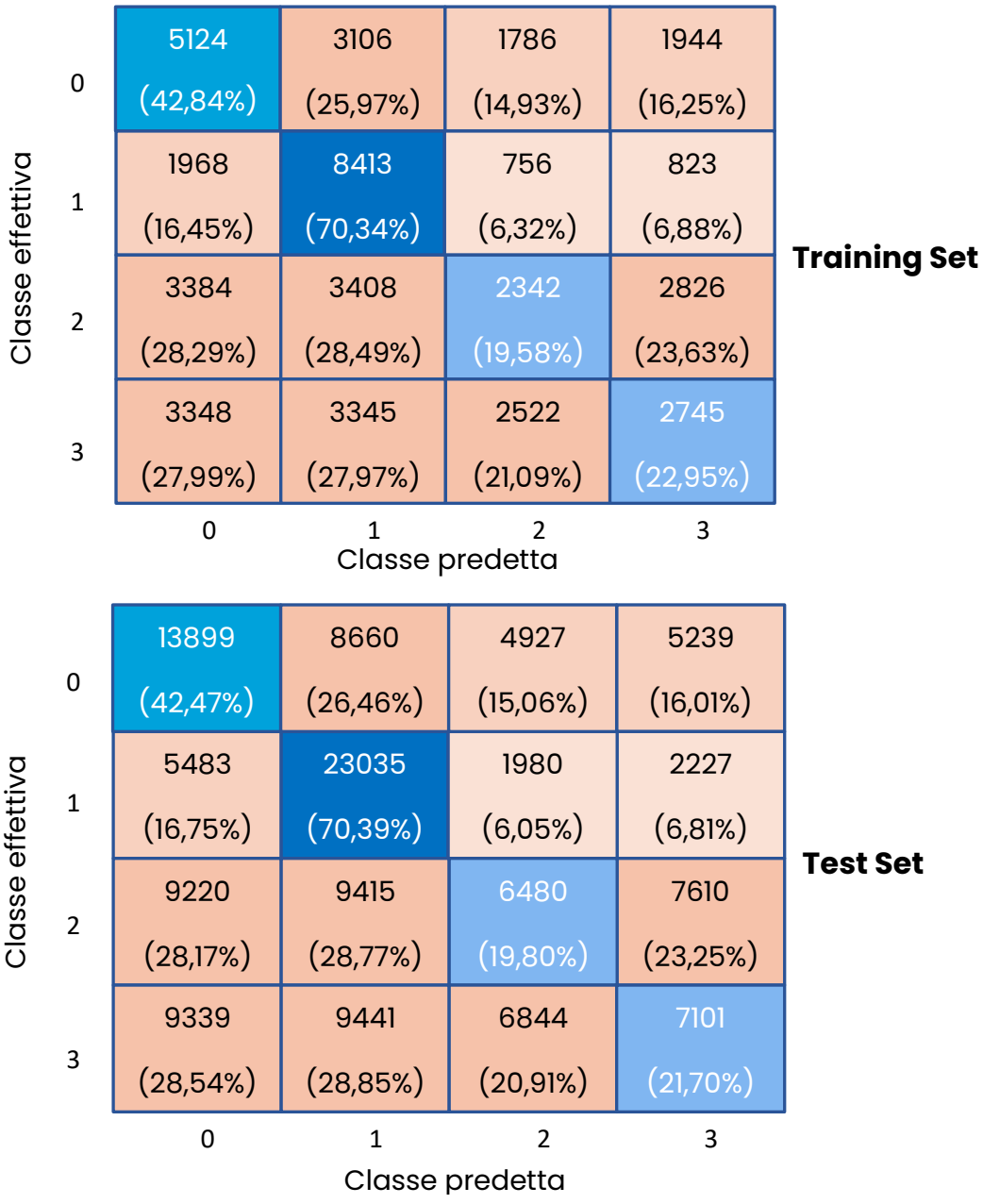

Figura 48. *Matrici di confusione del modello RF6.*

Anche in questo caso, essendo RF6 un modello di classificazione multiclasse, i risultati sono stati aggregati per essere ricondotti ad un problema di classificazione binaria. A tal fine, Sono stati utilizzati i due metodi di aggregazione introdotti nel sottocapitolo 3.1.3.

Il primo metodo di aggregazione permette di ottenere dei valori di accuratezza pari a $\mathbf{58,45\%}$ sul Training Set e al $\mathbf{58,31\%}$ sul Test Set. Se ne riportano le matrici di confusione nella **Figura 49**.

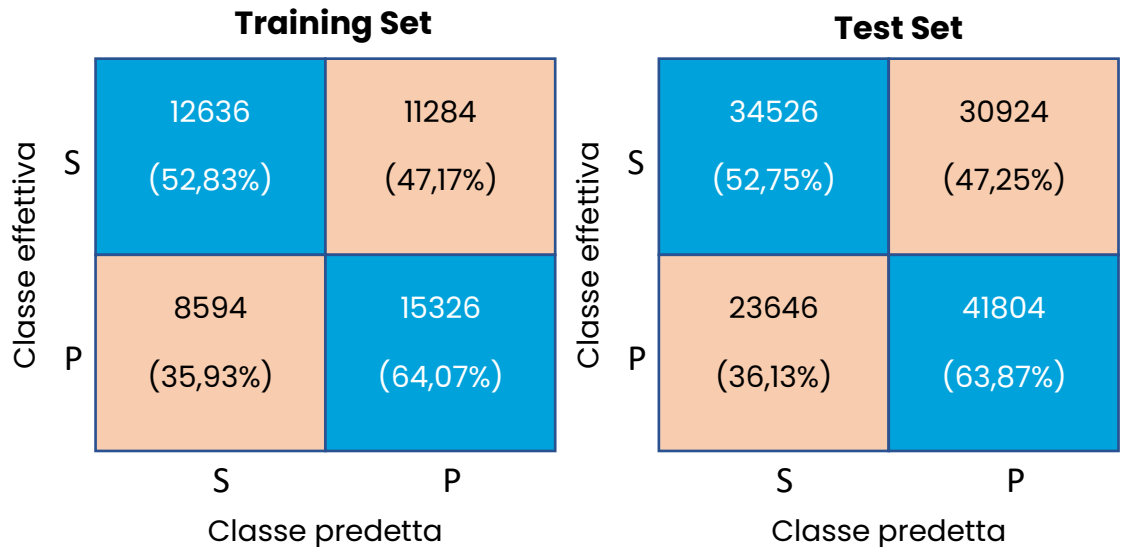

Figura 49. *Matrici confusionali del modello RF6 dopo l'aggregazione dei risultati tramite il metodo 1 (Paragrafo 3.1.3).*

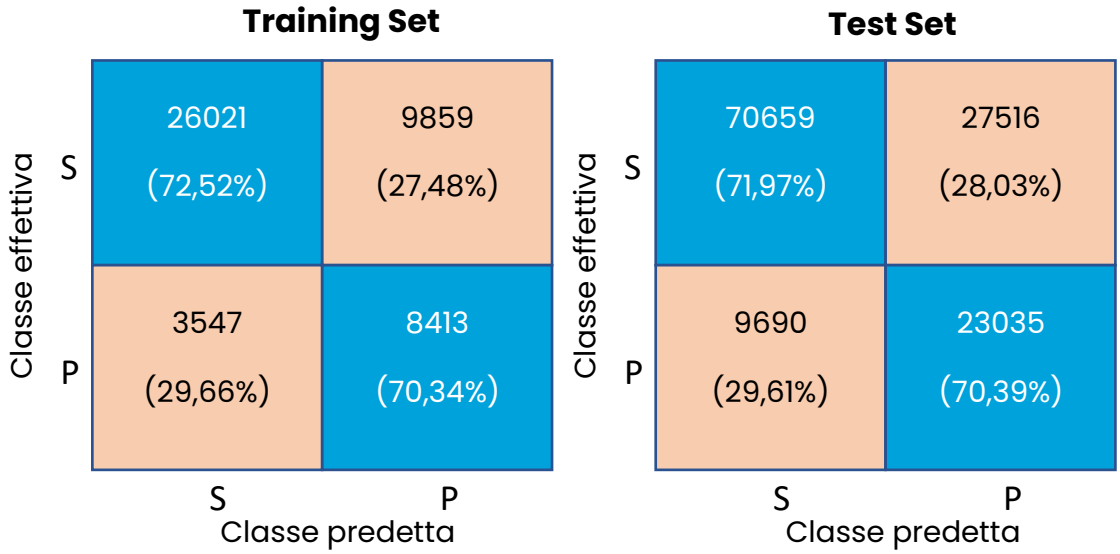

Figura 50. *Matrici confusionali del modello RF6 dopo l'aggregazione dei risultati tramite il metodo 2 (Paragrafo 3.1.3).*

Il secondo metodo di aggregazione permette di ottenere dei valori di accuratezza del $\mathbf{71,98\%}$ sul Training Set e del $\mathbf{71,58\%}$ sul Test Set. Le relative matrici di confusione sono rappresentate nella **Figura 50**.

3.1.7. Prestazioni a confronto

Si consideri il compito di classificazione binaria per la discriminazione di classe sana "S" e classe patologica "P". Si riportano le accuratezze dei vari metodi ricondotti a modelli binari nella **Tabella 22**.

Modello	***Composizione delle classi***		***Accuratezza***	
	ROI Classe «S»	*ROI Classe «P»*	*Training Set*	*Test Set*
RF1	«0»	«1»	**74,84%**	**74,50%**
RF2	«0» e «2»	«1»	73,28%	73,71%
RF3	«0» e «2»	«1» e «3»	**59,18%**	**58,98%**
	«0», «2» e «3»	«1»	73,01%	73,34%
RF4	«0»	«1»	**74,90%**	**74,86%**
RF5	«0» e «2»	«1»	71,88%	72,43%
RF6	«0» e «3»	«1» e «2»	**58,45%**	**58,31%**
	«0», «2» e «3»	«1»	71,98%	71,58%

Tabella 22. *Prestazioni complessive dei modelli RF nella classificazione binaria S-P delle ROI.*

Si nota come l'esclusione delle ROI sane estratte da immagini patologiche porti alla costruzione di modelli più accurati. Infatti, RF1 ed RF4 presentano le prestazioni migliori in termini di accuratezza (**Tabella 22**, riga 1 e riga 5). Le performance peggiori si sperimentano nel caso delle configurazioni RF3 ed RF6 (**Tabella 22**, riga 3 e riga 7) qualora i risultati siano stati aggregati tramite il metodo di aggregazione 1 (Vedi Paragrafo 3.1.3). Si consideri che il metodo di

aggregazione 1 si basa sulla tecnica utilizzata per la costruzione delle configurazioni RF3 ed RF6 che derivano da un Construction Set di tipo CS3 (Vedi Paragrafo 2.4.3). Nello specifico, le ROI sane estratte da immagini patologiche vengono divise in due cluster associati uno alla classe sana e l'altro alla classe patologica. Tuttavia, si può notare come questo criterio porti ad un peggioramento nelle prestazioni. Pertanto, il metodo proposto non è adatto al compito di rimediare efficacemente ad eventuali misclassificazioni commesse dal medico in fase di segmentazione manuale.

3.2. Classificazione dei pixel

I modelli allenati nella classificazione delle singole ROI vengono utilizzati per effettuare una discriminazione pixel per pixel attraverso la tecnica del *Majority Voting* (MV). La procedura è stata applicata ai soli modelli che eseguono un compito di classificazione binaria. Per le configurazioni RF3 ed RF6 è stato considerato il metodo di aggregazione 2 (Vedi Paragrafo 3.1.3) poiché presenta prestazioni migliori.

Considerando il fatto che la suddivisione in ROI è stata fatta in presenza di sovrapponimento parziale, uno stesso pixel all'interno di un'immagine può appartenere ad una o più ROI.

Pertanto, qualora un pixel appartenesse a più ROI classificate con etichette contrastanti, la determinazione della classe di tale pixel risulterebbe ambigua. Per ovviare a tale problematica è stata applicata la tecnica del MV pixel per pixel. Questa procedura, per come definita nel corrente studio, prevede che al pixel venga assegnata l'etichetta con il maggior numero di ricorrenze tra le ROI che contengono il pixel stesso.

Le prestazioni sono state valutate sull'intero Construction Set. I risultati ottenuti si riportano nella **Tabella 23**.

	Precisione	***Recall***	***Specificità***	***Accuratezza bilanciata***	***F1 score***
RF1	37,65%	**65,20%**	84,91%	**75,06%**	**47,73%**
RF2	37,49%	64,58%	84,96%	74,77%	47,44%
RF3	36,48%	62,57%	84,78%	73,67%	46,09%
RF4	**37,73%**	62,65%	**85,55%**	74,10%	47,10%
RF5	35,30%	63,25%	83,80%	73,53%	45,31%
RF6	34,91%	62,47%	83,75%	73,11%	44,79%

Tabella 23. *Confronto delle prestazioni dei modelli RF nella classificazione binaria pixel per pixel.*

Si può notare come le configurazioni RF1 (**Tabella 23**, riga 1) e RF4 (**Tabella 23**, riga 4) presentino le prestazioni migliori. Nello specifico, RF1 presenta i valori massimi di recall, accuratezza e F1-score, mentre RF2 quelli di precisione e specificità. Questo risultato conferma quanto visto nel Paragrafo 3.1.7 riguardo alle performance dei modelli nella classificazione delle ROI.

3.3. Effetto della Principal Component Analysis

Le configurazioni RF1 ed RF4 hanno portato all'ottenimento dei due classificatori dalle prestazioni migliori nella classificazione binaria pixel per pixel. Al fine di valutare l'effetto della Principal Component Analysis sulle performance ottenute, queste configurazioni sono state riutilizzate per allenare due modelli senza l'impiego di tale tecnica.

Per la configurazione RF1 è stato selezionato il modello che presenta un errore di classificazione pari a $0{,}2202$ (Vedi **Figura 51**).

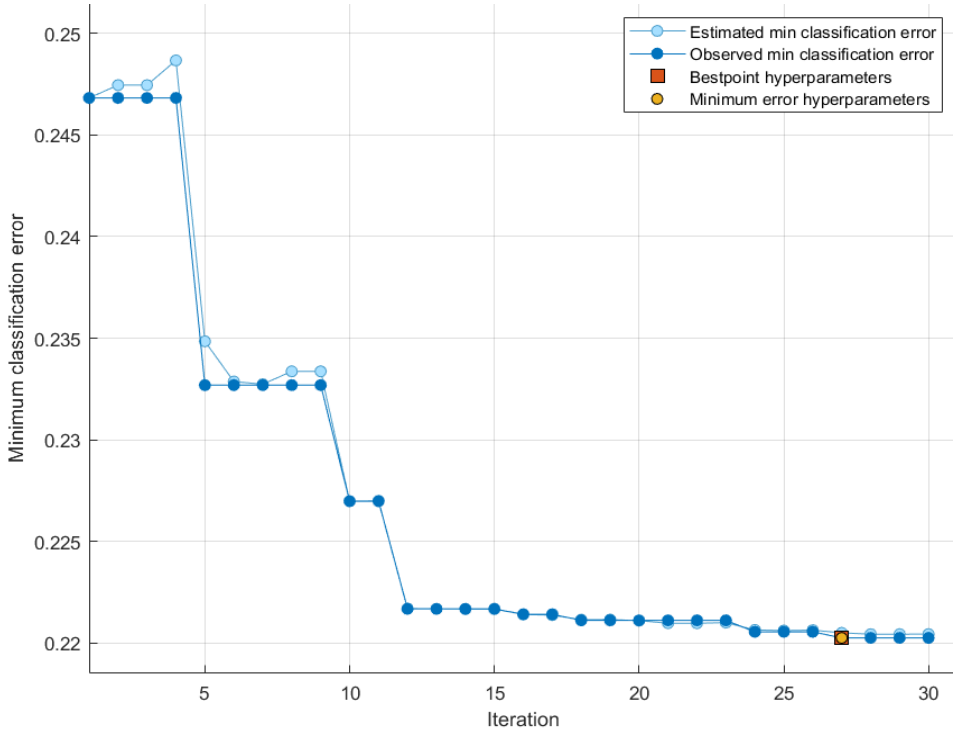

Figura 51. *Andamento dell'errore di classificazione del modello RF1 senza PCA durante l'ottimizzazione degli iperparametri.*

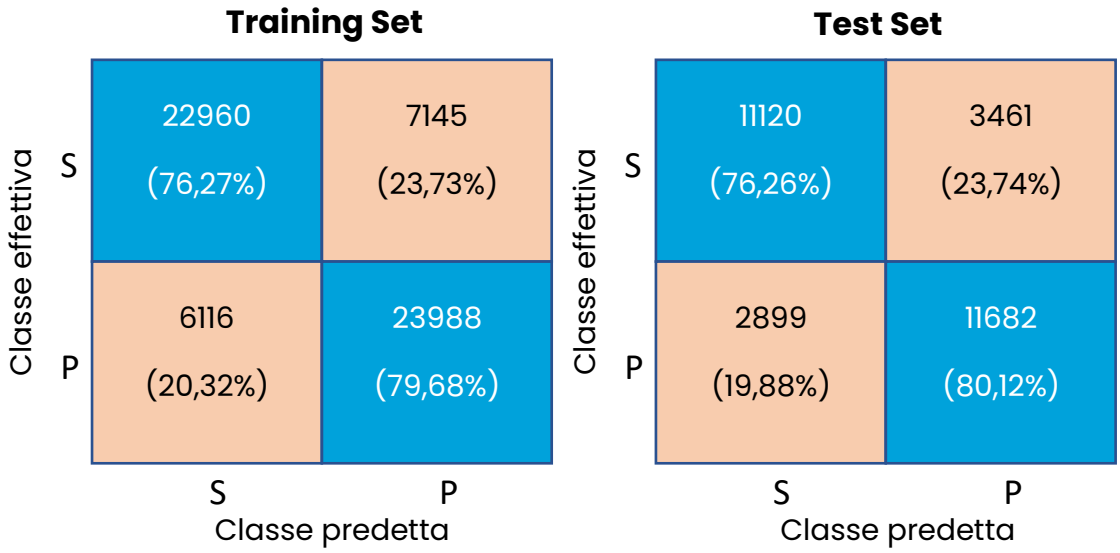

Figura 52. *Matrici di confusione del modello RF1 senza PCA.*

Le matrici di confusione in **Figura 52** riportano le prestazioni aggiornate su Training Set e sul Test Set, con accuratezze del $\mathbf{77,98\%}$ e del $\mathbf{78,19\%}$ rispettivamente. La costituzione dei due set è rappresentata nella **Tabella 16**.

Per la configurazione RF4 è stato selezionato il modello che presenta un errore di classificazione pari a $0{,}2309$ (Vedi **Figura 53**).

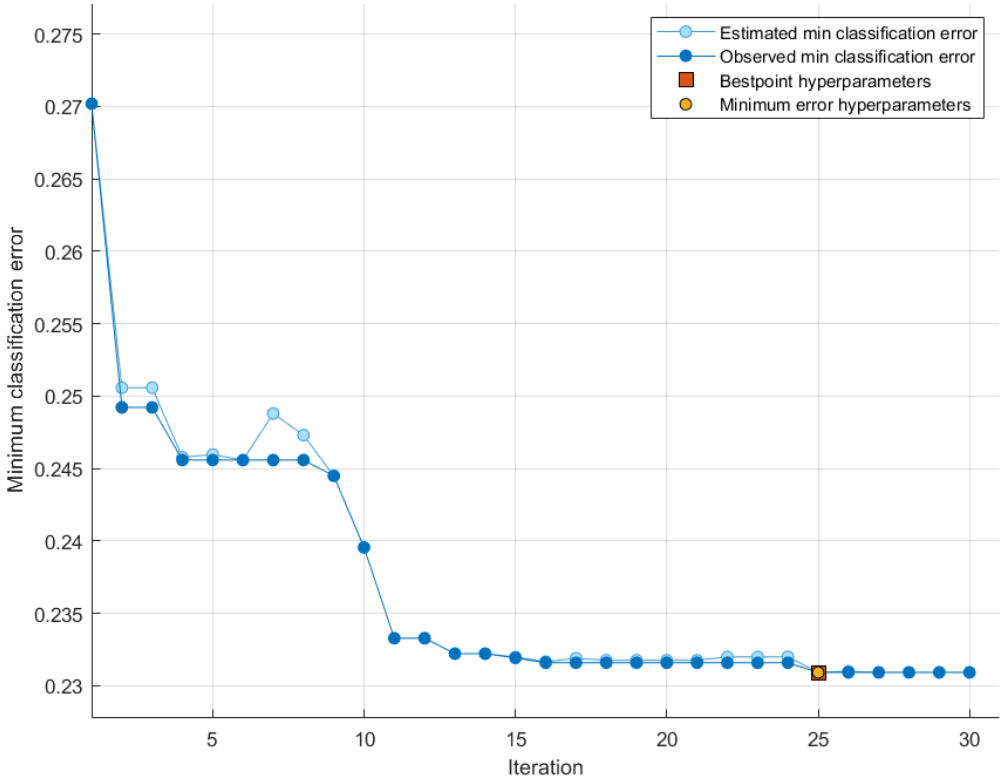

Figura 53. *Andamento dell'errore di classificazione del modello RF4 senza PCA durante l'ottimizzazione degli iperparametri.*

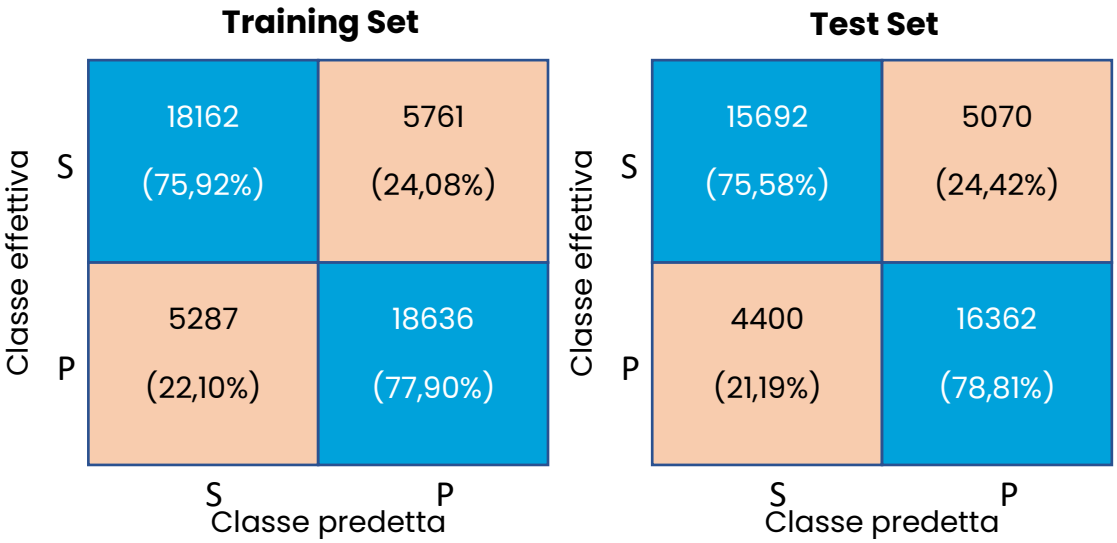

Figura 54. *Matrici di confusione del modello RF4 senza PCA.*

Le matrici di confusione in **Figura 54** riportano le prestazioni aggiornate su Training Set e sul Test Set, con accuratezze del $\mathbf{76,91\%}$ e del $\mathbf{77,19\%}$ rispettivamente. La costituzione dei due set è rappresentata nella **Tabella 19**.

Gli iperparametri relativi ai due modelli si riportano nella **Tabella 24**.

	Algoritmo	# max **partizioni**	# **alberi**	**Learning rate**	# **predittori**
RF1	LogitBoost	2840	416	0,30158	-
RF4	LogitBoost	131	75	0,19268	-

Tabella 24. *Iperparametri selezionati per i modelli RF1 ed RF4 senza PCA.*

Rivalutate le prestazioni dei due nuovi modelli, si effettua un confronto con le versioni allenate in presenza di PCA precedentemente descritte (Vedi **Tabella 25**).

	PCA	***Precisione***	***Recall***	***Specificità***	***Accuratezza bilanciata***	***F1 score***
RF1	Si	37,65%	65,20%	84,91%	75,06%	47,73%
	No	42,85%	66,36%	87,64%	77,00%	52,08%
RF4	Si	37,73%	62,65%	85,55%	74,10%	47,10%
	No	**46,71%**	**72,97%**	**88,37%**	**80,67%**	**56,96%**

Tabella 25. *Valutazione dell'effetto della PCA sulle prestazioni delle configurazioni RF1 ed RF4 nella classificazione binaria pixel per pixel.*

Si può notare come la rimozione della fase di costruzione delle *feature* tramite PCA porta ad un significativo miglioramento delle prestazioni. In particolare, il nuovo modello ottenuto a partire dalla configurazione RF4 risulta essere quello più prestante poiché massimizza tutti gli indicatori considerati (**Tabella 25**, riga 4).

3.4. Estrazione delle maschere di segmentazione automatica

Il modello RF4 allenato senza l'impiego della PCA è stato selezionato al fine di eseguire il compito finale di segmentazione automatica del tessuto metaplastico sulle immagini del Construction Set.

Le maschere ottenute sono delle immagini binarie in cui si rappresenta in bianco la presunta zona patologica.

Si riportano due esempi di confronti tra segmentazione manuale ed automatica nel caso di immagini patologiche nella **Figura 55**. Nel caso delle immagini sane non vi sono delle maschere manuali di riferimento ma ci si aspetta che l'output della segmentazione automatica coincida con una maschera completamente nera. Tuttavia, sono presenti errori di classificazione che portano a maschere automatiche contenenti piccoli oggetti rappresentati in bianco, simili a rumore sparso (Vedi **Figura 56**).

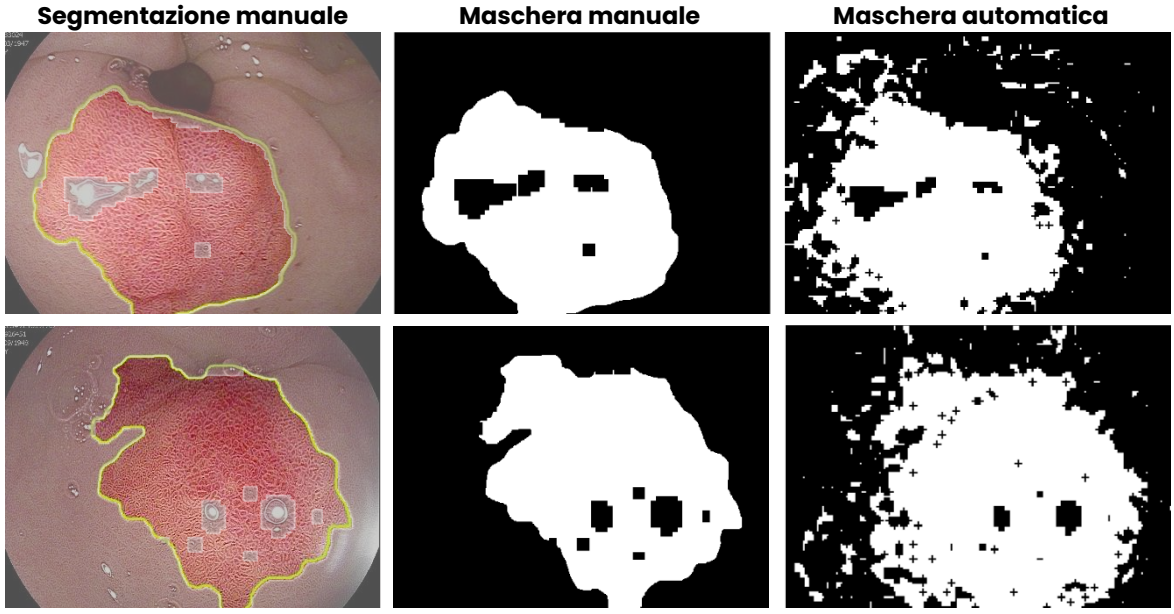

***Figura 55**. Immagini patologiche: segmentazione manuale ed automatica a confronto.*

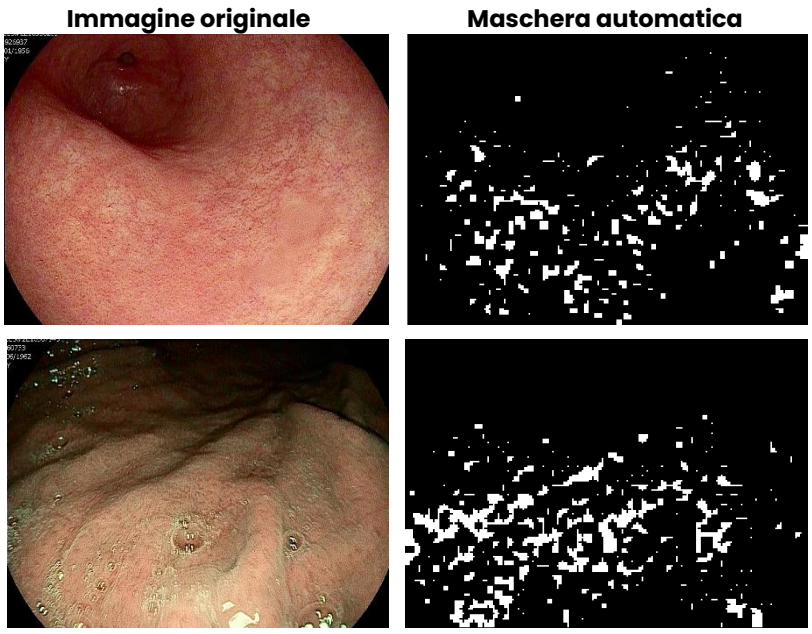

Figura 56. *Immagini sane: segmentazione automatica.*

Al fine di ottenere delle prestazioni qualitativamente e quantitativamente migliori nella segmentazione, le maschere automatiche sono state sottoposte a post-elaborazione con il principale obiettivo di ripulirle da output di classificazione rumorosa.

Nello specifico è stata imposta una soglia sulle minime dimensioni del singolo oggetto presente nell'immagine binaria. In particolare, vengono rimossi gli oggetti costituiti da un numero di pixel minore dell'1% del numero di pixel totali contenuti nell'immagine.

Le prestazioni sono state ricalcolate relativamente alla classificazione binaria pixel per pixel (Vedi **Tabella 26**, riga 2). Si nota, pertanto, come questa fase di post-elaborazione sia migliorativa per quasi la totalità degli indicatori ricavati.

	PCA	Mod.	Precisione	Recall	Specificità	Accuratezza bilanciata	F1 score
RF4	No	No	46,71%	**72,97%**	88,37%	80,67%	56,96%
	No	***Si***	**55,90%**	71,43%	**92,13%**	**81,78%**	**62,72%**

Tabella 26. *Confronto performance del modello RF4 senza PCA prima e dopo la post-elaborazione (Mod.) delle maschere automatiche.*

Dal punto di vista delle maschere delle immagini patologiche, dalla **Figura 57**, si può notare come vi sia un'effettiva rimozione del rumore di classificazione precedentemente presente attorno alla lesione correttamente riconosciuta.

Allo stesso modo, per quanto riguarda le immagini sane, si evidenzia una notevole riduzione delle misclassificazioni, talvolta fino all'ottenimento di una corretta classificazione dell'intera immagine (**Figura 58**, prima riga), altre volte con piccoli residui di aree erroneamente etichettate come patologiche (**Figura 58**, seconda riga).

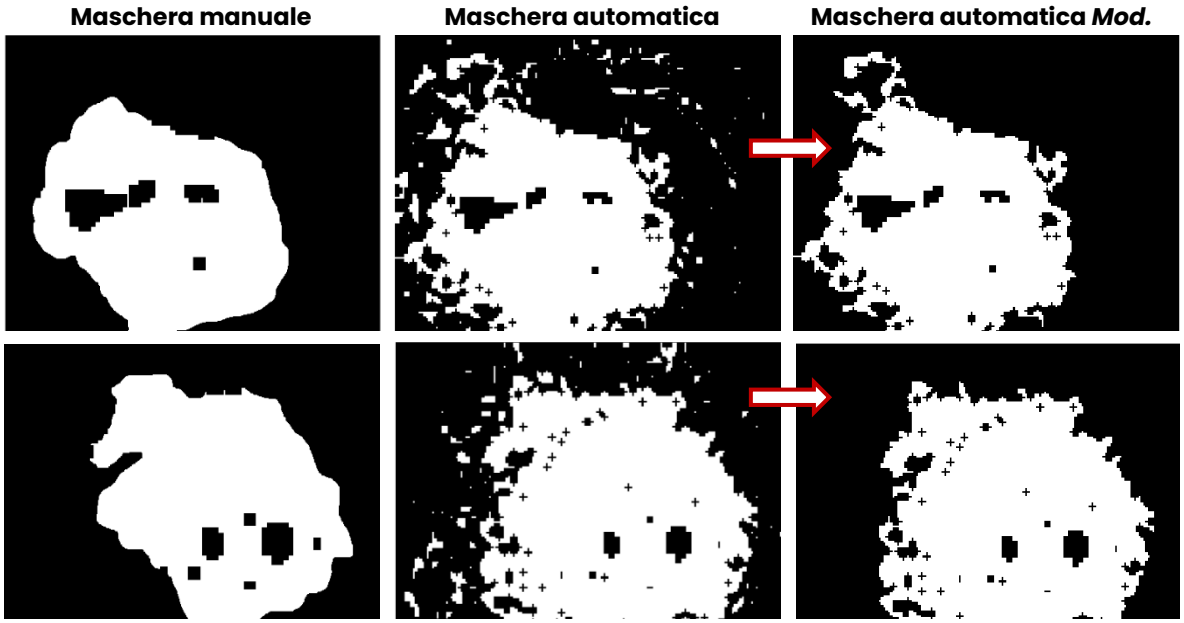

Figura 57. *Immagini patologiche: confronto maschere automatiche originali e post-elaborate (Mod.).*

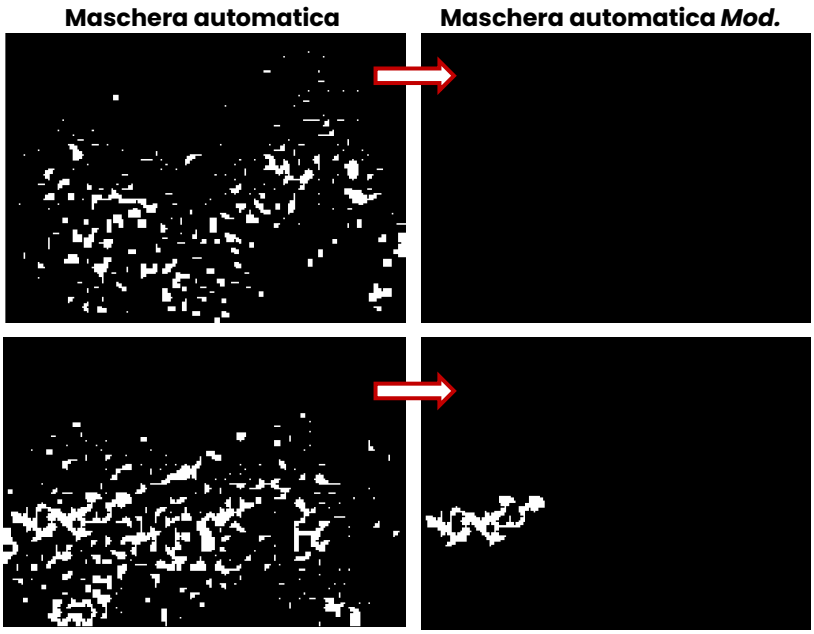

Figura 58. *Immagini sane: confronto maschere automatiche originali e post-elaborate (Mod.).*

Capitolo 4.
Conclusioni

Il principale obiettivo di questa tesi è valutare la possibilità di individuare in modo completamente automatico la presenza di metaplasie gastrointestinali a partire da immagini endoscopiche. Nello specifico, è stato valutato l'impiego di foreste di alberi decisionali allo scopo di discriminare le diverse ROI contenute nelle immagini, sulla base di diverse *feature* di texture.

Un primo punto critico esaminato in questo lavoro è la gestione delle ROI sane provenienti da immagini patologiche, contenenti uno o più focolai metaplastici. Infatti, tali ROI sono solo delle presunte ROI sane a seguito della segmentazione manuale condotta dal medico. L'esperto ha fatto in modo di delineare le lesioni metaplastiche con l'accortezza di racchiudere solo porzioni di tessuto che fossero riconducibili con alta probabilità a metaplasia. Questo ha fatto sì che le zone escluse da tale segmentazione potessero includere a loro volta un quantitativo non trascurabile di ROI patologiche.

Gli esperimenti effettuati sembrano portare a dei risultati migliori nel caso in cui le ROI in questione vengano omesse dai dati di *training* del classificatore, in modo da eliminare eventuali ambiguità in fase di costruzione dello stesso. Pertanto la configurazione del Construction Set di tipo CS1 descritta nel Paragrafo 2.4.1 risulta essere quella più efficace.

Un altro aspetto valutato nel corrente studio è stato l'effetto di una fase di *pre-processing* sulle immagini che fosse preliminare rispetto a quella di estrazione delle *feature*. Nello specifico, in questo lavoro sono stati implementati due passaggi di pre-elaborazione: l'equalizzazione dell'istogramma delle luminosità ed il filtraggio gaussiano.

Secondo quanto sperimentato attraverso la costruzione dei diversi modelli RF, si è notato che la presenza di queste due fasi di *pre-processing* permette di ottenere delle prestazioni migliori nella classificazione. Tale risultato può essere ricondotto al fatto che l'aumento globale del contrasto nelle immagini introdotto dall'equalizzazione e la riduzione del rumore dovuta al filtraggio hanno portano ad un riconoscimento più facilitato dei *pattern* presenti nelle acquisizioni.

Infine, è stato valutato l'effetto della Principal Component Analysis sulle prestazioni finali. È stato riscontrato che, se si omette la fase di applicazione della PCA sulle osservazioni, si ottiene un miglioramento significativo degli indicatori delle performance. Questo può essere dovuto alla possibile non linearità dei dati utilizzati per la classificazione.

Alla luce di tali osservazioni, il classificatore migliore risulta essere quello che mette insieme scelte di progetto coerenti con quanto analizzato. Si seleziona quindi il modello allenato a partire da un Construction Set in configurazione CS1 estratto da immagini sottoposte a *pre-processing* e sulle cui osservazioni non viene utilizzata la Principal Component Analysis.

Sulle maschere binarie ottenute attraverso l'impiego di tale classificatore si eliminano eventuali aree riconosciute come patologiche che siano più piccole dell'1% delle dimensioni complessive dell'immagine, in termini di numerosità di pixel.

Le prestazioni ottenute nella classificazione binaria pixel per pixel sono riassunte nella **Tabella 27**. In particolare, si ottengono un'accuratezza dell'$\mathbf{81,78\%}$ e un F1-score del $\mathbf{62,72\%}$.

	CS	***Pre-processing***	***PCA***	***Precisione***	***Recall***	***Specificità***	***Accuratezza bilanciata***	***F1 score***
RF4	CS1	Si	No	55,90%	71,43%	92,13%	**81,78%**	**62,72%**

Tabella 27. *Configurazione e prestazioni del modello finale.*

4.1. Possibili miglioramenti e sviluppi futuri

Al fine di migliorare le prestazioni ottenute si potrebbe pensare di valutare l'impiego di tipologie di classificatori diverse dalle foreste di alberi decisionali, quali, ad esempio, modelli di *Support Vector Machine* e *K-Nearest Neighbours*. In questo modo, si potrebbe selezionare il tipo di classificatore più adatto al compito di discriminazione in questione.

Allo stesso tempo, sarebbe opportuno analizzare diverse tecniche di selezione delle *feature* che possano individuare eventuali caratteristiche il cui apporto sia riconducibile a rumore in concomitanza ad un basso contributo informativo. Alternativamente, si potrebbero utilizzare tecniche di costruzione delle *feature* che abbiano un approccio di tipo non-lineare.

Un altro tentativo di miglioramento potrebbe interessare la fase di post-elaborazione delle maschere. Infatti, un'alternativa all'analisi della dimensionalità dei singoli oggetti etichettati come patologici, sarebbe quella di tentare un approccio basato sulla densità degli stessi nell'immagine binaria. Si nota, infatti, che quando si forniscono delle immagini sane, il classificatore tende a restituire delle maschere automatiche caratterizzate da piccole aree bianche sparse e distanti tra loro. Diversamente, per quanto riguarda le immagini patologiche, l'output spesso consiste in aree bianche più estese e compatte.

Il modello proposto in questo studio, se opportunamente sottoposto alle ulteriori ottimizzazioni suggerite, potrebbe essere implementato al fine di sviluppare uno strumento di supporto alla diagnosi endoscopica di metaplasie gastrointestinali.

A tale scopo, sarebbe utile calcolare un punteggio complessivo relativo all'immagine oggetto di indagine che sia indicativo della probabilità di presenza

di lesioni metaplastiche. Questo in concomitanza ad una sovrapposizione di una maschera binaria alla vista endoscopica analizzata, che possa fornire all'esperto una guida per l'individuazione di tali modificazioni.

Bibliografia

[1] P. Pimentel-Nunes, D. Libanio, J. Lage, D. Abrantes, M. Coimbra, G. Esposito e M. Dinis-Ribeiro, «A multicenter prospective study of the real-time use of narrow-band imaging in the diagnosis of premalignant gastric conditions and lesions,» *Endoscopy,* vol. 48, n. 08, pp. 723-730, 2016.

[2] M. M. Walker, «Is intestinal metaplasia of the stomach reversible?,» *Gut,* vol. 52, n. 1, pp. 1-4, 2003.

[3] S. J. Spechler, «Barrett esophagus and risk of esophageal cancer: a clinical review,» *Jama,* vol. 310, n. 6, pp. 627-636, 2013.

[4] N. R. Barrett, «Chronic peptic ulcer of oesophagus and "oesophagitis",» *British Journal of Surgery,* vol. 38, pp. 175-182, 1950.

[5] S. J. Spechler e R. K. Goya, «Barrett's esophagus,» *New England Journal of Medicine,* vol. 315, n. 6, pp. 362-371, 1986.

[6] B. J. Dicken, D. L. Bigam, C. Cass, J. R. Mackey, A. A. Joy e S. M. Hamilton, «Gastric adenocarcinoma: review and considerations for future directions,» *Annals of surgery,* vol. 241, n. 1, p. 27, 2005.

[7] Y. H. Park e N. Kim, «Review of atrophic gastritis and intestinal metaplasia as a premalignant lesion of gastric cancer,» *Journal of cancer prevention,* vol. 20, n. 1, p. 25, 2015.

[8] J. R. Jass e M. I. Filipe, «Intestinal metaplasia subtypes and cancer risk,» *Gastric carcinoma,* 1986.

[9] M. L. Caruso e R. Fiocca, «Gastriti croniche. Linee guida e criteri diagnostici minimi,» *Pathologica,* vol. 89, n. 6, pp. 592-597, 1997.

[10] W. K. Leung e J. J. Y. Sung, «Intestinal metaplasia and gastric carcinogenesis,» *Alimentary phamacology & therapeutics,* vol. 16, n. 7, pp. 1209-1216, 2002.

[11] S. J. Cho, I. J. Choi, M. C. Kook, B. H. Nam, C. G. Kim, J. Y. Lee, ... e Y. W. Kim, «Staging of intestinal- and diffuse-type gastric cancers with OLGA and

OLGIM staging systems,» *Alimentary pharmacology & therapeutics,* vol. 38, n. 10, pp. 1292-1302, 2013.

[12] P. Correa e M. B. Piazuelo, «The gastric precancerous cascade.,» *Journal of digestive diseases,* vol. 13, n. 1, pp. 2-9, 2012.

[13] R. M. Genta, M. Cassaro, E. R. Tschanz e M. Rugge, «Gastriti: Criteri classificativi,» *Giornale Italiano di Endoscopia Digestiva,* vol. 26, n. 4, pp. 257-263, 2003.

[14] R. J. Huang, A. Y. Choi, C. D. Truong, M. M. Yeh e J. H. Hwang, «Diagnosis and management of gastric intestinal metaplasia: current status and future directions,» *Gut and Liver,* vol. 13, n. 6, p. 596, 2019.

[15] PENTAX Medical, «PENTAX i-SCAN Functionality, Application and Technical Analysis,» 2013.

[16] L. G. Capelle, J. Haringsma, A. C. de Vries, E. W. Steyerberg, K. Biermann, H. van Dekken e E. J. Kuipers, «Narrow band imaging for the detection of gastric intestinal metaplasia and dysplasia during surveillance endoscopy,» *Digestive diseases,* vol. 55, n. 12, pp. 3442-3448, 2010.

[17] S. N. Hong, W. H. Choe, J. H. Lee, S. I. Kim, J. H. Kim, T. Y. Lee e C. S. Shim, «Prospective, randomized, back-to-back trial evaluating the usefulness of i-SCAN in screening colonscopy,» *Gastrointestinal Endoscopy,* vol. 75, n. 5, pp. 1011-1021, 2012.

[18] C. K. Lee, S. H. Lee e Y. Hwangbo, «Narrow-band imaging versus I-Scan for real-time histological prediction of diminutive colonic polyps: a prospective comparative study by using the simple unified endoscopic classification,» *Gastrointestinal Endoscopy,* vol. 74, n. 3, pp. 603-609, 2011.

[19] G. Cammarota, G. Ianiro, L. Sparano, R. La Mura, R. Ricci, L. M. Larocca e A. Gasbarrini, «Image-Enhanced Endoscopy with i-SCAN Technology for the Evaluation of Duodenal Villous Patterns,» *Digestive diseases and sciences,* vol. 58, n. 5, pp. 1287-1292, 2013.

[20] M. Leja, W. You, M. C. Camargo e H. Saito, «Implementation of gastric cancer screening-the global experience,» *Best practice & research Clinical gastroenterology,* vol. 28, n. 6, pp. 1093-1106, 2014.

[21] F. Van Der Sommen, S. Zinger, E. J. Schoon e P. H. De With, «Supportive automatic annotation of early esophageal cancer using local gabor and color features,» *Neurocomputing,* vol. 144, pp. 92-106, 2014.

[22] T. Kanesaka, T. C. Lee, N. Uedo, K. P. Lin, H. Z. Chen, J. Y. Lee e H. T. Chang, «Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging,» *Gastrointestinal endoscopy,* vol. 87, n. 5, pp. 1339-1344, 2018.

[23] R. Zhou, C. Yang, M. Q. H. Meng, G. Xu, C. Hu e B. Li, «Capsule endoscopy images classification by random forest and ferns,» in *2014 4th IEEE International Conference on Information Science and Technology*.

[24] M. Xu, W. Zhou, L. Wu, J. Zhang, J. Wang, G. Mu e H. Yu, «Artificial intelligence in the diagnosis of gastric precancerous conditions by image-enhanced endoscopy: a multicenter, diagnostic study (with video),» *Gastrointestinal Endoscopy,* vol. 94, n. 3, pp. 540-548, 2021.

[25] T. Yan, P. K. Wong, I. C. Choi, C. M. Vong e H. H. Yu, «Intelligent diagnosis of gastric intestinal metaplasia based on convolutional neural network and limited number of endoscopic images,» *Computers in Biology and Medicine,* vol. 126, p. 104026, 2020.

[26] MathWorks, «imbinarize. Binarize 2-D grayscale image or 3-D volume by thresholding,» *MathWorks Documentation for Image Processing Toolbox,* 2022.

[27] MathWorks, «graythresh. Global image threshold using Otsu's method,» *MathWorks Documentation for Image Processing Toolbox,* 2022.

[28] KDnuggets, «Who is your Golden Goose?: Cohort Analysis - KDnuggets,» 2022. [Online]. Available: https://www.kdnuggets.com/2019/05/golden-goose-cohort-analysis.html/2. [Consultato il giorno 20 Maggio 2022].

[29] J. E. Angus, «The probability integral transfor and related results,» *SIAM review,* vol. 36, n. 4, pp. 652-654, 1994.

[30] MathWorks, «histeq. Enhance contrast using histogram equalization,» *MathWorks documentation for Image Processing Toolbox,* Copyright 1993-2020 The MathWorks, Inc..

[31] MathWorks, «imgaussfilt. 2-D Gaussian filtering of images.,» *MathWorks documentation for Image Processing Toolbox,* Copyright 2014-2021 The MathWorks, Inc.

[32] MathWorks, «skewness. Skewness,» *MathWorks documentation for Statistics and Machine Learning Toolbox,* Copyright 1993-2018 The MathWorks, Inc.

[33] MathWorks, «kurtosis. Kurtosis,» *MathWorks documentation for Statistics and Machine Learning Toolbox,* Copyright 1993-2018 The MathWorks, Inc.

[34] R. C. Gonzalez, R. E. Woods e S. L. Eddins, «Digital Image Processing using MATLAB: entropy. Entropy of grayscale image.,» *MathWorks documentation for Image Processing Toolbox,* p. Chapter 11, Copyright 1993-2007 The MathWorks, Inc.

[35] M. Patio, B. Cramariuc, M. Gabbouj e A. Visa, «Rock texture retrieval using gray level co-occurrence matrix.,» *Proc. of 5th Nordic Signal Processing Symposium,* vol. 75, 2002.

[36] MathWorks, «graycoprops. Properties of gray-level co-occurrence matrix,» *MathWorks documentation for Image Processing Toolbox,* Copyright 2003-2020 The MathWorks, Inc.

[37] H. Zhang, C. L. Hung, G. Min, J. P. Guo, M. Liu e X. Hu, «GPU-accelerated GLRLM algorithm for feature extraction of MRI.,» *Scientific reports,* vol. 9, n. 1, pp. 1-13, 2019.

[38] G. Ciaburro, MATLAB for machine learning, Packt Publishing Ltd, 2017.

[39] MathWorks, «Hyperparameter Optimization in Classification Learner App,» *MathWorks Documentation.*

[40] MathWorks, «bayesopt. Select optimal machine learning hyperpatameters using Bayesian optimization,» *MathWorks Documentation for Statistics and Machine Learning Toolbox,* 2022.

[41] MathWorks, «Bayesian Optimization Algorithm,» *MatWorks Documentation,* 2022.

[42] MathWorks, «loss. Classification error,» *MathWorks Documentation for Statistics and Machine Learning Toolbox.*

[43] J. D. Rodriguez, A. Perez e J. A. Lozano, «Sensitivity analysis of k-fold cross validation in prediction error estimation,» *IEEE transactions on pattern analysis and machine intelligence,* vol. 32, n. 3, pp. 569-575, 2009.

[44] Y. Bengio e Y. Grandvalet, «No unbiased estimator of the variance of k-fold cross-validation,» *Journal of machine learning research,* pp. 1089-1105, 2004.

[45] J. Friedman, T. Hastie e R. Tibshirani, «Special invited paper. Additive logistic regression: As statistical view of boosting,» *Annals of statistics,* pp. 337-374, 2000.

[46] Y. D. Cai, K. Y. Feng, W. C. Lu e K. C. Chou, «Using LogitBoost classifier to predict protein structural classes,» *Journal of theoretical biology,* vol. 238, n. 1, pp. 172-176, 2006.

[47] MathWorks, «Choose Classifier Options,» *MathWorks Documentation.*

[48] J. P. Mower, «PREP-Mt: predictive RNA editor for plant mitochondrial genes,» *BMC bioinformatics,* vol. 6, n. 1, pp. 1-15, 2005.

[49] Y. Sasaki, «The truth of the F-measure,» *Teach tutor mater,* vol. 1, n. 5, pp. 1-5, 2007.

[50] MathWorks, «bestPoint. Best point in a Bayesian optimization according to a criterion,» *MathWorks Documentation for Statistics and Machine Learning Toolbox.*

[51] Y. Zhang, F. Li, F. Yuan, K. Zhang, L. Huo, Z. Dong, ... e L. Shen, «Diagnosing chronic atrophic gastritis by gastroscopy using artificial intelligence,» *Digestive and Liver Disease,* vol. 52, n. 5, pp. 566-572, 2020.

[52] J. G. Fox e T. C. Wang, «Dietary factors modulate Helicobacter-associated gastric cancer in rodent models,» *Toxicologic pathology,* vol. 42, n. 1, pp. 162-181, 2014.

[53] J. Lage, P. Pimentel-Nunes, P. C. Figueiredo, D. Libanio, I. Ribeiro, M. Jacome e M. Dinis-Ribeiro, «Light-NBI to identify high-risk phenotypes for gastric

adenocarcinoma: do we still need biopsies?,» *Scandinavian journal of gastroenterology,* vol. 51, n. 4, pp. 501-506, 2016.

[54] M. Di Stefano, E. Miceli e G. R. Corazza, «Anatomia funzionale dello stomaco,» 2010.

Ringraziamenti

Questo spazio lo dedico alle persone che, con il loro supporto, mi hanno aiutata a portare a compimento questo percorso.

Ringrazio la mia relatrice, la professoressa Visintin, per la sua diponibilità e tempestività nel fornirmi strumenti, migliorie e suggerimenti utili alla stesura dell'elaborato. Un grazie anche per l'attenzione dedicatami in fase di organizzazione dell'esame finale, consentendo a mia sorella diversamente abile di assistere alla discussione di questo progetto di tesi.

Grazie al mio correlatore, il professor Pagana, e a Valeria della Fondazione Links i quali sono sempre stati pronti a darmi le giuste indicazioni in ogni fase della realizzazione di questo lavoro.

Un ringraziamento speciale va al dottor Alessandro Adriani dell'A.O. Ordine Mauriziano di Torino, fonte di ispirazione di questo progetto e figura guida importante.