

UNIVERSITA' DEGLI STUDI DEL MOLISE
DIPARTIMENTO DI MEDICINA E SCIENZE DELLA SALUTE "V. TIBERIO"



Dottorato di ricerca in Medicina Traslazionale e Clinica

XXXVIII CICLO

Settore scientifico disciplinare Malattie dell'Apparato Visivo MEDS-17/A

TESI DI DOTTORATO

**Radiomica e Intelligenza Artificiale per l'Analisi del Materiale Iperriflettente Sottoretinico
nelle Immagini OCT della Retina**

*Subretinal Hyperreflective material Analysis with Radiomics and artificial intelligence Evaluation
in retinal OCT imaging (SHARE Study)*

Coordinatore

Ch.mo Prof. Giovanni Scapagnini

Relatore

Ch.mo Prof. Ciro Costagliola

Candidata

Marina Concilio

Matr. 176182

A.A. 2024/2025

INDICE

1. INTRODUZIONE

- 1.1 Le patologie retiniche maculari ed il materiale iper-riflettente sotto retinico
- 1.2 La Degenerazione Maculare legata all'età
 - 1.2.1 Epidemiologia
 - 1.2.2 Fisiopatologia
 - 1.2.3 Fattori di rischio
 - 1.2.4 La semeiologia della DMLE
- 1.3 Le lesioni vitelliformi
- 1.4 La diagnosi differenziale tra la degenerazione maculare legata all'età e le lesioni vitelliformi acquisite
- 1.5 L'intelligenza artificiale e la radiomica: spunti di applicazione nelle patologie retiniche

2. LO STUDIO

- 2.1 Modello di Residual self-attention Vision Transformer per l'individuazione delle lesioni vitelliformi e le drusen nella DMLE
 - 2.1.1 Materiali e metodi
 - 2.1.1.1 Dataset
 - 2.1.1.2 Data augmentation
 - 2.1.1.3 Architettura del modello: Vision Transformer e Residual Self-Attention
 - 2.1.1.4 Modelli di confronto
 - 2.1.1.5 Valutazione delle metriche
 - 2.1.1.6 Grad-CAM
 - 2.1.2 Risultati
- 2.2 Applicazione della radiomica all'OCT strutturale per la caratterizzazione delle lesioni vitelliformi acquisite: uno studio pilota
 - 2.2.1 Materiali e Metodi
 - 2.2.1.1 Estrazione delle caratteristiche radiomiche
 - 2.2.1.2 Pipeline del machine learning
 - 2.2.2 Risultati

3. DISCUSSIONE

Bibliografia

1. INTRODUZIONE

1.1 Le patologie retiniche maculari e il materiale iper-riflettente sotto retinico

Le patologie retiniche maculari rappresentano un gruppo eterogeneo di patologie che costituiscono una delle principali cause di ipovisione e cecità a livello globale. Il loro impatto epidemiologico è in costante aumento, parallelamente all'invecchiamento della popolazione e alla crescente prevalenza di condizioni sistemiche croniche, quali il diabete mellito. Tali patologie determinano un rilevante *burden* clinico e socioeconomico, incidendo in maniera significativa sulla qualità di vita dei pazienti e sui costi diretti e indiretti a carico dei sistemi sanitari nazionali.

L'esame strumentale fondamentale e imprescindibile per la diagnosi di una maculopatia è la tomografia a coerenza ottica (anche detta tomografia ottica computerizzata, nonché più comunemente OCT). L'OCT è una tecnica di imaging non invasiva, che permette una visualizzazione multistrato dei tessuti oculari. Sebbene ad oggi, grazie al progredire della tecnologia, l'OCT sia un esame di riferimento anche per lo studio del segmento anteriore dell'occhio, la sua applicazione cardine resta la valutazione del segmento posteriore e, più nello specifico, della macula e della papilla del nervo ottico [1]. Basato sul principio dell'interferometria a bassa coerenza, l'OCT è in grado di calcolare il tempo di ritardo e l'intensità della luce retrodiffusa dai tessuti a diverse profondità, così da poter elaborare immagini dettagliate ad alta risoluzione dei tessuti. Rispetto alla sua introduzione negli anni '90, le tecnologie più recenti, come lo *Spectral-Domain OCT* (SD-OCT) e lo *Swept-Source OCT* (SS-OCT), utilizzano una tecnica di analisi dello spettro dell'interferenza tramite un interferometro di Fourier, che consente una maggiore velocità di acquisizione e una risoluzione assiale dell'ordine di pochi micrometri [2,3]. In questo modo l'OCT permette di visualizzare con estrema precisione tutti gli strati retinici (Figura 1).

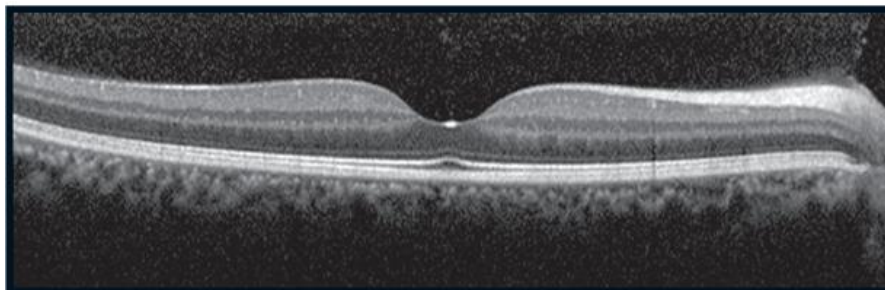


Figura 1 Scansione OCT strutturale di una macula con depressione foveale in esempio di occhio sano.

Un comune, ma non specifico, *biomarker* osservabile all'esame OCT è il materiale iperiflettente sottoretinico (comunemente noto come “*subretinal hyperreflective material*” – *SHRM*), la cui importanza è stata progressivamente sottolineata dalla letteratura scientifica.

Costituito da materiale di deposito localizzato nello spazio sottoretinico (i.e. tra l'epitelio pigmentato retinico (EPR) e la neuroretina) è associato a molteplici eziologie: la degenerazione maculare legata all'età (DMLE) in forma secca e/o umida (ossia complicata dalla presenza di una neovascolarizzazione), il materiale drusenoidale, le lesioni vitelliformi, le emorragie retiniche, le patologie infiammatorie, le patologie dell'interfaccia vitreo-retinica [4]. La presenza di *SHRM* all'esame OCT richiede una valutazione complessiva, con un approccio definito di "*multimodal imaging*" retinico al fine di definire una diagnosi e, soprattutto, di discriminare la presenza o meno di una neovascolarizzazione retinica. Ad oggi, grazie all'introduzione dell'OCT angiografia (OCTA), è possibile rilevare, nella maggior parte dei casi, la presenza o meno di un anomalo *net* vascolare, così da poter classificare lo *SHRM* nella forma neovascolare e non-neovascolare, differenziazione fondamentale per la scelta e il timing terapeutici, nonché la prognosi della patologia. Talvolta, però, i risultati degli esami OCT e OCTA non sono risolutivi, a causa di possibili artefatti, di una ridotta qualità delle immagini, di una scarsa compliance del paziente. In tal caso, è necessario effettuare degli esami invasivi, come l'angiografia con fluoresceina (FAG) e l'angiografia con il verde di indocianina (ICGA). Tali esami permettono lo studio della vascolarizzazione retinica e coroideale a seguito dell'iniezione intravenosa di coloranti vegetali naturali. Talvolta, è necessario far riferimento anche ad un'altra modalità aggiuntiva di *imaging* retinico, nota come *Multicolor imaging*, per poter discriminare un'emorragia retinica da un'essudazione retinica. È, quindi, chiaro come la presenza dello *SHRM* all'OCT, sia un biomarker importante che richiede l'inizio di un iter diagnostico e di diagnosi differenziale che implica l'utilizzo di diverse modalità di acquisizione delle immagini retiniche, il cosiddetto *multimodal imaging* [5]. Questo processo può risultare indaginoso, prolisso, non sempre prontamente disponibile, e soprattutto, legato ad un *expertise* da parte di chi eroga l'iter diagnostico.

In considerazione dell'elevata prevalenza dello *SHRM* come riscontro occasionale e dell'impatto che l'iter diagnostico può avere in termini di costi sanitari, difficoltà nella definizione diagnostica, ritardi nella diagnosi e conseguente scelta terapeutica, si è deciso di intraprendere uno studio clinico volto all'applicazione dell'intelligenza artificiale all'analisi di scansioni di OCT strutturale. L'obiettivo dello studio è stato quello di valutare l'efficacia e le prestazioni di diversi modelli di intelligenza artificiale nel discriminare alcune diagnosi differenziali e nel supportare il percorso diagnostico, con l'intento di ridurre il ricorso al multimodal imaging, approccio spesso *time-consuming* e non sempre prontamente disponibile in un singolo ambulatorio di retina medica.

Nello specifico, lo studio si concentra su due capitoli fondamentali della patologia maculare: la degenerazione maculare legata all'età e le lesioni vitelliformi, condizioni che condividono caratteristiche morfologiche sovrapponibili all'OCT ma che differiscono significativamente per prognosi, follow-up e strategie terapeutiche.

1.2 La Degenerazione maculare legata all'età

1.2.1 Epidemiologia

La degenerazione maculare legata all'età è una patologia cronica, degenerativa, acquisita della macula interessante gli individui di età superiore ai 50 anni. Rappresenta, ad oggi, la prima causa di cecità negli adulti a livello Mondiale, interessando 200 milioni di persone, con una distribuzione geografica disomogenea [6,7]. La prevalenza di cecità e di ipovisione, moderata e severa, legata alla DMLE, infatti, è elevata nella popolazione anziana e nei paesi con popolazioni più longeve. A partire dal 1999 al 2020, si è registrata una riduzione di circa il 30% della prevalenza di cecità legata alla DMLE, con un incremento dei casi di ipovisione. Questa tendenza non è legata ad una riduzione della prevalenza della patologia, quanto piuttosto all'introduzione di metodiche di prevenzione e di terapia più efficienti [8]. A partire dal 2004, infatti, l'introduzione delle terapie con anticorpi diretti contro il VEGF (*vascular endothelial growth factor*) come prima linea di trattamento della DMLE umida, ha rivoluzionato l'approccio terapeutico e la prognosi visiva dei pazienti con tale diagnosi [9]. Ad oggi, si stanno effettuando passi avanti anche nell'individuazione di possibili terapie per la DMLE secca nella sua forma atrofica avanzata, nota come atrofia a carta geografica [10, 11]. L'atrofia a carta geografica rappresenta uno stadio avanzato di DMLE secca, che comporta una progressiva, irreversibile morte dei fotorecettori degli strati retinici esterni, causando una profonda ipovisione centrale [12]. La DMLE secca rappresenta la forma più comune di DMLE, interessando circa l'80-90% dei pazienti con diagnosi di DMLE [13]. A livello globale l'atrofia a carta geografica interessa circa 5 milioni di persone nel Nord America ed è responsabile di circa il 20% dei casi di cecità legale [14].

1.2.2 Fisiopatologia

La DMLE è una patologia complessa, multifattoriale caratterizzata da suscettibilità genetica, infiammazione cronica, stress ossidativo, disfunzione mitocondriale e stretta correlazione con alcune varianti in CFH, ARMS2/HTRA1, e il fattore C3 del complemento, evidenziata dagli studi *genome-wide* [15,16].

La patogenesi della DMLE secca inizia con un ispessimento di uno specifico strato retinico, la membrana di Bruch (MB), per un accumulo di materiale ricco di lipidi e proteine, derivante da

un alterato metabolismo del turnover del materiale di scarto dei fotorecettori. L'accumulo di questo materiale altera il normale flusso e scambio di nutrienti e materiale di scarto tra la coroide (fonte vascolare principale) e la retina, compromettendo la funzione dell'epitelio pigmentato retinico (EPR). Di conseguenza, si verifica un aumento dello stress ossidativo all'interno delle cellule dell'EPR con progressivo accumulo di lipofusina (prodotto di una degradazione lisosomiale incompleta), che altera ulteriormente la funzione lisosomiale ed il metabolismo cellulare. Inoltre, in risposta allo stress cellulare, le cellule dell'EPR possono rilasciare vescicole, come gli esosomi, che si crede contribuiscano alla formazione del deposito sotto l'EPR. Infine, da analisi effettuate sulle drusen (i.e. materiale accumulato a livello sottoretinico), si è evidenziato la presenza di proteine derivanti dal siero, suggerendo che, in correlazione alla disfunzione della MB, sostanze dal circolo vascolare sistemico possono accumularsi nello spazio sotto l'EPR e favorire ulteriormente l'accumulo di materiale di deposito [17].

Per la patogenesi della DMLE vascolare, numerosi studi scientifici hanno dimostrato il ruolo centrale del VEGF. Quest'ultimo promuove l'angiogenesi, induce un aumento della permeabilità vascolare in risposta all'ipossia, e determina, così, fuoriuscita di fluido dal circolo vascolare locale ed accumulo intra- e/o sotto-retinico, con alterazione della normale architettura retinica e conseguente progressiva perdita dell'acuità visiva.

1.2.3 Fattori di rischio

La DMLE è una patologia poligenica e multifattoriale, caratterizzata da fattori di rischio modificabili e non modificabili [18]. Tra i fattori di rischio modificabili presentano un ruolo fondamentale:

- Il fumo, che aumenta il rischio di incidenza da 3 a 6 volte;
- Le patologie cardiovascolari e l'ipertensione;
- La disregolazione alimentare. Esistono, infatti, alcuni componenti che vengono introdotti con la dieta e che risultano indispensabili per la corretta anatomia e funzione della retina.

Ad oggi, è in discussione l'associazione tra aumentato rischio di DMLE e condizioni di dislipidemia, sia di ipercolesterolemia che aumento dei trigliceridi. In letteratura sono presenti studi con risultati contrastanti e che pongono l'attenzione sulla necessità di affrontare nuovi studi per stabilire il ruolo dei lipidi nella patogenesi della DMLE. Ciononostante, molteplici lavori scientifici hanno messo in evidenza come fattori genetici predisponenti siano legati al

metabolismo lipidico. Nel 1998 due studi hanno dimostrato, in maniera indipendente, come l'isoforma $\epsilon 4$ dell'Apolipoproteina E (ApoE) abbia un ruolo protettivo nei confronti della DMLE. L'ApoE è stata identificata con studi di immunoistochimica all'interno delle drusen di occhi di pazienti affetti da DMLE. È una proteina che svolge un ruolo fondamentale nel flusso di colesterolo a livello retinico e svolge anche una funzione di *chaperonne* nel citoplasma cellulare [19]. Recentemente, da una meta-analisi su studi di Genome-wide association study (GWAS), sono stati evidenziati altri loci di geni implicati nel metabolismo lipidico che risultano essere associati a DMLE: 3 loci di rischio (CYP26A1, C1R, andHIC1), un nuovo locus, CASP10, associato precedentemente ad elevate livelli di trigliceridi [20].

Da un punto di vista fisiopatologico, i lipidi retinici svolgono un ruolo chiave nell'omeostasi fotorecettoriale, nella modulazione delle vie infiammatorie e ossidative implicate nella DMLE. I fosfolipidi ricchi in DHA contribuiscono all'integrità dei segmenti esterni dei fotorecettori e modulano la produzione di mediatori pro-risolutivi (resolvine, protectine), mentre alterazioni nel metabolismo lipidico — incluse disfunzioni nel trasporto lipoproteico dell'EPR — sono associate all'accumulo di materiale drusenoide. Nonostante solide basi biochimiche e dati osservazionali favorevoli, gli studi d'intervento non hanno ancora dimostrato un chiaro beneficio clinico della supplementazione lipidica; tuttavia, il metabolismo dei lipidi rimane una traiettoria di ricerca promettente per strategie preventive e per la modulazione del microambiente retinico nella DMLE. Lo studio AREDS2, il più importante trial clinico multicentrico randomizzato nella DMLE (con l'AREDS1), ha valutato l'impatto dell'integrazione di luteina e zeaxantina e dell'aggiunta di acidi grassi polinsaturi a lunga catena (LC-PUFA) omega-3, in particolare DHA ed EPA, sulla progressione della DMLE. I risultati hanno evidenziato che la formula integrativa AREDS2 con i suoi componenti (ad alto dosaggio di antiossidanti - vitamina C 500 mg, vitamina E 400 UI - in combinazione con zinco e i componenti suddetti) riduce significativamente il rischio di progressione verso DMLE avanzata e di perdita visiva moderata o maggiore nei soggetti a rischio elevato [21].

1.2.4 La semeiologia della DMLE

La degenerazione maculare legata all'età è, più appropriatamente, definibile come uno spettro continuo di alterazioni degenerative maculari; tuttavia, per finalità cliniche e didattiche, viene tradizionalmente distinta in due forme principali: la forma non neovascolare (anche detta “secca”) e la forma neovascolare (o umida). La forma secca è caratterizzata dalla presenza di drusen e da alterazioni dell'EPR, mentre la forma umida è definita dallo sviluppo di neovascolarizzazioni coroideali e/o retiniche, spesso associate a essudazione, emorragie e fibrosi sottoretinica [22].

Le drusen rappresentano il segno clinico e strutturale patognomonico della DMLE non neovascolare (Figura 2). Esse comprendono diverse tipologie, classificabili in base alle dimensioni (piccole, intermedie, grandi), alla morfologia e alla localizzazione anatomica, come evidenziabile all'esame di OCT [23]. Indipendentemente dalla classificazione, le drusen sono costituite da materiale di accumulo extracellulare localizzato tra la membrana di Bruch e l'EPR e si presentano come strutture iper-riflettenti all'esame OCT strutturale.

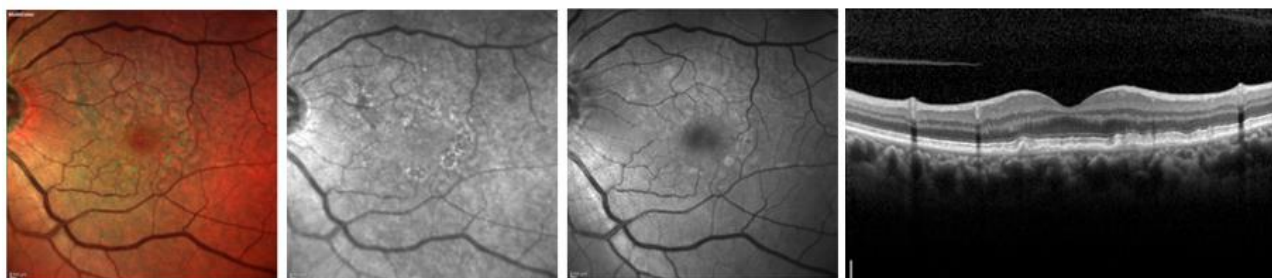


Figura 2 Esempio di multimodal imaging in caso di drusen maculari. Da sinistra verso destra: esame Multicolor presenta lesioni rotondeggianti giallastre, Near Infrared, autofluorescenza, OCT strutturale.

Le alterazioni pigmentarie comprendono sia fenomeni di iperpigmentazione sia aree di ipopigmentazione, espressione di disfunzione o perdita dell'EPR. La coesistenza di drusen e anomalie pigmentarie rappresenta un importante fattore prognostico negativo, associato a un aumentato rischio di progressione della forma secca verso stadi avanzati della malattia, quali l'atrofia a carte geografica e/o la DMLE neovascolare [24, 25]. Tale rischio giustifica, dal punto di vista clinico, la necessità di un monitoraggio strutturato e ravvicinato, al fine di intercettare precocemente i segni di progressione e impostare tempestivamente le strategie terapeutiche appropriate.

1.3 Le lesioni vitelliformi

Il termine lesioni vitelliformi (LV) identifica un gruppo eterogeneo di alterazioni retiniche caratterizzate dall'accumulo di materiale giallastro a disposizione cupoliforme, osservabile all'esame del fondo oculare e/o alla retinografia a colori. Dal punto di vista dell'*imaging* retinico multimodale, tali lesioni corrispondono a un'area di iper-autofluorescenza all'esame dell'autofluorescenza del fondo (AF) e alla presenza di SHRM, in genere in assenza di fluido sottoretinico, all'esame OCT. Istologicamente, il materiale vitelliforme è costituito prevalentemente da lipofuscina e da prodotti di degradazione dei segmenti esterni dei fotorecettori, espressione di una disfunzione primaria o secondaria dell'EPR. L'eziologia delle LV è complessa e multifattoriale: esse possono presentarsi in

forme ereditarie oppure, più frequentemente, in forme sporadiche o acquisite, interessando un ampio spettro di età [26].

Nei pazienti in età giovanile, la distrofia maculare di Best è la principale eziologia; una patologia ereditaria a trasmissione autosomica dominante, più comunemente associata a mutazioni del gene BEST1, coinvolto nei meccanismi di trasporto ionico e di omeostasi dell'EPR. Mutazioni del gene PRPH2 sono, invece, tipicamente correlate a un esordio più tardivo e a fenotipi clinici più variabili, talora sovrapponibili alle forme vitelliformi dell'adulto [27].

Nei soggetti adulti, generalmente di età superiore ai 60 anni, il riscontro di LV è relativamente comune e rientra nel quadro delle lesioni vitelliformi acquisite (LVa). In questi casi, le LV sono considerate un segno secondario a molteplici condizioni patologiche, con meccanismi fisiopatologici parzialmente sovrapponibili a quelli alla base dello SHRM. Le principali associazioni cliniche includono la DMLE, l'accumulo di materiale drusenoide, le neovascolarizzazioni coroideali e/o retiniche, la fibrosi sottoretinica, lo spettro delle pachicoroidi, le patologie dell'interfaccia vitreo-retinica con trazione maculare, nonché sindromi tossiche o paraneoplastiche e condizioni più rare quali le strie angioidi [28, 29].

Le LV sono caratterizzate da un'evoluzione dinamica nel tempo, generalmente lenta e progressiva. Nelle fasi iniziali si osserva un incremento delle dimensioni del materiale vitelliforme sottoretinico, cui segue una graduale fase di riassorbimento, con la comparsa, all'esame OCT, di una cavità otticamente vuota. Contestualmente, l'imaging OCT può documentare la presenza di alterazioni strutturali a carico degli strati retinici esterni e, in particolare, della membrana limitante esterna (ELM), della zona ellissoide (EZ), fino alla possibile evoluzione verso l'atrofia dell'EPR (Figura 3).

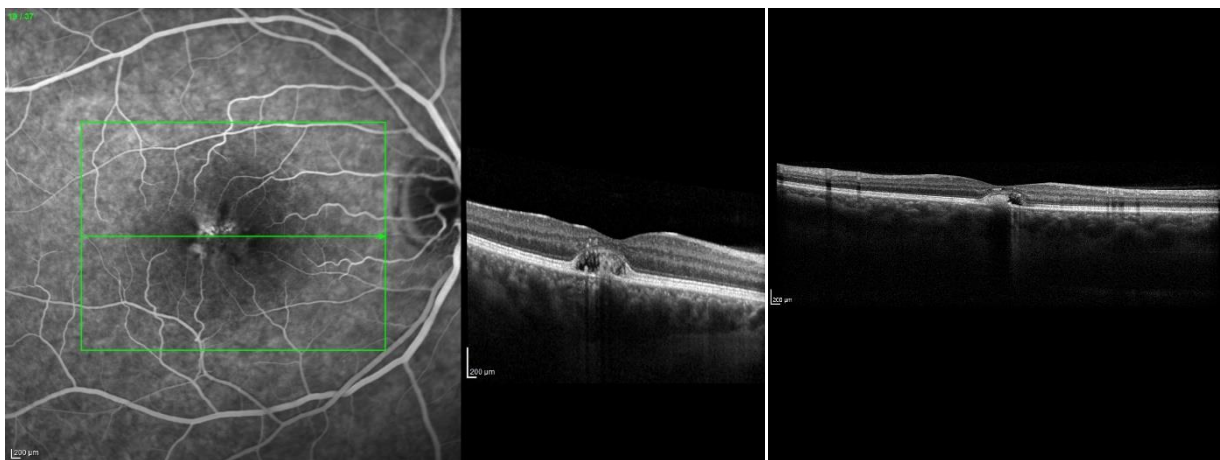


Figura 3 Esempio di due scansioni in OCT strutturale in presenza di lesione vitelliforme. L'esame OCT presenta normale spessore maculare con parziale perdita del fisiologico pit foveale. Strati retinici interni normoriflettenti e normorappresentati. Strati retinici esterni (ELM, giunzione IS/OS, EPR) parzialmente riconoscibili al di sopra del materiale vitelliforme, iperiflettente, in fase di riassorbimento, con adiacente spazio otticamente vuote. Pachicoroide.

Tali cambiamenti strutturali sono frequentemente associati alla comparsa di sintomi visivi, in particolare alla riduzione dell'acuità visiva [4].

La compresenza dell'insorgenza di sintomi visivi e l'evidenza all'*imaging* retinico delle variazioni morfologiche delle LV costituiscono elementi centrali per l'impostazione di una diagnosi differenziale accurata, soprattutto per distinguere forme neovascolari da forme non neovascolari. In questo contesto, l'*imaging* multimodale riveste un ruolo essenziale nell'interpretazione eziologica, nella diagnosi differenziale, nel monitoraggio e nella valutazione della risposta terapeutica delle patologie retiniche. Gli esami FAG e ICGA possono risultare utili nei casi diagnosticamente equivoci [30] (Figura 4).

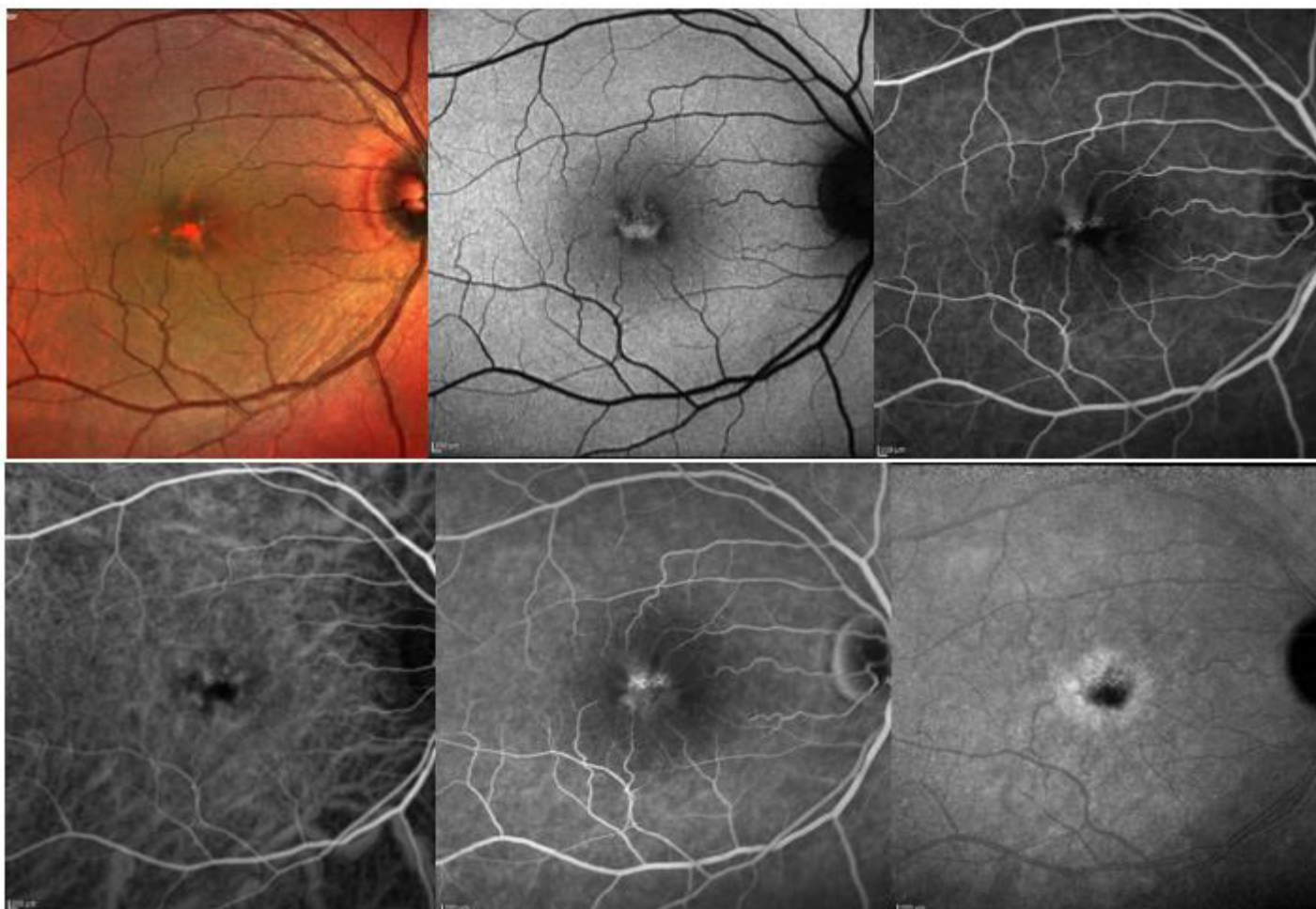


Figura 4 Esempio di multimodal imaging retinico in caso di LVA. Prima riga, da sinistra verso destra: lesione giallastra in regione maculare all'esame Multicolor; area iperautofluorescente all'esame dell'autofluorescenza; FAG tempi precoci: iperfluorescenza localizzata da pooling. Seconda riga, da sinistra verso destra: FAG tempi tardivi, lieve riduzione della fluorescenza dei tempi precoci. ICGA tempi precoci e tardivi lieve ipercianescenza tardivi che non aumenta di dimensioni nel tempo con ipocianescenza centrale per effetto mascheramento da presenza del materiale vitelliforme.

L'angiografia OCT rappresenta attualmente uno strumento non invasivo di grande utilità nella distinzione tra forme neovascolari e non; tuttavia, essa non è sempre dirimente (Figura 5).

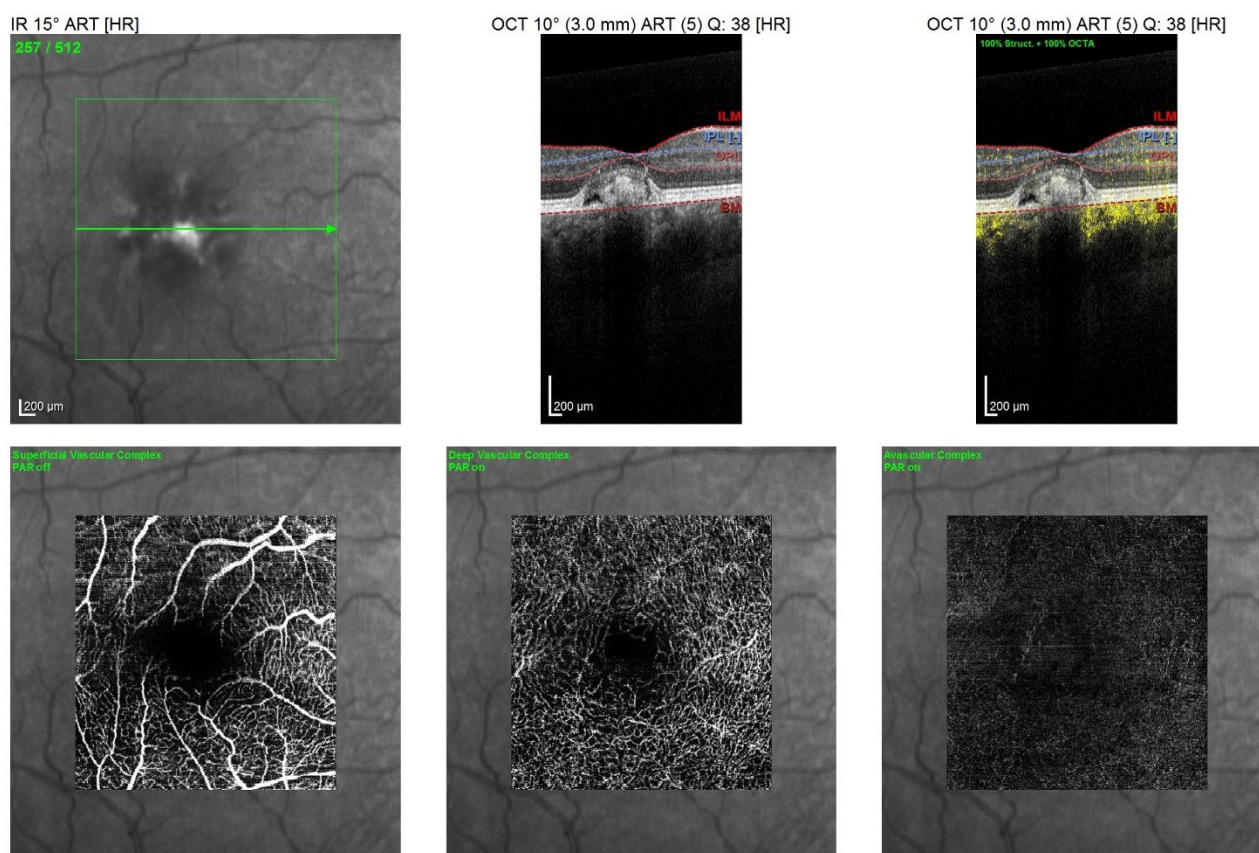


Figura 5 Esempio di OCT strutturale e OCTA in presenza di LVa non complicata da neovascolarizzazione. Assenza di anomalie vascolari.

L'analisi dell'*imaging* retinico presenta alcune limitazioni, tra cui la componente di soggettività interpretativa e la necessità di eseguire esami multipli, non sempre prontamente disponibili nella pratica clinica.

1.4 La diagnosi differenziale tra la degenerazione maculare legata all'età e le lesioni vitelliformi acquisite

La diagnosi differenziale tra la DMLE e le LVa rappresenta una sfida clinica rilevante, in particolare nei pazienti adulti e anziani, nei quali le LVa possono mimare le manifestazioni neovascolari della DMLE. Entrambe le condizioni possono presentare reperti sovrapponibili al fondo oculare e all'*imaging* OCT strutturale, rendendo necessario un approccio basato sul *multimodal imaging* e sull'inquadramento clinico complessivo. L'OCTA consente, nella maggior parte dei casi, di identificare la presenza di un *net* vascolare nelle forme umide di DMLE, assente

nelle LVA non complicate. Un corretto inquadramento diagnostico risulta fondamentale per evitare trattamenti inappropriati e per impostare un adeguato follow-up clinico-strumentale.

Negli ultimi anni, l'analisi delle immagini mediche con l'applicazione dell'intelligenza artificiale (IA) ha permesso di migliorare l'interpretazione dei dati, e quindi, la performance sulla diagnosi, la prognosi, la risposta al trattamento e la definizione degli intervalli di monitoraggio del paziente, divenendo di grande impatto e supporto per la gestione clinica.

1.5 L'intelligenza artificiale e la radiomica: spunti di applicazione nelle patologie retiniche

Recentemente, l'IA è stata ampiamente applicata all'analisi delle immagini in ambito oculistico, permettendo di automatizzare l'analisi di alcuni esami, come fotografie del fondo oculare, OCT e FAG, per supportare la diagnosi e il monitoraggio di alcune patologie oculari. Un ruolo centrale è svolto dal deep learning (DL), una classe avanzata di tecniche di apprendimento automatico basate su reti neurali profonde, capaci di apprendere rappresentazioni gerarchiche dei dati attraverso molteplici livelli di astrazione. A differenza dei metodi tradizionali di machine learning, il DL consente l'elaborazione diretta di dati ad alta dimensionalità senza la necessità di una ingegnerizzazione manuale delle caratteristiche (*feature engineering*), permettendo l'identificazione automatica di *patterns* complessi e strutture latenti, mediante trasformazioni non lineari e proiezioni in spazi di rappresentazione a dimensionalità ridotta. Tali proprietà hanno reso il DL particolarmente efficace in numerosi ambiti applicativi, tra cui la *computer vision*, l'elaborazione del linguaggio naturale e il riconoscimento vocale, con prestazioni superiori rispetto agli approcci convenzionali [31].

All'interno del paradigma del DL, le reti neurali convoluzionali (*Convolutional Neural Networks* - CNN) rappresentano l'architettura di riferimento per l'analisi delle immagini, grazie alla loro capacità di sfruttare la struttura spaziale locale dei dati visivi. L'architettura delle CNN si basa su strati convoluzionali che applicano filtri condivisi su regioni locali dell'immagine, consentendo l'estrazione gerarchica di caratteristiche visive che spaziano da elementi di basso livello, come bordi e *texture*, fino a rappresentazioni di alto livello, quali strutture anatomiche e alterazioni patologiche, a prescindere dal loro orientamento, posizione o scale [32,33].

Numerosi studi hanno dimostrato elevate prestazioni delle CNN nella segmentazione, rilevazione automatica e classificazione di patologie comuni quali la retinopatia diabetica, il glaucoma, la DMLE e la retinopatia del prematuro, con accuratezze comparabili o superiori a quelle di esperti clinici [34-36]. Nonostante il successo delle CNN, tali architetture presentano

alcuni limiti intrinseci, in particolare nella valutazione di relazioni globali a lungo raggio all'interno delle immagini. Al fine di superare questi limiti, sono stati recentemente introdotti i *Vision Transformer* (ViT), modelli che utilizzano dei meccanismi di *self-attention* per catturare dipendenze globali tra regioni distanti [37]. I ViT suddividono l'immagine in *patches* e apprendono rappresentazioni globali senza vincoli di località, mostrando risultati promettenti anche nell'imaging medico. In ambito oftalmologico, modelli basati su ViT e architetture ibride CNN–Transformer sono e stanno emergendo come alternative o complementi alle CNN tradizionali, dimostrando un potenziale miglioramento nella classificazione e nella segmentazione di immagini OCT e del fondo oculare, soprattutto in presenza di strutture patologiche complesse e variazioni morfologiche sottili [38]. Nonostante ciò, le CNN rimangono il *backbone* più utilizzato nelle applicazioni cliniche correnti, mentre i *Vision Transformer* rappresentano una direzione di ricerca promettente per lo sviluppo di sistemi di intelligenza artificiale più robusti e generalizzabili.

Parallelamente a questi modelli, di recente insorgenza in oculistica, è l'approccio radiomico. La radiomica è una tecnica avanzata di analisi quantitativa, ad alte prestazioni, delle immagini mediche, che sfrutta varie tecniche di estrazione di *biomarkers* e di applicazione di metodi di IA per creare modelli diagnostici e prognostici [39]. La radiomica consente l'estrazione quantitativa di un ampio numero di caratteristiche (*features*), non visibili all'occhio umano, permettendo una caratterizzazione oggettiva e riproducibile dei fenotipi patologici. Come evidenziato dalla recente letteratura, la radiomica presenta un potenziale promettente nel campo dell'oftalmologia, facilitando la diagnosi, lo screening precoce e la prognosi della risposta al trattamento; rappresenta, inoltre, un approccio prezioso per la definizione di *imaging biomarkers* per migliorare l'approccio clinico nel contesto di molteplici patologie, migliorando complessivamente la prognosi dei pazienti [40].

2 LO STUDIO

2.1 Modello di Residual self-attention Vision Transformer per l'individuazione delle lesioni vitelliformi e le drusen nella DMLE

Obiettivo dello studio: L'obiettivo principale è stato lo sviluppo e la validazione di un sistema automatico basato su tecniche di machine learning per la classificazione e il riconoscimento di immagini OCT, finalizzato alla distinzione tra drusen associate a DMLE e lesioni vitelliformi acquisite (LVA), nonché alla discriminazione di tali condizioni rispetto ad immagini OCT di regioni maculari sane. Più precisamente, gli obiettivi specifici dello studio possono essere definiti, come di seguito:

1. Progettare e implementare un modello innovativo di DL basato sul *Residual Self-Attention Vision Transformer* (RS-A ViT) per l'analisi automatica di immagini OCT, in grado di riconoscere e classificare in modo accurato LVA, drusen maculari e casi normali (controllo).
2. Valutare le prestazioni del modello proposto al fine di determinarne l'efficacia nella classificazione delle principali alterazioni retiniche considerate.
3. Confrontare le prestazioni del modello RS-A ViT con quelle di architetture di DL consolidate, includendo modelli basati su CNN e Vision Transformer convenzionali, per verificarne il potenziale miglioramento rispetto agli approcci pre-esistenti.
4. Analizzare l'impatto dell'introduzione del modulo di *Residual Self-Attention* rispetto al meccanismo di *Self-Attention* standard, per dimostrarne il contributo nel preservare caratteristiche discriminative rilevanti per la classificazione di alterazioni retiniche sottili.
5. Garantire l'interpretabilità clinica del modello mediante l'applicazione di tecniche di *explainable artificial intelligence* per identificare le regioni delle immagini OCT maggiormente influenti nel processo decisionale del classificatore.
6. Valutare il potenziale clinico del sistema sviluppato come strumento di supporto alla pratica oftalmologica, con l'obiettivo di favorire una diagnosi più precoce, più accurata e contribuire, così, ad un possibile approccio di prevenzione con impatto sulla prognosi della patologia.

2.1.1 Materiali e Metodi

Si tratta di uno studio retrospettivo multicentrico finalizzato allo sviluppo e alla validazione di un modello di deep learning per la classificazione automatica di immagini OCT, con particolare riferimento alla distinzione tra LVA, drusen associate a DMLE e casi di macule sane.

2.1.1.1 Dataset

Le immagini OCT relative a pazienti affetti da LVA sono state raccolte presso due centri oftalmologici terziari: la clinica oculistica dell'Università Federico II di Napoli e il dipartimento di oculistica generale e pediatrica dell'Università di Medicina di Lublino in Polonia. Lo studio è stato condotto secondo i canoni della Dichiarazione di Helsinki, approvato dal comitato etico dell'Università di Lublino (identificativo: KE-0254/260/12/2022). Ogni partecipante incluso nello studio ha firmato un consenso informato.

Tutti i dati disponibili sono stati anonimizzati, in conformità alle normative vigenti in materia di protezione dei dati personali. Le scansioni delle LVA sono state acquisite con il medesimo dispositivo OCT (AngioVue OCT-A system - Optovue Inc., Fremont, CA, USA), secondo i protocolli clinici di routine dei centri coinvolti. Sono state considerate scansioni 6x6 mm B-scan volumetriche maculari centrate sulla fovea. Le immagini sono state analizzate, così da escludere quelle con una qualità non adatta per la successiva analisi, a causa di possibili artefatti da movimento durante la precedente acquisizione o per la presenza di immagini non correttamente a fuoco. Sono stati esclusi, inoltre, i casi con patologie retiniche concomitanti, precedenti interventi chirurgici oculari o immagini incomplete. Complessivamente, 125 immagini di LVA sono state considerate qualificabili per lo studio. Le immagini selezionate sono state sottoposte ad una fase di pre-processamento, comprendente: ridimensionamento delle immagini alla dimensione richiesta dall'architettura di rete, normalizzazione dell'intensità dei pixel, standardizzazione del formato di input. Nello specifico, le immagini OCT sono state processate mediante l'applicazione di due filtri, gaussiano e mediano, con l'obiettivo di ridurre il rumore di tipo *speckle*, possibile nelle immagini OCT, preservando le caratteristiche chiave delle immagini. Successivamente, è stata effettuata una selezione manuale di una regione di interesse (ROI) rettangolare comprendente l'area maculare; nelle immagini con LVA, tale procedura ha consentito di escludere le aree periferiche non rilevanti per il riconoscimento della patologia e, quindi, ai fini dell'analisi. Le immagini risultanti sono state ridimensionate a una risoluzione uniforme fissata di 224×224 pixel, in accordo con i requisiti di input delle reti neurali di classificazione utilizzate. Per ridurre la variabilità legata alle differenti condizioni di acquisizione e normalizzare i valori di intensità, i pixel di ciascuna immagine sono stati scalati nell'intervallo $[0,1]$ sulla base dei rispettivi valori minimo e massimo.

Il database delle immagini OCT dei casi controllo sani e dei casi con drusen in DMLE sono stati costruiti a partire dalle immagini rese disponibili dai due centri-studio e integrate da una coorte aggiuntiva proveniente da due dataset pubblici. In particolare, sono stati utilizzati i dataset NEH e UCSD (University of California San Diego), che hanno permesso di garantire un'adequata rappresentazione delle due classi, migliorando la solidità metodologica e la riproducibilità dello studio. Il dataset finale comprende 125 immagini di LVA, 577 immagini di drusen e 578 immagini di casi normali. L'intero campione è stato suddiviso in set di addestramento, validazione e test secondo una proporzione del 70%, 15% e 15%, rispettivamente, mantenendo l'equilibrio tra le classi. I dettagli relativi alla composizione e alla suddivisione del dataset sono rappresentati in Figura 6.

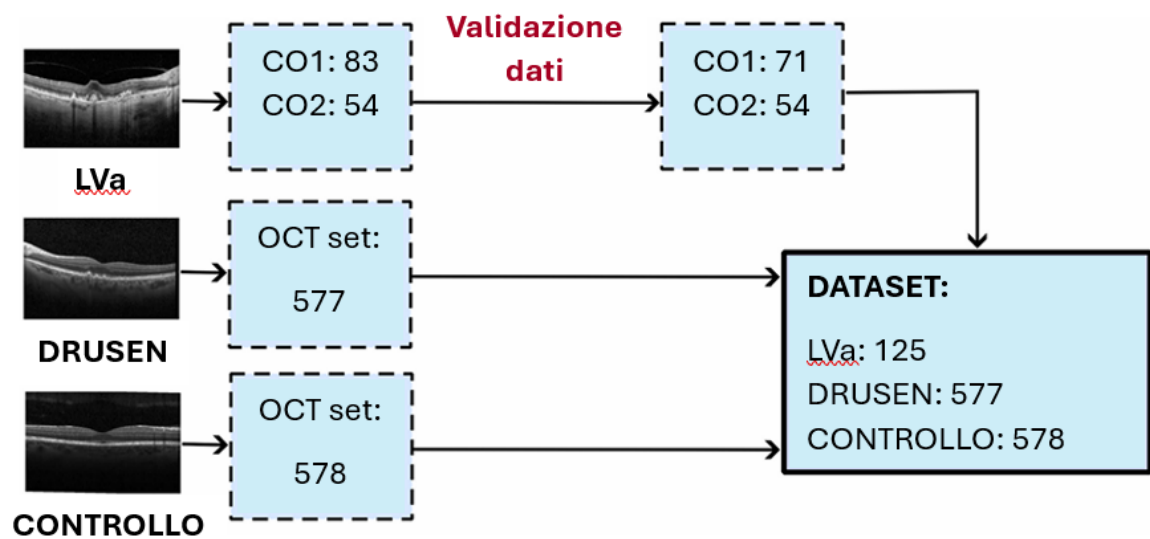


Figura 6 Struttura del dataset. Il dataset delle LVA è stato costruito a partire dalle immagini incluse nei due centri-studio: CO1 e CO2. Le immagini delle drusen e del controllo sono state integrate con un dataset pubblico. CO1: clinica oculistica Università Federico II; CO2: clinica oculistica Università di Lublino.

2.1.1.2 Data augmentation

Come parte integrante della fase di *pre-processing* è stata adottata la strategia di *data augmentation*. Quest'ultima è ampiamente utilizzata nella classificazione di immagini per migliorare le prestazioni e la capacità di generalizzazione dei modelli di apprendimento automatico e di DL. Tale approccio consiste nell'ampliare artificialmente il dataset di addestramento attraverso la generazione di nuovi campioni ottenuti, mediante l'applicazione di trasformazioni controllate alle immagini originali. L'aumento della variabilità dei dati consente di ridurre il rischio di *overfitting*, di rendere il modello più robusto alle comuni variazioni presenti nelle immagini del mondo reale, quali cambiamenti di illuminazione, scala, orientamento e deformazioni locali. Inoltre, la *data augmentation* permette di incrementare la numerosità del dataset senza la necessità di acquisire ulteriori immagini, migliorando così la capacità del modello di generalizzare correttamente su dati non utilizzati in fase di addestramento. In particolare, sono state applicate operazioni di rotazione, *scaling*, *crop and resize* casuale, variazioni fotometriche (*color jittering*), aggiunta di rumore gaussiano, deformazioni elastiche, tecniche di *random erasing* e *cutout*, ribaltamenti orizzontali e verticali. L'impiego combinato di tali trasformazioni ha consentito di aumentare la diversità dei campioni e di rendere il modello più stabile. La struttura finale del dataset ottenuta dopo l'applicazione delle procedure di *data augmentation* è riportata in Tabella 1.

Set	Classe	N. dati pre-augmentation	N. dati post-augmentation
Training	LVa	87	400
	Drusen	404	450
	Controllo	404	450
Validazione	LVa	19	-
	Drusen	86	-
	Controllo	87	-
Test	LVa	19	-
	Drusen	87	-
	Controllo	87	-

Tabella 1 Struttura dettagliata dei sottoinsiemi di dati utilizzati nello studio, con indicazione del numero di immagini per ciascuna classe (AVL, drusen e casi normali) nei set di addestramento, validazione e test, prima e dopo l'applicazione della procedura di data augmentation.

2.1.1.3 Architettura del modello: Vision Transformer e Residual Self-Attention

Il Vision Transformer rappresenta un'evoluzione nel campo della visione artificiale, adattando l'architettura dei *transformers*, originariamente sviluppata per l'elaborazione del linguaggio naturale, ai compiti di riconoscimento di immagini. Tale adattamento si basa sulla rappresentazione delle immagini come sequenze di *patches embeddings*, ossia sequenze di vettori, ottenute suddividendo l'immagine di input in regioni di dimensione fissa, successivamente elaborate attraverso una serie di blocchi *encoder* basati su meccanismi di *Multi-Head Self-Attention* (MHSA) e reti *feed-forward*. Nel presente studio è stata adottata un'architettura Vision Transformer con meccanismi di Residual Self-Attention (RS-A ViT).

Il modello ViT utilizzato è progettato per elaborare immagini OCT RGB di dimensioni 224×224 pixel. In fase iniziale una convoluzione bidimensionale (Conv2D) suddivide l'immagine in *patches* non sovrapposte di dimensione 16×16 pixel, ciascuna delle quali viene proiettata in uno spazio vettoriale a dimensione fissa, generando una sequenza di *token* di input per i blocchi *encoder*. Ciascun *encoder* è costituito da moduli di attenzione *multi-head*, reti *feed-forward*, normalizzazione dei *layers* e connessioni residue, che consentono una propagazione stabile del gradiente e un apprendimento efficace delle relazioni globali tra le diverse regioni dell'immagine.

Nel corso dell'addestramento, sono stati presi in considerazione diversi iperparametri, tra cui il numero di *encoder*, la dimensionalità dei vettori, la dimensione delle reti *feed-forward*, la dimensione delle *patches* e il tasso di *dropout*. L'ottimizzazione di tali parametri è stata condotta monitorando le prestazioni sul set di validazione, al fine di raggiungere un compromesso ottimale tra capacità del modello e generalizzazione. La ricerca degli iperparametri è stata effettuata mediante una combinazione di campionamento casuale e ottimizzazione bayesiana, che ha permesso di

esplorare in modo efficiente lo spazio delle configurazioni, privilegiando progressivamente quelle con migliori prestazioni e scartando rapidamente le soluzioni meno efficaci.

Il meccanismo di *Self-Attention* rappresenta il nucleo funzionale del Vision Transformer, consentendo al modello di attribuire pesi differenti alle diverse regioni dell'immagine in funzione della loro rilevanza per il compito di classificazione. L'uso della MHSA permette di catturare simultaneamente molteplici relazioni spaziali e semantiche, mentre l'integrazione di connessioni residue all'interno dei moduli di attenzione favorisce una migliore dinamica di apprendimento, preservando le informazioni di input e migliorando la stabilità del modello. Attraverso un'attenta ottimizzazione di parametri, è stata individuata una configurazione finale in grado di ridurre l'overfitting e migliorare in modo consistente le metriche di classificazione sull'intero dataset.

Inoltre, in questo studio, è stata adottata una strategia di *Residual Self-Attention* per favorire la propagazione delle informazioni e incrementare la diversità delle caratteristiche estratte.

In particolare, è stata introdotta una connessione di tipo *skip* tra strati MHSA consecutivi, che consente alle informazioni apprese nei livelli più superficiali di essere propagate e accumulate nei livelli più profondi della rete. A differenza delle connessioni residue tradizionali, il meccanismo di *residual attention* agisce direttamente sulle matrici *query* (Q) e *key* (K), responsabili della modellazione delle relazioni tra le *patches embeddings*; ciò permette al modello di preservare le relazioni già apprese tra le diverse regioni dell'immagine, integrandole con nuove informazioni man mano che la profondità della rete aumenta.

Formalmente, il meccanismo proposto modifica il calcolo dei punteggi di attenzione all'interno del MHSA, aggregando i punteggi di attenzione dello strato corrente con quelli provenienti dagli strati precedenti, anziché calcolarli in modo indipendente a ciascun livello. Questo approccio consente una più efficace trasmissione delle dipendenze spaziali lungo la profondità del modello, migliorando la stabilità dell'apprendimento e le prestazioni complessive nella fase di classificazione. La struttura del meccanismo di *Residual Self-Attention* è illustrata in Figura 7.

Complessivamente, la struttura di tale modello rispetto ai modelli già noti, migliora la capacità di catturare caratteristiche sia locali che globali, particolarmente rilevanti per la discriminazione tra drusen e lesioni vitelliformi. Le formulazioni matematiche dettagliate del modello proposto non sono riportate nella presente tesi in quanto eccessivamente tecniche

rispetto agli obiettivi medico-clinici del lavoro; esse sono tuttavia descritte in modo esaustivo nell'articolo di riferimento [33].

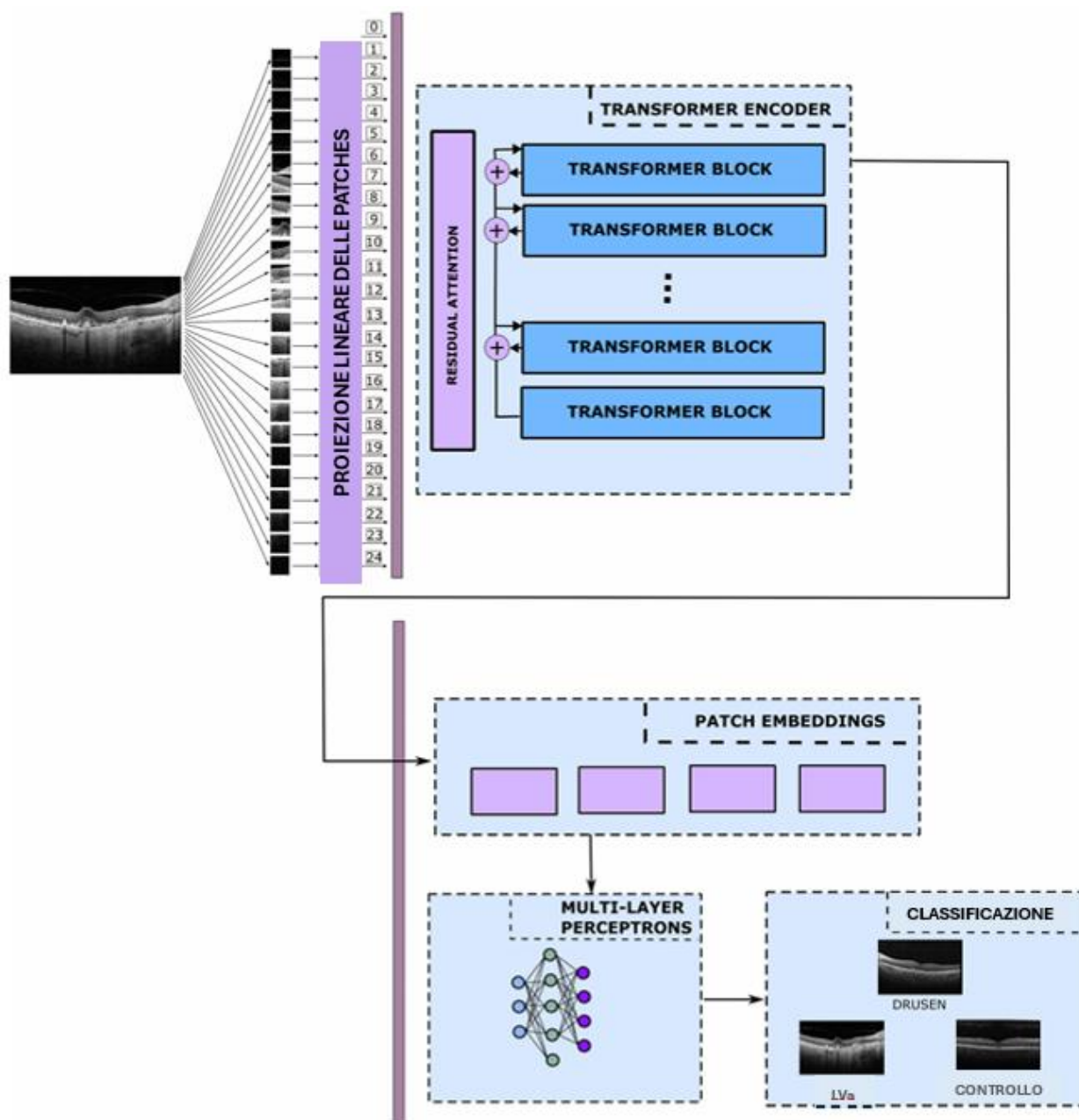


Figura 7 Struttura generale del modello Residual Self-Attention Vision Transformer

2.1.1.4 Modelli di confronto

Al fine di valutare l'efficacia del modello RS-A ViT, le sue prestazioni sono state confrontate con quelle di diverse architetture di DL consolidate, ben nota per la loro capacità di discriminare condizione di DMLE in esame OCT. Tra questi, si è fatto riferimento a: EfficientNet, InceptionV3, ResNet50 e VGG16. Tutti i modelli di confronto sono stati addestrati utilizzando lo stesso dataset,

le medesime procedure di *pre-processing* e analoghe strategie di *data augmentation*, così da garantire un confronto equo.

2.1.1.5 Valutazione delle metriche

Le prestazioni dei modelli sono state valutate sul set di test mediante metriche standard di classificazione, includendo: accuratezza, precisione, specificità, *recall*, F1-score. Il parametro *recall* valuta la capacità del modello di individuare correttamente i campioni positivi, rapportando i veri positivi alla somma di veri positivi e falsi negativi. L’F1-score, definito come media armonica tra precisione e *recall*, fornisce una misura bilanciata delle prestazioni del classificatore, risultando particolarmente utile in presenza di classi sbilanciate. Inoltre, sono state analizzate le matrici di confusione per una valutazione dettagliata delle predizioni corrette e degli errori di classificazione.

2.1.1.6 Grad-CAM

Per migliorare l’interpretabilità clinica del modello, è stata applicata la *Gradient-weighted Class Activation Mapping* (Grad-CAM), una tecnica di interpretabilità progettata per evidenziare le regioni di un’immagine che contribuiscono maggiormente alla predizione di una specifica classe da parte di una rete neurale convoluzionale. Il metodo si basa sul calcolo del gradiente del punteggio associato alla classe di interesse rispetto alle mappe di attivazione dell’ultimo strato convoluzionale, consentendo di stimare l’importanza di ciascuna mappa di caratteristiche nel processo decisionale del modello. La tecnica Grad-CAM ha l’obiettivo di rendere più prontamente visualizzabili le regioni delle immagini OCT maggiormente rilevanti per il processo decisionale della rete neurale, permettendo così di verificare che il modello elaborato focalizzi l’attenzione su strutture retiniche clinicamente significative [41]. L’immagine generata è di facile interpretabilità essendo associata ad una mappa di colori, in cui le aree più calde (ossia nella scala del colore rosso) mostrano maggiore influenza sulla classificazione.

2.1.2 Risultati

Gli esperimenti sono stati condotti mediante una suddivisione casuale del dataset in insiemi di addestramento, validazione e test con proporzioni pari rispettivamente al 70%, 15% e 15%, garantendo la presenza di tutte le classi in ciascun sottoinsieme. Il set di validazione è stato utilizzato per valutare gli iperparametri, ma anche per definire un check-point di controllo per verificare la stabilità del modello e, in caso di risultati migliorabili, per

interrompere precocemente il training. Il confronto tra sei architetture di DL evidenzia come il modello RS-A ViT proposto raggiunga le migliori prestazioni complessive (Figura 8).

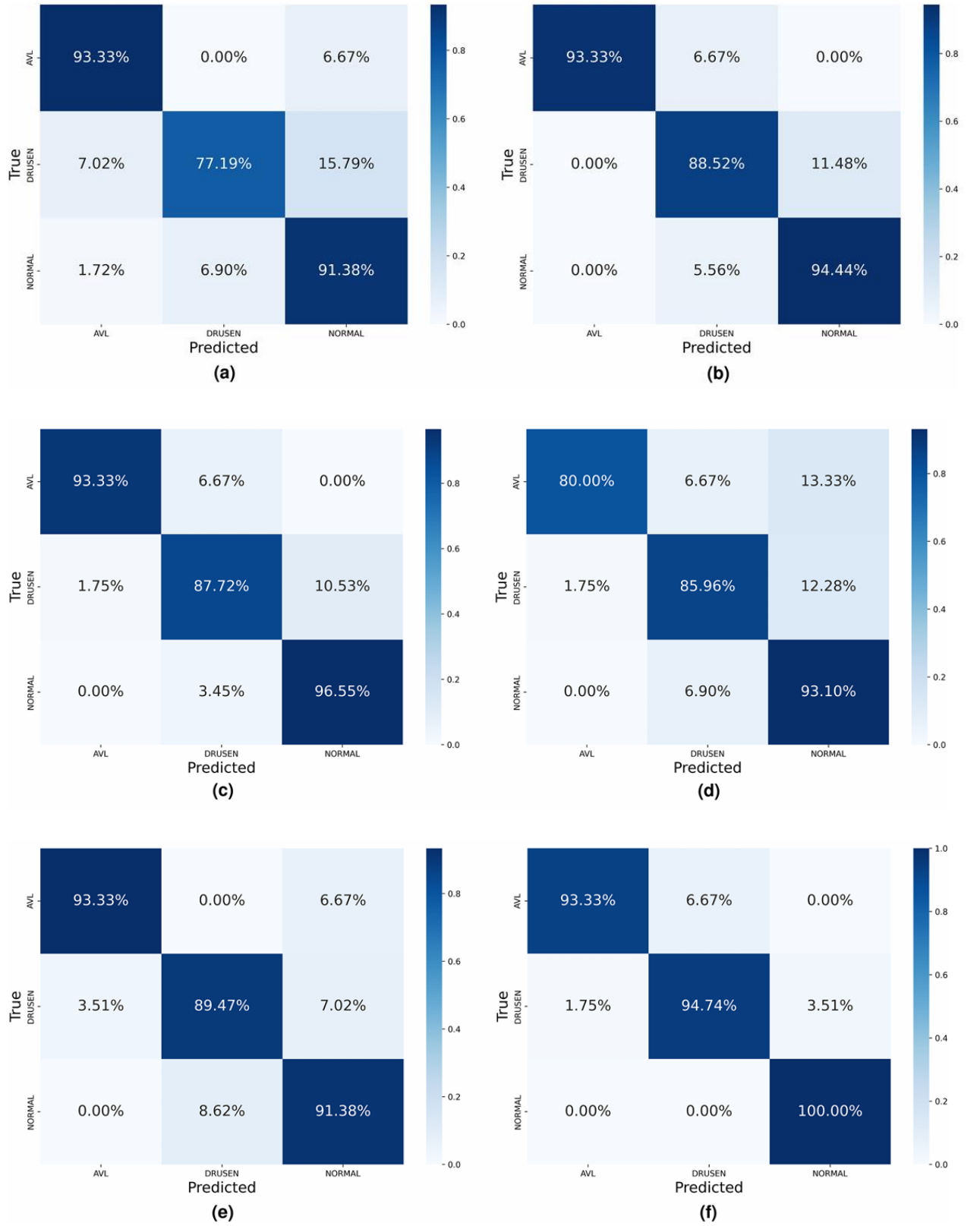


Figura 8 Matrici di confusione dei modelli: EfficientNet (a), InceptionV3 (b), ResNet50 (c), VGG16 (d), ViT (e), RS-A ViT (f).

In particolare, sebbene tutti i modelli presentino un'elevata accuratezza nel riconoscere i casi normali (ossia di controllo), l'RS-A ViT è l'unico che raggiunge il 100%. Tutti i modelli, eccetto il VGG16, presentano un'elevata capacità di riconoscere le LVa (93,33%). Nel caso delle drusen, la miglior performance nel riconoscimento è ottenuta dal modello ViT. Complessivamente, la massima capacità di classificazione è raggiunta dal modello RS-A ViT, superando gli altri modelli fino ad un massimo del 5,27% in termini di accuratezza.

Le metriche di precisione e *recall* rappresentano indicatori chiave per una valutazione approfondita delle prestazioni del modello, in particolare in presenza di dataset sbilanciati. L'analisi congiunta di tali metriche consente di valutare, in modo robusto, l'efficacia del classificatore nel riconoscimento delle classi di interesse. Sebbene tutti i modelli considerati abbiano ottenuto valori elevati di precisione e recall per le diverse classi, il modello proposto RS-A ViT si distingue per un livello prestazionale costantemente alto e bilanciato, con valori compresi tra il 93,33% e il 100% e una differenza minima tra le due metriche, inferiore al 3,45%. L'*F1-score* conferma ulteriormente tale risultato; nonostante InceptionV3 mostri il miglior F1-score per la patologia LVa, l'RS-A ViT risulta complessivamente superiore nel rilevamento delle drusen e nella corretta identificazione dei casi sani, evidenziando una maggiore robustezza e affidabilità del modello. Le metriche sono riportate dettagliatamente in Tabella 2.

Modello	Classe	Accuratezza	Precisione	Recall	F1-score	Specificità
EfficientNet	LVa	93,33	73,68	93,33	82,35	95,65
	Drusen	77,19	91,67	77,19	83,81	94,52
	Controllo	91,38	84,13	91,38	87,60	86,11
InceptionV3	LVa	93,33	100,00	93,33	96,55	99,13
	Drusen	88,52	93,10	88,52	90,76	94,20
	Controllo	94,44	87,93	94,44	91,07	90,80
ResNet50	LVa	93,33	93,33	93,33	93,33	99,13
	Drusen	87,72	94,34	87,72	90,91	95,89
	Controllo	96,55	90,32	96,55	93,33	91,67
VGG16	LVa	80,00	92,31	80,00	85,71	86,14
	Drusen	85,96	90,74	85,96	88,29	93,15
	Controllo	93,10	85,71	93,10	89,26	87,50
	LVa	93,33	87,50	93,33	90,32	98,26

ViT	Drusen	89,47	91,07	89,47	90,27	93,15
	Controllo	91,38	91,38	91,38	91,38	93,06
RS-A ViT	LVa	93,33	93,33	93,33	93,33	99,13
	Drusen	94,74	98,18	94,74	96,43	98,63
	Controllo	100,00	96,67	100,00	98,31	97,22

Tabella 2 Metriche ottenute per ogni modello, espresse in percentuale (%). I valori statisticamente significativi sono evidenziati in grassetto

I modelli tradizionali, quali EfficientNet, InceptionV3 e ResNet, necessitano di un numero maggiore di *epochs* (circa 80) per raggiungere la convergenza ottimale, mentre le architetture basate su Transformer, in particolare i modelli ViT e RS-A ViT, raggiungono prestazioni elevate già dopo 50 *epochs* di addestramento. Sebbene durante il training siano presenti temporanee fluttuazioni dell'accuratezza, in tutti i casi si osserva un andamento complessivamente crescente. Il modello RS-A ViT mostra la curva di accuratezza più regolare, indice di una maggiore stabilità del processo di apprendimento. Analogamente, la funzione di perdita (*validation loss*) diminuisce progressivamente per tutte le architetture considerate, segnalando un miglioramento continuo delle capacità predittive; tuttavia, il modello proposto presenta le minori oscillazioni a conferma della sua robustezza (Figure 9, 10). Nel complesso, questi risultati dimostrano che il modello RS-A ViT è particolarmente adatto alla classificazione delle patologie retiniche, come LVa e drusen, e alla loro distinzione dai casi sani (Figure 9 e 10).

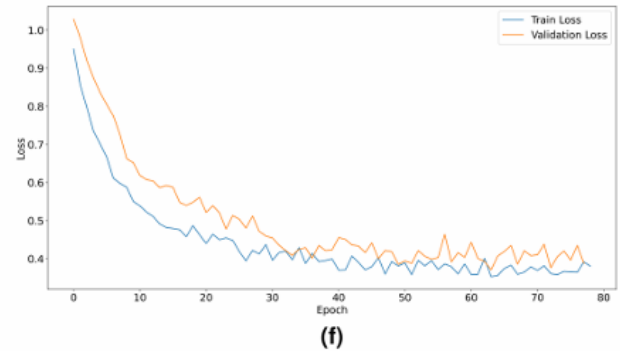
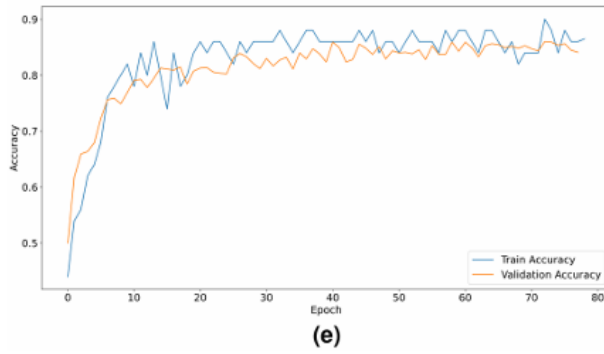
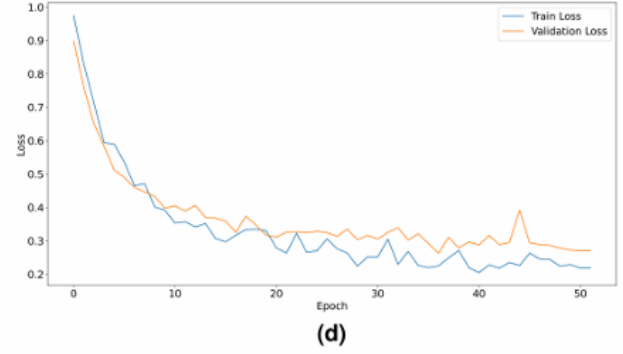
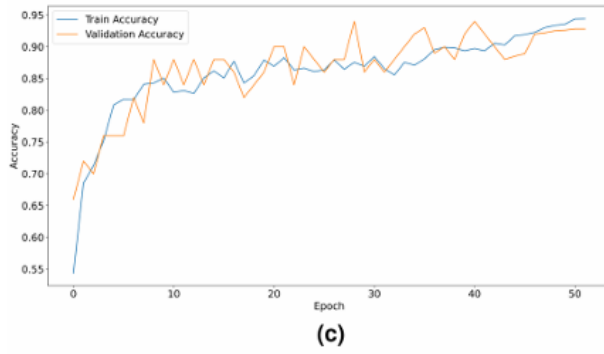
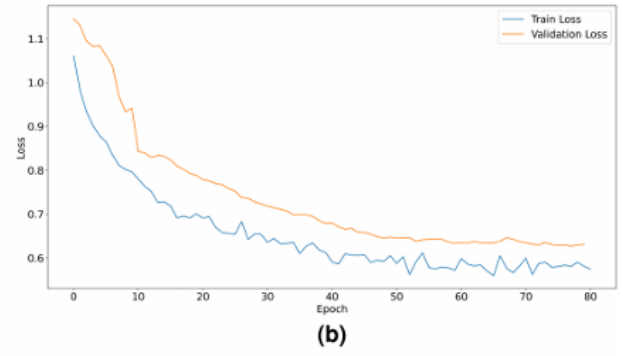
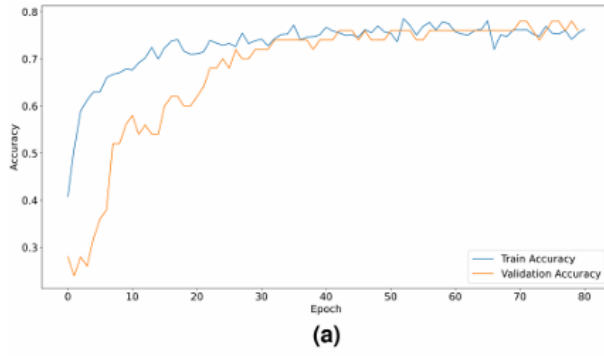


Figura 9 Modelli di addestramento e validazione per: EfficientNet accuratezza (a) e loss function (b); InceptionV3 accuratezza (c) e loss function (d); ResNet50 accuratezza (e) e loss function (f).

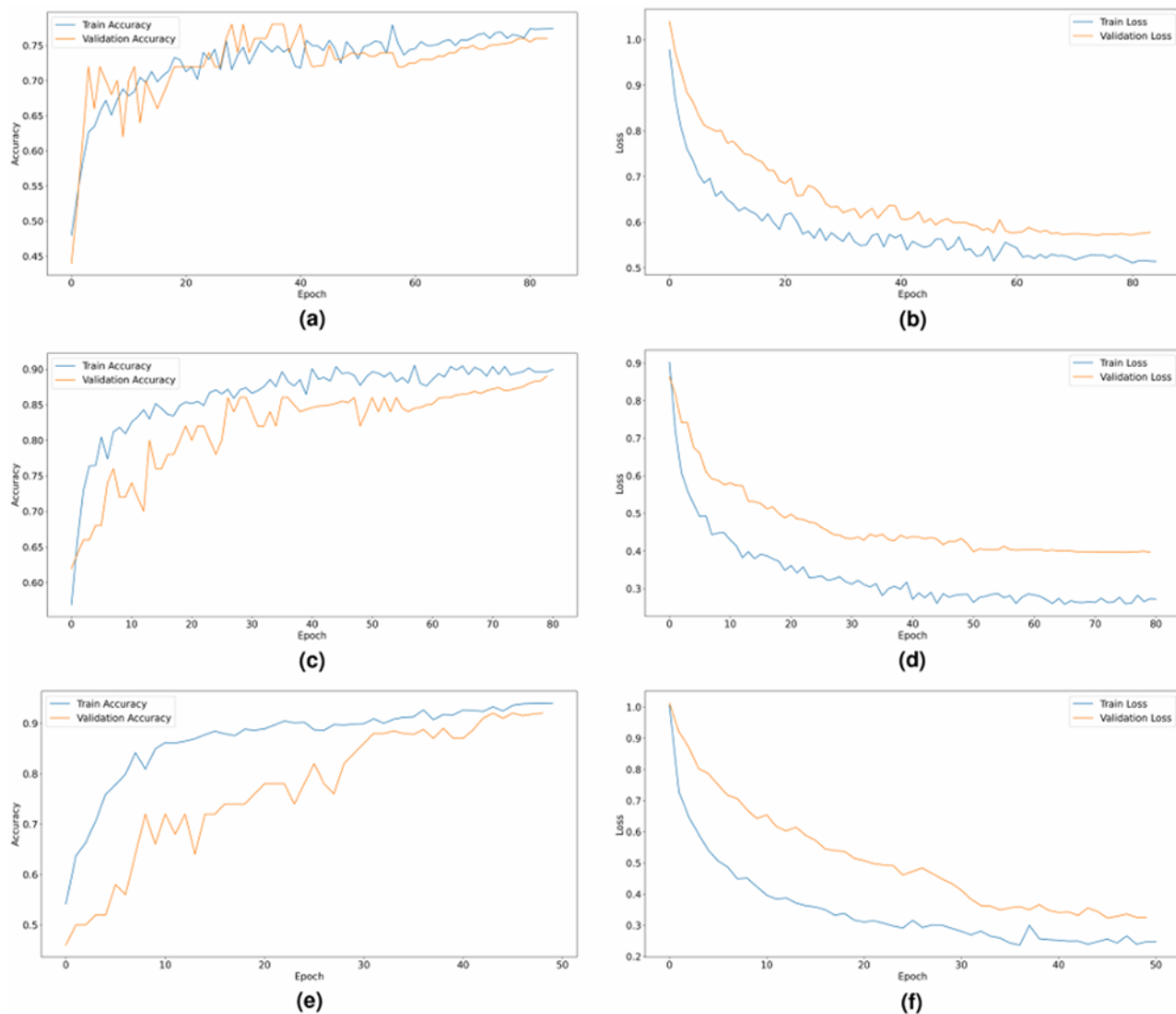


Figura 10 Modelli di addestramento e validazione per: VGG16 accuratezza (a) e loss function (b); ViT accuratezza (c) e loss function (d); RS-A ViT accuratezza (e) e loss function (f).

Infine, l'analisi delle visualizzazioni Grad-CAM dei vari modelli mostrano differenze significative nel modo in cui le diverse architetture analizzano le immagini OCT affette da LVa e drusen. Nel caso delle LVa (Figura 11), i modelli basati su CNN (EfficientNet, InceptionV3, ResNet50 e VGG16) tendono a generare mappe di attivazione più estese, vale a dire che spesso includono la lesione principale e la regione retinica circostante. Ad esempio, EfficientNet e InceptionV3 presentano una forte attenzione sulla lesione centrale (espresso dal colore rosso nelle immagini), accompagnata, però, da un'attivazione diffusa nelle regioni circostanti. Al contrario, i modelli basati su Transformer, in particolare il Vision Transformer e il Residual Self-Attention Vision Transformer (RS-A ViT), evidenziano una focalizzazione più precisa sulla regione centrale dell'LVa. In particolare, l'RS-A ViT concentra l'attivazione quasi esclusivamente sul deposito

vitelliforme, suggerendo una migliore capacità di discriminare le caratteristiche patologiche rilevanti e di ridurre l'influenza delle informazioni di sfondo.

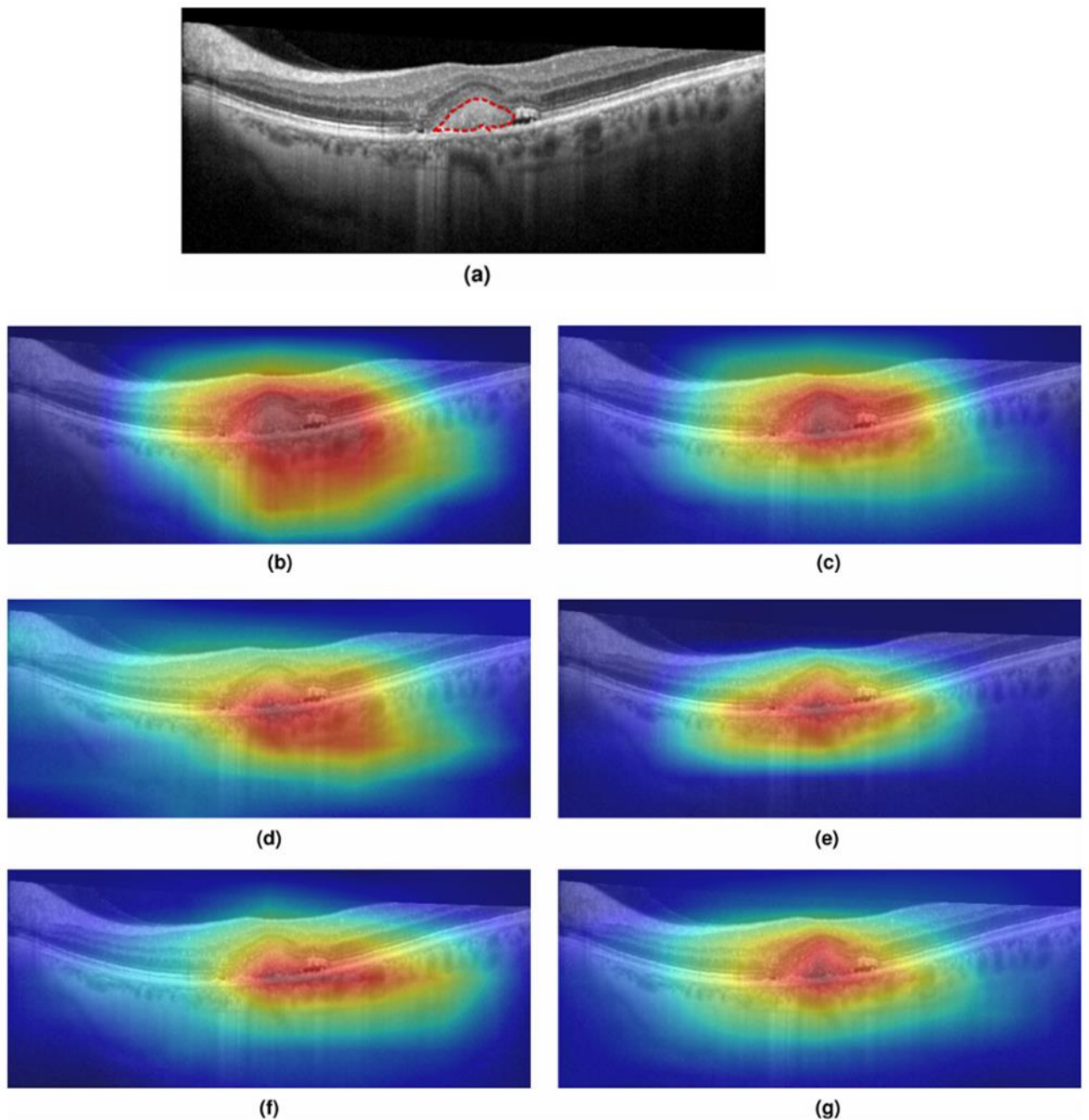
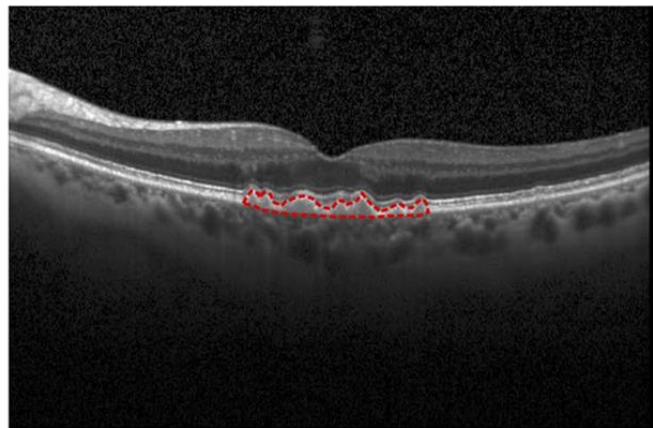


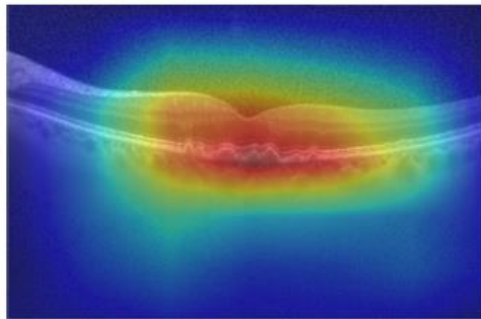
Figura 11 Scansione SD-OCT di LVA (a); visualizzazione con tecnica di interpretabilità Grad-CAM per i modelli: EfficientNet (b), InceptionV3 (c), ResNet (d), VGG16 (e), ViT (f), RS-A ViT (g).

Un comportamento analogo si osserva nel caso dei drusen (Figura 12). Sebbene tutti i modelli identifichino correttamente l'area patologica, le CNN producono mappe di attivazione più ampie e diffuse, che si estendono alle regioni retiniche adiacenti alla regione di interesse. I modelli Transformer, con il Vision Transformer, mostrano una localizzazione più accurata che restringe l'attenzione al deposito e l'RS-A ViT raffina ulteriormente tale focalizzazione,

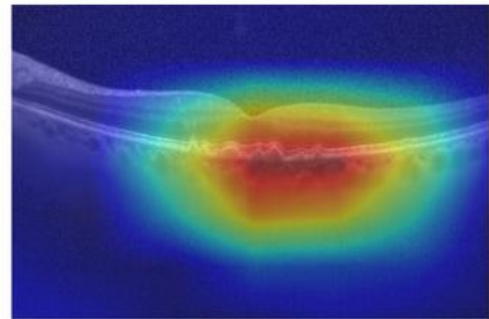
evidenziando prevalentemente la banda orizzontale delle drusen con minima attivazione esterna.



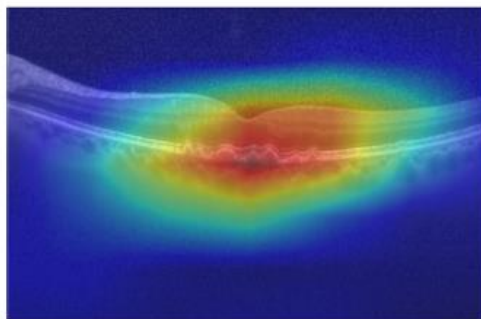
(a)



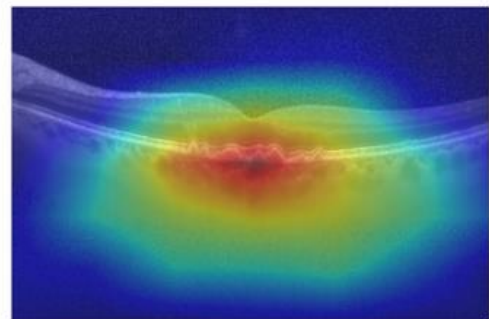
(b)



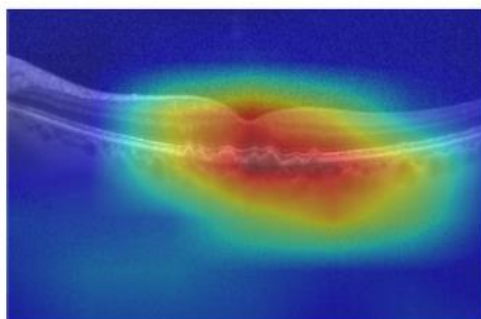
(c)



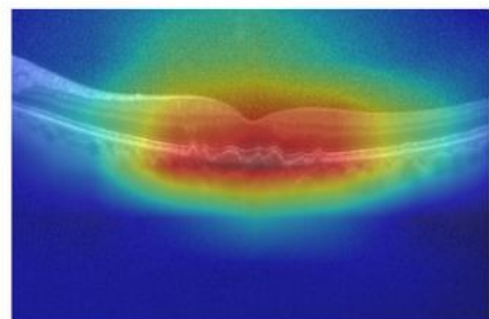
(d)



(e)



(f)



(g)

Figura 12 Scansione SD-OCT di drusen (a); visualizzazione con tecnica di interpretabilità Grad-CAM per i modelli: EfficientNet (b), InceptionV3 (c), ResNet (d), VGG16 (e), ViT (f), RS-A ViT (g).

L'analisi comparativa con i modelli già conosciuti conferma il potenziale elevato degli algoritmi Vision Transformer nella diagnosi delle patologie retiniche. In questo contesto, l'RS-A ViT si colloca tra i modelli con le migliori prestazioni, superando diversi modelli. È, tuttavia, importante sottolineare che le prestazioni dei modelli di deep learning dipendono non solo dall'architettura, ma anche dalla dimensione e dal bilanciamento del dataset. Un aspetto particolarmente rilevante di questo studio riguarda l'efficacia del modello proposto in presenza di un numero limitato di immagini. Nonostante il dataset ridotto (1280 immagini), l'RS-A ViT raggiunge un'accuratezza media del 96,92%, dimostrando un'elevata capacità di generalizzazione. Inoltre, lo studio si concentra sulle lesioni vitelliformi acquisite, una patologia relativamente rara e spesso misdiagnosticata con drusen confluenti, rendendo il compito di classificazione particolarmente *challenging*. Le elevate prestazioni ottenute in questo contesto evidenziano la robustezza del modello.

Le mappe Grad-CAM confermano, ulteriormente, che l'RS-A ViT è in grado di focalizzare l'attenzione in modo più selettivo sulle regioni clinicamente rilevanti. Il miglioramento nella gestione e nel riutilizzo delle mappe di attenzione intermedie consente al modello di preservare dettagli fini che potrebbero andare persi nei Vision Transformer standard. Nel complesso, i risultati dimostrano che il modello Residual Self-Attention Vision Transformer rappresenta una soluzione accurata, stabile e interpretabile per il riconoscimento delle patologie retiniche considerate e la loro distinzione dai casi sani.

2.2 Applicazione della radiomica all'OCT strutturale per la caratterizzazione delle lesioni vitelliformi acquisite: uno studio pilota

Obiettivo dello studio: Valutare la capacità di un modello radiomico di discriminare le caratteristiche intrinseche delle lesioni vitelliformi acquisite (LVa), all'esame OCT strutturale, per distinguere gli spazi otticamente vuoti dalla presenza di fluido sottoretinico, migliorando la precisione diagnostica e supportando una precoce identificazione di possibili neovascolarizzazioni maculari.

2.2.1 Materiali e metodi

Si tratta di uno studio retrospettivo, osservazionale, monocentrico, condotto presso l'ambulatorio di retina medica della clinica Oculistica dell'Università Federico II di Napoli, includendo tutti pazienti che hanno effettuato almeno un accesso ambulatoriale presso la suddetta struttura nel periodo compreso tra marzo 2022 e gennaio 2025 con una diagnosi di

LVa e/o corioretinopatia sierosa centrale acuta non complicata (CRSCa). Lo studio è stato condotto secondo i canoni della Dichiarazione di Helsinki, approvato dal comitato etico locale, sotto il cappello di un precedente progetto di studio includente le patologie retiniche maculari (ID 06/19), registrato sul sito Clinicaltrial.gov (<https://clinicaltrials.gov/study/NCT07074041>). Tutti i pazienti inclusi hanno espresso il loro consenso alla partecipazione tramite firma del consenso informato.

Criteri di esclusione: età inferiore ai 18 anni, lunghezza bulbare assiale superiore a 26 mm, presenza di cataratta clinicamente significativa, presenza di concomitanti patologie oculari e /o sistemiche impattanti (es. glaucoma, uveiti, retinopatia diabetica), patologie note per essere cause di lesioni vitelliformi secondarie e/o fluido retinico (DMLE, neovascolarizzazioni, patologie dell'interfaccia vitreo-retinica con effetto trazionale a livello maculare, sindromi tossiche e/o paraneoplastiche), distrofie retiniche ereditarie. Tutti i pazienti, come da protocollo standard dell'ambulatorio di retina medica, sono stati sottoposti ad esami di *multimodal imaging* retinico per definire la diagnosi, includendo: esame *multicolor*, esame dell'autofluorescenza retinica, esame SD-OCT, OCTA, possibili esami aggiuntivi quali FAG e ICGA, nel sospetto di una possibile neovascolarizzazione maculare, dubbia in esame OCTA. Tutti gli esami sono stati effettuati con il medesimo strumento: Heidelberg Spectralis, (Heidelberg Engineering), da operatori esperti in condizioni di illuminazione standardizzata, previa corretta centratura del paziente e messa a fuoco ottimale del sistema di acquisizione secondo modalità manuale. Per l'esame SD-OCT maculare è stato utilizzato un protocollo di scansione volumetrica B-scan con un'area di acquisizione pari a 6×6 mm, centrata sulla fovea. Durante l'acquisizione, la funzione di *eye-tracking* automatico era attiva allo scopo di ridurre gli artefatti da movimento e migliorare la riproducibilità delle scansioni. Tutte le immagini sono state revisionate per il controllo di qualità. Le scansioni considerate di qualità insufficiente sono state escluse dall'analisi, in particolare quelle caratterizzate da un indice di intensità del segnale (*signal strength index*) ≤ 15 [42]. Le immagini OCT valide sono state archiviate in formato digitale ed utilizzate per le successive analisi quantitative e qualitative.

Per creare il modello radiomico, tutti i pazienti sono stati stratificati in due classi rispetto alla diagnosi retinica:

- classe 1 – pazienti con diagnosi di LVa
- classe 2 - pazienti con diagnosi CRSCa. Questa seconda classe di pazienti è stata necessaria per definire una condizione di riferimento “presenza di fluido retinico”, essenziale al fine di effettuare il training del modello nell'individuare, riconoscere e differenziare il fluido retinico dallo spazio vuoto. La CRSCa è una patologia idiopatica caratterizzata da *leakage* focale o multifocale a livello dell'EPR, generalmente associato alla presenza di un distacco sieroso del neuroepitelio. Comunemente unilaterale e asimmetrico, interessa prevalentemente soggetti di sesso

maschile in una fascia di età tra i 30 ed i 50 anni. Il decorso naturale della CRSCa è generalmente benigno con la risoluzione del distacco del neuroepitelio in un intervento di tempo di circa tre mesi dalla comparsa. Pertanto, in fase acuta, corrispondente ai primi tre mesi dalla comparsa dei sintomi, l'OCT è caratterizzato dalla presenza di fluido retinico, e corrisponde così ad un valido esempio di diagnosi “fluido” per il training del modello.

2.1.1.1 Estrazione delle caratteristiche radiomiche

Il *workflow* radiomico richiede l'applicazione di molteplici passaggi tecnici alle immagini da analizzare. Come fase preliminare, tutte le scansioni OCT selezionate, a partire dal protocollo volumetrico di acquisizione, sono state convertite da un formato RGB (*red-green-blue*) ad un formato di scala di grigi, utilizzando la funzione dedicata del pacchetto Python PILLOW (versione 11.1) e applicando le impostazioni di default. Tale conversione è stata eseguita per garantire l'uniformità dei dati di input per la successiva analisi radiomica. Successivamente, ogni immagine è stata segmentata manualmente tramite delineazione accurata delle lesioni, in maniera indipendente, da tre lettori esperti (MC, EG, MS), per definire la regione di interesse (ROI) di studio. A tale scopo è stato utilizzato il software di segmentazione ITK-SNAP (versione 3.8) (Figura 13).

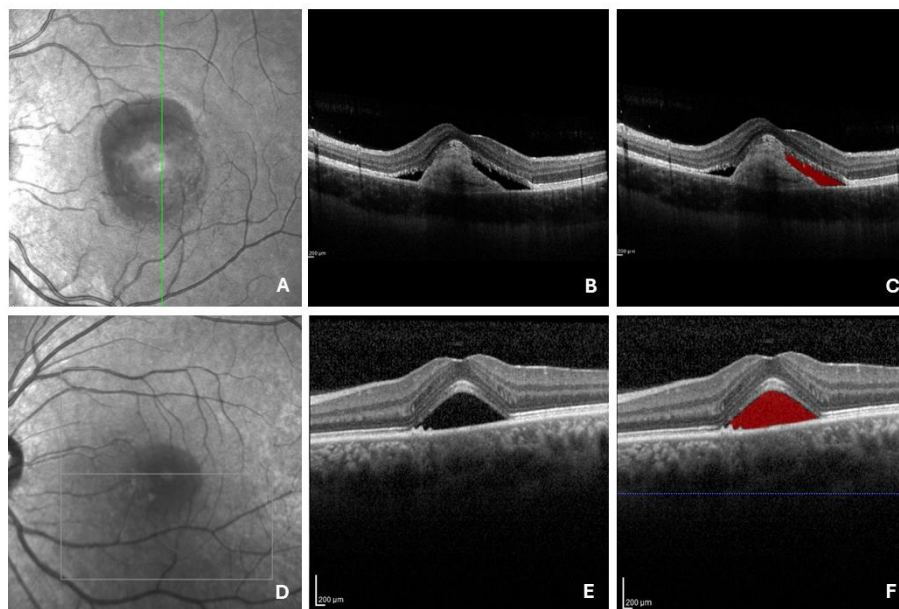


Figura 13 Esempio di segmentazione manuale di una lesione vitelliforme acquisita (A-C) e di una condizione di corioretinopatia sierosa centrale acuta (D-F), in esame OCT strutturale. **A:** en face OCT; **B:** corrispondente immagine B scan in OCT strutturale; **C:** regione di interesse (evidenziata in rosso) individuante lo spazio otticamente vuoto nel caso di lesione vitelliforme con materiale parzialmente riassorbito, segmentata con il software di segmentazione ITK-SNAP. **D:** en face OCT; **E:** corrispondente immagine B scan in OCT strutturale; **F:** regione di interesse (evidenziata in rosso) individuante il fluido sottoretinico in caso di corioretinopatia sierosa centrale, segmentata con il software di segmentazione ITK-SNAP

L'impiego di segmentazioni multiple, effettuata in maniera autonoma da tre diversi lettori, è stato finalizzato alla valutazione della robustezza e riproducibilità delle caratteristiche radiomiche rispetto alla variabilità inter-osservatore, elemento essenziale nei *workflow* radiomici. Ogni ROI individuata su singola immagine OCT ha definito una maschera; le maschere delle lesioni così ottenute sono state successivamente accoppiate alle corrispondenti immagini OCT in scala di grigi per consentire l'estrazione delle caratteristiche radiomiche mediante il software PyRadiomics (versione 3.0.1), in maniera conforme agli standard dell'*Image Biomarker Standardization Initiative* (IBSI, (<https://pyradiomics.readthedocs.io/en/latest/index.html>)). Tale software consente l'esecuzione integrata delle fasi di *pre-processing* delle immagini e di estrazione dei *biomarkers* quantitativi. Il software ha effettuato una sequenza di operazioni con l'obiettivo di standardizzare l'intensità dei pixel delle immagini OCT, prima dell'estrazione delle caratteristiche radiomiche, riducendo la variabilità non biologica e rendendo i biomarcatori quantitativi comparabili tra immagini e soggetti diversi. La sequenza è specificata qui di seguito:

1) ricampionamento spaziale mediante un *kernel* di 3×3 pixel; consiste nell'applicazione di una griglia uniforme che armonizza la risoluzione spaziale locale dell'immagine. Si tratta di un passaggio fondamentale per minimizzare le differenze legate alla densità di campionamento e per garantire che le caratteristiche della *texture* vengano calcolate su strutture spaziali confrontabili tra immagini diverse. In termini pratici, il ricampionamento assicura che ciascun voxel (o pixel, nel caso bidimensionale) rappresenti una porzione spaziale omogenea dell'immagine, migliorando la riproducibilità delle *features* radiomiche.

2) normalizzazione dei livelli di grigio; è basata sulla distribuzione statistica delle intensità dell'immagine, limitando i valori inclusi all'intervallo definito dalla media ± 3 deviazioni standard. Questa operazione consente di attenuare l'influenza di valori estremi o *outlier*, spesso dovuti a rumore o artefatti di acquisizione, preservando al contempo la variabilità informativa rilevante per l'analisi radiomica.

3) fase di *rescaling* dell'intensità (con fattore pari a 100) ha lo scopo di amplificare numericamente i valori dei livelli di grigio, rendendoli più adatti alle successive elaborazioni matematiche.

4) traslazione dei voxel (+300) permette di evitare valori negativi e di posizionare l'intera distribuzione delle intensità in un intervallo positivo e controllato. La combinazione di queste due ultime operazioni determina una distribuzione attesa dei livelli di grigio compresa tra 0 e 600, garantendo coerenza e stabilità numerica durante il calcolo delle matrici della *texture*.

5) discretizzazione dei livelli di grigio (ampiezza di bin fissa pari a 3) rappresenta un passaggio cruciale per l'estrazione delle caratteristiche radiomiche di ordine superiore. Tale procedura trasforma i valori continui di intensità in un numero finito di livelli discreti, riducendo la sensibilità al rumore e migliorando la robustezza delle matrici statistiche (come GLCM, GLSZM, GLRLM e GLDM). L'utilizzo di un'ampiezza di bin fissa consente di mantenere una rappresentazione coerente delle *texture* tra immagini diverse, facilitando il confronto quantitativo delle caratteristiche estratte.

6) In aggiunta alle immagini originali pre-processate, sono state incluse nel processo di estrazione anche immagini filtrate, al fine di migliorare il rapporto segnale-rumore delle caratteristiche radiomiche. In particolare, sono stati applicati filtri di *Laplacian of Gauss* (LoG) per l'enfaticizzazione dei bordi (valori di sigma pari a 3, 4 e 5) e una decomposizione *wavelet* bidimensionale, comprendente tutte le combinazioni di filtri passa-alto e passa-basso lungo gli assi x e y. Infine, da ciascuna ROI sono state estratte caratteristiche morfologiche bidimensionali (shape2D), caratteristiche di primo ordine e caratteristiche di *textures* di ordine superiore. Queste ultime si basano su matrici statistiche standardizzate implementate in PyRadiomics, tra cui la Gray-Level Co-occurrence Matrix (GLCM), che descrive la distribuzione delle coppie di livelli di grigio, la Gray-Level Size Zone Matrix (GLSZM), che quantifica zone di pixel con identico valore di intensità, la Gray-Level Run Length Matrix (GLRLM), che cattura la lunghezza delle sequenze di pixel con uguale intensità, e la Gray-Level Dependence Matrix (GLDM), che valuta la dipendenza dei livelli di grigio rispetto a un pixel centrale entro una distanza predefinita. Una descrizione dettagliata delle classi di caratteristiche e delle relative definizioni matematiche è disponibile nella documentazione ufficiale di PyRadiomics (<https://pyradiomics.readthedocs.io/en/latest/features.html>).

2.1.1.2 Pipeline di Machine Learning

Tutte le immagini analizzate e processate sono state suddivise in modalità randomizzata tra addestramento (*training*) e test set, secondo un rapporto 75:25. Il training set è stato utilizzato per le fasi di selezione delle caratteristiche e di addestramento del modello, mentre il test set è stato impiegato nella fase di inferenza. Tutte le trasformazioni dei dati sono state basate esclusivamente sui dati di training e successivamente applicate al test set, al fine di prevenire rischi di *information leakage*. Innanzitutto, è stata valutata la robustezza delle caratteristiche radiomiche rispetto alla variabilità inter-osservatore nella segmentazione manuale mediante il calcolo dell'*Intraclass Correlation Coefficient* (ICC), utilizzando un

modello a due vie, con accordo assoluto e misura singola del lettore. Nello specifico, è stata adottata una soglia di $ICC \geq 0.75$ per identificare le caratteristiche considerabili riproducibili e, pertanto, idonee alle successive fasi di analisi. Successivamente, è stata applicata una procedura di *scaling* (o ridimensionamento) per eliminare potenziali fonti di bias derivanti dalle differenti scale di valori dei *biomarkers* e, così, normalizzare i dati. A tal fine, è stato utilizzato il metodo di ridimensionamento Min–Max (intervallo 0–1), adattato esclusivamente sul training set e successivamente applicato sia ai dati di training sia a quelli di test. Le caratteristiche presumibilmente non informative sono state eliminate mediante l'applicazione di un filtro di varianza, con soglia fissata al valore di *cut-off* di 0.01, mentre la ridondanza tra caratteristiche è stata affrontata calcolando la matrice di correlazione di Pearson (valore assoluto) tra tutte le coppie di variabili residue; una soglia di 0.8 è stata utilizzata per identificare un'elevata correlazione *inter-features*, determinando la rimozione di una delle caratteristiche ridondanti (Figura 14).

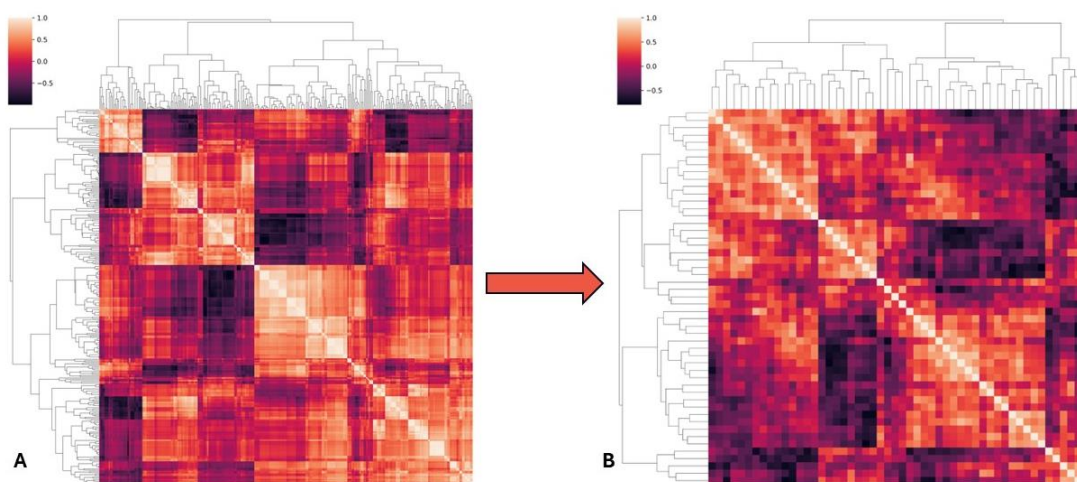


Figura 14 Clustermap della correlazione tra le caratteristiche radiomiche prima (A) e dopo (B) la selezione delle variabili con coefficiente di correlazione $r > 0,80$.

Lo sbilanciamento tra le classi è stato gestito esclusivamente sul training set mediante una procedura di *oversampling* della classe minoritaria, utilizzando la *Synthetic Minority Over-sampling Technique* (SMOTE), con parametro k pari a 5. L'*oversampling* è un processo che consiste nel generare nuovi dati, simili alla classe meno rappresentata, così da superare il problema della “quantità relativa dei dati”. La fase finale di selezione delle caratteristiche è stata condotta mediante l'impiego del *WEKA subset evaluator*, che consente di bilanciare l'informatività delle caratteristiche e la loro ridondanza relativa sulla base della correlazione con le etichette di classe.

Il dataset finale così ottenuto è stato utilizzato per l'addestramento di un modello di regressione logistica con regolarizzazione di tipo *ridge*, le cui prestazioni sono state valutate sul training set mediante cross-validazione e, successivamente, sul test set indipendente. Le comuni

metriche di accuratezza sono state calcolate a partire dalla matrice di confusione risultante. L'analisi è stata condotta utilizzando il pacchetto R *irr* (versione 0.84.1), il pacchetto Python *scikit-learn* (versione 1.6) e la piattaforma di *data mining* *WEKA* (versione 3.9.6).

2.1.2 Risultati

Sulla base dei criteri di inclusione ed esclusione, 99 pazienti sono risultati candidabili e, pertanto, inclusi nello studio. A seguito della suddivisione randomizzata delle immagini, 74 pazienti sono stati inclusi nel training set e 25 nel data test. A partire dai dati disponibili, applicando il *workflow* precedentemente descritto, un totale di 520 *features* radiomiche sono state estratte. Il processo di selezione delle *features* è stato articolato in molteplici fasi. Innanzitutto, l'analisi della riproducibilità ha permesso di identificare 277 *features* stabili sulla base dell'ICC. Parallelamente, l'analisi della varianza ha individuato 277 *features* informative e l'analisi di correlazione ha permesso di evidenziare 51 *features* non altamente correlate tra loro. Il flusso del processo di selezione delle *features* è illustrato in Figura 15.

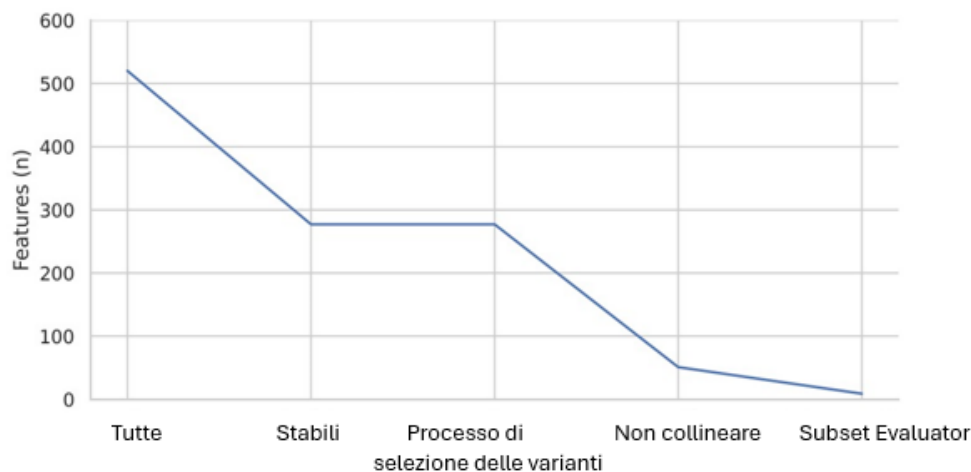


Figura 15 Grafico illustrante la riduzione del numero di parametri (asse delle ordinate) durante le diverse fasi del processo di selezione delle *features* (asse delle ascisse).

L'algoritmo *SubsetEvaluator* ha selezionato automaticamente un sottoinsieme finale di nove *features*, individuate come la combinazione più efficace per l'addestramento del modello. Sulla base di tali *features*, il modello di regressione logistica ha raggiunto un'accuratezza dell'86,5% sul set di training, con un valore predittivo positivo (VPP) e un valore predittivo negativo (VPN) entrambi pari all'87%. L'area sotto la curva ROC (*Receiver Operating Characteristic, Area Under the Curve, AUC*) è risultata pari a 0,89, mentre l'area sotto la curva *Precision-Recall* (PRC) è stata pari a 0,86. Nel dataset di test, il modello ha mantenuto un'accuratezza elevata dell'80% (sebbene inferiore rispetto all'accuratezza del

training set), un NPV del 92% e un PPV del 67%. Sia l'AUC che l'area sotto la curva PRC hanno raggiunto valore elevato, pari a 0,92 (Tabella 3).

	Accuratezza	VPP	VPN	ROC area	PRC area
Set di addestramento	86.7	87	87	0.89	0.86
Test set	80	67	92	0.92	0.92

Tabella 3 Metriche dettagliate di accuratezza del modello nei set di addestramento e di test. Sono riportati accuratezza (%), valore predittivo positivo (VPP), valore predittivo negativo (VPN), espressi in percentuale (%) area sotto la curva ROC e area sotto la curva PRC.

3 DISCUSSIONE

A livello mondiale, le patologie retiniche sono una causa significativa di ipovisione e cecità in tutte le fasce d'età, negli adulti (con maggiore frequenza associata alla DMLE e alla retinopatia diabetica) e nei bambini (patologie retiniche ereditarie e retinopatia del prematuro). Diagnosticare correttamente queste patologie necessita spesso di un percorso *time-consuming*, che include un approccio di *multimodal imaging* e molteplici visite con specialistici retinici, non sempre prontamente disponibili e, pertanto, causante un possibile ritardo del potenziale trattamento *sight-saving*.

L'applicazione dell'intelligenza artificiale nel campo delle patologie retiniche sta emergendo come uno strumento di supporto clinico di grande rilevanza, in particolare in contesti diagnostici poco frequenti, complessi e ad elevata variabilità interpretativa. Come evidenziato dalla letteratura recente, l'IA ha dimostrato elevate prestazioni nella diagnosi, classificazione e prognosi di numerose malattie retiniche, soprattutto attraverso l'analisi automatizzata di immagini OCT e lo sviluppo di modelli sempre più sensibili e specifici, grazie alla collaborazione multidisciplinare tra clinici e ingegneri biomedici ed informatici [43]. In questo scenario rapidamente in evoluzione, l'integrazione di approcci avanzati di analisi di immagini, come i modelli di deep learning e la radiomica, consente di estrarre biomarcatori quantitativi, non percepibili all'osservazione clinica convenzionale, migliorando la precisione diagnostica, la stratificazione del rischio e la possibile prognosi legata al trattamento. Nello studio proposto, tale potenziale risulta particolarmente significativo per condizioni spesso misdiagnosticate, come le lesioni vitelliformi acquisite (LVa), frequentemente confuse con soft drusen confluenti, e per quelle condizioni in cui, per i cambiamenti dinamici del materiale vitelliforme, si evidenzia all'esame OCT strutturale la presenza di spazi otticamente vuoti, in cui la distinzione tra materiale sottoretinico riassorbito e vero fluido retinico, associato a complicanze neovascolari, rappresenta una sfida clinica rilevante.

Il modello proposto basato su Vision Transformer e, in particolare, su architetture con meccanismi di *self-attention* residua, dimostra di essere in grado di raggiungere elevate prestazioni in presenza di LVA, drusen e soggetti sani, anche su dataset limitati. In particolare, il modello proposto presenta un'accuratezza media di classificazione del 96,92%, dimostrando la sua capacità di apprendere da ridotti, ma ultra-specialistici datasets, e di generalizzare la conoscenza su nuove immagini, caratteristica utile in condizione cliniche a bassa epidemiologia, con possibilità di collezionare dati spesso limitata. Lo studio proposto focalizza l'attenzione sulla capacità diagnostica in caso di lesioni relativamente rare quali le LVA, spesso confuse con le drusen. In questo caso, RS-A ViT mostra un'elevata accuratezza nel discriminare le due condizioni, dimostrandosi un modello robusto. Inoltre, attraverso l'applicazione della tecnica di interpretabilità clinica di Grad-CAM, il modello RS-A ViT si dimostra più preciso, rispetto alle architetture di confronto, nel focalizzare l'attenzione sulle aree retiniche di interesse al fine della diagnosi. Complessivamente, sebbene i modelli addestrati su ampi dataset possano raggiungere gradi di accuratezza comparabili, l'efficacia del modello RS-A ViT si basa sull'abilità di gestire poche, ma ultraspecializzate immagini cliniche, per raggiungere risultati con elevata accuratezza. Inoltre, nonostante livelli complessivi di accuratezza simili, il modello proposto presenta valori di precisione e sensibilità superiori nel caso di LVA, la classe meno rappresentata, indicando la sua capacità di superare problemi legati ai dataset non bilanciati. RS-A ViT si dimostra come il modello che tende a confondere meno le classi presentate, soprattutto nel confronto tra drusen e controlli sani, che rappresentato le principali fonti di errore per gli altri modelli.

Per l'analisi dei cambiamenti dinamici delle lesioni vitelliformi è stato adottato un approccio di analisi delle immagini differente rispetto al modello RS-A ViT. Il riassorbimento del materiale vitelliforme, soprattutto quando associato a sintomi visivi quali la riduzione dell'acuità visiva, rappresenta una condizione clinicamente complessa da interpretare. Al fine di indagare caratteristiche non direttamente osservabili ad occhio nudo, è stata applicata l'analisi radiomica a scansioni OCT strutturali di LVA con materiale parzialmente riassorbito, confrontandole con immagini di corioretinopatia sierosa centrale, utilizzata come classe di controllo per addestrare il modello all'individuazione, al riconoscimento e alla discriminazione del "fluido vero". Tale distinzione consente una valutazione indiretta della possibile presenza di neovascolarizzazioni, permettendo al modello di identificare *patterns di imaging* potenzialmente associati a tali condizioni senza la necessità di ulteriori esami diagnostici aggiuntivi e/o invasivi, oltre l'esame OCT stesso.

Nello specifico, il modello radiomico proposto ha mostrato prestazioni robuste sia nel set di addestramento sia nella coorte di validazione esterna, raggiungendo un'accuratezza dell'86,5%, con valori bilanciati di valore predittivo positivo e negativo (entrambi pari all'87%). Gli elevati valori di AUC (0,89) e di PRC (0,86) supportano la capacità discriminativa globale del modello e la sua robustezza in presenza di dataset non bilanciati. Sebbene nel test set l'accuratezza si riduca lievemente (80%), tale risultato risulta coerente con la variabilità intrinseca dei dati non visti e conferma la capacità di generalizzazione del modello. È particolarmente rilevante l'elevato valore predittivo negativo (92%), indicativo di una forte capacità di escludere falsi negativi, che è caratteristica auspicabile in contesti di screening. Al contrario, il valore predittivo positivo più contenuto (67%) suggerisce una tendenza ai falsi positivi, limitando l'utilizzo del modello come strumento diagnostico autonomo, ma confermandone l'elevato potenziale come supporto decisionale clinico.

Pur presentando un elevato potenziale applicativo, lo studio presenta alcuni limiti: il disegno retrospettivo, la dimensione contenuta del campione e la necessità di segmentazione manuale della regione di interesse (ROI). Per mitigare quest'ultima criticità, ciascuna immagine è stata segmentata indipendentemente da tre lettori esperti e la robustezza delle caratteristiche radiomiche è stata quantificata mediante il coefficiente di correlazione intraclasse (Intraclass Correlation Coefficient, ICC) al fine di valutarne la riproducibilità inter-osservatore.

Studi futuri, idealmente prospettici e longitudinali, con l'implementazione di procedure di segmentazione semi-automatica o completamente automatizzata, potranno contribuire a incrementare ulteriormente la riproducibilità e l'affidabilità predittiva del modello. Inoltre, un possibile sviluppo volto a migliorare le prestazioni diagnostiche, in particolare in termini di valore predittivo positivo (VPP), potrebbe consistere nell'integrazione delle scansioni OCT con le corrispondenti immagini di autofluorescenza, immagini di rilevante significato clinico in presenza di lesioni vitelliformi.

Nonostante le limitazioni evidenziate, la solidità dello studio si fonda su due elementi cardine: la strutturazione metodologicamente rigorosa del modello radiomico e l'analisi statistica, condotte in modo sistematico e conforme alle linee guida internazionali della European Society of Medical Imaging Informatics [44].

In conclusione, i risultati ottenuti corroborano il potenziale e promettente ruolo dell'intelligenza artificiale applicata all'analisi delle immagini retiniche, evidenziandone la capacità di estrarre informazioni quantitative non immediatamente apprezzabili all'osservazione clinica convenzionale e di identificare pattern nascosti ad alto contenuto informativo. Tali evidenze delineano un contributo concreto al supporto della medicina di precisione in oftalmologia, intesa

come diagnosi precoce, stratificazione del rischio, personalizzazione dei percorsi diagnostico-terapeutici e l'ottimizzazione dei follow-up.

Nel contesto della crescente complessità delle patologie retiniche e dell'elevato volume di dati clinici sottoforma di immagini, l'adozione di soluzioni automatizzate, riproducibili e standardizzate appare sempre più necessaria. Tali strumenti devono essere concepiti come sistemi di supporto alla decisione clinica, pienamente integrati nel percorso assistenziale del paziente, con funzione complementare e non sostitutiva del giudizio medici, al fine di migliorare l'accuratezza diagnostica, l'efficienza gestionale, e pertanto, complessivamente, la prognosi clinica.

Bibliografia

1. Mirzayev I, Gündüz AK, Aydın Ellialtıoğlu P, Gündüz ÖÖ. Clinical applications of anterior segment swept-source optical coherence tomography: A systematic review. *Photodiagnosis Photodyn Ther*. 2023 Jun;42:103334. doi: 10.1016/j.pdpdt.2023.103334.
2. Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, Hee MR, Flotte T, Gregory K, Puliafito CA, et al. Optical coherence tomography. *Science*. 1991 Nov 22;254(5035):1178-81. doi: 10.1126/science.1957169.
3. Drexler W, Fujimoto JG. Optical coherence tomography: technology and applications. *Biomed Opt Express*. 2008;29(6):1304–1330.
4. Dansingani KK, Tan ACS, Gilani F, Phasukkijwatana N, Novais E, Querques L, Waheed NK, Duker JS, Querques G, Yannuzzi LA, Sarraf D, Freund KB. Subretinal Hyperreflective Material Imaged With Optical Coherence Tomography Angiography. *Am J Ophthalmol*. 2016 Sep;169:235-248. doi: 10.1016/j.ajo.2016.06.031.
5. Feo A, Stradiotto E, Sacconi R, Menean M, Querques G, Romano MR. Subretinal hyperreflective material in retinal and chorioretinal disorders: A comprehensive review. *Surv Ophthalmol*. 2024 May-Jun;69(3):362-377. doi: 10.1016/j.survophthal.2023.10.013.
6. Wong WL, Su X, Li X, Cheung CM, Klein R, Cheng CY, Wong TY. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health*. 2014 Feb;2(2):e106-16. doi: 10.1016/S2214-109X(13)70145-1.
7. Vision Loss Expert Group of the Global Burden of Disease Study; GBD 2019 Blindness and Vision Impairment Collaborators. Global estimates on the number of people blind or visually impaired by age-related macular degeneration: a meta-analysis from 2000 to 2020. *Eye (Lond)*. 2024 Aug;38(11):2070-2082. doi: 10.1038/s41433-024-03050-z.
8. GBD 2019 Blindness and Vision Impairment Collaborators; Vision Loss Expert Group of the Global Burden of Disease Study. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *Lancet Glob Health*. 2021 Feb;9(2):e144-e160. doi: 10.1016/S2214-109X(20)30489-7.
9. Rinaldi M, Cennamo G, Concilio M, Corvino G, Riccardo A, Nubi R, Costagliola C. Aflibercept 8 mg in Neovascular AMD-A Fast-Drying Anti-VEGF Drug: A Prospective Morpho-Functional Pilot Study. *Ophthalmol Ther*. 2025 Dec 12. doi: 10.1007/s40123-025-01296-8.

10. Chiang A, Davis M, Stevens W, Garg S, Sheidow T, Jones DL, Intorcchia M, McKeown A, Schmidt-Erfurth UM, Sarda SP, Bauman CR. Visual Acuity and Quality of Life Outcomes With Pegcetacoplan Treatment: A Post Hoc Analysis From the OAKS and DERBY Trials. *Am J Ophthalmol*. 2025 Sep;277:471-481. doi: 10.1016/j.ajo.2025.05.023.
11. Holz FG, Le Mer Y, Muqit MMK, Hattenbach LO, Cusumano A, Grisanti S, Kodjikian L, Pileri MA, Matonti F, Souied E, Stanzel BV, Szurman P, Weber M, Bartz-Schmidt KU, Eter N, Delyfer MN, Girmens JF, van Overdam KA, Wolf A, Hornig R, Corazzol M, Brodie F, Olmos de Koo L, Palanker D, Sahel JA. Subretinal Photovoltaic Implant to Restore Vision in Geographic Atrophy Due to AMD. *N Engl J Med*. 2025 Oct 20;10.1056/NEJMoa2501396. doi: 10.1056/NEJMoa2501396.
12. Fleckenstein M, Mitchell P, Freund KB, Sadda S, Holz FG, Brittain C, Henry EC, Ferrara D. The Progression of Geographic Atrophy Secondary to Age-Related Macular Degeneration. *Ophthalmology*. 2018 Mar;125(3):369-390. doi: 10.1016/j.ophtha.2017.08.038.
13. Intonti S, Olivieri C, Reibaldi M, Borrelli E, Curcio C, Conedera FM. Translational Molecular and Fluid Biomarkers for Age-Related Macular Degeneration: Practical Insights from Animal Models and Humans. *Biomolecules*. 2025 Nov 8;15(11):1571. doi: 10.3390/biom15111571.
14. Holz FG, Strauss EC, Schmitz-Valckenberg S, van Lookeren Campagne M. Geographic atrophy: clinical features and potential therapeutic approaches. *Ophthalmology*. 2014 May;121(5):1079-91. doi: 10.1016/j.ophtha.2013.11.023.
15. Fritsche LG, Igl W, Bailey JN, Grassmann F, Sengupta S, Bragg-Gresham JL et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet*. 2016 Feb;48(2):134-43. doi: 10.1038/ng.3448.
16. Hageman GS, Gehrs K, Lejnine S, Bansal AT, Deangelis MM, Guymer RH, Baird PN, Allikmets R, Deciu C, Oeth P, Perlee LT. Clinical validation of a genetic model to estimate the risk of developing choroidal neovascular age-related macular degeneration. *Hum Genomics*. 2011 Jul;5(5):420-40. doi: 10.1186/1479-7364-5-5-420.
17. Bowes Rickman C, Farsiu S, Toth CA, Klingeborn M. Dry age-related macular degeneration: mechanisms, therapeutic targets, and imaging. *Invest Ophthalmol Vis Sci*. 2013 Dec 13;54(14):ORSF68-80. doi: 10.1167/iovs.13-12757.
18. Babaker R, Alzimami L, Al Ameer A, Almutairi M, Alam Aldeen R, Alshatti H, Al-Johani N, Al Taisan A. Risk factors for age-related macular degeneration: Updated systematic

- review and meta-analysis. *Medicine (Baltimore)*. 2025 Feb 21;104(8):e41599. doi: 10.1097/MD.00000000000041599.
19. Leveziel N, Puche N, Zerbib J, Benlian P, Coscas G, Soubrane G, Souied E. Génétique de la dégénérescence maculaire liée à l'âge [Genetic factors associated with age-related macular degeneration]. *Med Sci (Paris)*. 2010 May;26(5):509-15. French. doi: 10.1051/medsci/2010265509.
 20. He W, Han X, Ong JS, Wu Y, Hewitt AW, Mackey DA, Gharahkhani P, MacGregor S. Genome-Wide Meta-analysis Identifies Risk Loci and Improves Disease Prediction of Age-Related Macular Degeneration. *Ophthalmology*. 2024 Jan;131(1):16-29. doi: 10.1016/j.ophtha.2023.08.023.
 21. AREDS2 Research Group; Chew EY, Clemons T, SanGiovanni JP, Danis R, Domalpally A, McBee W, Sperduto R, Ferris FL. The Age-Related Eye Disease Study 2 (AREDS2): study design and baseline characteristics (AREDS2 report number 1). *Ophthalmology*. 2012 Nov;119(11):2282-9. doi: 10.1016/j.ophtha.2012.05.027.
 22. Ferris FL 3rd, Wilkinson CP, Bird A, Chakravarthy U, Chew E, Csaky K, Sadda SR; Beckman Initiative for Macular Research Classification Committee. Clinical classification of age-related macular degeneration. *Ophthalmology*. 2013 Apr;120(4):844-51. doi: 10.1016/j.ophtha.2012.10.036.
 23. Spaide RF, Curcio CA. Drusen characterization with multimodal imaging. *Retina*. 2010 Oct;30(9):1441-54. doi: 10.1097/IAE.0b013e3181ee5ce8.
 24. Sadda SR, Guymer R, Holz FG, Schmitz-Valckenberg S, Curcio CA, Bird AC et al. Consensus Definition for Atrophy Associated with Age-Related Macular Degeneration on OCT: Classification of Atrophy Report 3. *Ophthalmology*. 2018 Apr;125(4):537-548. doi: 10.1016/j.ophtha.2017.09.028.
 25. Heesterbeek TJ, Lorés-Motta L, Hoyng CB, Lechanteur YTE, den Hollander AI. Risk factors for progression of age-related macular degeneration. *Ophthalmic Physiol Opt*. 2020 Mar;40(2):140-170. doi: 10.1111/opo.12675.
 26. Freund KB, Laud K, Lima LH, Spaide RF, Zweifel S, Yannuzzi LA. Acquired Vitelliform Lesions: correlation of clinical findings and multiple imaging analyses. *Retina*. 2011 Jan;31(1):13-25. doi: 10.1097/IAE.0b013e3181ea48ba.
 27. Crincoli E, Zhao Z, Querques G, Sacconi R, Carlà MM, Giannuzzi F et al. Deep learning to distinguish Best vitelliform macular dystrophy (BVMD) from adult-onset vitelliform macular degeneration (AVMD). *Sci Rep*. 2022 Jul 26;12(1):12745. doi: 10.1038/s41598-022-16980-z.

28. Nipp GE, Lee T, Sarici K, Malek G, Hadziahmetovic M. Adult-onset foveomacular vitelliform dystrophy: epidemiology, pathophysiology, imaging, and prognosis. *Front Ophthalmol (Lausanne)*. 2023 Aug 10;3:1237788. doi: 10.3389/fopht.2023.1237788.
29. Feo A, Bousquet E, Faes L, Ramtohul P, Sacconi R, Rissotto F et al. VITELLIFORM LESIONS ASSOCIATED WITH ANGIOID STREAKS: Long-Term Follow-Up of a Rarely Described Phenotype. *Retina*. 2025 Apr 11. doi: 10.1097/IAE.0000000000004485.
30. Balaratnasingam C, Hoang QV, Inoue M, Curcio CA, Dolz-Marco R, Yannuzzi NA, Dhrami-Gavazi E, Yannuzzi LA, Freund KB. Clinical Characteristics, Choroidal Neovascularization, and Predictors of Visual Outcomes in Acquired Vitelliform Lesions. *Am J Ophthalmol*. 2016 Dec;172:28-38. doi: 10.1016/j.ajo.2016.09.008.
31. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015 May 28;521(7553):436-44. doi: 10.1038/nature14539.
32. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017 Dec;42:60-88. doi: 10.1016/j.media.2017.07.005
33. Powroznik P, Skublewska-Paszkowska M, Nowomiejska K, Gajda-Deryło B, Brinkmann M, Concilio M, Toro MD, Rejdak R. Residual self-attention vision transformer for detecting acquired vitelliform lesions and age-related macular drusen. *Sci Rep*. 2025 May 16;15(1):17107. doi: 10.1038/s41598-025-02299-y.
34. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019 Feb;103(2):167-175. doi: 10.1136/bjophthalmol-2018-313173.
35. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018 Sep;24(9):1342-1350. doi: 10.1038/s41591-018-0107-6.
36. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016 Dec 13;316(22):2402-2410. doi: 10.1001/jama.2016.17216.
37. Dosovitskiy, A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale 2020. *arXiv* 2020, arXiv:2010.11929.

38. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows 2021. arXiv 2021, arXiv:2103.14030.
39. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016 Feb;278(2):563-77. doi: 10.1148/radiol.2015151169.
40. Zhang H, Zhang H, Jiang M, Li J, Li J, Zhou H, Song X, Fan X. Radiomics in ophthalmology: a systematic review. *Eur Radiol*. 2025 Jan;35(1):542-557. doi: 10.1007/s00330-024-10911-4.
41. Kumaran S, Y., Jeya, J. J., Khan, S. B., Alzahrani, S. & Alojail, M. Explainable lung cancer classification with ensemble transfer learning of vgg16, resnet50 and inceptionv3 using grad-cam. *BMC Med. Imaging* 24, 176 (2024).
42. Huang Y, Gangaputra S, Lee KE, Narkar AR, Klein R, Klein BE, Meuer SM, Danis RP. Signal quality assessment of retinal optical coherence tomography images. *Invest Ophthalmol Vis Sci*. 2012 Apr 24;53(4):2133-41. doi: 10.1167/iovs.11-8755
43. Daich Varela M, Sen S, De Guimaraes TAC, Kabiri N, Pontikos N, Balaskas K, Michaelides M. Artificial intelligence in retinal disease: clinical application, challenges, and future directions. *Graefes Arch Clin Exp Ophthalmol*. 2023 Nov;261(11):3283-3297. doi: 10.1007/s00417-023-06052-x.
44. Santinha J, Pinto Dos Santos D, Laqua F et al. ESR Essentials: radiomics-practice recommendations by the European Society of Medical Imaging Informatics. *Eur Radiol*. 2025 Mar;35(3):1122-1132.doi: 10.1007/s00330-024-11093-9.