



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

**Corso di Laurea Magistrale in
Bioingegneria**

Tesi di Laurea

Analisi automatica di immagini della retina acquisite con oftalmoscopio basato su smartphone

Relatore:

Prof.ssa Maria Pia SACCOMANI

Correlatori:

Fabio SCARPA

Alessia COLONNA

Laureanda:

Alexa BERTO

n° matr.1155358 / LMIBM

Anno Accademico 2018/2019

*Questa tesi è stata fatta in collaborazione con:
D-EYE S.r.l., Padova
www.d-eyecare.com*



*“A chi crede nei sogni, basta
un gradino per raggiungere le stelle.”*

*Ringrazio la Prof.ssa Maria Pia Saccomani per la disponibilità e la grande
umanità con la quale ha saputo incoraggiarmi e consigliarmi nella scelta della
tesi.*

*Porgo un ringraziamento particolare all'Ing. Fabio Scarpa che mi ha
accompagnato, con grande professionalità e dedizione, passo passo in questo
percorso. I suoi preziosi insegnamenti e la sua positività sono stati per me di
fondamentale importanza.*

*Ringrazio l'Ing. Alberto Scarpa, CEO di D-EYE, per avermi dato la possibilità di
realizzare un progetto di tesi su una tecnologia altamente innovativa e per il
tempo che mi ha dedicato, nonostante i numerosi impegni in azienda.*

INDICE

SOMMARIO	I
CAPITOLO 1	1
L' APPARATO VISIVO	1
1.1 Anatomia dell'occhio.....	1
1.2 Morfologia della retina	4
1.3 Principali patologie diagnosticabili tramite l'analisi del fondo oculare.....	11
1.3.1 Glaucoma	11
1.3.2 Le malattie della retina.....	14
1.3.2.1 Degenerazione Maculare correlata all'età	15
1.3.2.2 Retinite pigmentosa	16
1.3.2.3 Vasculopatie retiniche.....	17
1.3.2.4 Retinopatia ipertensiva.....	18
1.3.2.5 Occlusione dell'arteria e della vena centrale della retina	18
1.3.2.6 Retinopatia diabetica.....	19
1.3.2.7 Retinopatia della prematurità.....	23
CAPITOLO 2	25
TECNOLOGIE PER L' ANALISI DEL FONDO OCULARE	25
2.1 Esame del fondo oculare	25
2.2. Oftalmoscopio.....	27
2.2.1 Cenni storici	28
2.2.2 Oftalmoscopio diretto	34
2.2.3 Oftalmoscopio indiretto	35
2.3 Lampada a fessura o biomicroscopio.....	37
2.4 Fundus camera	37
2.5 D-EYE.....	39
CAPITOLO 3	43
DEEP LEARNING	43
3.1 Introduzione.....	43
3.2 Reti Neurali Artificiali (ANN: Artificial Neural Networks)	45

3.2.1. Modello di Neurone artificiale.....	45
3.2.2. Architettura di una rete neurale artificiale	48
3.3 <i>Apprendimento supervisionato (supervised learning)</i>	51
3.4 <i>Classificazione</i>	55
3.5 <i>Reti Neurali Convoluzionali (CNN: Convolutional Neural Networks)</i>	57
3.5.1 Architettura CNN.....	59
3.5.1.1 Convolutional Layer	61
3.5.1.2 ReLU Layer	63
3.5.1.3 Pooling Layer.....	64
3.6 <i>Segmentazione Semantica</i>	66
3.7 <i>Applicazioni cliniche del Deep Learning</i>	68
3.7.1 Deep learning nell'analisi di immagini in oftalmologia	70
CAPITOLO 4	73
ALGORITMO PER L'ANALISI AUTOMATICA.....	73
4.1 <i>Dataset</i>	74
4.2 <i>CNN – selezione dei frame</i>	77
4.2.1 Pre-processing.....	77
4.2.2 Architettura della CNN	78
4.2.3 Iperparametri.....	81
4.2.4 Fase di testing della CNN	85
4.3 <i>U-NET – segmentazione disco e cup</i>	86
4.3.1 Segmentazione del disco ottico.....	86
4.3.2 Selezione dei frame.....	88
4.3.3 Segmentazione della cup.....	88
CAPITOLO 5	91
RISULTATI	91
5.1 <i>Selezione frame</i>	93
5.2 <i>Segmentazione disco e cup</i>	94
5.2.1 Segmentazione del disco ottico.....	94
5.2.2 Selezione dei frame.....	95
5.2.3 Segmentazione della cup.....	95
5.3 <i>Glaucoma</i>	95
CAPITOLO 6	99
CONCLUSIONI	99
BIBLIOGRAFIA	101
RINGRAZIAMENTI.....	105

Sommario

L'esame del fundus oculi eseguito mediante oftalmoscopio consente di osservare ed esaminare le strutture interne dell'apparato visivo, in particolare la papilla ottica, la regione maculare e la porzione periferica della retina, consentendone lo studio dei caratteri morfologici e funzionali. Questo test diagnostico permette di portare alla luce tutta una serie di patologie che possono coinvolgere primariamente o secondariamente (come conseguenza di malattie sistemiche) l'occhio. Sempre più diffuse a causa dell'invecchiamento della popolazione, alcune di queste malattie se non diagnosticate in tempo e curate in maniera adeguata possono causare seri danni alla vista e in alcuni casi, portare ad ipovisione e cecità. Primo fra tutti, il glaucoma provoca lesioni irreversibili e rappresenta la terza causa principale di cecità nel mondo. Negli stadi iniziali può essere asintomatico e ciò comporta il pericolo che non venga effettuata una diagnosi tempestiva e ci si accorga della presenza della malattia soltanto in una fase avanzata del suo decorso, quando i danni hanno ormai raggiunto una notevole gravità e si sono manifestate importanti perdite del campo visivo.

L'oftalmoscopia smartphone rappresenta una risorsa fondamentale per la diagnosi precoce di tutte quelle patologie che come il glaucoma progrediscono in modo silente e per le quali una terapia tempestiva, mirata ed efficace permetterebbe di ridurre il rischio di gravi complicanze.

Il progetto di tesi consiste nello sviluppo di un algoritmo per l'analisi automatica di immagini retiniche acquisite con il dispositivo ottico sviluppato da D-EYE, start-up del gruppo M31 con sede a Padova (www.d-eyecare.com) [1].

Questo sistema viene agganciato alla fotocamera di uno smartphone trasformandolo in un oftalmoscopio digitale diretto in grado di registrare video e immagini ad alta definizione del fondo oculare. La semplicità di questo strumento consente anche a personale non medico di utilizzarlo e lo rende adatto ad effettuare un ampio screening sulla popolazione.

Gli obiettivi dell’algoritmo sviluppato, in stretta collaborazione con l’azienda, sono quelli di migliorare la qualità dell’immagine acquisita con l’oftalmoscopia smartphone e di individuare la possibile presenza di segni patologici. Più in particolare, l’applicativo seleziona i frame di interesse clinico ed esegue la segmentazione del disco ottico. Uno studio di fattibilità preliminare è stato fatto con un dataset molto limitato per la segmentazione della cup ed il calcolo del rapporto cup-to-disc verticale (VCDR) [2], fornendo informazioni utili sulla patologia glaucomatosa.

L’algoritmo proposto, basato sull’utilizzo di reti neurali convoluzionali, è stato implementato in ambiente Matlab (The Mathworks Inc., Natick, MA, USA, release 2019a) mediante il pacchetto software “Deep Learning Toolbox”.

Il Deep Learning costituisce una sezione del Machine Learning, a sua volta branca dell’Intelligenza Artificiale. Questo approccio, che si ispira alla struttura e al funzionamento del cervello, è caratterizzato da modelli computazionali capaci di imparare ad eseguire un determinato task direttamente dai dati senza essere programmati per la sua esecuzione.

Negli ultimi anni, questo ambito di ricerca ha suscitato un interesse crescente, poiché ha permesso di raggiungere performance nettamente superiori rispetto ad altri metodi e di superare talvolta le prestazioni ottenute dall’uomo.

Le reti neurali convoluzionali sono sempre più utilizzate nel campo di elaborazione delle immagini e sono stati ottenuti risultati interessanti anche in medicina [3] e, in particolare, in oftalmologia [4, 5, 6].

In relazione a questo progetto di tesi, sono state ottenute buone performance ad ogni fase dell’algoritmo, permettendo di dimostrare l’affidabilità dell’analisi automatica di video del fundus retinico acquisiti con oftalmoscopia smartphone.

Un’analisi automatizzata pone l’operatore in una posizione proattiva rispetto alla gestione dei video e, soprattutto nel caso in cui non si tratti di un medico oftalmologo, può fornire un valido supporto decisionale utile ad incrementare l’efficienza del servizio. Ciò, in particolare, potrebbe essere cruciale in un contesto di telemedicina per facilitare la trasmissione a distanza in tempo reale di immagini e dati di interesse clinico e rendere immediata l’interpretazione e la consultazione da parte di specialisti.

Capitolo 1

L'apparato visivo

1.1 Anatomia dell'occhio

L'apparato visivo è formato dal bulbo oculare con la sua componente nervosa, nervo ottico e vie ottiche intracerebrali, e dagli annessi oculari, che comprendono cavità orbitaria, corpo adiposo dell'orbita, palpebre, sopracciglia, congiuntiva, fascia del bulbo, muscoli extra-oculari e apparato lacrimale.

Da un punto di vista funzionale e per sommi capi, questo apparato si è evoluto in modo che la radiazione luminosa proveniente dall'esterno, attraversata la parete anteriore dell'occhio, la cornea, venga focalizzata sul versante più interno della parete posteriore del globo oculare, la retina; quest'ultima converte l'impulso elettromagnetico in impulsi nervosi reindirizzandoli attraverso il nervo ottico all'encefalo.

Il **bulbo oculare**, organo pari e dalla morfologia grossomodo sferica (diametro antero-posteriore, detto anche asse ottico, e diametro trasversale di circa 23,5 mm, diametro cranio-caudale di circa 23,5-24 mm), è alloggiato nella cavità orbitaria, protetto dalle palpebre e da altri annessi oculari, ed è costituito da tre tuniche che dall'esterno all'interno sono: la tunica fibrosa, la tunica vascolare e la tunica nervosa [7].

La **tunica fibrosa** è formata anteriormente dalla cornea e nei suoi cinque/sesti posteriori dalla sclera.

La **cornea** è una membrana trasparente con funzione di focalizzazione, rappresentando il principale mezzo diottrico dell'occhio (il secondo mezzo diottrico è il cristallino, un ulteriore ruolo diottrico è attribuibile al film lacrimale). La sua trasparenza deriva dalla struttura microscopica, in cui antero-posteriormente ci sono

CAPITOLO 1. L'APPARATO VISIVO

i seguenti strati: epitelio corneale, membrana limitante anteriore di Bowmann, stroma, membrana limitante posteriore di Descemet ed endotelio.

La disposizione in file allineate delle fibrille di collagene dello stroma è ciò che garantisce la trasparenza corneale. L'endotelio corneale sostiene tale peculiarità trasportando acqua dalla cornea alla camera anteriore. La cornea è complessivamente una struttura avascolare, il cui nutrimento dipende dall'umor acqueo e dal film lacrimale.

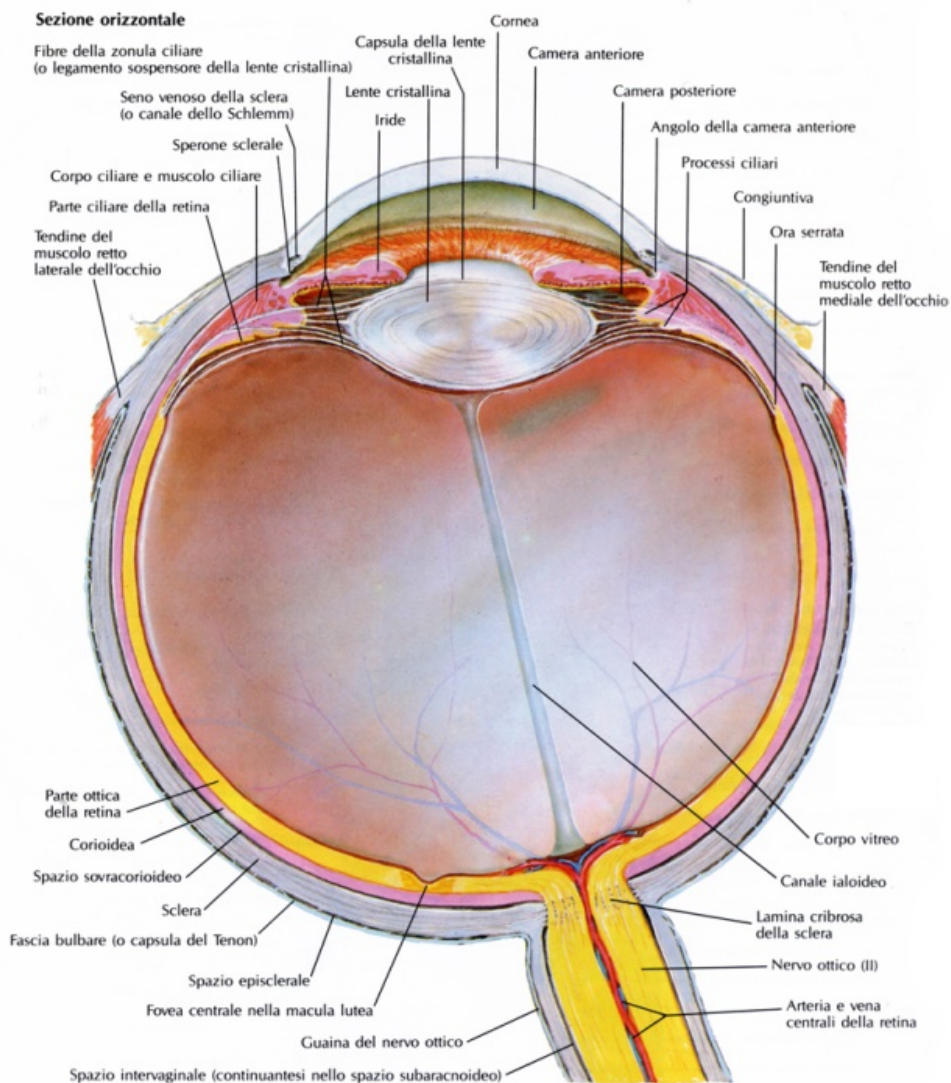


Figura 1.1 Sezione del bulbo oculare secondo un piano trasversale. Esso è costituito da tre tuniche che dall'esterno all'interno sono: fibrosa, vascolare e nervosa. Al polo posteriore si trova nervo ottico e arteria e vena centrali della retina; anteriormente cornea, iride, corpo ciliare e cristallino.

La **sclera**, membrana di tessuto connettivo denso e di colorito bianco madreperlaceo, ha funzione di trofismo e protezione. La sua opacità aiuta a formare la camera oscura entro cui avviene la trasmissione delle immagini. Sul suo versante posteriore, nel punto di uscita delle strutture vascolo-nervoso (vasi e nervi ciliari, nonché nervo ottico) dal globo oculare, si trova la lamina cribrosa.

La giunzione tra cornea e sclera è definita **limbus corneo-sclerale**: la cornea affonda sotto il margine della sclera come il vetro di un orologio rispetto alla cassa (la cornea ha un raggio di curvatura inferiore della sclera). Questa giunzione è importante perché sede del trabecolato corneo-sclerale, una struttura di deflusso dell'umor acqueo verso il canale dello Schlemm, un canale simil-venoso fenestrato di fondamentale importanza nella regolazione della pressione intra-oculare.

La **tunica vascolare o uvea** è formata anteriormente dal corpo ciliare e dall'iride e posteriormente dalla coroide.

L'**iride** è una membrana circolare, convessa anteriormente, pigmentata e con aspetto spugnoso, situata tra la cornea ed il cristallino; essa è centrata da un forame tondeggiante, la pupilla, il quale, dilatandosi e contraendosi, modula la quantità di luce che penetra nell'occhio.

Il **corpo ciliare** è formato da due porzioni: la pars plicata, anteriore, e la pars plana, posteriore. Mentre la pars plicata è costituita dai processi ciliari, strutture che fanno parte del ligamento sospensore del cristallino, la porzione posteriore è rivestita internamente da due strati che si continuano rispettivamente con l'epitelio pigmentato della retina e la retina neurosensoriale. Funzionalmente, il corpo ciliare partecipa alla formazione di umor acqueo ed è responsabile dell'adattamento refrattivo del cristallino.

La **coroide** si estende dal margine della papilla ottica fino all'**ora serrata** (punto di transizione tra la superficie posteriore del corpo ciliare e dell'iride e la retina visiva). È una struttura polistratificata dedicata alla nutrizione dell'epitelio pigmentato e degli strati più esterni della retina per mezzo della lamina coriocabillare, uno strato costituito da una fitta rete di capillare anastomizzati tra loro. Sul versante più interno presenta la lamina basale o membrana di Bruch, che si trova in diretta continuità con la membrana basale dell'epitelio pigmentato retinico.

Oltre a funzione trofica, la coroide, che contiene una certa quantità di pigmenti, coadiuva con l'epitelio pigmentato nell'impedire la riflessione dei raggi luminosi

Il **cristallino** è un organo a forma di lente biconvessa, circa 2 mm dietro la cornea. Privo di vascolarizzazione e innervazione, che lo rende otticamente perfetto, possiede la capacità di modificare la propria forma che gli permette di mantenere una corretta focalizzazione sulla retina dei i raggi luminosi.

La **tunica nervosa** è formata anteriormente dalla superficie posteriore del corpo ciliare e dell'iride, posteriormente dalla **retina** visiva. La zona di transizione tra queste due parti, come già detto, è chiamata ora serrata.

1.2 Morfologia della retina

La retina è una membrana molto sottile, polistratificata, sensibile agli stimoli luminosi che è in grado di convertire in impulsi bioelettrici, che riveste circa i due terzi posteriori del bulbo oculare, tranne che a livello della testa del nervo ottico (papilla ottica).

Al centro della retina, situata sull'asse visivo, si trova la **macula lutea**, area di circa 1500 micron di diametro la cui colorazione deriva da un pigmento giallo della famiglia dei carotenoidi. Da un punto di vista fisiologico, qui si trova la sede della visione centrale. La macula è a sua volta centrata dalla **fovea**, una piccola infossatura di circa 200-500 micron di diametro. La macula e, in particolare, la fovea, sono la sede della visione distinta, in quanto qui si concentra la maggiore quantità di raggi luminosi; inoltre la fovea rappresenta la regione retinica con maggiore acuità visiva grazie all'attività, pressoché esclusiva, di un tipo di fotorecettori, i coni: quando un oggetto attrae la nostra attenzione, gli occhi si muovono in modo tale che il fascio luminoso proveniente da tale oggetto converga sulla fovea.

Per inciso, l'acuità visiva rappresenta l'inverso delle dimensioni angolari minime che uno stimolo visivo deve possedere per provocare nel soggetto osservatore una risposta che ne segnali la corretta percezione. Quanto più piccole sono le dimensioni dello stimolo percepibile, tanto più elevata è l'acutezza visiva.

L'occhio normale riesce a distinguere due punti di un oggetto quando essi sono osservati sotto un angolo visuale di almeno $1' = 1/60$ di grado ($3 \cdot 10^{-4}$ rad). Questo valore definisce l'acuità visiva di un occhio normale e viene posto uguale a 10/10. Il potere accomodativo dell'occhio è tale da consentire la formazione sulla retina di immagini nitide di oggetti che si trovano a distanze variabili da circa 15 cm, definito punto prossimo, fino all'infinito, detto punto remoto; la visione più confortevole da

1.2 MORFOLOGIA DELLA RETINA

vicino, con i muscoli ciliari solo debolmente contratti, viene definita visione distinta e si ha per distanze di circa 25 cm.

Tornando alle strutture sopradescritte, all'esame oftalmoscopico queste appaiono come in figura sotto:

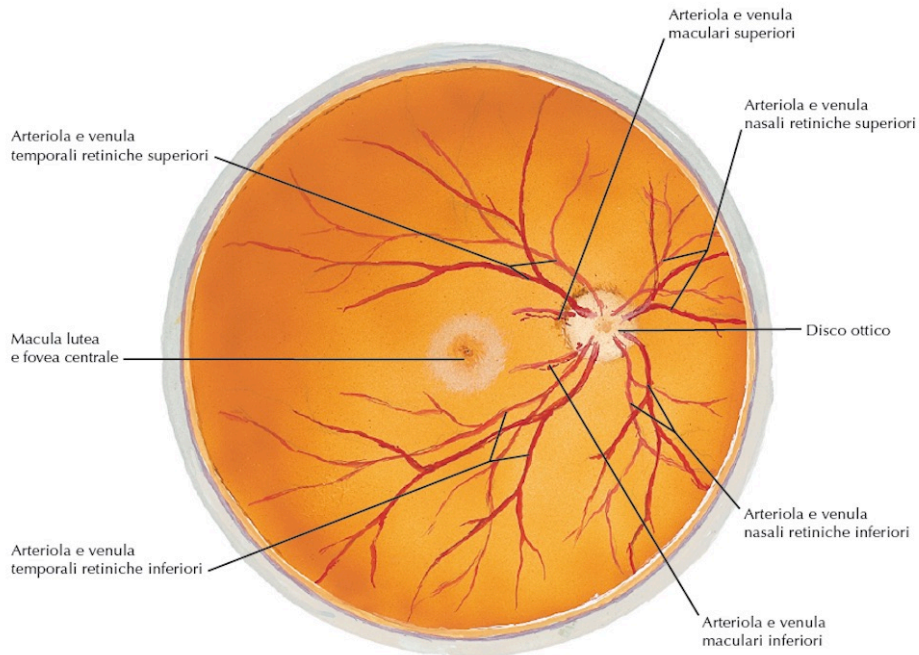


Figura 1.2 Rappresentazione del fondus oculi di un occhio destro. Centralmente è apprezzabile la macula, centrata dalla fovea, mentre medialmente e poco superiormente si trova la testa del nervo ottico. L'arteria centrale della retina, seguita dalla vena centrale della retina, si suddivide in rami superiori ed inferiori, che a loro volta si ramificano in rami temporali e nasali.

Ad una distanza di circa 3 mm sul versante nasale e circa 1 mm superiormente alla macula è visibile la **papilla**, o **testa del nervo ottico**, punto di uscita del nervo ottico; si tratta di una area ovalare di circa 1,5 mm di diametro, lievemente depressa centralmente e di colorito roseo, attraversata anche dall'arteria e dalla vena centrale della retina [8].

A livello della papilla ottica, tutti gli strati cessano tranne le fibre nervose, che si riuniscono a formare il nervo ottico. Questa regione è insensibile alla luce costituendo il punto cieco o macchia cieca della retina.

CAPITOLO 1. L'APPARATO VISIVO

Il nervo ottico emerge dalla parte posteriore del bulbo oculare, è il secondo nervo cranico ed è costituito principalmente dagli assoni delle cellule gangliari della retina che, attraversata la lamina cribrosa, vengono rivestiti dalla guaina mielinica.

Strutturalmente la retina è un tessuto nervoso, formato da un foglietto pluristratificato, la pars optica retinae, avvolto esternamente dall'epitelio pigmentato retinico.

L'epitelio pigmentato, che è in contiguità esternamente con la membrana di Bruch che lo separa dalla coroide, è un foglietto monostratificato di cellule densamente pigmentate, interconnesse da giunzioni strette a formare la barriera emato-retinica esterna. La barriera emato-retinica esterna impedisce l'ingresso di fluidi dalla coroide verso la retina.

L'epitelio pigmentato riduce la riflessione dei raggi luminosi, che renderebbe sfumate le immagini visive, e partecipa al metabolismo della vitamina A, fondamentale per la trasformazione dell'impulso luminoso in impulso nervoso.

La retina neurosensoriale è la porzione più interna della retina ed è costituita da una catena di tre neuroni con un apparato stromale gliale di sostegno [9]. Si tratta di una struttura formata da otto strati che dall'esterno all'interno sono:

1 *Strato dei fotorecettori*: i fotorecettori sono i coni e i bastoncelli e sono gli unici neuroni direttamente sensibili alla luce. Più precisamente questo strato contiene i segmenti esterni delle cellule neurosensoriali, i cui nuclei vanno a costituire lo strato nucleare esterno.

I bastoncelli, nell'ordine di $1-1,3 \times 10^8$, contengono la rodopsina, un pigmento visivo, e sono distribuiti su tutta la superficie retinica; sono sensibili agli stimoli luminosi più deboli e di conseguenza sono implicati nella visione crepuscolare in bianco e in nero. In quanto gli impulsi trasmessi da più cellule dei bastoncelli convergono sulla stessa cellula bipolare, la seconda cellula nervosa, non permettono una visione discriminata.

I coni, nell'ordine di $5-7 \times 10^6$, sono per lo più localizzati nella macula e richiedono per essere stimolati di una maggiore luminosità. Poiché ciascuno cono entra in sinapsi con una unica cellula bipolare che a sua volta entra in contatto con una determinata cellula gangliare, questi fotorecettori permettono la discriminazione visiva (visione distinta). Potendo contenere 3 diversi pigmenti visivi, cianolabile, clorolabile e eritrolabile, vengono distinti altrettanti tipi di coni, i quali risultano selettivamente sensibili a raggi luminosi con lunghezza d'onda rispettivamente del blu-violetto (437 nm), del verde (533 nm) e del rosso (564 nm); questo fenomeno permette la visione

cromatica e la percezione di un determinato colore dipenderà dalla stimolazione risultante tra i 3 differenti tipi di coni.

Coni e bastoncelli poggiano su una struttura definita membrana limitante esterna; non è una vera membrana e risulta formata dalla porzione esterna delle cellule di Muller. Queste cellule sono cellule gliali costituenti l'impalcatura della retina; i loro processi formano le due membrane limitanti della retina.

- 2 *Strato nucleare esterno*: contiene i nuclei delle cellule fotorecettoriali.
- 3 *Strato plessiforme esterno*: in questa sede si realizzano le sinapsi tra fotorecettori, cellule bipolari ed alcuni neuroni di associazione, le cellule orizzontali.
- 4 *Strato nucleare interno*: è formato dai nuclei delle cellule bipolari, di associazione (orizzontali e amacrine) e di Muller.
- 5 *Strato plessiforme interno*: in questo strato si formano le sinapsi tra cellule bipolari, cellule gangliari e di associazione amacrine.
- 6 *Strato delle cellule gangliari*: si tratta di cellule multipolari, grandi, costituenti i neuroni di secondo ordine della via ottica che con i loro assoni formano lo strato successivo nonché le fibre del nervo ottico. Questo strato è monostratificato tranne che a livello maculare, ove si ispessisce fino a otto file di cellule, e nella fovea, ove è assente,
- 7 *Strato delle fibre nervose*: contiene gli assoni delle cellule gangliari che convergendo da tutta la retina si riuniscono a formare il nervo ottico.
- 8 *Membrana limitante interna*: in questo strato si ha l'interconnessione delle espansioni terminali interni delle cellule di Muller, andando a separare la retina dal vitreo.

La pluristratificazione termina a livello dell'ora serrata, oltre la quale la retina diviene una struttura a due strati privi di funzione visiva, disposti a rivestire il corpo ciliare e la superficie posteriore dell'iride.

Quindi la retina ha uno spessore variabile a seconda della porzione considerata: nella periferia ha uno spessore di circa 140 μm , a livello maculare 230 μm , mentre nella fovea (porzione più centrale della macula) è più sottile, circa 100 μm , in quanto in questa sede ho meno strati a comporre la retina neurosensoriale.

Una ulteriore variazione dello spessore retinico avviene a livello della macula lutea, dove gli strati gangliari sono ispessiti per far in modo alla fovea di essere priva di terminazioni nervose; a questo ultimo livello infatti tutti gli strati più interni sono praticamente assenti e quello fotorecettoriale è costituito esclusivamente da coni.

CAPITOLO 1. L'APPARATO VISIVO

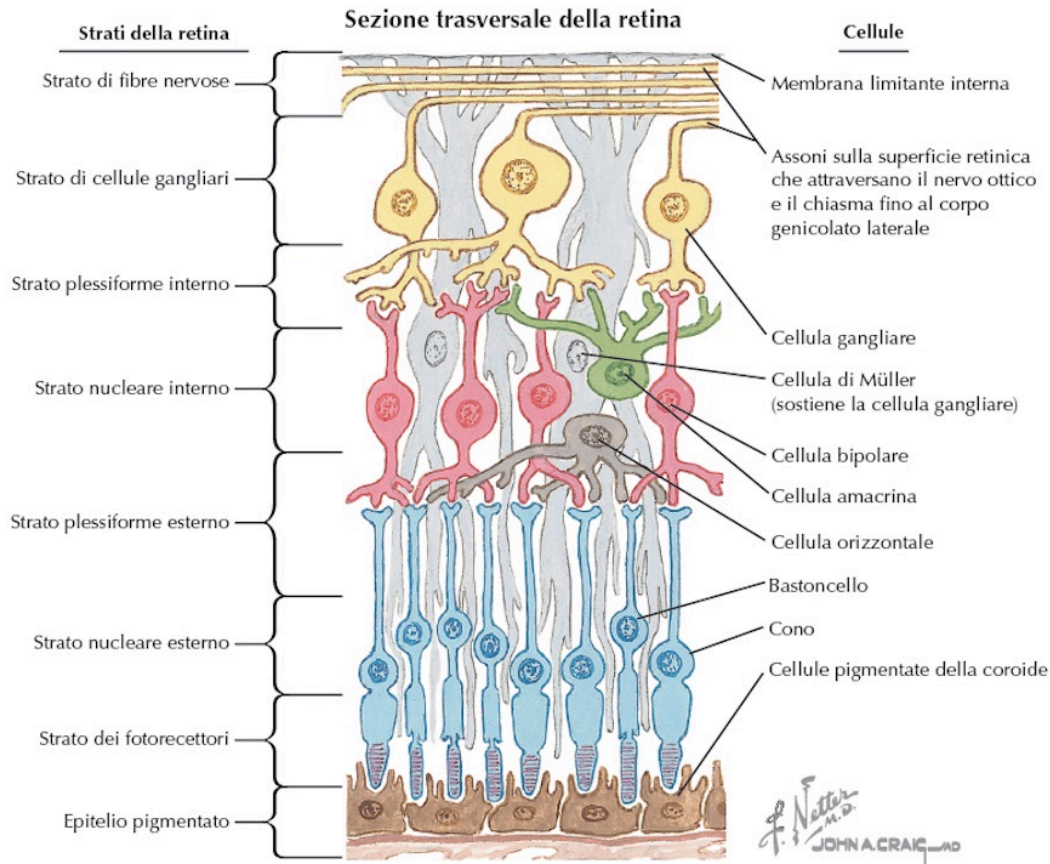


Figura 1.3 Sezione trasversale della retina neurosensoriale e dell'epitelio pigmentato retinico. Procedendo dall'esterno del bulbo oculare verso l'interno, più in profondità, si susseguono una serie di strati in cui si organizzano le cellule neuronali e gliali che compongono la retina: fotorecettori, cellule bipolari, cellule gangliari, cellule di associazione, cellule amacrine e cellule di Muller.

Riassumendo, nella retina si riconoscono 5 tipi di neuroni: i fotorecettori, le cellule bipolari, le cellule gangliari, le cellule amacrine e le cellule orizzontali.

I fotorecettori instaurano sinapsi con le cellule bipolari, che a loro volta fanno sinapsi con le cellule gangliari, i cui assoni convergono, formano il nervo ottico e lasciano il bulbo oculare. Le cellule orizzontali e le cellule amacrine interconnettono rispettivamente i fotorecettori nello strato plessiforme esterno e le cellule gangliari nello strato plessiforme interno.

Le cellule neuronali sono supportate da elementi gliali, le cellule di Muller. Poiché lo strato dei fotorecettori poggia sull'epitelio pigmentato, i raggi luminosi devono attraversare interamente la retina, superando gli altri neuroni che la compongono, prima di arrivare a coni e bastoncelli.

1.2 MORFOLOGIA DELLA RETINA

La vascolarizzazione della retina ed il suo nutrimento dipendono da un duplice sistema arterioso, entrambi tributari dell'arteria oftalmica, primo ramo emesso dall'arteria carotide interna nel suo ramo intracranico.

I due terzi interni della retina, quindi lo strato delle cellule gangliari e dei neuroni bipolari vengono irrorati da due plessi vascolari arteriosi situati rispettivamente nello strato delle fibre nervose e nello strato nucleare interno. Questi originano dalle diramazioni dell'arteria centrale della retina, ramo dell'arteria oftalmica.

L'arteria centrale della retina, penetrata nel bulbo oculare accompagnata dalla vena centrale della retina, si divide nelle due branche superiore e inferiore, che a loro volta si dividono in un ramo temporale (laterale) ed in un ramo nasale (mediale).

I capillari retinici presentano un endotelio con giunzioni strette formando la barriera emato-retinica interna. Gli strati più esterni della retina, sostanzialmente lo strato dei fotorecettori e l'epitelio pigmentato, sono riforniti dai vasi della lamina coriocalpillare, che derivano dalle arterie ciliari brevi e lunghe, anch'esse tributarie dell'arteria oftalmica.

Una eccezione a questo sistema è la regione maculare che, essendo priva di vasi retinici, dipende dalla rete coriocalpillare per la sua nutrizione; ciò comporta che in caso di distacco di retina posteriore si abbia un danno irreparabile.

Il ritorno venoso retinico è assicurato da quattro rami venosi che diretti verso la papilla confluiscono a formare la vena cava centrale della retina prima di uscire dal globo oculare.

All'esame oftalmoscopico la fovea viene riconosciuta per il suo colore più scuro rispetto alla retina circostante e per la pressoché totale assenza di vasi sanguigni. In corso di tale esame, una qualunque variazione dell'aspetto del disco ottico può avere un importante significato clinico: ad esempio un aumento della sua depressione centrale può essere indicativo della perdita di assoni delle cellule gangliari, viceversa, se il disco ottico dovesse apparire più spesso che di norma protrudendo verso il vitreo, ciò potrebbe essere una spia di edema papillare secondario ad esempio a ipertensione endocranica.

Le arteriole e le venule che si trovano negli strati superficiali della retina più interni, in prossimità del vitreo, possono essere studiate in oftalmoscopia. Alla luce di questo e in considerazione che le arteriole sono indagabili in modo non invasivo solo in questo distretto corporeo, l'esame oftalmoscopico assume notevole importanza nella diagnosi e nella valutazione delle vasculopatie sistemiche come il diabete mellito e l'ipertensione.

CAPITOLO 1. L'APPARATO VISIVO

L'anatomia dell'occhio sopradescritta determina la divisione del bulbo oculare in tre zone, che in senso antero-posteriore sono:

- la **camera anteriore**, spazio compreso tra la cornea e il diaframma irideo;
- la **camera posteriore** delimitata dalla parte posteriore dell'iride e dei corpi ciliari e dal cristallino; si tratta di una camera molto più piccola dell'anteriore ma ha un'importanza cruciale nell'omeostasi elettrolitica dell'interno dell'occhio.
- la terza camera, la **camera del corpo vitreo**, è delimitata dalla faccia posteriore dei corpi ciliari e del cristallino e dalla superficie retinica.

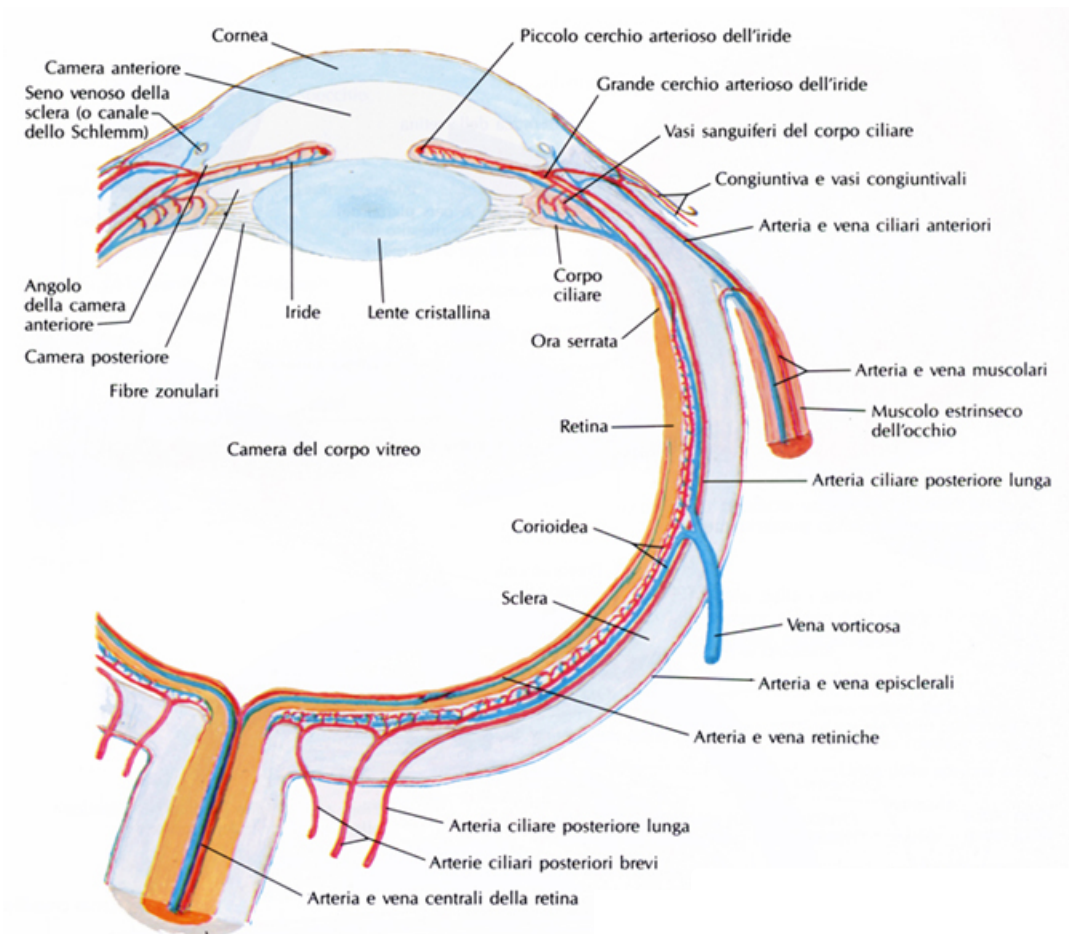


Figura 1.4 Immagine schematica della vascolarizzazione retinica. I due terzi anteriore della retina vengono alimentati da due plessi capillari che sono tributari dell'arteria centrale della retina; il terzo posteriore retinico invece riceve apporto ematico dal plesso capillare situato nella lamina corio-capillare, originante dalle arterie ciliari posteriori. Entrambi gli apporti ematici fanno capo all'arteria oftalmica (non mostrata in figura). In piani anteriori sono ben evidenti la camera anteriore, la camera posteriore e la camera del corpo vitreo.

1.3 PRINCIPALI PATOLOGIE DIAGNOSTICABILI TRAMITE L'ANALISI DEL FONDO OCULARE

Le prime due camere contengono umor acqueo, un fluido prodotto dal corpo ciliare, trasparente, che fornisce le sostanze nutritive necessarie alla cornea ed al cristallino. La terza camera contiene umor vitreo, una sostanza gelatinosa con funzione trofica. La pressione dei liquidi endo-oculari determina la consistenza del bulbo oculare, e generalmente varia tra i 10 e i 20 mmHg.

Questi ultimi concetti, insieme con una rapida rappresentazione della vascolarizzazione retinica, sono evidenti nell'immagine in *Figura 1.4*.

Ai fini di una miglior comprensione della clinica delle malattie retiniche, va ricordato che queste in genere non provocano dolore, in quanto la retina non contiene fibre dolorifiche, ma si accompagnano a disturbi della funzione visiva.

1.3 Principali patologie diagnosticabili tramite l'analisi del fondo oculare

1.3.1 Glaucoma

Il termine glaucoma fa da cappello a un gruppo di malattie che presentano tre caratteristiche in comune: un aumento variabile della pressione intra-oculare, una peculiare modificazione della papilla ottica, definita escavazione glaucomatosa, e la presenza di alterazioni ingravescenti del campo visivo.

Nel mondo il glaucoma è la terza causa di cecità; la malattia colpisce per lo più soggetti oltre la IV decade di vita: un terzo della popolazione over-65 anni presenta ipertensione oculare e il 10% di questo gruppo sviluppa il glaucoma.

Fisiologicamente l'umor acqueo viene prodotto dal corpo ciliare e il suo deflusso avviene mediante il canale dello Schlemm a livello dell'angolo della camera anteriore (sotteso da una linea tangente il piano irideo e da una linea tangente alla cornea); tale angolo in condizioni normali è di circa 40°.

In un soggetto sano la pressione endo-oculare oscilla nel corso della giornata tra 3 e 6 mmHg con un picco nelle prime ore del mattino. I pazienti glaucomatosi hanno uno squilibrio tra i fenomeni di produzione e riassorbimento tale per cui la loro pressione intra-oculare è superiore a 10 mmHg.

CAPITOLO 1. L'APPARATO VISIVO

Da un punto di vista eziopatogenetico, il glaucoma è suddivisibile in glaucoma primario, da alterazione dei meccanismi di deflusso dell'umor acqueo e glaucoma secondario da cause acquisite secondarie che meccanicamente portano ad un innalzamento della pressione endoculare.

Il glaucoma primario si suddivide ulteriormente nelle forme:

- ad angolo aperto, principale forma di glaucoma cronico, con una pressione endo-oculare che di rado supera i 30 mmHg;
- ad angolo stretto, che presenta in genere un esordio acuto, ove si arriva anche a 60-70 mmHg.

L'aumento della pressione intra-oculare induce primariamente un danno di tipo progressivo a livello della papilla del nervo ottico: se negli stadi iniziali si può osservare atrofia degli assoni con conseguente aumento del diametro e della profondità della escavazione fisiologica della testa del nervo ottico, nelle fasi avanzate si può realizzare il quadro dell'atrofia ottica glaucomatosa.

Il glaucoma, indipendente dai fattori patogenetici, si può perciò configurare come una neuropatia ottica, lentamente progressiva, associata ad un aumento della pressione intraoculare [10].

La porzione del disco ottico più sensibile al danno causato da una elevata pressione intra-oculare è costituita dagli assoni delle regioni infero-temporali e supero-temporale del nervo ottico. Queste sono le prime strutture assoniche ad essere danneggiate, provocando nelle fasi iniziali la comparsa di un tipico difetto visivo perimetrico, detto difetto fascicolare, cioè uno scotoma arcuato a partenza dalla macchia cieca.

Se, nel tempo, la pressione endo-oculare dovesse persistere elevata, ulteriori fibre andrebbero distrutte. Questo indurrebbe una ulteriore riduzione del contorno nervoso della papilla ottica, la comparsa di un maggiore infossamento patologico e la progressiva riduzione del campo visivo a circondare tutta l'area della fissazione centrale.

Nella *Figura 1.5* [11] viene esemplificata la normale anatomia e l'alterazione patologica della papilla ottica che si realizza nei pazienti con glaucoma.

Trattandosi di una delle principali cause di disabilità visiva, la diagnosi precoce è di importanza cruciale.

Esistono alcune difficoltà diagnostiche. Innanzitutto esiste un infossamento fisiologico, stabile nel tempo, in alcuni soggetti sani che presentano un

1.3 PRINCIPALI PATOLOGIE DIAGNOSTICABILI TRAMITE L'ANALISI DEL FONDO OCULARE

infossamento particolarmente largo e profondo che deve essere differenziato dall'infossamento patologico tipico del glaucoma, che aumenta nel tempo.

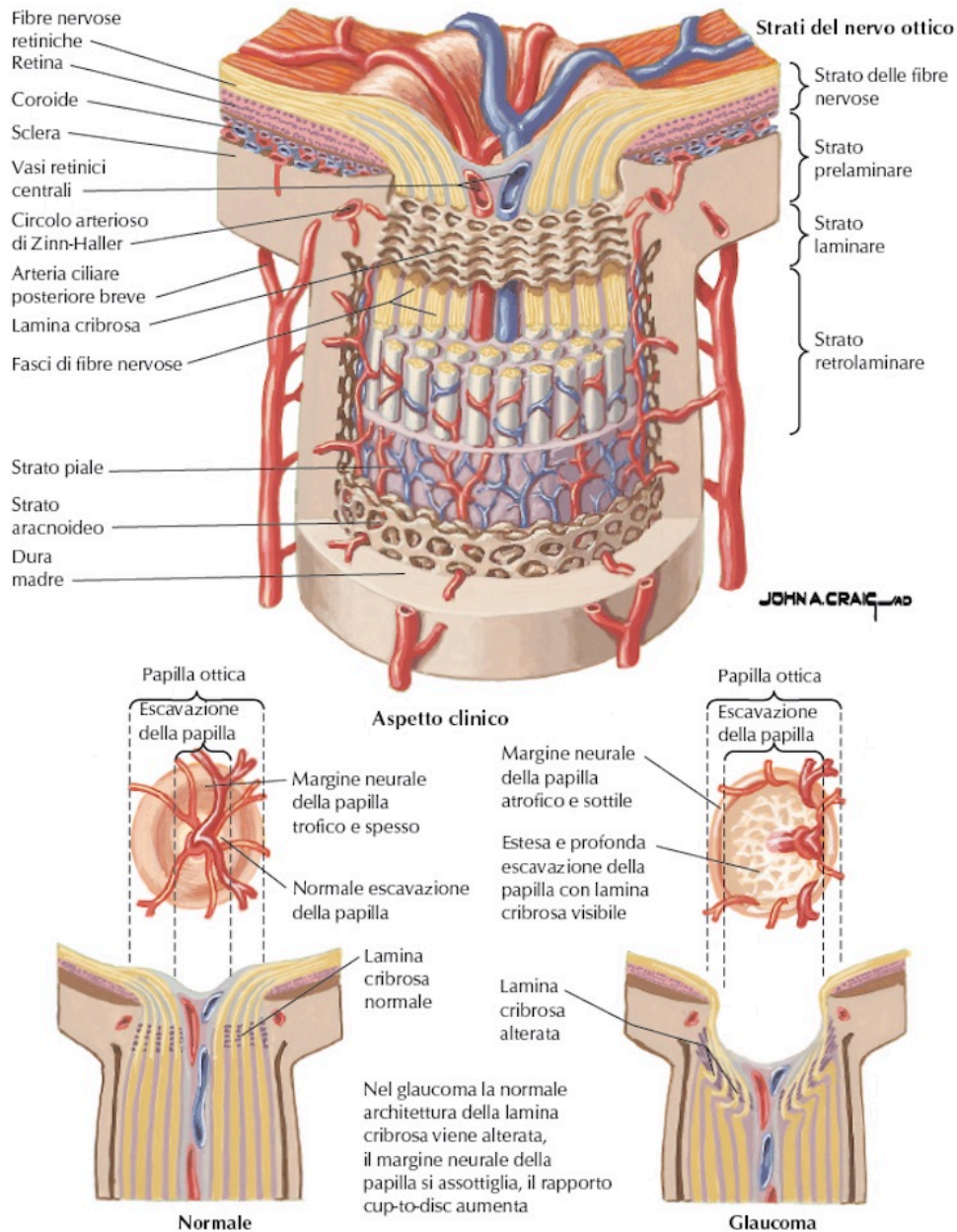


Figura 1.5 Sezione trasversale e dettaglio sul piano frontale della papilla ottica. A sinistra, in condizioni di normalità il margine neurale della papilla è trofico e si apprezza una fisiologica escavazione della papilla stessa; a destra, nei soggetti con glaucoma il margine neurale della papilla si assottiglia, la lamina cribrosa viene distorta, l'escavazione della papilla si fa più marcata e il rapporto cup-to-disc, cioè il rapporto tra il diametro dell'escavazione ed il diametro totale del disco ottico, aumenta.

CAPITOLO 1. L'APPARATO VISIVO

In secondo luogo, il danno al nervo è inizialmente asintomatico e la perdita del campo visivo periferico si realizza più tardivamente; purtroppo perdere il campo visivo periferico con risparmio dell'acuità visiva foveale fa sì che non ci si renda conto della perdita visiva fino a stadi molto avanzati.

Se generalmente il glaucoma è indolore, con scarsi sintomi e segni, il glaucoma ad angolo stretto si presenta con occhio rosso e dolente, annebbiamento della vista, edema corneale, nausea e talora vomito e pressione intraoculare molto elevata (anche 60-70 mmHg). Ciò dipende dall'improvviso e importante aumento della pressione intraoculare. I tre segni di malattia evidenziabili restano gli stessi: aumento della pressione intra-oculare, un aumento dell'escavazione del disco ottico e la presenza dei tipici difetti fascicolari del campo visivo.

In ultimo esiste una forma di glaucoma a bassa tensione, con presenza dei tipici danni (due dei tre segni), anche in presenza di una pressione intra-oculare normale o inferiore della norma.

Per questi motivi la visita oftalmica dei pazienti affetti da glaucoma comprende la valutazione del rapporto tra infossamento e papilla ottica all'oftalmoscopio e la misurazione della pressione intraoculare è di vitale importanza.

1.3.2 Le malattie della retina

Per retinopatia si intende una qualunque malattia della retina, generalmente secondaria a un disordine sistemico. Da un punto di vista eziopatogenetico, una retinopatia può insorgere nel contesto sia di patologie infiammatorie che di processi di tipo degenerativo-distrofico.

Tra le forme più diffuse ci sono la retinopatia ipertensiva, la retinopatia diabetica e la retinopatia del pretermine.

Le retinopatie su base infiammatoria spesso interessano contemporaneamente la retina neurosensoriale, l'epitelio pigmentato e la coroide; ne consegue lo sviluppo di edema maculare e/o distacco retinico essudativo, che comportano una notevole riduzione dell'acuità visiva.

Nel novero delle forme degenerative vi è la degenerazione maculare senile (o correlata all'età). Anche in queste forme il coinvolgimento retinico e coroidale avviene contemporaneamente nella quasi totalità dei casi.

1.3.2.1 Degenerazione Maculare correlata all'età

È una malattia degenerativa che colpisce la macula ed è la causa più frequente di perdita visiva bilaterale, centrale e graduale fra le persone anziane.

La malattia può manifestarsi in due forme: atrofica o neovascolare (o essudativa) [12].

- Nella Degenerazione Maculare Atrofica, la più comune (circa 85% dei casi), si ha inizialmente lo sviluppo di alcune alterazioni dell'epitelio pigmentato che possono proseguire fino alla comparsa di atrofia completa dell'epitelio stesso, nonché delle cellule neurosensoriali maculari della retina sovrastante.

Il processo non essudativo comincia con l'accumulo di depositi extracellulari, chiamati drusen, al di sotto dell'epitelio retinico. All'esame oftalmoscopico questi depositi appaiono come piccole lesioni distinte, pleomorfe e di colore giallo, maggiormente concentrate all'interno della macula.

Con la progressione della malattia tali alterazioni aumentano di numero e di dimensioni, tendendo a confluire tra loro. L'esito finale del processo patologico è il distacco focale dell'epitelio pigmentato della retina, che tende a divenire atrofico.

Clinicamente si ha la progressiva riduzione della visione centrale per interferenza con la funzione dei fotorecettori.

Tale forma può rimanere come tale oppure evolvere nella forma essudativa.

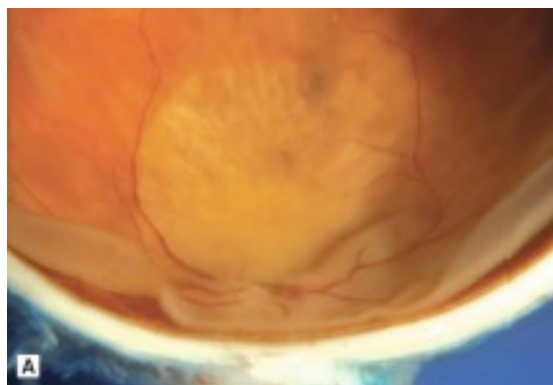


Figura 1.6 Degenerazione maculare correlata all'età, forma atrofica. Si può osservare una estesa area di atrofia dell'epitelio pigmentoso retinico in sede polare posteriore del bulbo oculare.

- La Degenerazione Maculare Essudativa colpisce una minoranza di pazienti, sia come processo patologico ex novo che, più frequentemente, come evoluzione di una forma atrofica. Infatti la degenerazione dell'epitelio pigmentato e dello strato dei fotorecettori provoca delle fessurazioni nella lamina di Bruch, attraverso le quali esili vasi neoformati di provenienza coroideale infiltrano lo spazio virtuale al di sotto dell'epitelio pigmentato della retina.

Questa forma è definita essudativa perché le cellule endoteliali parietali di questi vasi anomali non presentano giunzioni strette, permettendo il passaggio di componenti plasmatiche nel tessuto retinico; si avranno edema retinico e, in una fase più avanzata della malattia, essudati duri (agglomerati di lipidi e mucopolisaccaridi nella retina).

Clinicamente, l'infiltrazione vascolare provoca un sollevamento della retina e dell'epitelio pigmentato con distorsione della vista (metamorfopsia) e comparsa di una visione sfuocata.

Infine, negli stadi finali, il susseguirsi di emorragie ripetute può indurre lo sviluppo di fibrosi, con formazione di una cicatrice tondeggianti in regione maculare (cicatrice disciforme).

Le membrane neo-vascolari sono pressoché irriconoscibili all'esame del fondo dell'occhio in quanto sub-retiniche.

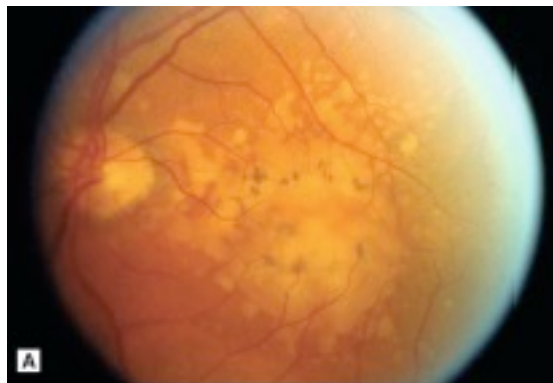


Figura 1.7 Degenerazione maculare correlata all'età, forma essudativa. Sono evidenti drusen confluenti a coprire la retina.

1.3.2.2 Retinite pigmentosa

Si tratta della forma più comune di distrofia retinica. In realtà è un termine generico che fa da cappello a diverse condizioni patologiche, sporadiche o associate a

1.3 PRINCIPALI PATOLOGIE DIAGNOSTICABILI TRAMITE L'ANALISI DEL FONDO OCULARE

ereditarietà autosomica, che interessano primariamente la retina e l'epitelio pigmentato [13].

La retinite pigmentosa in senso lato si manifesta clinicamente con una progressiva cecità notturna e, in stadi più avanzati, da una riduzione del campo visivo con scotoma ad anello (campo visivo tubulare).

Con l'esame oftalmoscopico si possono apprezzare alcuni depositi irregolari neri di pigmento ammassato alla periferia della retina, da cui il nome della malattia.

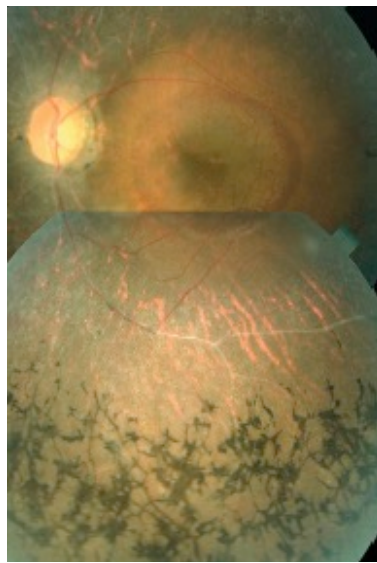


Figura 1.8 Retinite pigmentosa. La papilla ottica è pallida; sono evidenti pigmenti scuri nella periferia retinica

1.3.2.3 Vasculopatie retiniche

In generale, si tratta di un gruppo eterogeneo di patologie, con differente natura, che determinano occlusione del microcircolo retinico e conseguentemente l'ischemia retinica.

Il coinvolgimento retinico in corso di vasculopatia può essere sia di tipo macro-vascolare che micro-vascolare.

La retina risponde allo stato di ischemia con l'induzione di una neo-vascularizzazione che può essere epiretinica o epipapillare. Un po' come avviene per la Degenerazione Maculare Essudativa, la neo-vascularizzazione è un processo non fisiologico che porta con sé delle complicanze: l'emorragia vitreale, il distacco retinico retrattivo e il glaucoma neo-vascolare.

1.3.2.4 Retinopatia ipertensiva

Secondaria all'ipertensione arteriosa grave, da un punto di vista anatomicopatologico si caratterizza in una prima fase per una tipica modificazione del calibro e del decorso dei vasi arteriosi retinici, che tendono ad essere più rettilinei; i primi sintomi sono evidenti solo con alterazioni strutturali di un certo rilievo e constano di visione annebbiata e distorta. Con il progredire della malattia si assiste ad un aumento del riflesso assiale, all'instaurarsi di una fibrosi parietale e alla comparsa di segni di schiacciamento dell'arteria sulla vena [14].

In fasi più avanzate si verifica una dilatazione venosa, un aumento della tortuosità dei vasi, e la comparsa di segni indiretti di ischemia retinica: micro- e macroaneurismi, emorragie retiniche, noduli cotonosi (piccoli infarti dello strato delle fibre gangliari) ed essudati duri.

Fenomeni analoghi si possono realizzare anche a livello della coroide e del nervo ottico, con edema papillare, segno di ipertensione cerebrale.

Sebbene le prime manifestazioni della retinopatia ipertensiva siano simili a quelle para-fisiologiche dell'arteriolosclerosi retinica, che è per lo più correlata all'età, la comparsa delle manifestazioni più tardive permette di porre diagnosi di vasculopatia secondaria ad ipertensione grave.

In corso di esame oftalmoscopico si può valutare il calibro e il decorso dei vasi arteriosi e venosi, nonché la presenza di micro-emorragie e edema.

1.3.2.5 Occlusione dell'arteria e della vena centrale della retina

L'occlusione dell'arteria centrale della retina è un evento acuto, secondario per lo più a embolismo cardiaco o carotideo e in una minoranza di casi a fenomeni aterosclerotici complicati o a patologie infiammatorie, come le arteriti [15].

Spesso l'occlusione completa è preceduta da fenomeni transitori, che fungono da allarme, come l'amaurosi fugace, espressione generalmente di fenomeni microembolici.

Clinicamente si può avere una perdita completa della vista omolateralmente alla sede della malattia.

All'esame oftalmoscopico si osserva una retina pallida, bianco-giallastra e con vasi arteriosi esili per l'edema ischemico, in cui spicca la fovea di colore rosso ciliegia, per la coroide sottostante.

1.3 PRINCIPALI PATOLOGIE DIAGNOSTICABILI TRAMITE L'ANALISI DEL FONDO OCULARE

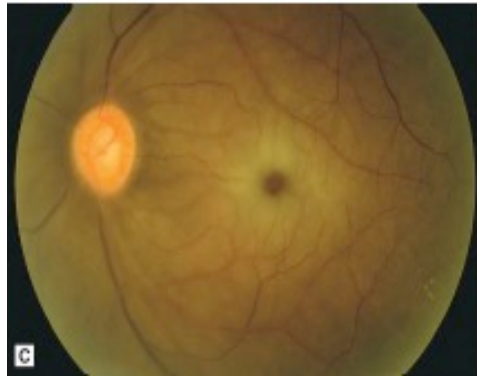


Figura 1.9 Occlusione dell'arteria centrale della retina: la fovea appare come un punto di colore rosso ciliegia nel contesto di una retina ischemica, opaca e pallida.

Anche l'occlusione della vena centrale della retina può comportare una rapida perdita dell'acuità visiva. La patologia è in genere secondaria a vasculopatia retinica di tipo arteriolosclerotico-ipertensivo, sebbene nei giovani sia documentata una natura del processo di tipo infiammatorio.

Oftalmoscopicamente si possono osservare vasi venosi congesti, edema della papilla ed emorragie retiniche.

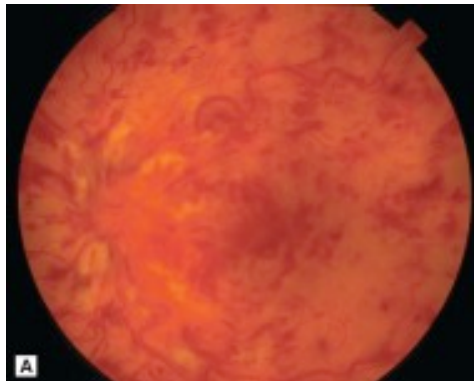


Figura 1.10 Occlusione della vena centrale della retina. Sono presenti molte emorragie, superficiali e profonde. I vasi venosi appaiono dilatati e tortuosi, mentre la papilla ottica, edematosa, è circondata da essudati molli.

1.3.2.6 Retinopatia diabetica

La retinopatia diabetica è una complicanza cronica di tipo micro-vascolare del diabete, per lo più associata ad un mal controllato livello glicemico [16].

CAPITOLO 1. L'APPARATO VISIVO

Si tratta di un disturbo che, raro fino all'introduzione del trattamento insulinico, attualmente è una delle principali cause di cecità e di ipovisione.

Una metanalisi del 2012 [17] suggerisce come prevalenza globale della retinopatia diabetica e della retinopatia diabetica vision-threatening nei pazienti affetti da diabete sia rispettivamente del 35% e del 10%.

Una prima considerazione è che la retinopatia diabetica, sebbene impieghi anni per svilupparsi, si manifesta ineluttabilmente, con un danno d'organo evidenziabile in circa l'80% dei soggetti con esordio del diabete da almeno 15 anni. Essendo questa condizione correlata al livello degli zuccheri ematici, è dimostrato che mantenere un buon controllo glicemico rallenti la progressione del danno funzionale all'apparato visivo.

Alla luce di questo emerge l'esigenza di uno screening precoce per identificare anche le più lievi manifestazioni organico-funzionali della malattia.

Da un punto di vista del danno d'organo, che precede di anni il danno funzionale, nel novero della retinopatia diabetica vengono riconosciute un certo numero di alterazioni, variamente associate tra loro, a carico dei capillari retinici:

- microaneurismi, cioè piccole ectasie capillari;
- aumento della permeabilità vasale; questo determina la compromissione della barriera emato-retinica, la diffusione di liquido plasmatico intra-retina e conseguente l'instaurarsi di edema retinico;
- fenomeni occlusivi;
- emorragie, le più comuni sono le petecchie emorragiche, localizzate negli strati più interni della retina, e le emorragie a fiamma, che si realizzano a livello dello strato delle fibre nervose;
- essudati duri, secondari all'aumento della permeabilità capillare e associati all'edema retinico, si formano per la deposizione della fibrina e dei lipidi fuoriusciti dal microcircolo nel contesto della retina (il termine "duro" deriva dalla morfologia dei loro margini che sono netti, duri per l'appunto);
- essudati molli, detti noduli cotonosi.

Vengono distinte due forme di retinopatia diabetica: retinopatia diabetica non proliferante e retinopatia diabetica proliferante.

La retinopatia diabetica non proliferante solitamente compare verso la fine della prima o l'inizio della seconda decade di malattia ed è caratterizzata dalla presenza di fenomeni occlusivi e aumento della permeabilità vasale a carico dei capillari retinici. Ne consegue la formazione di microaneurismi vascolari in presenza di aree

1.3 PRINCIPALI PATOLOGIE DIAGNOSTICABILI TRAMITE L'ANALISI DEL FONDO OCULARE

retiniche scarsamente perfuse; inoltre i vasi fragili e permeabili facilitano la comparsa di emorragie retiniche puntiformi (petecchie) e di essudati duri. Queste lesioni, tipiche della fase iniziale della malattia retinica diabete-correlata, costituiscono la cosiddetta retinopatia diabetica background.

Entrando nello specifico, vengono distinti tre stadi di retinopatia diabetica non proliferante:

- forma lieve: caratterizzata dalla presenza di almeno un microaneurisma;
- forma moderata: devono essere presenti emorragie o microaneurismi in maniera superiore rispetto ad una foto standard presente nelle linee guida;
- forma grave: emorragie e/o microaneurismi in tutti i quadranti retinici; le vene in uno o più quadranti devono avere aspetto a corona di rosario (il profilo della parete non è regolare ma ha piccole stenosi).

Nella forma grave le anomalie vascolari precedono la proliferazione dei vasi, tipica della forma proliferativa. Le emorragie sono un segno indiretto ma fortemente suggestivo di neoangiogenesi.

Un esame utile per localizzare i neovasi è l'esame angiografico con fluorescina: poiché il colorante non è trattenuto dai vasi anomali, oltre a definire la topografia di queste strutture, esso si diffonde in corrispondenza di aree retiniche ischemiche, permettendo una valutazione.



Figura 1.11 Fondo oculare di un paziente con retinopatia diabetica background. Si può osservare la presenza di emorragie, essudati duri, essudati molli ed edema retinico.

La forma non proliferante può persistere come tale o evolvere verso la forma più grave di retinopatia diabetica proliferante, con comparsa di una neovascolarizzazione epiretinica o epicapillare (verso il vitreo) secondaria alla

CAPITOLO 1. L'APPARATO VISIVO

ischemia retinica. Le alterazioni descritte nella forma non proliferante grave, associate al processo oclusivo dei capillari retinici, comportano un processo flogistico cronico importante realizzando un diffuso danno capillare, con ispessimento della membrana basale dell'endotelio, sclerosi parietale e perdita dei periciti peri-capillari.

Questa cascata di eventi determina l'alterazione del flusso ematico retinico che favoriscono l'ipoperfusione retinica.

Con l'avanzare del danno organico si vengono a creare regioni francamente ischemiche dove si evidenziano lesioni con aspetto cotonoso, gli essudati molli ("cotton wools").

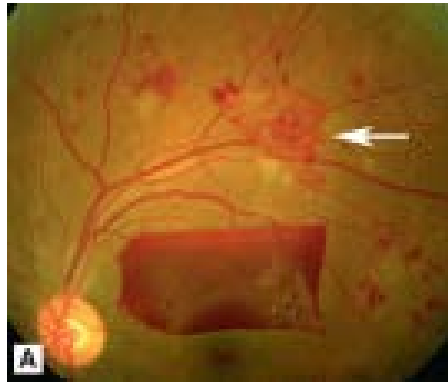


Figura 1.12 Fondo oculare di un paziente con retinopatia diabetica proliferativa: la freccia indica una sede di neovascolarizzazione.

I vasi neoformati possono localizzarsi in corrispondenza del nervo ottico e/o della macula e possono andare facilmente incontro a rottura, con conseguente emorragia nel vitreo. Il tessuto fibrovascolare che avvolge i neo-vasi e la fibrosi del vitro, secondaria ad emorragie ripetute, determinano una azione retraente sulla retina, fino a provocare distacco retinico e/o lacerazione retinica.

Nell'ambito delle due forme di retinopatia diabetica può esserci un coinvolgimento prevalente della regione maculare con emorragie, essudati duri ed edema che caratterizzano il quadro della maculopatia diabetica.

Una grave complicanza della retinopatia diabetica con estesa ischemia è la comparsa di un glaucoma secondario di tipo neovascolare.

Da un punto di vista clinico, la cecità è il risultato di una retinopatia diabetica progressiva con edema maculare clinicamente significativo; infatti se in una prima fase le lesioni retiniche possono realizzarsi senza comportare alterazioni del visus, ogniqualvolta venga interessata la macula, sede della visione distinta, inevitabilmente si ha l'instaurarsi di un deficit visivo.

1.3 PRINCIPALI PATOLOGIE DIAGNOSTICABILI TRAMITE L'ANALISI DEL FONDO OCULARE

L'esame regolare del fondo oculare dilatato è indispensabile nel paziente affetto da diabete.

All'esame oftalmoscopico solo i microaneurismi più grandi possono essere visti, in quanto il riflesso rosso della retina maschera quelli di diametro minore, e appaiono come minuti punti rossi. Gli essudati duri appaiono come chiazze di colore giallastro a margini netti, mentre gli essudati duri appaiono come aree nodulari, a margini sfumati, bianco-grigiastre. Bisogna osservare che sebbene la presenza di essudati indichi che ci sia stato edema, ciò non indica che si tratti di un edema attualmente presente.

Anche la fluoroangiografia, che può evidenziare una diffusione del colorante, segno di una fase proliferativa, non necessariamente dimostra edema retinico.

Nelle forme avanzate e nella forma proliferativa vengono visualizzati le diffuse emorragie retiniche e i segni di neovascolarizzazione.

Nell'ambito dello screening, in considerazione del fatto che il diabete mellito di tipo II viene diagnosticato generalmente quando la malattia diabetica persiste da molto tempo, una valutazione iniziale con monitoraggio periodico è fortemente raccomandata. Nei pazienti con DM di tipo I non è consigliato uno screening precoce in quanto sappiamo che la retinopatia diabetica richiede un certo numero di anni prima di evidenziarsi clinicamente; a questa categoria di pazienti viene consigliato un follow-up correlato al loro controllo glicemico.

Una particolarità è rappresentata dalle donne incinte, alle quali è consigliata un'appropriata visita oftalmica perché nel corso della gravidanza il pattern endocrinologico della donna cambia.

1.3.2.7 Retinopatia della prematurità

Nel novero delle vasculopatie retiniche rientra la retinopatia della prematurità, una retinopatia bilaterale, di natura vascolare e di origine iatrogena, secondaria ad un eccesso di ossigenazione oculare dei bambini prematuri con comparsa di neovascolarizzazione periferica. La sua evoluzione clinica non necessariamente prevede una restitutio ad integrum e si può verificare una proliferazione fibrovascolare e/o distacco di retina con grave compromissione della vista. Fortunatamente con il miglioramento dell'assistenza medica neonatale si sta riducendo l'incidenza di questa patologia.

Capitolo 2

Tecnologie per l'analisi del fondo oculare

Gli occhi rappresentano il canale primario di comunicazione: circa l'80% di tutte le informazioni che giungono dall'ambiente esterno al nostro cervello sono fornite dal senso della vista.

Le patologie che affliggono l'apparato visivo, sia primarie che secondarie a malattia sistemica, possono potenzialmente portare ad una menomazione visiva, che nei casi più gravi si concretizza in ipovisione o cecità. Purtroppo queste condizioni sono spesso sottodiagnosticate in quanto si manifestano con una blanda sintomatologia o sono addirittura asintomatiche; in tale ottica solo uno screening precoce razionale e controlli periodici mirati possono offrire una diagnosi rapida ed in tempi utili al processo terapeutico.

Prevenzione e diagnosi precoce si configurano quindi come un fondamentale ausilio medico per identificare, ed eventualmente monitorare, le malattie dell'occhio in una fase di danno organico prima che funzionale, in modo da impedirne l'aggravarsi nel tempo e ridurre l'influenza sulla qualità di vita.

Il riconoscimento di eventuali danni retinici e/o del nervo ottico, che sono le sedi più frequentemente colpite da affezioni oculari invalidanti come la retinopatia diabetica, il glaucoma, le maculopatie e il distacco della retina, può essere ottenuto mediante l'esame del fondo oculare.

2.1 Esame del fondo oculare

L'indagine oftalmoscopica è parte integrante dell'esame obiettivo degli occhi in quanto necessaria per una valutazione delle strutture del bulbo oculare poste dietro al cristallino, il corpo vitreo e la retina.

CAPITOLO 2. TECNOLOGIE PER L'ANALISI DEL FONDO OCULARE

L'importanza di poter indagare le modificazioni e le alterazioni della retina e della sua vascolarizzazione risiede nel fatto che, oltre allo studio della malattia retinica di per sé e quindi le patologie degenerative o distrofiche, si possono valutare i cambiamenti delle strutture dell'occhio posteriore secondarie a certe malattie sistemiche.

Il fondo dell'occhio è infatti l'unica parte del corpo umano dove è possibile osservare direttamente il micro-circolo e i primi segni di una vasculopatia sistemica. Per esempio, l'esame del fondo oculare offre la possibilità di rilevare quelle variazioni di calibro e di decorso che si riscontrano nei vasi della retina nei soggetti con ipertensione arteriosa, arteriosclerosi o diabete mellito.

Oltre a ciò, dato che in corso di esame oftalmoscopico può essere visualizzato anche il nervo ottico, o meglio la papilla ottica, questa indagine può diagnosticare un aumento della pressione endocranica che generalmente si manifesta proprio con un rigonfiamento del disco ottico, definito papilledema.

Non di minore importanza, l'indagine del fondo oculare è un ausilio nella diagnosi dei tumori della retina.

Tradizionalmente l'esame del fundus oculi prevede l'utilizzo di una serie di dispositivi: l'oftalmoscopio diretto, l'oftalmoscopio indiretto, la lampada a fessura (o biomicroscopio) e la fundus camera. Si tratta di dispositivi, descritti in seguito, che permettono all'oftalmologo di vedere la retina e le sue strutture, ma solo l'ultimo è un vero e proprio sistema di imaging del fondo oculare ad alta risoluzione.

L'esame viene condotto in un ambiente scarsamente illuminato dopo aver indotto una dilatazione della pupilla mediante instillazione di speciali colliri midriatici.



Figura 2.1 Fotografia di un normale fondo oculare destro.



Figura 2.2 Fotografia di un normale fondo oculare sinistro.

La midriasi, che comporta un temporaneo annebbiamento della vista accompagnato da un senso di abbagliamento, si instaura circa un quarto d'ora dopo la somministrazione del farmaco, raggiunge un acme in circa 45-60 minuti e termina completamente nell'arco di 5-6 ore. Di conseguenza si tratta di un esame che va accuratamente programmato.

Sono stati documentati, seppur raramente, casi di reazione avversa al collirio. Ovviamente gli effetti sono correlati alla concentrazione utilizzata e al quantitativo di collirio instillato nell'occhio.

Va precisato che l'esame del fondo oculare è possibile anche senza il ricorso a un collirio midriatico, ma allo stato dell'arte si ritiene che la midriasi consenta una valutazione più efficace.

Generalmente non ha controindicazioni, tuttavia, bisogna prestare particolare attenzione nei pazienti affetti da glaucoma, specie in quelli con ostacolato deflusso dell'umor acqueo, e nei pazienti con camera anteriore dell'occhio poco profonda. In tali casi l'esame viene effettuato con la pupilla non dilatata, esplorando in tal modo quasi esclusivamente il polo posteriore della retina.

In corso di esame oftalmoscopico si può ricorrere all'utilizzo di alcuni filtri che permettono di magnificare selettivamente le specifiche strutture dell'occhio. Un filtro rosso privo, per esempio, maschera in parte le strutture vasali e rende meglio identificabili le fibre nervose: ciò migliora l'osservazione di piccole emorragie, minuti buchi retinici o iniziali degenerazioni maculari.

Un filtro blu cobalto associato alla somministrazione di fluoresceina aiuta nell'identificazione di eventuali alterazioni della cornea.

Con una luce polarizzata, che minimizza gli effetti che provengono dai riflessi della cornea, si è in grado, riducendo l'ingrandimento, di visualizzare le strutture del segmento anteriore.

2.2. Oftalmoscopio

L'oftalmoscopio è uno strumento diagnostico specialistico usato dal medico oculista per osservare ed esaminare, attraverso il foro pupillare, le strutture interne dell'apparato visivo, in particolare la retina e la testa del nervo ottico, consentendone lo studio dei caratteri morfologici e funzionali.

2.2.1 Cenni storici

L'invenzione dell'oftalmoscopio può essere considerata il più importante progresso nella storia dell'oftalmologia.

Le sue origini si fondano nell'antica osservazione che gli occhi degli animali crepuscolari assumono un aspetto luminoso nella penombra. Nel mondo antico esisteva la credenza che vi fosse uno sviluppo spontaneo di luce all'interno dell'occhio per consentire agli animali di vedere al buio.

Sulla base di queste osservazioni, nel corso dei secoli furono elaborate diverse teorie della visione.

Edmé Mariotte (1620-1684) osservò che l'occhio di un cane era luminoso, ma erroneamente pensò che ciò fosse dovuto al fatto che la sua coroide era bianca mentre nell'uomo era scura [18].

Nel 1704, Jean Méry (1645-1722), chirurgo all'ospedale Hôtel-Dieu di Parigi, scoprì che immergendo la testa di un gatto sott'acqua poteva essere apprezzata la luminosità dell'occhio [19]. In particolare, nel corso dell'esperimento, venne osservato che l'immersione induce una midriasi pupillare non-riducibile con l'illuminazione, che permette di identificare all'interno del globo oculare i vasi retinici, la testa del nervo ottico e la coroide. Méry fu la prima persona a studiare il fondo dell'occhio in-vivo, ma non fu in grado di interpretare in maniera esaustiva il fenomeno ottico che lo aveva permesso. Egli dedusse, in accordo con le teorie di quei tempi, che le sue osservazioni erano state rese possibili da tre fattori: la midriasi, la luce intensa e l'irregolarità che caratterizza la superficie della cornea [20].

Cinque anni più tardi, il fisico Philippe De La Hire (1640-1718), in un discorso all'Académie des Sciences a Parigi, mise in discussione le fallaci conclusioni tratte da Méry e fornì una corretta spiegazione del fenomeno ottico, dimostrando che la visione del fondo dell'occhio era dovuta all'abolizione della rifrazione sulla superficie convessa della cornea e alla sostituzione della curvatura della cornea con la superficie piana del liquido [21].

Per più di un secolo la possibilità di esaminare nel dettaglio le strutture retiniche sembrò non interessare più il mondo accademico e la scienza e la luminosità spontanea rimase (un fenomeno) inspiegabile.

A Montauban, una cittadina dei Pirenei Francesi, nel 1810, Bénédict Prévost (1755-1819), studioso di origine svizzera, notando come gli occhi degli animali posti nella completa oscurità non brillassero, riuscì a dare un'interpretazione alla

luminescenza oculare: “non è la luce che procede dall’occhio ad un oggetto che consente all’occhio di percepire quell’oggetto, ma la luce che arriva nell’occhio da esso”. In altri termini, la luminescenza non era altro che il riflesso di una luce esterna [18-22].

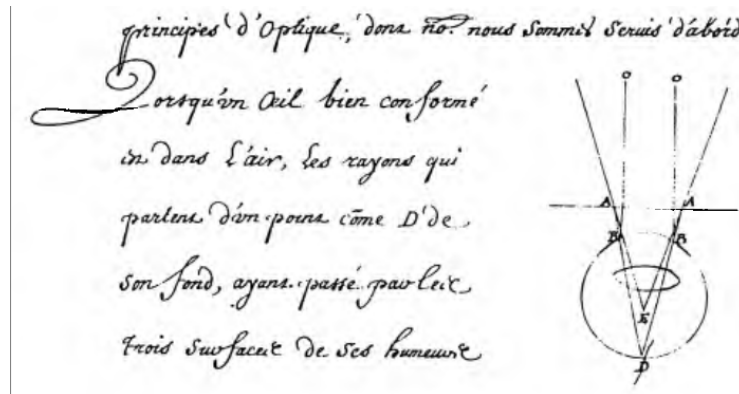


Figura 2.3 Diagramma con cui La Hire spiega la visualizzazione del fondo del gatto sommerso in acqua. Estratto da *Proceedings of the Royal Academy of Sciences for the year 1709 — Session of 20th March 1709*.

Nella stessa epoca, il fisico bavarese Franz von Gruithuisen (1774-1852) attestò che quel riflesso è la manifestazione visiva del *tapetum lucidum*, una membrana iridescente posta subito dietro la retina di molti animali, in grado di aumentare la capacità visiva in condizioni di bassa luminosità [20-23].

A sostegno di questa tesi, nel 1821, il biologo svedese Karl Asmund Rudolphi (1771-1819) dimostrò come, per vedere il riflesso oculare illuminando gli occhi di un gatto decapitato, l’osservatore dovesse guardare nell’occhio lungo la direzione del fascio irradiante [20].

La comprovazione dell’esistenza di strutture riflettenti nel fondo dell’occhio di diversi animali spostò l’attenzione degli studiosi al fondo dell’occhio umano.

Le prime osservazioni furono fatte nel 1823 da Johannes Evangelista Purkinje (1787-1869), un fisiologo dell’attuale Repubblica Ceca. In quell’anno egli rese pubblica la descrizione del suo esperimento di oftalmoscopia: posizionando una candela dietro alla schiena del paziente e utilizzando una lente per occhiali concava, veniva riflesso dall’occhio della persona in esame un bagliore arancione, molto simile al riflesso oculare descritto sugli animali [18-19-23-24].

Il suo lavoro, sebbene cruciale, non fu riconosciuto per molti anni e numerosi altri autori pubblicarono ulteriori studi su questo argomento. Tra questi, quasi contemporaneamente, il filosofo tedesco Ernst Wilhelm Ritter von Brücke (1819-



Figura 2.4 Johannes Evangelista Purkinje.



Figura 2.5 Charles Babbage.

1892), in un quotidiano pubblicato nel 1945, e William Cumming (1812-1886), l'anno successivo in Inghilterra, descrissero chiaramente le condizioni alle quali il riflesso luminoso diventava visibile: l'asse dell'illuminazione e dell'osservazione dovevano coincidere [19-20-24]. Tuttavia, i loro esperimenti per visualizzare nel dettaglio il fondo oculare rimasero infruttuosi.

Il primo, purtroppo sterile, tentativo di mettere a punto uno strumento per un'analisi accurata del fondo dell'occhio umano è del 1847, ad opera di Charles Babbage, motivato probabilmente a scoprire di più sulla diplopia monoculare che lo affliggeva. Il matematico inglese, facendo ricorso ad uno specchio con un foro al centro e ad una sorgente di luce coassiale allo sguardo dell'osservatore, riuscì ad osservare la retina [18-19-24]. Le immagini ottenute però risultarono poco nitide e il suo risultato non fu pubblicato.

Per questo motivo, spesso, viene accreditato come inventore dell'oftalmoscopio Hermann Von Helmholtz (1821-1894), uno scienziato tedesco che nel 1850, mentre preparava una lezione sulla fisiologia oculare e ignaro del lavoro di Babbage, ideò un prototipo analogo alla strumentazione già realizzata dal collega, ma migliorato [18-19].

Helmholtz aveva capito che i raggi luminosi emessi dall'occhio formavano immagini ottiche che se messe a fuoco avrebbero permesso di vedere nel dettaglio le strutture retiniche. Il suo "oftalmoscopio" era costituito da 3 componenti essenziali: una candela come fonte di illuminazione, uno specchio per dirigere la luce verso l'occhio e un dispositivo per focalizzare l'immagine [19].

La prima comunicazione di Helmholtz in merito al suo strumento avvenne il 6 dicembre 1850 alla Physical Society di Berlino [18].



Figura 2.6 Hermann von Helmholtz.

“Questa invenzione era veramente a portata di mano, non esigendo altre conoscenze che quelle acquisite al ginnasio, e io trovo ora quasi da ridere sul fatto che tutti, me compreso, avessimo la mente così offuscata da non capirci nulla. Si tratta di una combinazione di vetri che permette di illuminare attraverso la pupilla il fondo oscuro dell’occhio e di vedere così tutti i dettagli della retina ... ingrandita venti volte come se la si guardasse con una lente di ingrandimento. Si vedono le arterie, le vene e l’entrata nell’occhio del nervo ottico ...” [24].

In queste parole, tratte da una lettera che Helmholtz indirizzò al padre poco prima del Natale del 1850, lo scienziato sottolineava la semplicità e quasi ovvietà della scoperta.

Il primo annuncio pubblico del nuovo strumento, che definì un *“Augenspiegel”*, ovvero uno “specchio per gli occhi”, avvenne l’11 novembre 1851 alla Society for Scientific Medicine di Königsberg, una società fondata pochi giorni prima. La parola “oftalmoscopio” fu introdotta per la prima volta, da Maressal de Marsilly, nel 1852 [18-24].

Nello stesso anno Helmholtz pubblicò una sua monografia *“Beschreibung eines Augen-Spiegels”* in cui descriveva il principio fisico del suo dispositivo: per visualizzare nel dettaglio le strutture retiniche è necessario che la fonte luminosa esterna e l’osservatore si trovino lungo lo stesso asse, senza costituire però un intralcio reciproco. Per ovviare a questo problema interpose tra l’osservatore e il soggetto esaminato tre lamine di vetro semiriflettente, disposte ad angolo per riflettere la quantità massima di luce nell’occhio. Queste lenti fungevano da specchi

CAPITOLO 2. TECNOLOGIE PER L'ANALISI DEL FONDO OCULARE

per riflettere la luce, ma essendo trasparenti consentivano anche di far passare una parte dei raggi incidenti, rendendo visibile la retina [18-25].

Per porre rimedio alle difficoltà nel mettere a fuoco le strutture interne, montò sul lato dell'osservatore un supporto per una lente concava.

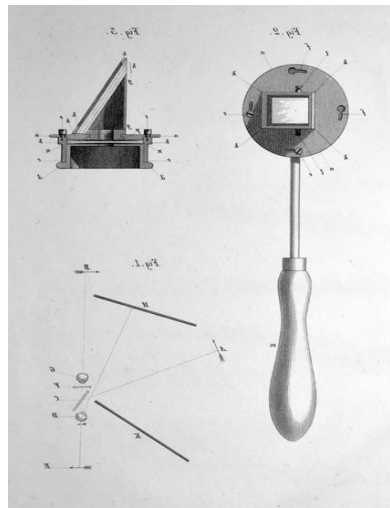


Figura 2.7 Incisione presa da da Beschreibung eines Augen-Spiegels di Helmholtz.



Figura 2.8 Primo modello dell'oftalmoscopio di Helmholtz.

Tra i principali difetti dell'oftalmoscopio di Helmholtz c'erano l'inadeguatezza dell'illuminazione e la scomodità in termini di tempo per cambiare le lenti a seconda degli errori di rifrazione del paziente e dell'osservatore [20].

Tra le numerose modifiche dell'oftalmoscopio, le più significative furono apportate alla fonte di illuminazione, che passò da una lampada nuda a lampade a petrolio, a gas ed elettricità. A determinare un cambiamento radicale nella costruzione dell'oftalmoscopio, aumentando l'efficacia della fonte di illuminazione fu lo sviluppo da parte di Thomas Edison, nel 1879, della lampadina ad incandescenza [19].

Un miglioramento del metodo di sostituzione delle lenti fu ideato dal produttore di strumenti dell'università di Königsberg, Egbert Rekow, che nel 1852 modificò lo strumento con due dischi metallici rotanti sovrapposti, ciascuno contenente una serie di lenti correttive concave e convesse di diverso potere ottico [19-20]. Ad oggi, dopo oltre 150 anni, questa modifica è tuttora presente nella maggior parte degli oftalmoscopi portatili.

Quello introdotto da Helmholtz era un oftalmoscopio diretto; il primo oftalmoscopio monoculare indiretto, caratterizzato da uno specchio concavo perforato al centro, fu sviluppato nel 1852 da Theodor Ruetz [18-24].

L'avvento dell'oftalmoscopio rivoluzionò la pratica dell'oftalmologia: aprì prospettive nuove e inimmaginabili fino ad allora, consentendo enormi progressi nella conoscenza della patologia e della fisiopatologia oculare.

Nel 1853 Earnest Adolph Coccius (1825-1890) fornì la prima descrizione del distacco di retina e della retinite pigmentosa [18]. Richard Liebreich (1830-1917) riconobbe per primo, nel 1855, l'occlusione venosa retinica e nell'anno 1863 pubblicò il primo atlante di oftalmologia [19].

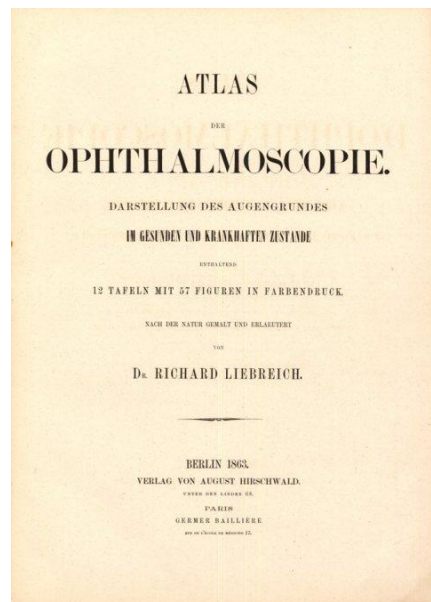


Figura 2.9 Primo atlante di oftalmoscopia di Richard Liebreich.

Albrecht von Graefe (1828-1872), direttore della clinica oculistica dell'Università di Berlino, fu senza dubbio il più grande oculista del XIX secolo; introdusse una tecnica per l'estrazione della cataratta e descrisse per primo le alterazioni del campo visivo nel glaucoma, proponendone l'iridectomia per la cura chirurgica [18].

2.2.2 Oftalmoscopio diretto

L'oftalmoscopio diretto è uno strumento semplice e di impiego comune nella pratica dell'esame fundoscopico.

Nel corso del tempo si è evoluto continuando a ricevere migliorie innovative, ma il principio di funzionamento è lo stesso di quello realizzato da Helmholtz a metà dell'Ottocento: un fascio di luce viene diretto verso l'occhio del paziente e, attraversata la pupilla, va ad illuminare il fondo dell'occhio; qui si riflette sulla retina, tornando poi indietro sulla stessa linea del fascio di andata per essere raccolto dall'occhio dell'osservatore. In altri termini, i raggi luminosi prodotti devono essere coassiali all'asse dell'osservatore.



Figura 2.10 Oftalmoscopio diretto.

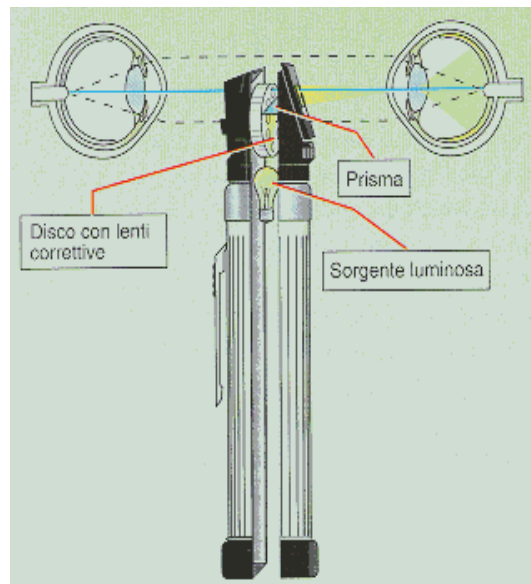


Figura 2.11 Schema di funzionamento di un oftalmoscopio diretto

I moderni oftalmoscopi diretti sono dispositivi palmari di piccole dimensioni, portatili e maneggevoli, del tutto simili ad una torcia elettrica.

Sono apparecchi dotati di una fonte di illuminazione propria, costituita da una piccola lampadina ad intensa luminosità, incorporata nella testina. La radiazione illuminante prodotta viene inviata all'occhio osservato del paziente grazie ad un sistema di prismi o specchi, proiettando un fascio sulla stessa linea di visione dell'osservatore.

L'oftalmoscopio viene tenuto dall'oculista con la mano corrispondente all'occhio che sta osservando, ovvero esamina l'occhio destro del paziente con il proprio occhio destro ed il sinistro con il proprio occhio sinistro.

La testina ottica dello strumento contiene un disco con lenti sferiche positive e negative (Disco di Rekoss), selezionabili ruotando una ghiera, che possono essere interposte sull'asse di osservazione per correggere i difetti refrattivi di osservatore e/o osservato modificando la messa a fuoco.

Sull'indicatore delle lenti appare un numero, che corrisponde alle diottrie della lente inserita: in nero sono scritte le diottrie positive delle lenti convergenti, generalmente utilizzate per correggere l'ipermetropia; in rosso sono scritte le diottrie negative delle lenti divergenti, utilizzate per correggere la miopia. In caso di emmetropia la retina del paziente può essere visualizzata in modo nitido, senza l'inserimento di alcun tipo di lente.

Questo sistema consente di ottenere un'immagine dritta, con i lati non invertiti del fondo oculare, ingrandita di circa 15 volte (qualora sia osservatore che osservato siano emmetropi) e una buona risoluzione nei dettagli.

Gli oftalmoscopi diretti sono caratterizzati da un campo di osservazione tipicamente monoculare e piuttosto ristretto, in quanto le aree retiniche periferiche risultano inaccessibili. Pertanto, al fine di consentire una più accurata valutazione da parte dell'oculista, spesso viene indotta una midriasi farmacologica. Quando l'esame viene effettuato a pupilla stretta può risultare particolarmente utile osservare l'occhio del paziente in una stanza scarsamente illuminata.

Dunque i principali limiti di questo dispositivo sono due:

- non consente uno studio della periferia retinica
- opacità vitreali o del cristallino nell'occhio osservato possono riflettere una certa quota del raggio luminoso incidente, inficiando l'esame del fondo oculare.

2.2.3 Oftalmoscopio indiretto

L'oftalmoscopio indiretto attuale consiste in un caschetto, dotato di un potente sistema di illuminazione coassiale e di una lente biconvessa *ad alto potere positivo*

CAPITOLO 2. TECNOLOGIE PER L'ANALISI DEL FONDO OCULARE

(rende l'occhio osservato fortemente miope), che l'oftalmologo indossa e tiene pochi centimetri davanti all'occhio del soggetto esaminato.

Per effettuare l'esame del fondo oculare tramite oftalmoscopia indiretta è necessario ottenere una sufficiente dilatazione della pupilla mediante l'instillazione di colliri miotici.

Il dispositivo produce un fascio luminoso diretto nell'occhio osservato; dalla retina di quest'ultimo la radiazione riflessa si dirigerà, attraversata la lente, verso l'occhio osservatore.



Figura 2.12 Oftalmoscopia indiretta.

L'ingrandimento fornito è modesto, 2-4 X, e l'immagine è capovolta e con i lati invertiti, ma l'angolo illuminato, rispetto all'oftalmoscopia diretta, è molto più ampio, consentendo un'ispezione dell'intero fundus fino alla periferia retinica.

Questa tecnica, inoltre, garantisce una maggiore profondità di campo in quanto l'immagine, osservata con entrambi gli occhi, è stereoscopica permettendo di apprezzare la profondità.

Infine questo metodo indiretto permette una valutazione meno condizionata da eventuali vizi refrattivi e opacità del sistema diottrico.

Nel sospetto di una lesione retinica come il distacco di retina, per esaminare il fondo oculare si ricorre preferenzialmente all'oftalmoscopia indiretta in quanto permette di ispezionare anche la periferia retinica, porzione del fondo dove più frequentemente si riscontrano questo tipo di lesioni. Inoltre, grazie alla visualizzazione tridimensionale peculiare di questo oftalmoscopio, se ne può rilevare con maggiore precisione dimensioni e profondità.

2.3 Lampada a fessura o biomicroscopio

L'esame del fundus oculi può essere eseguito anche tramite la lampada a fessura, che sebbene generalmente utilizzata per lo studio delle camere anteriori, può essere utilizzata per lo studio del corpo vitreo e del fondo oculare utilizzando lenti addizionali.

2.4 Fundus camera

L'oftalmoscopio diretto, tuttora di ampio utilizzo, ha una bassa sensibilità ed i risultati dell'indagine dipendono fortemente dall'esperienza del clinico. L'introduzione della fundus camera, attorno alla metà del ventesimo secolo, ha portato ad un grande miglioramento nella pratica oftalmica, permettendo di ottenere immagini della retina di qualità sempre più elevata e di documentare una varietà di difetti oculari con relativa facilità.



Figura 2.13 Fundus camera.

La fundus camera presenta gli stessi principi di funzionamento dell'oftalmoscopio ed è del tutto simile per il tipo di informazione ricavabile; tuttavia grazie all'utilizzo di tecnologie digitali consente di ottenere una maggiore risoluzione per l'immagine in output, che diviene non solo visualizzabile in tempo reale, ma può essere anche facilmente catturata ed archiviata.

CAPITOLO 2. TECNOLOGIE PER L'ANALISI DEL FONDO OCULARE

Questa strumentazione permette di effettuare, principalmente dopo instillazione di colliri midriatici, una vera e propria fotografia del fondo dell'occhio in alta definizione e dunque con un livello di massimo dettaglio, essenziale per un esame rapido ed efficace per l'individuazione di eventuali patologie.

La fundus camera sta emergendo oltremodo come strumento fondamentale nel monitoraggio del decorso naturale di certe affezioni oculari, quali la retinopatia diabetica, e follow-up degli effetti nel tempo di eventuali trattamenti e terapie. Il medico oculista, infatti, con un esame del tutto non invasivo, ha la possibilità di valutare, ingrandire e stampare l'immagine, rendendo possibile un confronto dei referti recenti con quelli passati.

Le fotografie retiniche ottenute possono inoltre essere utilizzate come valido strumento didattico per formare oftalmologi sempre più preparati, oltre che illustrare ai pazienti le caratteristiche della patologia retinica di cui sono affetti.

Il primo prototipo di fundus camera è stato sviluppato da Frederick Dimmer all'inizio del Ventesimo secolo. Dimmer inoltre introdusse la fotografia del fundus oculi in un libro di testo di oftalmologia e pubblicò il primo atlante fotografico in bianco e nero. Il dispositivo, che occupava un intero piano di lavoro, a causa delle sue dimensioni notevoli non venne mai prodotto in serie. La prima effettiva fundus camera fu progettata nel 1925 da J.W Nordenson e un anno dopo venne realizzata e messa in commercio dalla ditta Carl Zeiss [27].



Figura 2.14 Brochure della Reflex free Retinal Camera after Nordenson Zeiss, Price \$ 768 f.o.b. N.Y.



Figura 2.15 Fundus Camera Zeiss progettata da J.W. Nordenson.

Da allora le funzionalità della fundus camera sono notevolmente migliorate con device non-mydiatric, illuminazione elettronica, allineamento automatico degli occhi e acquisizione digitale ad alta risoluzione. Le più attuali e innovative fundus camera non miatrici digitali integrano tutte le funzioni necessarie per l'esame retinico, permettendo di catturare nel dettaglio anche le patologie retiniche più periferiche e sono caratterizzati da un efficiente software per la gestione clinica e l'archiviazione dati.

2.5 D-EYE

D-EYE è una startup italiana del segmento digital health, fondata nel 2014, che ha immesso sul mercato un oftalmoscopio basato su smartphone, utile soprattutto per lo screening precoce di alcune patologie retiniche. Il dispositivo consiste di un piccolo sistema ottico che si applica magneticamente allo smartphone in corrispondenza della fotocamera, che consente a quest'ultima di visualizzare la retina. Al dispositivo fisico è associata una piattaforma software, sia app che cloud, che permette di registrare e gestire i dati, e fornire servizi di telemedicina. Attualmente è disponibile solo per smartphone Apple (Apple Inc., Cupertino, USA).

Si tratta di uno strumento totalmente portatile ad un costo relativamente basso e ad orientamento internazionale.



Figura 2.16 Dispositivo D-EYE.

La macchina fotografica retinica D-EYE è in grado di registrare foto e video ad alta definizione del fondo oculare e la relativa app permette di gestire i dati acquisiti e

CAPITOLO 2. TECNOLOGIE PER L'ANALISI DEL FONDO OCULARE

memorizzarli nella cartella clinica del paziente. Disponendo di una connessione internet, consente inoltre di condividerli con medici specialisti di tutto il mondo per una valutazione e un teleconsulto in tempo reale.

La vera rivoluzione introdotta da D-EYE nell'imaging retinico consiste nel migliorare l'accessibilità alla diagnosi delle patologie dell'occhio, sfruttando le grandi potenzialità di uno strumento di uso comune come il cellulare. Consente una diagnosi avanzata, scambiando informazioni, servizi e utilità in modo intuitivo, efficace e ad un costo contenuto.

L'esame del fondo oculare eseguito con oftalmoscopio digitale diretto D-EYE, inoltre, non necessitando di dilatazione farmacologica della pupilla, può essere eseguito anche da un operatore non medico. Le caratteristiche di versatilità e facilità di impiego consentono a questo dispositivo di poter essere utilizzato anche in aree non sviluppate, dove la popolazione non ha accesso a strutture mediche specialistiche, e quindi ad esami diagnostici effettuati con strumenti più complessi e costosi.



Figura 2.17 Foto scattata durante il programma di telemedicina rurale nel sud-est asiatico.

L'età della popolazione sta aumentando e sono sempre più frequenti casi di patologie che dovrebbero essere diagnosticate precocemente per essere curate tempestivamente e in maniera opportuna. Molte persone nel mondo perdono la vista per malattie che sono curabili, ma che non vengono diagnosticate in tempo utile. Dispositivi come quello di D-EYE rivestono un ruolo emergente ma sempre più cruciale nella fase di screening fornendo un metodo di lavoro che consente di esaminare un elevato numero di pazienti, ma di portare negli ospedali solo quelli che necessitano di analisi più approfondite, con strumenti diagnostici che per

l'elevato costo sono spesso reperibili solo presso le grandi strutture ospedaliere e che richiedono maggior tempo per l'assistenza medica.

Il basso costo, la portabilità e la facilità di utilizzo rendono D-EYE uno strumento utile per lo screening su larga scala, diffuso su tutto il territorio, e quindi per la diagnosi precoce delle malattie retiniche, superando i limiti imposti da apparecchiature costose e spesso ingombranti come le moderne fundus camera.

Non da meno, il supporto clinico fornito da questo apparecchio si estende anche nella valutazione dell'evoluzione della patologia nota pre-terapia e nel follow-up post-terapia.

È uno strumento portatile e pertanto ideale per pazienti costretti a letto o con mobilità limitata.

Capitolo 3

Deep Learning

3.1 Introduzione

L' *Intelligenza Artificiale* (spesso si scrive *AI*, acronimo dell'inglese *Artificial Intelligence*) è un ramo della *Computer Science* che studia metodologie e tecniche per lo sviluppo di macchine (hardware e software) in grado di risolvere problemi o riprodurre compiti e attività tipici dell'intelligenza umana, in modo del tutto autonomo.

È un campo fiorente che sta facendo grandi passi in avanti con molte applicazioni pratiche e argomenti di ricerca attiva. In questi ultimi decenni ha avuto un forte impatto nella società moderna, entrando sempre più a fare parte della vita di tutti i giorni e cambiando completamente le nostre abitudini.

In modo inconscio l'essere umano analizza costantemente quello che lo circonda, facendo predizioni su tutto ciò che vede. Per questo motivo quando vediamo qualcosa siamo in grado di dargli un significato e di etichettarlo in base a ciò che abbiamo appreso in passato. Nonostante esistano dei compiti, tipicamente di tipo percettivo, di cui non conosciamo il dettaglio su come vengano eseguiti dal nostro cervello, oggi possediamo alcune conoscenze relative alla corrispondenza tra le aree cerebrali e le parti del corpo da esse controllate, o da cui ricevono input sensoriale.

Nell'ambito delle *Neuroscienze Computazionali*, sono attive a questo proposito linee di ricerca volte allo studio di modelli realistici del singolo neurone e della dinamica di reti neurali, dove l'interazione è mediata dai potenziali d'azione emessi dai neuroni stessi; il punto di forza dell'*Intelligenza Artificiale* è stata proprio la scoperta di modelli matematici sempre più sofisticati in grado di emulare queste specifiche funzionalità.

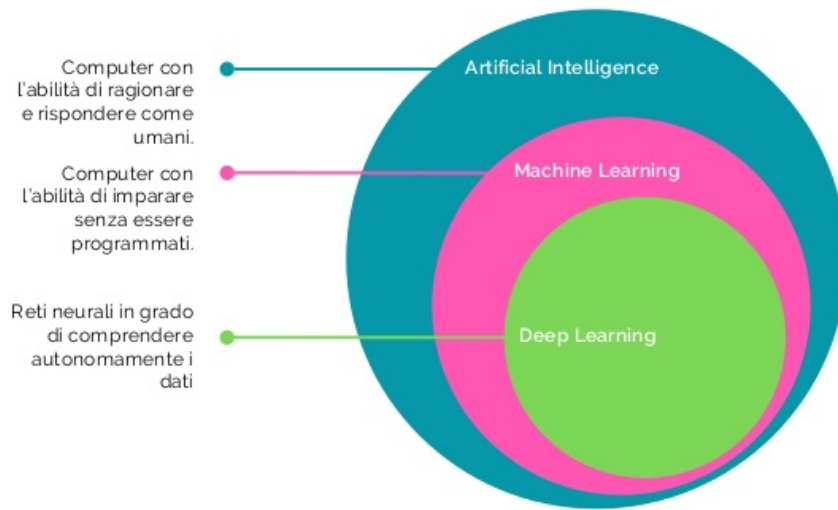


Figura 3.1 L'intelligenza artificiale comprende il Machine Learning e il Machine Learning include a sua volta il Deep Learning.

L' *Intelligenza Artificiale* contemporanea si basa su *Machine Learning* o *Deep Learning*.

Il *Machine Learning (ML)* è un sottoinsieme delle tecniche di *Intelligenza Artificiale* e sviluppa algoritmi di apprendimento automatico, principalmente metodi statistici avanzati, capaci di migliorare il proprio output con l'esperienza. Raccogliendo conoscenze dall'esperienza, questo approccio evita la necessità per gli operatori umani di specificare esplicitamente tutte le conoscenze di cui il computer ha bisogno.

Il *Deep Learning (DL)* è un sottoinsieme del *ML*, che si basa su un particolare tipo di apprendimento automatico, caratterizzato da un modello a più livelli. Ogni livello più profondo prende in input i dati di output del livello precedente, estraendo via via, all'aumentare della profondità, sempre più informazioni.

Questo approccio, che si ispira al modo in cui il nostro cervello elabora le informazioni, rispondendo agli stimoli esterni, permette ai modelli computazionali di imparare a svolgere un'attività che per l'uomo è naturale, ovvero quella di imparare tramite esempi [28].

3.2 Reti Neurali Artificiali (ANN: Artificial Neural Networks)

Le reti neurali artificiali sono dei modelli matematici che mirano a simulare la struttura ed il funzionamento dei sistemi nervosi biologici, ed in particolare i processi computazionali che avvengono all'interno del cervello umano durante l'apprendimento ed il riconoscimento. Questi modelli matematici sono di fondamentale importanza per insegnare ad una macchina come apprendere dai dati attraverso l'esperienza, mediante un processo di autoadattamento.

Lo sviluppo di reti neurali artificiali si basa sull'ipotesi che la capacità dei sistemi biologici di affrontare compiti anche molto complessi risieda nell'architettura fisica del sistema nervoso.

3.2.1. Modello di Neurone artificiale

Come possiamo vedere nella *Figura 3.2* un neurone biologico è composto principalmente da 3 componenti:

- il *corpo cellulare (soma)*, che è la componente da cui si dipartono gli altri costituenti del neurone e rappresenta l'unità di calcolo;
- l'*assone*, prolungamento citoplasmatico a ramificazione multipla, che costituisce la linea di trasmissione in uscita, conducendo il segnale in direzione centrifuga verso i successivi neuroni;
- i *dendriti*, anch'essi prolungamenti a ramificazione multipla *del soma*, costituiscono le principali strutture recettrici del neurone.

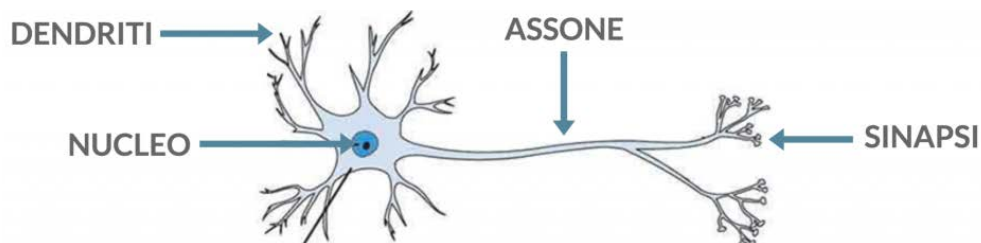


Figura 3.2 Modello schematico del neurone biologico.

I dendriti di ogni neurone ricevono segnali da migliaia di punti di contatto, detti sinapsi, che provengono da altri neuroni.

La comunicazione tra le sinapsi di due neuroni avviene attraverso lo scambio di sostanze chimiche, chiamate *neurotrasmettitori*, che possono essere eccitatorie o inibitorie, cioè possono rispettivamente promuovere la creazione di un impulso nervoso nel neurone ricevente o inibire l'impulso.

Il corpo cellulare esegue una somma pesata dei segnali elettrici in ingresso e se il risultato supera un certo valore di soglia, allora il neurone si attiva e produce un potenziale di azione che si propaga lungo l'assone fino a raggiungere le sinapsi.

Partendo da questa conoscenza di base del funzionamento del sistema nervoso, sono stati sviluppati dei modelli di neuroni artificiali, che costituiscono l'unità di elaborazione fondamentale delle AAN.

Un neurone artificiale, come si può evincere dallo schema a bocchi in *Figura 3.3*, è costituito generalmente da molti canali di ingresso x_i ed un'unica uscita y .

Le componenti essenziali sono:

- un set di *sinapsi* o *link di collegamento* a ciascuno dei quali è associato un peso w_i , costituito da un numero reale, che riproduce le sinapsi neurali. Se il peso è positivo il canale si dice eccitatorio, se è negativo inibitorio;
- un *modulo sommatore*, che effettua la somma pesata degli ingressi
- una *funzione ingresso-uscita* o *funzione di attivazione* f , che determina il segnale in uscita da trasmettere alle unità successive;
- un termine costante b , detto *bias*, che ha l'effetto di aumentare o diminuire l'input della funzione di attivazione, a seconda che esso sia rispettivamente positivo o negativo.

Riassumendo, in termini matematici, l'uscita del neurone si ottiene come [29]:

$$a = \sum_{i=1}^n w_i x_i \quad (3.1)$$

$$y = f(a + b) \quad (3.2)$$

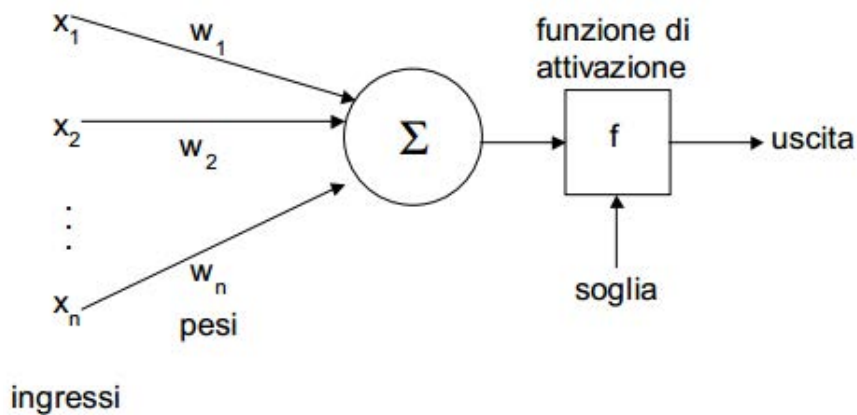
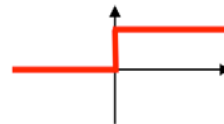


Figura 3.3 Modello schematico del neurone artificiale: gli ingressi (x) sono pesati (w) e sommati. Se il risultato supera il valore di soglia (f) della funzione di attivazione, il neurone si accende, altrimenti rimane spento.

Ci sono varie funzioni di attivazione, tutte definite nel campo Reale e con un dominio ben definito. Quelle più comunemente utilizzate sono [29]:

- Funzione a gradino

$$f(a) = \begin{cases} 1 & \text{se } a \geq 0 \\ 0 & \text{se } a < 0 \end{cases} \quad (3.3)$$



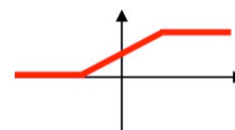
- Funzione lineare

$$f(a) = a \quad (3.4)$$



- Funzione lineare a tratti

$$f(a) = \begin{cases} 0 & \text{se } a \leq -0.5 \\ a + 0.5 & \text{se } -0.5 < a < 0.5 \\ 1 & \text{se } a \geq 0.5 \end{cases} \quad (3.5)$$



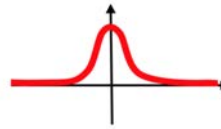
- Funzione sigmoidea

$$f(a) = \frac{1}{1 + \exp(-a)} \quad (3.6)$$



- Funzione Gaussiana

$$f(a) = \begin{cases} 0 & \text{se } a \leq -1 \\ 1 & \text{se } a = 0 \\ 0 & \text{se } a \geq 1 \end{cases} \quad (3.7)$$



La scelta del calcolo della funzione di attivazione prende spunto dall'analogia con i neuroni biologici e un insieme di questi neuroni artificiali può essere organizzata in una rete neurale artificiale simile a quella che gestisce il sistema nervoso umano.

3.2.2 Architettura di una rete neurale artificiale

L'origine concettuale delle reti neurali si trova nella modellizzazione astratta di questo sistema di comunicazione tra neuroni. Si tratta di un sistema dinamico che ha la topologia di un grafo, in cui i nodi rappresentano i neuroni artificiali e gli archi i pesi sinaptici. I nodi sono in genere disposti lungo linee parallele, detti *strati*.

Il più semplice modello di rete neurale è formato da tre strati di neuroni: il primo è lo strato di unità di ingresso (*input layer*), che ricevono l'informazione grezza dall'ambiente; ad esso è collegato uno strato di unità nascoste (*hidden layer*), che possono comunicare solamente con le unità all'interno della rete; l'ultimo strato è quello delle unità di uscita (*output layer*), che fornisce il risultato finale della rete. Questo modello di rete neurale, essendo costituita da un solo strato di unità nascoste viene indicata come *single layer neural network*.

3.2 RETI NEURALI ARTIFICIALI (ANN)

Quando uno stimolo viene applicato ai nodi di ingresso della rete, i segnali viaggiano in parallelo lungo le connessioni attraverso i nodi interni fino ai nodi di uscita, la cui attivazione fornisce in output la risposta finale della rete neurale.

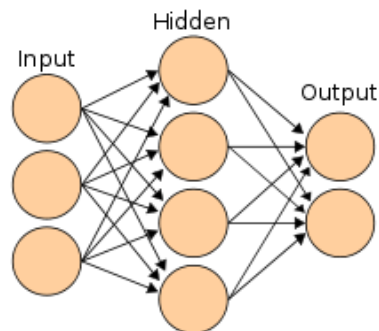


Figura 3.4 Struttura di una rete neurale a singolo strato.

Il rapporto tra i dati in ingresso e quelli in uscita è determinato da:

- **tipologia di unità elementari**, definita dalla complessità della struttura interna e dalla funzione di attivazione usata per ogni neurone;
- **struttura formale della rete**, ovvero numero, orientamento e direzione dei nodi;
- **valori dei parametri interni** associati ai neuroni e alle relative interconnessioni, che vengono calcolati mediante opportuni algoritmi di apprendimento.

Quando ad una rete neurale a singolo strato vengono aggiunti ulteriori strati di unità nascoste si parla di rete neurale multistrato (*multi-layer neural network*). In una ordinaria rete neurale stratificata ogni neurone di un layer è connesso a tutti i neuroni del layer precedente (si parla di architettura *fully connected*).

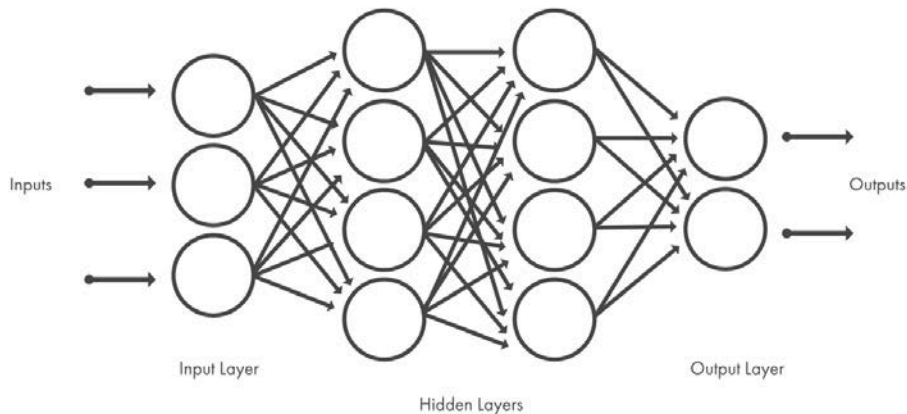


Figura 3.5 Struttura di una rete neurale multistrato.

Proprio da considerazioni di natura neuroscientifica è nata l'idea di utilizzare reti neurali anche molto profonde, con molteplici livelli di unità nascoste, che agiscono in parallelo.

Il nostro cervello è caratterizzato da un'organizzazione gerarchica, in cui ogni livello di questa gerarchia corrisponde ad una zona della corteccia cerebrale; tutte le aree cerebrali contribuiscono insieme alla realizzazione di un compito consentendo di imparare prima concetti semplici e poi di combinarli per interpretare concetti sempre più astratti. Sulla base di questa intuizione sono state introdotte le *Deep Neural Network (DNN)*, caratterizzate da centinaia o migliaia di unità, ad altissima distribuzione e in grado di lavorare parallelamente per imparare a generalizzare.

Il termine *Deep, profondo*, dà il nome all'intero ambito, *Deep Learning*, e si riferisce generalmente al numero di layer nascosti tra i livelli di input ed output, che possono essere fino a 150 [28]. Le *DNN* maggiormente utilizzate consistono di un numero di livelli compresi tra 7 e 50, in quanto reti più profonde hanno dimostrato di poter garantire prestazioni leggermente migliori, a scapito però dell'efficienza.

Una volta stabilite tutte le caratteristiche della rete neurale, topologia, numero e tipo di neuroni, collegamenti, ecc. occorre determinare i pesi delle connessioni in modo da rendere la rete funzionale. Questa operazione prende il nome di *addestramento* della rete neurale. Il campo di apprendimento di una rete neurale è molto ampio e in seguito ci limiteremo a considerare solamente il cosiddetto apprendimento supervisionato, di interesse per lo sviluppo di questa tesi.

3.3 Apprendimento supervisionato (supervised learning)

Si riferisce generalmente alle tecniche che mirano ad addestrare una rete neurale a risolvere determinati problemi sulla base di informazioni che gli vengono presentate come rappresentative di quello che deve apprendere.

Alla rete viene passato un insieme di esempi per l'addestramento (*training set*), composti da una coppia di dati ingresso-uscita: per ogni *input* viene fornita al sistema l'uscita desiderata, *target*. L'obiettivo finale del modello è quello di stimare la funzione incognita che lega le variabili di input e di output in modo tale che, al presentarsi di un esempio sconosciuto in precedenza, possa fornire il risultato corretto. Il successo dell'apprendimento consiste proprio nella capacità della rete, dopo essere stata addestrata su un numero limitato di esempi, di essere in grado di produrre una risposta appropriata a dei dati in ingresso che non ha mai visto in precedenza (*test set*), ma che presentano comunque una certa somiglianza con gli esempi presentati durante il training.

Questo problema viene tipicamente formulato come un problema di minimizzazione di una funzione costo, chiamata anche funzione di perdita (*loss function*), che misura la discrepanza tra i valori predetti dal modello (\hat{y}_i) ed i target (y_i) [31]. Da notare che i valori predetti dal modello sono ottenuti in base al valore dei pesi w_i , contenuti in ciascun neurone, quindi la funzione costo è derivabile rispetto ai pesi.

Sono possibili molteplici scelte per la funzione costo, ma la maggior parte dei sistemi utilizza la funzione di scarto quadratico [32]:

$$L(y, \hat{y}) = \frac{1}{2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.8)$$

dove n è la dimensione sia del vettore target sia del vettore dei valori predetti. La loss function può essere minimizzata attraverso una procedura di stima dell'impatto che hanno su di essa piccole variazioni dei pesi. Tipicamente a tale scopo viene utilizzata la tecnica di discesa del gradiente (*Gradient Descent*) mediante *back propagation* [31-32].

L'idea alla base di questa procedura è quella di calcolare il gradiente della funzione costo allo stato generico dei pesi attuale, dopodiché si calcolano i pesi che portano la funzione costo nella direzione opposta al gradiente, scendendo verso un minimo locale.

Come si può vedere in *Figura 3.6*, una generica funzione costo presenta un punto di minimo assoluto, in corrispondenza del punto in cui si crea l'inarcamento a valle. L'obiettivo dell'algoritmo di minimizzazione è "discendere" la funzione costo, partendo da un punto iniziale determinato in maniera casuale, per giungere fino al suo valore di minimo assoluto.

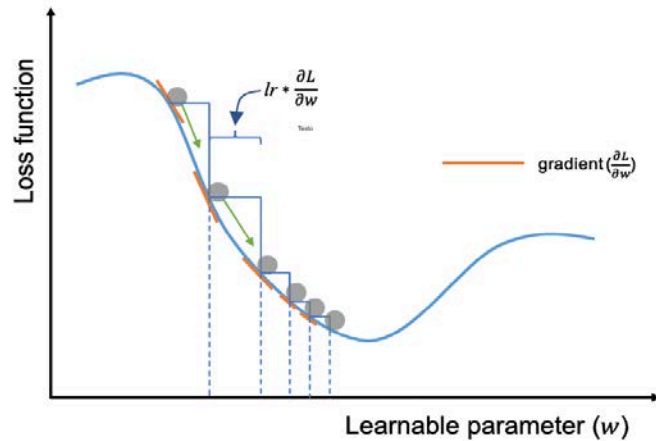


Figura 3.6 Grafico funzione di perdita con discesa del gradiente.

La direzione verso la quale la funzione costo scende più rapidamente in un punto è data dall'opposto del suo gradiente in quel punto. Una volta calcolata la direzione per arrivare al minimo, si deve scegliere quanto velocemente scendere. Il parametro che controlla la velocità di discesa lungo il gradiente è chiamato *Learning Rate (lr)* ed è un valore compreso tra 0 e 1.

Questo parametro viene impostato dal programmatore e deve essere scelto in modo molto accurato. Infatti, incrementarlo troppo porta a non raggiungere mai il minimo: l'algoritmo non riesce a scendere a valle, rimbalzando lungo le discese. Viceversa, scegliendo un valore troppo piccolo i tempi per arrivare alla soluzione richiesta potrebbero diventare molto lunghi. Si tratta di un approccio iterativo. Ripetendo più volte questa operazione, ci si muove in discesa, verso il minimo della funzione costo.

Questa procedura può essere riassunta in due fasi:

- **propagazione:** in una prima fase un vettore di input viene dato in pasto alla rete, viene propagato in avanti (*forward step*) attraversando l'intera rete, strato su strato, fino allo strato di uscita. A questo punto si recuperano gli output predetti e tramite la loss function viene calcolato l'*errore di predizione* rispetto ai target. Questo errore viene quindi utilizzato per il

3.3 APPRENDIMENTO SUPERVISIONATO (SUPERVISED LEARNING)

calcolo del gradiente della loss function rispetto ai pesi, che viene propagato all'indietro nella rete fino a quando ogni neurone ottiene il suo valore di gradiente (da cui *back propagation*). Il gradiente indica di quanto l'errore cresce o decresce se i pesi vengono incrementati di una quantità infinitesima.

- **aggiornamento:** in questa seconda fase i valori del gradiente calcolati sono usati per calcolare i nuovi pesi della rete:

$$w_i = w_i - lrd \frac{L}{dw_i} \quad (3.9)$$

Il procedimento è reiterato finché l'errore tra target e output predetto diviene il più piccolo possibile e viene ottenuta la combinazione ottimale dei pesi contenuti in ciascun neurone. A quel punto la rete è addestrata per risolvere correttamente il problema richiesto.

È evidente che lo sforzo necessario per addestrare la rete cresce significativamente con l'aumento del numero di parametri in gioco e soprattutto con la complessità della struttura della rete (altamente non lineare).

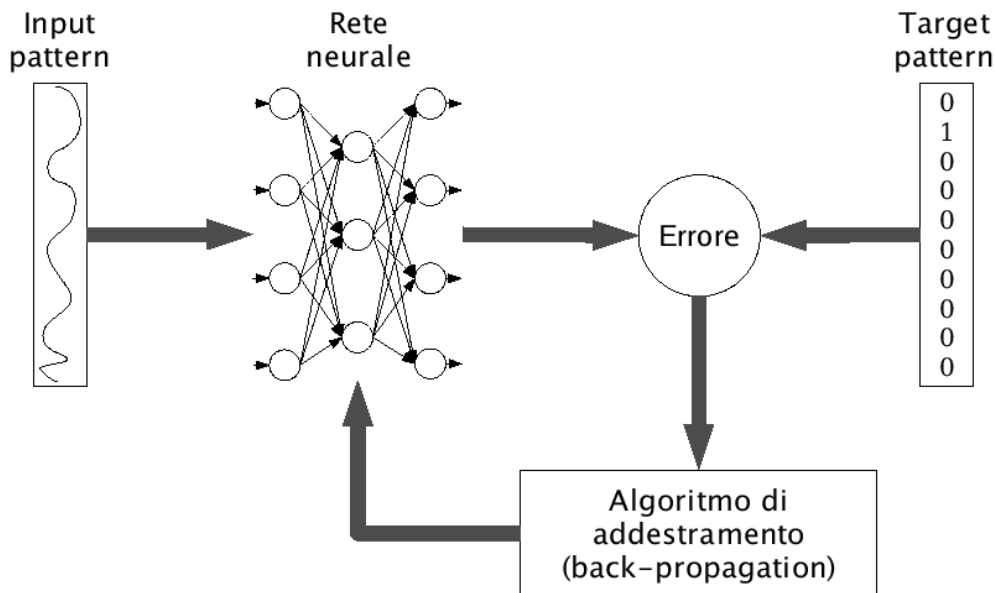


Figura 3.7 Procedura di addestramento con algoritmo di back propagation.

CAPITOLO 3. DEEP LEARNING

La scelta del modello di rete neurale deve essere valutata sulla base dei dati a disposizione e del tipo di applicazione per la quale la rete viene progettata. Modelli troppo semplici potrebbero portare ad un addestramento troppo difficile da portare a termine; situazione detta di *underfitting*, in cui la rete fatica ad apprendere informazioni, perché il numero di parametri del modello non è sufficiente per memorizzare le caratteristiche del training set.

Al contrario, modelli troppo complessi potrebbero portare la rete a dipendere troppo dai dati di esempio, perdendo la capacità di generalizzare. Questa situazione è nota come *overfitting*: a causa di un eccessivo numero di parametri la rete non impara, ma “memorizza”, specializzandosi così nella gestione delle sole informazioni contenute nel training set.

Una delle tecniche più semplici e diffuse per monitorare le performance della rete ed evitare *overfitting* è quella di conservare parte dei dati di apprendimento per la validazione, *validation set*. Solitamente l’80% del training set viene utilizzato per il training e il restante 20% per la validazione [31].

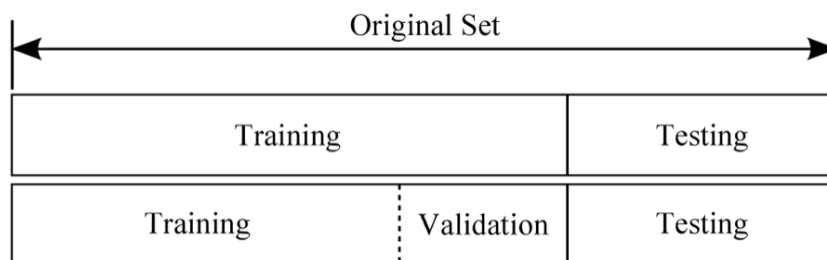


Figura 3.8 Suddivisione del dataset originale in training set, validation set e testing set.

Quando la rete addestrata raggiunge bassi livelli di performance sui dati di convalida è necessario modificare il modello predittore. Si ripete il procedimento fino al raggiungimento di performance soddisfacenti. Poiché i dati del validation set sono indipendenti dai dati del training set, le prestazioni rilevate sono una buona misura della capacità di generalizzazione della rete stessa. Il modello risultante viene infine testato sul test set per determinarne l’accuratezza.

3.4 Classificazione

Uno dei campi in cui trovano maggiore applicazione le reti neurali è quello della classificazione, un tipo di apprendimento automatico supervisionato in cui un algoritmo “impara” a classificare nuove osservazioni da esempi etichettati [28].

Anche nel lavoro esposto in questa tesi una rete sarà addestrata ai fini di classificazione, in particolare di immagini retiniche.

Classificare significa specificare a quale tra k categorie appartiene un dato di ingresso. Si raggruppano i dati di input sulla base di caratteristiche comuni in k classi disgiunte, a ciascuna delle quali viene assegnata un’*etichetta* o *label*. Il training data sarà quindi costituito da coppie $\{input, label\}$.

Il sistema in fase di addestramento elabora gli esempi in input, descritti da un vettore x e tenendo conto delle loro caratteristiche crea una funzione $f: \mathbb{R}^n \rightarrow \{1, \dots, k\}$ che associa a ciascun input la sua specifica classe di appartenenza k .

Dopo la fase di addestramento il classificatore, a partire dal riconoscimento di determinate feature, sarà in grado di operare anche su oggetti mai visti prima, attribuendogli una classe di appartenenza o label definito a priori. L’accuratezza di un classificatore viene misurata valutando la percentuale di dati per i quali il modello elabora un output corretto.

Questo lavoro di tesi si basa, nello specifico, su un problema di classificazione binaria, in cui i dati di input vengono classificati come appartenenti ad una di due categorie predefinite.

Ciò che gioca il ruolo da protagonista in un processo di classificazione è l’estrazione delle feature. Nelle tradizionali tecniche di *Machine Learning* queste devono essere valutate manualmente e preliminarmente all’addestramento. Questo approccio ha però due grandi limiti: richiede una significativa quantità di tempo e l’accuratezza del classificatore è fortemente dipendente dall’esperienza del programmatore, raggiungendo spesso livelli di performance non soddisfacenti. A tal proposito, uno degli impieghi principali del *Deep Learning* si basa sulla creazione di algoritmi di apprendimento specializzati nell’extrapolazione automatica di feature.

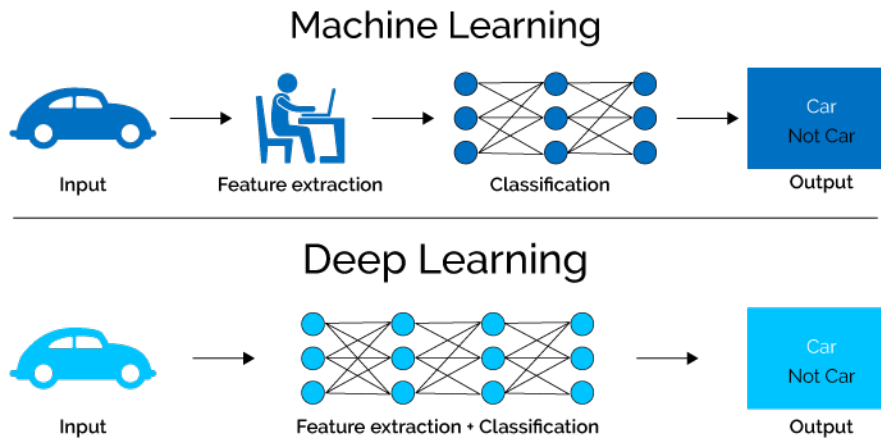


Figura 3.9 Differenza tra Machine Learning e Deep Learning nel processo di estrazione di feature in un problema di classificazione. Gli algoritmi di Deep Learning sono in grado di identificare informazioni di interesse autonomamente, senza ricevere in ingresso valori predefiniti dall'operatore.

Così come si insegna ad un bambino a riconoscere degli animali a partire da ben determinate *feature* (forma, colore, dimensione, ecc...), allo stesso modo possiamo addestrare una rete ad interpretare ed estrarre informazioni utili da un oggetto per classificarlo correttamente. Il modello impara a svolgere attività di classificazione direttamente dai dati, che possono essere immagini, video, suoni o un testo scritto e l'attività di estrazione delle caratteristiche diventa un processo interno alla rete svolta dagli strati nascosti.

I modelli di *DL* possono raggiungere una precisione allo stato dell'arte, superando talvolta in alcune attività le prestazioni ottenute dall'essere umano.

Si può inoltre notare dal grafico in *Figura 3.10* come le performance di algoritmi di Deep Learning aumentino all'aumentare dei dati di training. Viceversa, metodi di Machine Learning, una volta raggiunto un livello di performance, anche aggiungendo dati di training, non consentono ulteriori sviluppi [28].

Il raggiungimento di alti livelli di accuratezza richiede quindi una grande quantità di dati etichettati e questo per contro comporta lunghi tempi per l'addestramento.

Questo limite può essere superato solamente tramite l'utilizzo di adeguati sistemi hardware, in grado di elaborare rapidamente i dati. Il vero passo in avanti in questo senso è stato fatto con lo sviluppo negli ultimi anni di GPU, schede grafiche ad architettura parallela dotate di notevole potenza elaborativa, che hanno consentito

di accelerare i tempi di addestramento in modo sostanziale, da diverse settimane a poche ore.

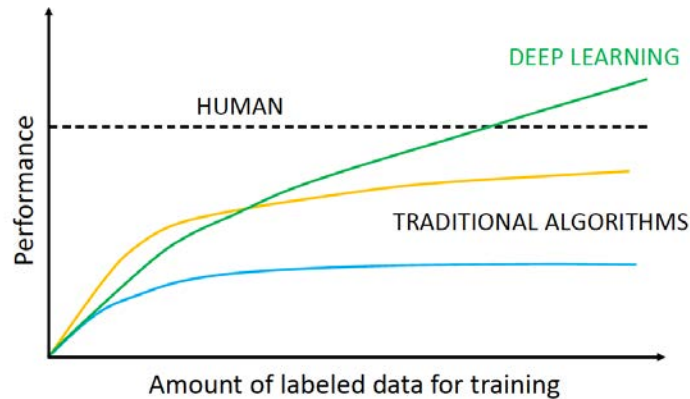


Figura 3.10 Grafico che mostra le differenze di performance tra algoritmi di Deep Learning e tradizionali algoritmi di Machine Learning.

Ciò ha permesso di soddisfare le aspettative di molti utenti, portando ad una rapida diffusione del *DL* nell'ambito di molte discipline.

Tra le applicazioni di maggior successo troviamo il riconoscimento di oggetti e la visione artificiale, con task che includono *classificazione* e *segmentazione* di immagini, quest'ultima verrà descritta nei prossimi capitoli.

L'architettura di rete neurale più comunemente utilizzata per risolvere questo tipo di problemi sono le *Convolutional Neural Network*, che verranno presentate nel prossimo paragrafo.

3.5 Reti Neurali Convoluzionali (CNN: Convolutional Neural Networks)

Le *Convolutional Neural Network* (*ConvNet* o *CNN*) sono una tipologia di reti neurali, esplicitamente progettata per *l'image processing* [34-35].

Tramite una *CNN* un algoritmo di *DL* è in grado di classificare cosa mostra un'immagine e di identificare con buona probabilità il suo contenuto. Sono di fatto delle reti neurali artificiali, costituite da neuroni connessi fra loro tramite rami pesati; tutto quanto detto in precedenza sull'addestramento, *backward propagation* e aggiornamento dei pesi vale anche in questo contesto. Tuttavia le *CNN* si basano sulla specifica assunzione che l'input sia costituito da immagini (matrici di 2 o più dimensioni) e questa ipotesi consente di codificare un'architettura adatta e vantaggiosa all'elaborazione di tale tipologia di dati.

CAPITOLO 3. DEEP LEARNING

Se si lavora con immagini ad elevate risoluzioni, una tradizionale rete fully connected, in cui ogni neurone è connesso a tutti i neuroni del livello successivo, porta ad un'esplosione combinatoria del numero di parametri contenuti nella rete [34].

Le principali differenze di una *CNN* rispetto alle ordinarie reti neurali che permettono di ovviare a questo problema riguardano [36]:

- il *processing*: i neuroni sono connessi solamente ad una piccola porzione dell'input (output dello strato precedente). Pertanto, ogni neurone esegue un'elaborazione locale, ripetuta su tutto l'input tramite l'operazione di convoluzione (descritta in seguito). Questo comporta una forte riduzione del numero di connessioni;
- i *pesi*: sono condivisi a gruppi. Ogni peso corrisponde ad una connessione. Ne consegue che il numero di entrambi risulta notevolmente ridotto.

Questa architettura si ispira alla struttura e al funzionamento della corteccia visiva dell'essere umano, in cui:

- i singoli neuroni hanno un piccolo campo recettivo (*receptive field*), ovvero rispondono solamente a stimoli provenienti da una porzione limitata del campo visivo e processano aspetti molto semplici degli stimoli visivi, per esempio i bordi delle superfici;
- i campi recettivi di neuroni differenti sono parzialmente sovrapposti in modo da ricoprire l'intero campo visivo e l'attività simultanea di tutti i neuroni permette di elaborare aspetti progressivamente più complessi degli stimoli visivi.

Similmente, le connessioni locali e i pesi condivisi fanno sì che porzioni diverse di un'immagine vengano processate allo stesso modo. Questo è importante perché permette di rilevare localmente informazioni ricorrenti (come bordi, spigoli o porzioni di oggetti) e globalmente, dall'attività di tutti i neuroni, forme via via più complesse fino a definire in modo univoco gli oggetti.

Il nome "rete convoluzionale" deriva dal fatto che la rete durante il suo funzionamento si serve dell'operazione matematica *convoluzione*. Questa rappresenta una buona approssimazione della risposta di un singolo neurone agli stimoli che hanno luogo nel suo campo recettivo.

La convoluzione è operazione tra due funzioni a valori reali; date queste due funzioni x e w , il loro prodotto di convoluzione può essere definito come [35]:

$$s(t) = (x * w)(t) = \int x(a)w(t - a)da \quad (3.10)$$

dove, in ambito di reti convoluzionali, x si riferisce all'input, generalmente un vettore multidimensionale di dati e w al *kernel* o *filtro*, un array multidimensionale di parametri impiegati dall'algoritmo di apprendimento (generalmente una matrice di poche righe e colonne). Tale filtro, nel contesto di rete neurale, corrisponde al neurone, e gli elementi di w ai pesi.

Le immagini elaborate dai computer sono digitali, e quindi discretizzate nello spazio, con le singole unità rappresentate da pixel ed il cui valore è quantizzato in livelli di grigio o di ciascun colore primario. L'integrale della formula precedente può quindi essere sostituito da una sommatoria. Se si ha come input un'immagine bidimensionale I , si può usare un kernel bidimensionale W :

$$S(i, j) = (I * W)(i, j) = \sum_m \sum_n I(m, n)W(i - m, j - n) \quad (3.11)$$

L'uscita $S(i, j)$ è rappresentata da una mappa di caratteristiche, *feature map* [35-36]. I kernel di convoluzione possono essere interpretati come feature detector, cioè come degli strumenti che ci permettono di rilevare i dettagli locali di un'immagine. Se prendiamo un'immagine di input e ad essa applichiamo un determinato kernel di convoluzione, otteniamo un'immagine risultante in cui la funzionalità che vogliamo analizzare viene isolata: ad esempio se vogliamo evidenziare le zone scure in una foto, allora avremo un'immagine risultante che sarà di fatto una mappa dei toni scuri. Analizzeremo più approfonditamente la struttura ed il funzionamento di una CNN nei prossimi paragrafi.

3.5.1 Architettura CNN

In genere le reti convoluzionali presentano un'architettura gerarchica organizzata nel seguente modo:

- *livello di input*: direttamente connesso ai pixel dell'immagine;
- *livelli per l'estrazione automatica di feature*: numerosi hidden layer eseguono operazioni che modificano i dati con lo scopo di apprendere le loro specifiche feature. Eliminano la necessità di estrazione manuale da parte dell'utente. Tre dei layer più comuni sono:
 1. *Convolutional Layer*: applica alle immagini filtri convoluzionali, che attivano determinate feature;

2. *ReLU Layer*: ricoprono il ruolo della funzione di attivazione dei neuroni. Solo le feature attivate vengono passate al layer successivo;
 3. *Pooling Layer*: riduce la dimensione dell'immagine, eseguendo un sotto-campionamento non lineare, preservando le informazioni più importanti. Ne deriva una progressiva riduzione del numero dei parametri che la rete deve apprendere.
- *livelli per la classificazione*: dopo avere appreso le feature in vari layer, la rete passa alla classificazione. Possono essere presenti più *Fully Connected Layer*, l'ultimo dei quali genera un vettore di dimensione k , dove k è il numero di classi che la rete è in grado di distinguere. Questo vettore contiene, per ogni classe, la probabilità che ciascuna immagine ha di appartenere ad essa.
L'output dell'ultimo layer convoluzionale per essere utilizzato come input ad uno strato fully connected, deve essere "appiattito", ovvero trasformato da tridimensionale a bidimensionale da un layer intermedio (*Flatten Layer*). Il layer finale è solitamente un *SoftMax Layer* che normalizza l'output del fully connected layer tramite la funzione probabilistica *SoftMax*.
 - *livello di output*: utilizza le probabilità calcolate dal softmax layer per assegnare l'input in maniera esclusiva alla classe di appartenenza.

Tipicamente, in una CNN ad una serie di strati convoluzionali e strati di ReLU vengono fatti seguire strati di pooling, riducendo via via la dimensione della matrice, ma aumentando il livello di astrazione.

Supponiamo che l'input della CNN sia un'immagine rappresentante un'auto. Si passa da filtri elementari che identificano caratteristiche semplici come curve, linee

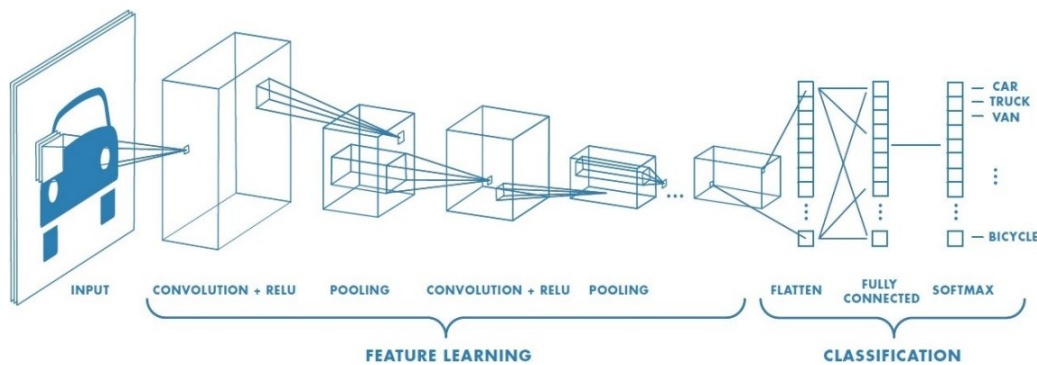


Figura 3.11 Esempio di architettura di una rete convoluzionale. Ad ogni immagine vengono applicati filtri e l'output di ogni immagine convolta viene utilizzato come input per il layer successivo [28].

verticali, linee orizzontali, a filtri sempre più sofisticati [36], in grado ad esempio di riconoscere i fanali, le ruote, il parabrezza... fino all'ultimo livello, che è in grado di distinguere un'automobile da un camion.

La rete prende come input l'immagine e viene processata livello per livello; ogni livello produce un output che viene preso in input dal livello successivo, fino ad arrivare a quello finale che restituisce una predizione sulla probabilità che l'immagine appartenga ad una classe piuttosto che ad un'altra.

Nei successivi paragrafi verrà analizzata la funzione degli strati più comunemente impiegati nella progettazione di una *CNN*.

3.5.1.1 Convolutional Layer

Come si può intuire il *Convolutional Layer* rappresenta l'elemento principale delle *CNN*. Alla base di questo strato c'è l'operazione di convoluzione, cioè l'applicazione di filtri digitali (maschere di 2 o più dimensioni, i cui elementi corrispondono ai pesi).

L'immagine da analizzare è rappresentata come un insieme di pixel $M \times N \times Z$, dove M indica la larghezza, N l'altezza e Z la profondità (ad esempio, per un'immagine a colori Z è uguale a 3, cioè i tre canali *RGB*). I filtri sono spazialmente piccoli lungo le dimensioni di larghezza ed altezza, generalmente sono matrici 3×3 o 5×5 , ma si estendono su tutta la profondità del volume di input.

Il filtro viene fatto scorrere lungo la larghezza e l'altezza del volume di input, calcolando nelle varie posizioni il prodotto scalare fra la maschera e la porzione dell'input coperta. Man mano che il filtro si sposta lungo l'area dell'input, produce in output una mappa di attivazione che rappresenta la risposta relativa a quello specifico filtro in ogni posizione spaziale.

In ogni layer convoluzionale è presente un set di filtri (ogni filtro corrisponde ad un neurone, ed i suoi elementi ai pesi) e ognuno di essi genera una diversa mappa di attivazione; impilando tutte queste mappe lungo la dimensione di profondità si ottiene il volume di output completo.

Gli output degli strati di convoluzione risultano così organizzati in feature map, dove ogni regione locale è ottenuta dalla rispettiva regione locale del volume di input. L'estensione spaziale di tale regione (dimensione del filtro) è proprio il campo recettivo del neurone di cui abbiamo parlato in precedenza. Queste congiunzioni locali tra le feature dei vari strati permettono di rilevare gli stessi

pattern in differenti parti dell'immagine e questo è un comportamento desiderato, perché banalmente, se fossero presenti due cerchi, uno nella parte destra di un'immagine e uno nella parte sinistra, entrambi dovrebbero essere riconosciuti e processati dall'algoritmo alla stessa maniera [36].

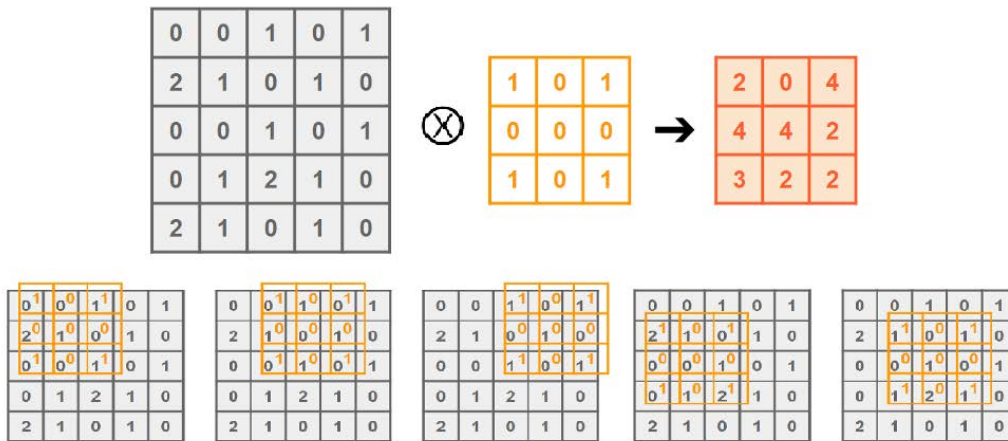


Figura 3.12 Esempio numerico di convoluzione bidimensionale con stride 1, in assenza di zero-padding. La matrice in input ha dimensione 5x5, il filtro ha dimensione 3x3.

Le dimensioni del volume di output di un certo livello convoluzionale sono influenzate da tre importanti parametri, ed appartengono agli *iper-parametri* [34]:

1. La *profondità (depth)*: controlla il numero di neuroni nel layer convoluzionale, ovvero il numero di filtri dello strato di convoluzione. Lungo la dimensione della profondità ciascuno strato di neuroni impara a rilevare caratteristiche dell'ingresso differenti. Ci si può riferire ad un insieme di neuroni che sono connessi agli stessi dati di input come ad una *colonna profondità (depth column)*.
2. Il passo (*stride*): serve per specificare come allocare la colonna profondità lungo le dimensioni spaziali (larghezza, altezza) dell'input. In altre parole corrisponde al numero di pixel di cui viene spostato il filtro sulla matrice in input. All'aumentare dello stride si riduce la dimensione della feature map nel volume di output.
3. Lo *zero padding* è un'operazione che permette di regolare le dimensioni spaziali del volume di output e consiste nell'aggiungere spazialmente valori uguali a 0 ai bordi del volume di input.

Data la dimensione spaziale del volume in input (W_{in}), la dimensione spaziale del filtro (F), lo stride (S) e l'ammontare di zero-padding aggiunto ai bordi (P) è

possibile ottenere la dimensione spaziale del volume di uscita W_{out} tramite la seguente relazione:

$$W_{out} = \frac{(W_{in} - F + 2P)}{S} + 1 \quad (3.12)$$

La capacità di una *CNN* può variare in base al numero di layer convoluzionali che contiene. Solitamente partendo dall'input layer fino all'output layer troviamo una serie di questi layer che permettono di rilevare via via feature sempre più complesse; pertanto più numerosi sono questi strati e più feature dettagliate la rete riesce ad elaborare.

3.5.1.2 ReLU Layer

Lo strato *ReLU* (*Rectified Linear Units*) è molto comune in una *CNN* e viene utilizzato più volte all'interno della stessa rete, molto spesso dopo ciascun layer di convoluzione.

Questo tipo di strato incrementa la proprietà di non linearità della rete. Prende in input un tensore e fornisce in output un tensore di pari dimensioni, ottenuto applicando a tutti i valori del volume di input la funzione di attivazione *ReLU* [31]:

$$ReLU(x) = \max(0, x) = \begin{cases} x & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} \quad (3.13)$$

In semplici parole questo livello annulla tutti i valori negativi e lascia invece invariati i valori positivi.

La funzione di attivazione *ReLU* viene preferita alle funzioni di attivazione introdotte in precedenza perché rispetto ad esse permette di ridurre i tempi di addestramento in modo considerevole senza impattare significativamente sull'accuratezza dei risultati.

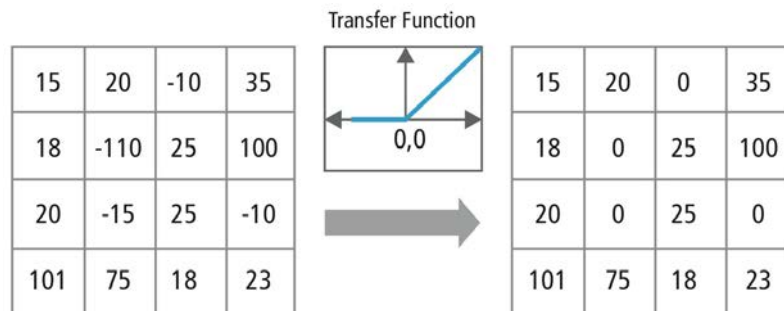


Figura 3.13 Applicazione funzione di attivazione *ReLU* ad una matrice 4x4.

3.5.1.3 Pooling Layer

Nell'architettura tipica di una *CNN* è pratica comune inserire tra due successivi strati convoluzionali uno strato di *pooling*.

Questo riduce la dimensione spaziale dell'immagine (e quindi anche la sua risoluzione), attraverso un'operazione di *downsampling* non lineare che aggrega con una statistica riassuntiva le informazioni di pixel vicini.

Da notare che questo livello opera indipendentemente su ogni feature map e di conseguenza la profondità del volume di input rimane inalterata (cambiano solamente altezza e larghezza).

Gli operatori di aggregazione più utilizzati sono la *media (AVE)* e il *massimo (MAX)* e si parla rispettivamente *Mean Pooling* e *Max Pooling*.

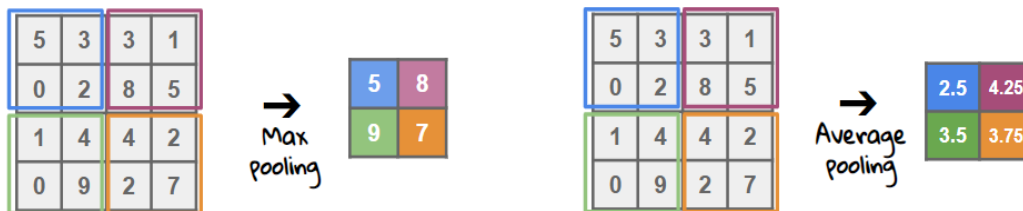


Figura 3.14 Sotto-campionamento eseguito con l'operatore di media e massimo su una matrice in input di dimensione 4x4. Il filtro ha dimensione 2x2. L'output è rappresentato da una matrice 2x2 con i valori max e medi rispettivamente delle regioni filtrate. L'applicazione locale è evidenziata con diversi colori.

Per eseguire il ridimensionamento l'immagine in input viene divisa in un insieme di sottomatrici non sovrapposte e, per ciascuna di queste regioni si prende la media dei pixel nel primo caso o il valore massimo nel secondo caso.

L'obiettivo è quello di rendere invariante la rete rispetto a piccole trasformazioni, distorsioni, traslazioni dell'input, mantenendo al tempo stesso le informazioni significative per la discriminazione dei pattern (per esempio una piccola distorsione di input non cambierà l'output del pooling, perché prendiamo il valore massimo / medio in una regione). Alla base di questo ragionamento c'è l'intuizione che una volta individuata una certa caratteristica specifica nell'immagine, la sua posizione esatta non è importante quanto la sua posizione relativa rispetto alle altre caratteristiche [36].

Questo serve a risolvere due problematiche introdotte dal layer convoluzionale:

3.5 RETI NEURALI CONVOLUZIONALI (CNN)

- *Complessità*: analizzando piccole parti dell'input con numerose zone di overlapping, il volume di output aumenta molto rispetto a quello di input e livello dopo livello la dimensionalità diventerebbe intrattabile;
- *Overfitting*: un layer convoluzionale genera delle informazioni troppo relazionate al training set e questo di conseguenza potrebbe portare ad una condizione progressiva di sovradattamento.

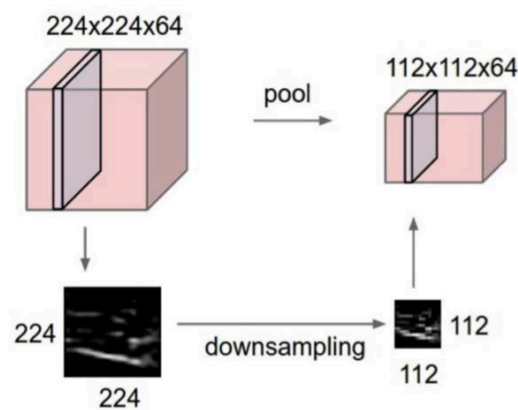


Figura 3.15 sotto-campionamento eseguito da un pooling layer.

Al contrario, come preannunciato uno strato di pooling fonde semanticamente feature simili, preservando solamente le informazioni salienti ai fini dell'analisi e questo consente di ridurre il numero di parametri all'interno della rete, quindi di alleviare il carico computazionale e di prevenire la suddetta situazione di overfitting [34-35].

Anche i pooling layer possiedono dei parametri settabili:

- L'estensione spaziale della finestra (F) che di volta in volta viene considerata per il downsampling;
- Il parametro di stride (S) che quantifica lo spostamento della finestra;

Generalmente, invece, non si ricorre allo zero-padding in quanto non apporta miglioramenti.

La dimensione spaziale del volume di uscita può essere ancora calcolata con l'equazione (3.12), considerando $P=0$ [34].

3.6 Segmentazione Semantica

Nei problemi di classificazione, generalmente è presente un'immagine con un singolo oggetto come focus ed il compito è dire cosa rappresenta l'immagine.

Talvolta può essere utile estendere il livello di comprensione dell'immagine e in questo senso uno dei più importanti task della Computer Vision è rappresentato dalla segmentazione semantica: si tratta di un algoritmo che invece di classificare a livello dell'intera immagine, classifica a livello di pixel.

Rispetto a quanto visto finora per la classificazione, l'output previsto dalla segmentazione semantica non sono solo le etichette, ma un'immagine ad alta risoluzione, generalmente delle stesse dimensioni dell'immagine in input, in cui ogni pixel è classificato in una determinata categoria.

Poiché prevediamo questa attività per ciascun pixel, spesso si parla di *previsione densa*.

Anche per questo tipo di problemi l'architettura più utilizzata è costituita dalle *CNN*.

Le tradizionali *CNN* sono però nella loro forma intrinseca delle reti di classificazione. Prendono in input un'immagine e assumendo che in essa sia presente un solo oggetto classificano l'intera immagine in una tra le varie categorie predefinite.

Abbiamo visto che mediante l'inserimento periodico di livelli di downsampling all'interno della rete, l'immagine viene concentrata passo dopo passo in poco spazio, mantenendo le informazioni utili a comprenderne solamente il contesto complessivo. La struttura ordinaria di una *CNN* consente perciò di analizzare l'immagine nella sua interezza, rendendo chiaro "cosa" rappresenta, ma scendendo

in profondità, perde via via la cognizione di "dove" le informazioni si trovano rispetto all'input. In ambito di segmentazione semantica è necessario ripristinare le informazioni di localizzazione e ciò può essere conseguito aggiungendo un'implementazione inversa alla *CNN* [38].

Tipicamente, l'architettura di una *CNN* per eseguire task di segmentazione semantica contiene due percorsi:

1. *Encoder*: il primo percorso è quello di “contrazione”, costituito da una tradizionale pila di strati convoluzionali e di pooling, che ha il compito di ridurre la dimensionalità con le operazioni precedentemente descritte al fine di acquisire il contesto d’insieme nell’immagine;
2. *Decoder*: il secondo percorso è detto di “espansione simmetrica” e viene utilizzato per recuperare la localizzazione spaziale tramite i *transposed convolutional layer*, a volte impropriamente appellati livelli di deconvoluzione. Questi livelli agiscono in maniera inversa rispetto al pooling ed hanno come risultato la crescita spaziale degli input. Convertono un’immagine a bassa risoluzione in un’immagine ad alta risoluzione.

È utile osservare che affinché l’immagine finale abbia dimensioni comparabili con quelle dell’immagine in input, è necessario che i processi di downsampling e upsampling [35] vengano eseguiti lo stesso numero di volte.

Queste tecniche hanno permesso al campo della Computer Vision di avanzare rapidamente negli ultimi anni, trovando applicazione sempre più diffusa in molti settori: progettazione di veicoli a guida autonoma, segmentazione di immagini biomediche, Geo-Sensing, agricoltura di precisione, segmentazione facciale e molti altri.

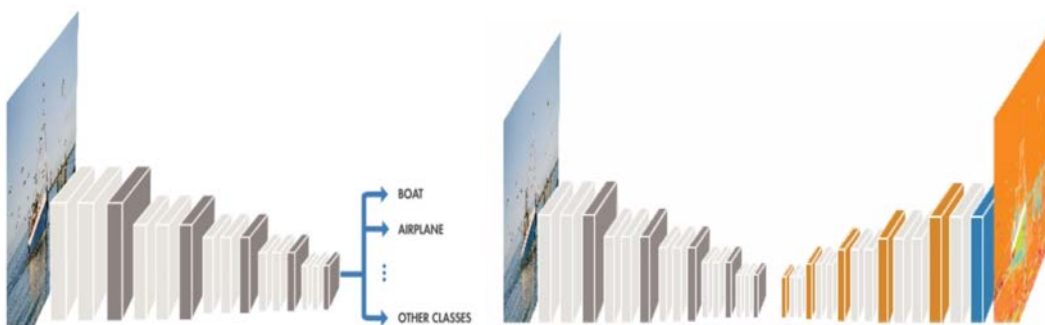


Figura 3.16 Architettura di una rete per la classificazione a sinistra e per la segmentazione semantica a destra.

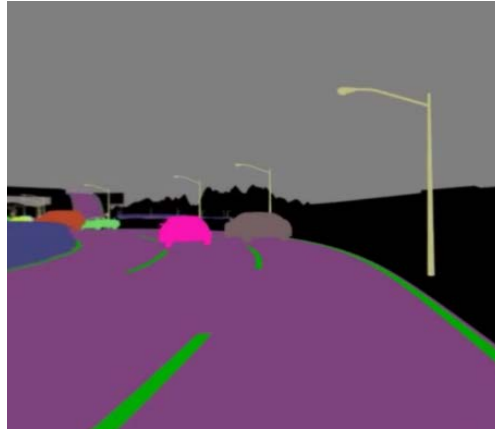


Figura 3.17 Applicazione della segmentazione semantica nella realizzazione di veicoli a guida autonoma: a ciascun pixel viene associata un'etichetta, permettendo di riconoscere all'interno dell'immagine categorie distinte di oggetti: veicoli, marciapiedi, lampioni, i segni di corsia ed altri elementi della strada.

In questo progetto di tesi è stato sviluppato un algoritmo per la segmentazione semantica di immagini del fundus oculi, mirato a fornire valido supporto alla diagnosi del glaucoma. La struttura che è stata scelta per la rete è la *U-Net*, sviluppata per la prima volta da *Olaf Ronneberger* per la *Bio Image Segmentation* [39]. Maggiori dettagli a riguardo verranno forniti nel prossimo capitolo.

3.7 Applicazioni cliniche del Deep Learning

Uno studio pubblicato recentemente ha rilevato come esistano più di 300 applicazioni di *DL* di tipo biomedicale [41].

Ad ulteriore riprova dell'interesse crescente per questa tecnologia, ad oggi sono in corso molti progetti che ne prevedono l'utilizzo, soprattutto nello screening, in dispositivi di assistenza alla diagnosi e, più recentemente, nella radiomica.

3.7 APPLICAZIONI CLINICHE DEL DEEP LEARNING

Tra i principali settori di applicazione: la mammografia, gli studi morfo-funzionali cerebrali, l'esame del fondo oculare e i software per l'assistenza radiologica, ad esempio per riconoscimento dei noduli polmonari o la segmentazione delle lesioni epatiche.

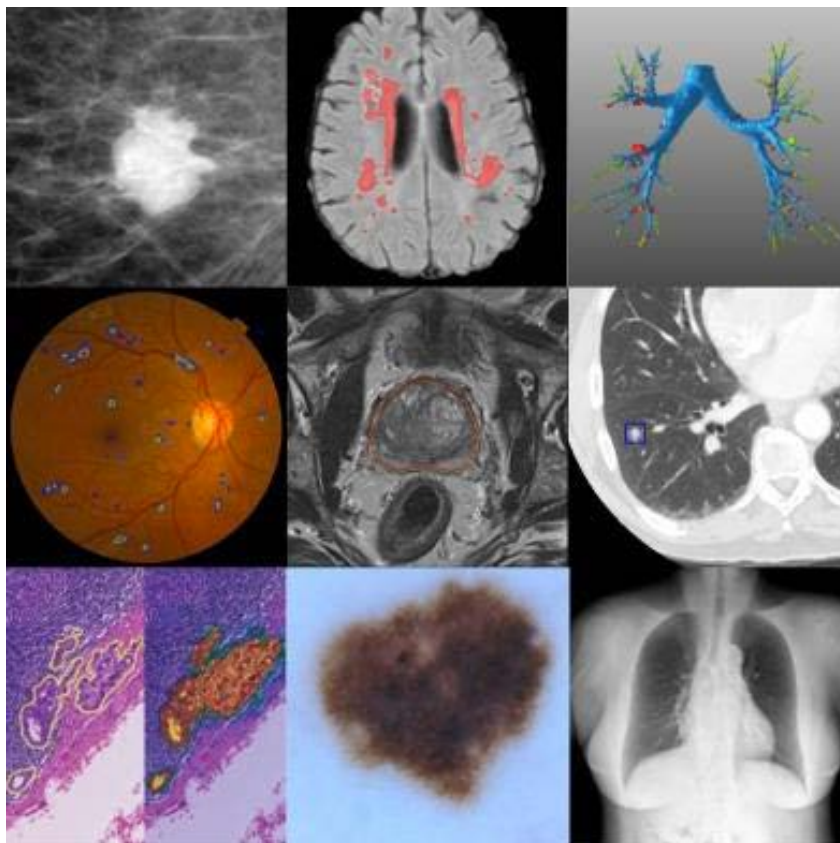


Figura 3.18 Esempi di applicazione del DL in ambito biomedicale

Queste applicazioni condividono la possibilità di offrire alla macchina una banca dati di immagini da cui l'algoritmo possa estrapolare delle caratteristiche ben definite di patologia che in un secondo momento sarà in grado di riconoscere o escludere in corso di analisi biomedicale, definendo un pattern come patologico o meno.

Una cosa interessante è il soggetto dell'analisi: le aree di applicazione anatomica della nuova tecnologia sono molto differenziate e pressoché tutti i distretti corporei sono stati studiati come possibilità di applicazione del *DL*.

Da un punto di vista operativo il *DL* ha dimostrato una grande efficacia nella *classificazione*, nella *segmentazione* e nel *riconoscimento*.

La capacità di *classificazione* in questo campo consiste nel predire una variabile diagnostica avendo molte immagini come input e spesso si concretizza nel

determinare la presenza o l'assenza di una malattia; un classico esempio l'imaging retinico o la patologia polmonare in CT.

Oltre che di immagine o esame, la classificazione può essere di oggetto anatomico o lesione.

In questo ultimo caso, in genere ci si focalizza su una piccola parte dell'immagine, che, dopo essere stata identificata, viene a sua volta classificata.

L'abilità di *riconoscere* permette la localizzazione di organo, di regione, di confini, oppure di oggetto o di lesione. Una grande accuratezza diagnostica, in particolare, è stata evidenziata nel riconoscimento precoce di lesioni polmonari.

La localizzazione di oggetti anatomici di un organo o dei suoi confini è stato un passo importante per permettere lo sviluppo delle capacità di segmentazione.

La *segmentazione*, che può essere di organo, di struttura o di lesione, ha dimostrato grandi possibilità, ad esempio nell'identificazione dei pixel o voxel che costituiscono aree di interesse, in particolare le lesioni neoplastiche, permettendo un'analisi quantitativa dei parametri clinici in correlazione a volume e forme.

3.7.1 Deep learning nell'analisi di immagini in oftalmologia

Numerosi studi hanno riportato l'applicabilità di classificazione, riconoscimento e segmentazione in oftalmologia, interessando uno spettro ampio di patologie [43]:

- segmentazione e riconoscimento di anomalie retiniche;
- segmentazione dei vasi retinici e del disco ottico;
- classificazione della cataratta con la lampada a fessura;
- riconoscimento del glaucoma, della retinopatia diabetica, della degenerazione maculare correlata all'età;
- segmentazione con tomografia a coerenza ottica dei vasi retinici.

Pertanto il *DL* ha dimostrato una straordinaria versatilità nella valutazione di alcune delle più frequenti patologie che affliggono l'apparato visivo: glaucoma, degenerazione maculare correlata all'età; retinopatia diabetica e retinopatia del prematuro [6].

Inoltre mediante l'analisi di certe caratteristiche oculari si è dimostrato che è possibile predire età, sesso e status di fumatore, andando addirittura a determinare l'eventuale esistenza di fattori di rischio cardiovascolari. In tale ottica olistica, il *DL* è stato anche applicato con successo nella diagnostica di alcune patologie

3.7 APPLICAZIONI CLINICHE DEL DEEP LEARNING

sistemiche, come la neuropatia diabetica (mediante una valutazione dei nervi corneali) o il Parkinson.

Lo straordinario sviluppo nel settore dell'oftalmologia del *DL* ha raggiunto un apice nell'aprile 2018, quando la Food and Drug Administration (FDA) ha permesso ad una compagnia di immettere nel mercato un dispositivo diagnostico AI-powered per un utilizzo in oftalmologia, con software dedicato al riconoscimento della retinopatia diabetica, ed il cui output costituisce diagnosi clinica.

Questo è stato il primo dispositivo con applicazione, autorizzata, nello screening decisionale clinico senza la necessità di una validazione del medico.

Le potenzialità del *DL* in oftalmologia non si limitano allo screening ma si estendono anche alla diagnosi e al management delle patologie oculari; infatti la nuova tecnologia nel contesto diagnostico oftalmologico fornisce una rappresentazione digitale dei pattern patologici individuati in corso di esame, fornendo informazioni con importanti risvolti clinico-terapeutici.

Un aspetto importante è che il medico non dovrà avere un ruolo marginale nell'avvento della nuova era digitale, in quanto potrà avvalersi delle nuove tecnologie di aiuto alla diagnosi e concentrarsi nel suo ruolo di essere medico-persona, maggiormente attenta all'interazione e alla gestione con i pazienti.

Capitolo 4

Algoritmo per l'analisi automatica

Il progetto di tesi, svolto in collaborazione con D-EYE, consiste nello sviluppo di un algoritmo per l'*analisi automatica di immagini retiniche* acquisite con questo dispositivo ottico, finora mai sviluppato, con l'obiettivo di migliorare il sistema in termini di praticità ed efficienza.

In maggior dettaglio, il lavoro svolto si è articolato in due diverse fasi, in ciascuna delle quali si è sperimentato un diverso approccio alle *CNN*, *Classificazione* e *Segmentazione semantica*.

Nella prima fase è stato affrontato il problema della qualità delle immagini retiniche acquisite dal dispositivo. È stato sviluppato un algoritmo in grado di estrapolare autonomamente dai video registrati solamente i frame utili o, in altri termini, di interesse clinico.

Più precisamente, è stata implementata una *Convolutional Neural Network* con l'obiettivo di *classificare* le immagini in input in due categorie banalmente nominate come “buoni” o “cattivi” a seconda che in esse sia visibile o meno la retina.

Spesso infatti le registrazioni video includono numerosi frame di scarsa utilità, ad esempio troppo sfocati, o addirittura inutili, come volti di persone o dettagli della stanza in cui si è svolto l'esame; inoltre, in considerazione del fatto che l'esame del fondo dell'occhio prevede che il fascio luminoso penetri attraverso la pupilla, anche la motilità oculare o un battito di ciglia possono inficiare sulla qualità dell'immagine.

In un secondo momento, è stato realizzato un algoritmo basato su reti convoluzionali a forma di U, anch'esso completamente automatizzato, per l'analisi dei frame cosiddetti “buoni”. L'applicativo sviluppato è in grado di *segmentare*

il disco ottico e selezionare i fotogrammi più rilevanti (ovvero quei frame in cui il disco ottico è ben visibile e messo a fuoco). Inoltre, seppur su un numero molto limitato di immagini, è stata realizzata una *segmentazione della cup*. Il rapporto tra la grandezza del disco e della cup costituisce un'informazione utile per la diagnosi

del Glaucoma, e l'analisi automatica è quindi in grado di fornire un parametro direttamente collegato alla presenza di questa patologia.

L'aspetto più interessante di questo progetto è che mediante i metodi di *Deep Learning* è stato possibile eseguire un'analisi dei video acquisiti in maniera del tutto automatica, senza necessità di intervento da parte di un utente esterno.

L'automazione permette di avere un risultato oggettivo e riproducibile, e, soprattutto nel caso in cui l'operatore arruolato non sia un medico specialista, ridurre il rischio di errore operatore-correlato e di evitare una lunga analisi manuale dei frame, incrementando così l'efficienza del sistema diagnostico con riduzione di tempi e costi.

Di seguito verrà esposta brevemente la metodologia adoperata. La prima parte del capitolo sarà dedicata all'organizzazione dei dati utilizzati; verranno poi descritti il pre-processing delle immagini e l'architettura delle CNN implementate. Infine verranno riportati alcuni rilevanti aspetti relativi alle procedure di training e al settaggio degli iperparametri.

4.1 Dataset

L'operazione preliminare per l'addestramento delle reti convoluzionali è stata la creazione di adeguati *dataset*. Come già accennato ogni addestramento è stato effettuato con un approccio di apprendimento supervisionato, quindi l'accortezza è stata quella di etichettare manualmente ogni frame in funzione del proprio contenuto, a livello di immagine per la messa a punto del classificatore, a livello di pixel per la segmentazione.

In questo progetto abbiamo utilizzato 2 diversi dataset:

- Il **dataset 1** comprende video di fundus oculi acquisiti con un adattatore D-EYE collegato ad iPhone 5 e 6 (Apple Inc., Cupertino, USA). Ciascun video ha una durata di circa 14 secondi, 30 fotogrammi al secondo e una risoluzione di 1080x1920 pixel. Sono stati acquisiti in diversi centri clinici e ogni video è stato reso anonimo eliminando qualunque informazione relativa al paziente.

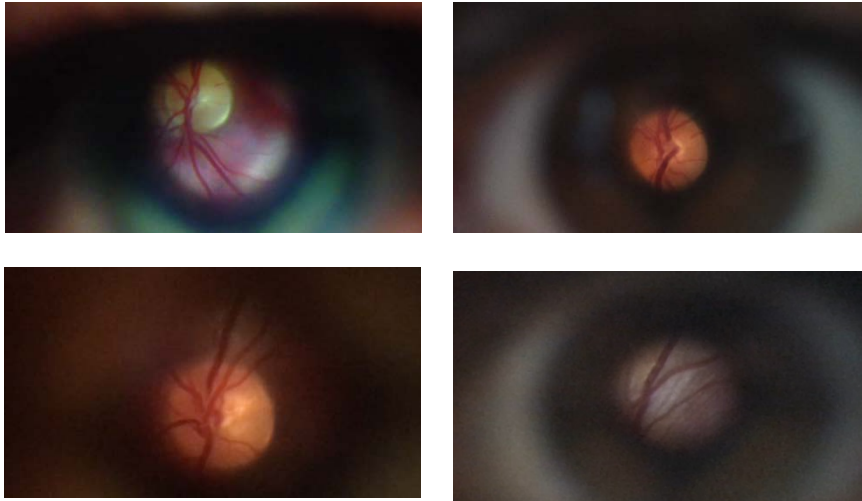


Figura 4.1 Esempi di immagini rappresentative appartenenti al dataset 1.

Questo dataset, a seconda del tipo di addestramento da conseguire, è stato suddiviso in 3 diversi modi:

- Il *subset 1* comprende 66 video del fundus oculi. Da questi video sono stati selezionati manualmente 1874 frame, in particolare 829 frame “buoni” in cui la retina è visibile e 1045 “cattivi” in cui invece la retina non è rilevabile oppure è vistosamente sfocata.
- Il *subset 2* comprende 30 video del fundus oculi, acquisiti su soggetti sani. Da questi video sono stati selezionati manualmente 1000 frame. La *Figura 4.2* mostra dei frame rappresentativi. Nei frame selezionati, il disco può essere interamente visibile (c), parzialmente visibile (b) o non visibile (a). Su ciascun fotogramma è stata eseguita la segmentazione manuale del disco ottico (se presente).
- Il *subset 3* comprende 5 video di fundus oculi acquisiti su soggetti sani e 5 video acquisiti su soggetti con glaucoma. Per ciascun video, è stato selezionato manualmente il frame migliore (ovvero il fotogramma in cui il disco ottico è più visibile) e ancora manualmente sono stati segmentati disco ottico e cup.

CAPITOLO 4. ALGORITMO PER L'ANALISI AUTOMATICA

- Il **dataset 2** è il dataset *Retinal fundus Images for Glaucoma Analysis (RIGA)* [44] ed è costituito da 750 immagini acquisite con fundus camera midiatriche e non midiatriche, con diverse risoluzioni.

In particolare, il dataset RIGA è composto da 460 immagini del dataset MESSIDOR, 195 immagini del dataset Bin Rushed, 95 immagini del dataset Magrabi. Abbiamo escluso un'immagine di questo dataset perché contenente annotazioni manuali. Ogni immagine è stata analizzata manualmente da 6 esperti di Glaucoma che hanno eseguito la segmentazione di disco ottico e cup.

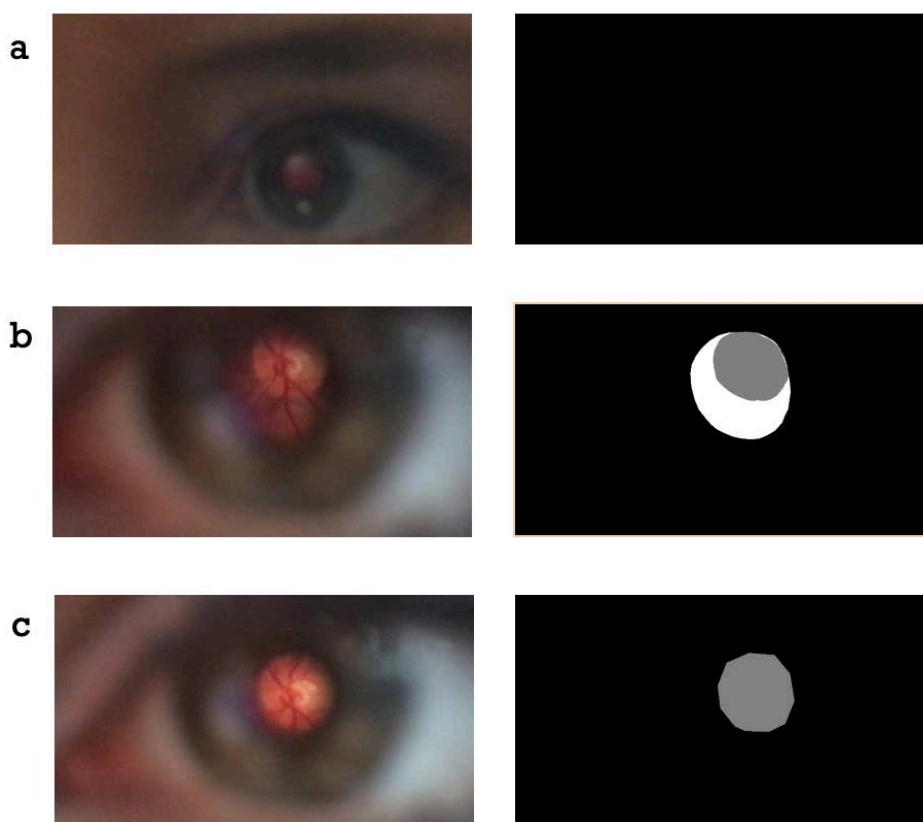


Figura 4.2 Esempi rappresentativi di frame acquisiti appartenenti al subset 2 e relativa segmentazione manuale, in bianco la retina (non utilizzata) e in grigio il disco ottico.

4.2 CNN – selezione dei frame

In questa sezione verranno delineati i principali passi seguiti per l'implementazione della *CNN* realizzata per risolvere il problema di classificazione automatica dei frame di interesse clinico.

Per lo sviluppo del classificatore è stato utilizzato il *subset I*. Prima di procedere con l'analisi, le immagini sono state suddivise in due gruppi:

- *training set* per l'addestramento della rete;
- *test set* per determinare l'accuratezza del modello.

Nel campo specifico delle *CNN* il training set richiesto presenta requisiti particolarmente stringenti relativamente alla cardinalità di dati necessari all'apprendimento. Pertanto solamente nel caso in cui il training set sia abbastanza numeroso è possibile utilizzarne una parte per il *validation set*, utile a stimare la generalizzazione del modello durante l'allenamento.

Nel nostro caso, disponendo di una mole di dati limitata, il tentativo di partizionare ulteriormente il training set portava il modello a perdere la capacità di classificazione, quindi in fase di sperimentazione questa idea è stata abbandonata.

Il codice di implementazione per definire la *CNN*, allenarla e testarla, seguendo un approccio supervised, prevede quattro importanti passi:

1. la definizione delle operazioni per gestire l'input e facilitare l'apprendimento (*pre-processing*);
2. la definizione dell'architettura della rete convoluzionale;
3. il settaggio (*tuning*) degli iperparametri per la fase di training;
4. stima delle performance del modello attraverso la fase di testing.

4.2.1 Pre-processing

Le difficoltà che emergono nell'implementare un efficiente algoritmo di classificazione possono essere ridotte ricorrendo ad una fase di **pre-processing** delle immagini da elaborare. Le *CNN* comunque essendo progettate per ricavare i dati significativi direttamente dalle immagini in input richiedono una quantità di pre-processing molto limitata.

Per processare i dati non esiste un approccio standard, ma per questioni legate all'ottimizzazione dell'addestramento è bene che i frame da passare in ingresso alla rete siano ridimensionati. Da considerare che se da una parte avere delle immagini ridimensionate porta ad una velocità di esecuzione molto rapida, dall'altra comporta una riduzione della qualità delle stesse. Quindi, sulla base di queste osservazioni, previa opportuna valutazione, è stato applicato ad ogni immagine un resize tale che la dimensione finale risulta essere 135x240. Si è notato essere un buon compromesso perché pur non comportando una rilevante perdita di informazioni ha permesso di alleviare in modo significativo il peso computazionale. Inoltre, con lo scopo di standardizzare tutte le raffigurazioni, si è imposto un orientamento orizzontale per tutti i frame da elaborare.

4.2.2 Architettura della CNN

Una volta definita la natura delle immagini di input, si è definita l'**architettura** della rete come segue:

```
layers = [  
    imageInputLayer([135 240 3])  
  
    convolution2dLayer(3,8,'Padding','same')  
    batchNormalizationLayer  
    reluLayer  
  
    maxPooling2dLayer(2,'Stride',2)  
  
    convolution2dLayer(3,16,'Padding','same')  
    batchNormalizationLayer  
    reluLayer  
  
    maxPooling2dLayer(2,'Stride',2)  
  
    convolution2dLayer(3,32,'Padding','same')  
    batchNormalizationLayer  
    reluLayer  
  
    maxPooling2dLayer(2,'Stride',2)  
  
    convolution2dLayer(3,32,'Padding','same')  
    batchNormalizationLayer  
    reluLayer
```

```
fullyConnectedLayer(2)
softmaxLayer('Name','Prob')
classificationLayer];
```

Figura 4.3 Codice MATLAB usato per definire l'architettura della CNN.

Si tratta di un'architettura CNN caratterizzata da una struttura molto semplice, ma allo stesso tempo molto profonda: contiene 19 livelli.

Il primo strato, *imageInputLayer*, specifica la dimensione delle immagini in ingresso e consente di effettuare una normalizzazione delle stesse. Più precisamente i valori 135x240x3 indicano altezza, larghezza e numero di canali del volume di input. Nel nostro caso i frame sono immagini a colori RGB per cui i canali sono 3 ed il valore di ogni pixel è codificato in double precision con valori compresi tra 0 e 1.

Successivamente sono state inserite 4 unità convoluzionali, intervallate da uno strato di *Max Pooling*.

Ciascuna unità convoluzionale è costituita da 3 diversi strati disposti nel seguente ordine:

```
convolution2dLayer
batchNormalizationLayer
reluLayer
```

Figura 4.4 Strati che costituiscono una unità convoluzionale della CNN implementata.

I *Convolutional Layer* rappresentano la componente principale della CNN. Il loro compito consiste nell'estrazione di feature locali dall'immagine attraverso l'applicazione di filtri apprendibili.

Ogni filtro di dimensione 3x3 viene fatto scorrere sulla matrice in input, compiendo l'operazione di convoluzione e determinando in output una *activation map*. Livello dopo livello è stato applicato un numero crescente di filtri: il primo modulo usa 8 filtri, il successivo strato 16 e gli ultimi due 32.

Di default lo *stride* è stato impostato a 1, ovvero il filtro si sposta lungo le dimensioni spaziali della matrice di input di un pixel alla volta. Per quanto riguarda lo zero-padding, invece, è stato specificato il valore 'Same' in modo tale che l'output preservi le dimensioni della matrice di input.

Per ridurre i tempi dell'addestramento e apportare un generale miglioramento delle performance è stato poi interposto tra uno strato convoluzionale e lo strato di ReLU un layer di *Batch Normalization*. In questo modo oltre all'input della rete si

normalizza anche l'output di ogni layer convoluzionale. Ciò porta beneficio al training in quanto riduce il range dinamico dei dati in input ad un range fisso consentendo alla rete di estrarre feature in maniera più robusta e veloce.

Tipicamente l'addestramento delle reti neurali viene effettuato mediante mini-batch, cioè i dati di training vengono divisi in modo casuale in mini-gruppi ed ogni mini-gruppo viene considerato a turno in una diversa iterazione.

Il *batchNormalizationLayer* in primo luogo normalizza le attivazioni di ciascun canale x_i sottraendo la media (μ_B) del mini-batch e dividendo per la sua deviazione standard (σ_B^2):

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (4.1)$$

dove ϵ migliora la stabilità numerica quando la varianza del mini-lotto è molto piccola.

In un secondo momento va a shiftare l'input di un offset β e lo ridimensiona con un fattore di scala γ . L'uscita normalizzata può essere scritta come:

$$y_i = \gamma \hat{x}_i + \beta \quad (4.2)$$

Si noti che γ e β vengono appresi durante l'allenamento insieme ai parametri originali della rete e nella fase di test, la rete non calcolerà più i valori di media e deviazione standard, ma userà quelli appresi durante l'addestramento.

Lo strato successivo è rappresentato da un *reluLayer*, composto da neuroni che applicano la funzione di attivazione $f(x) = \max(0, x)$.

Questi livelli annullano tutti i valori negativi, aumentando la non linearità della rete, senza però influenzare i campi recettivi dei livelli di convoluzione.

I *Max Pooling Layer* inseriti tra le varie unità di convoluzione partizionano l'immagine in input in un insieme di quadrati, nel nostro caso di dimensione 2x2 (specificata dal primo argomento) e per ciascuna di queste regioni restituiscono come output il valor massimo. Per quanto riguarda lo *Stride*, è stato impostato a 2.

Lo scopo di questo strato è quello di ridurre progressivamente le dimensioni spaziali delle immagini, in modo da diminuire il numero di parametri, quindi la complessità computazionale della rete e prevenire allo stesso tempo il verificarsi di overfitting.

In coda all'ultimo di questi strati è presente un *fullyConnectedLayer*, la cui principale funzione è quella di raggruppare le feature di alto livello identificate dai livelli precedenti per poter assegnare l'immagine ad una specifica categoria.

Esegue un prodotto tra la matrice di ingresso e la matrice dei pesi e aggiunge i bias di bilancio, generando in uscita un vettore k- dimensionale, dove k è il numero di classi che la rete sarà in grado di distinguere; in questo caso sono due 2, 'buoni' e 'cattivi'.

Il penultimo strato della rete è un *softmaxLayer*, che normalizza l'uscita del *Fully Connected Layer*. Questo livello fornisce in output un valore compreso tra 0 e 1 per ogni classe, che rappresenta la probabilità di appartenenza dell'immagine in input a quella specifica classe. Più precisamente, la funzione implementata da questo strato calcola la distribuzione di probabilità condizionata dei suoi valori in input y_i come segue:

$$S(y_i) = \frac{e^{y_i}}{\sum_i e^{y_i}} \quad (4.3)$$

Le probabilità fornite dalla funzione di attivazione softmax vengono sfruttate dal layer finale, che è un *classificationLayer*, per calcolare la perdita di *cross-entropy* e assegnare l'immagine ad una delle due classi.

È stata scelta questa semplice architettura per questioni legate ai tempi ed ai costi computazionali di elaborazione delle immagini. L'obiettivo finale sarà infatti implementare CNN che, una volta addestrate, possono essere eseguite direttamente sullo smartphone, ottenendo comunque buone performance nella classificazione.

4.2.3 Iperparametri

Dopo aver effettuato l'organizzazione dei dati e aver progettato l'architettura si è proseguito con il training della rete.

La principale sfida nell'addestramento di una CNN è rappresentata dalla scelta degli **iperparametri**, le variabili che regolano l'allenamento stesso.

L'individuazione di un'opportuna configurazione per questi parametri è cruciale nel progetto di una rete in quanto ne determina in larga misura le prestazioni e l'efficienza, sia in termini di accuratezza di classificazione che di richiesta di risorse computazionali durante la fase di addestramento.

CAPITOLO 4. ALGORITMO PER L'ANALISI AUTOMATICA

L'impostazione di adeguati valori è un'operazione legata all'intuizione e all'esperienza degli sviluppatori; non ha solide basi matematiche, ma dipende dallo specifico set di dati a disposizione e dal tipo di attività, pertanto richiede un'abbondante quantità di sperimentazioni.

Dopo aver testato varie combinazioni, gli iperparametri per il training (definiti in fase di implementazione dalla funzione *trainingOptions*), sono stati impostati nel seguente modo:

```
options = trainingOptions('sgdm', ...
    'InitialLearnRate',0.01, ...
    'LearnRateDropFactor',0.5, ...
    'LearnRateDropPeriod',5, ...
    'LearnRateSchedule','piecewise', ...
    'MaxEpochs',50, ...
    'Shuffle','every-epoch', ...
    'Verbose',false, ...
    'MiniBatchSize',50, ...
    'Plots','training-progress');
```

Figura 4.5 Codice MATLAB usato per il tuning degli iperparametri.

Per l'aggiornamento dei pesi della rete è stato usato l'algoritmo di *discesa stocastica del gradiente con momento*, indicato con 'sgdm' (*Stochastic Gradient Descent with Momentum*).

Si tratta di una variante semplificata dell'algoritmo di discesa del gradiente, che differisce da esso nella cardinalità di dati utilizzati per valutare il gradiente della funzione di perdita, prima di effettuare l'aggiornamento dei parametri; invece di calcolare il gradiente utilizzando tutti i dati a disposizione, ne fa una stima prendendo in considerazione solo un sottoinsieme (mini-batch) di esempi del training set.

Nel nostro caso si è impostata la dimensione del mini-batch a 50 ('*MiniBatchSize*'); significa che il numero di sample propagati attraverso la rete per calcolare il gradiente e aggiornare i pesi in una singola iterazione sono 50.

Il principale problema del normale approccio di discesa del gradiente è che non è scalabile, ovvero nel caso si debba processare una grande quantità di dati esso richiede lunghi tempi per il calcolo, diventando inefficiente. Può quindi risultare conveniente elaborare piccole quantità di dati per volta, anzi questa diventa una scelta obbligata qualora le dimensioni del training set siano tali da non poter essere gestite dalla memoria.

L'appellativo 'stocastico' deriva dal fatto che gli aggiornamenti dei parametri calcolati utilizzando un mini-batch sono una stima rumorosa dell'aggiornamento che risulterebbe dall'utilizzo del set di dati completo.

Se il training set è abbastanza eterogeneo ed è ordinato in maniera randomica, il mini-batch comunque approssima abbastanza bene l'intero training set e di conseguenza l'approssimazione è più verosimile al calcolo dell'intero gradiente. Questo metodo permette quindi di ottenere un buon compromesso accuratezza-velocità di calcolo.

L'aggiunta del *momento* all'aggiornamento dei parametri è un metodo che aiuta ad accelerare la discesa del gradiente verso la direzione corretta, riducendo l'effetto di eventuali oscillazioni e cambi di direzione repentini. Di default il momento è stato impostato a 0.9.

Il parametro '*MaxEpochs*', controlla il numero massimo di epoche, dove un'epoca indica il passaggio di tutto il training set nella rete.

L'addestramento si compie epoca dopo epoca finché non viene raggiunta una condizione di arresto, che idealmente è l'uguaglianza tra output predetto e reale, ma più tipicamente è appunto un numero predefinito di epoche.

Nella scelta di questo numero c'è il rischio di sceglierne uno troppo basso e quindi andare in underfitting oppure, dall'altra parte settare un valore eccessivo che può condurre il modello ad adattarsi eccessivamente ai dati di training, andando incontro al problema di overfitting [31].

Nello specifico caso, si è impostato un numero massimo di epoche per l'allenamento pari a 50.

È necessario che ad ogni epoca dell'addestramento i dati vengono rimescolati. Questo perché se la dimensione del mini-batch non fosse un sottomultiplo del numero di immagini del training set alcune immagini verrebbero sistematicamente escluse dalla fase di training. L'opzione '*shuffle*', impostata su '*every-epoch*' previene questa eventualità.

La '*LearnRateSchedule*' è un'opzione che permette di regolare il tasso di apprendimento, ovvero la velocità con cui l'algoritmo aggiorna i parametri.

È bene ribadire che la scelta del *learning rate* risulta fondamentale ai fini dell'efficienza dell'algoritmo di addestramento: un learning rate troppo grande potrebbe rendere instabile l'algoritmo, causando fluttuazioni attorno al punto di minimo o portando addirittura alla divergenza; mentre un learning rate troppo

CAPITOLO 4. ALGORITMO PER L'ANALISI AUTOMATICA

piccolo potrebbe rallentare parecchio il processo di convergenza, rendendo necessario un numero di epoche più elevato per raggiungere il punto di minimo.

Non è detto che il learning rate debba essere un parametro fisso durante tutta la fase di training. È anzi una pratica comune durante l'addestramento quella di far diminuire il suo valore ad intervalli regolari.

Anche nel nostro caso è stato impostato (tramite l'opzione *'piecewise'*) un learning rate variabile lungo il processo di elaborazione. Si parte da un valore iniziale 0.01 (*'InitialLearnRate'*) e viene decrementato ogni 5 epoche (*'LearnRateDropPeriod'*) di un fattore pari a 0.5 (*'LearnRateDropFactor'*). In questo modo l'algoritmo durante il training cambia progressivamente i pesi di un valore più piccolo avvicinandosi al punto di minimo, affinando così sempre più le capacità di apprendimento della rete.

Durante la fase di sperimentazione sono stati effettuati ripetuti cicli di training provando differenti combinazioni di iperparametri; quella appena presentata è risultata essere la migliore per gestire il problema, ovvero quella che ha consentito di ottimizzare la capacità di apprendimento e ottenere le migliori prestazioni in termini di accuratezza.

Per quanto riguarda i parametri qualitativi delle prestazioni della *CNN*, è possibile visualizzare e monitorare l'avanzamento del training (*'training-progress'*) ad ogni iterazione tramite l'istruzione *'plot'*.

Una volta avviato l'allenamento viene generata una figura con due grafici, uno relativo alla *training accuracy* e uno alla *training loss*, definite rispettivamente come l'accuratezza di classificazione e la perdita di energia calcolate su ogni mini-batch del training set.

A titolo di esempio in **Figura 4.6** sono state riportate le curve ottenute in una delle prove effettuate durante la fase di tuning degli iperparametri.

Il normale e corretto andamento di questi grafici dovrebbe essere quello in cui all'aumentare delle iterazioni la *training-accuracy* sale, mentre la funzione di *loss* diminuisce il proprio valore.

Qualora si disponga di un set di validazione, l'ambiente Matlab oltre alle metriche di addestramento permette di visualizzare anche quelle di convalida, consentendo di analizzare visivamente l'eventuale overfitting della rete.

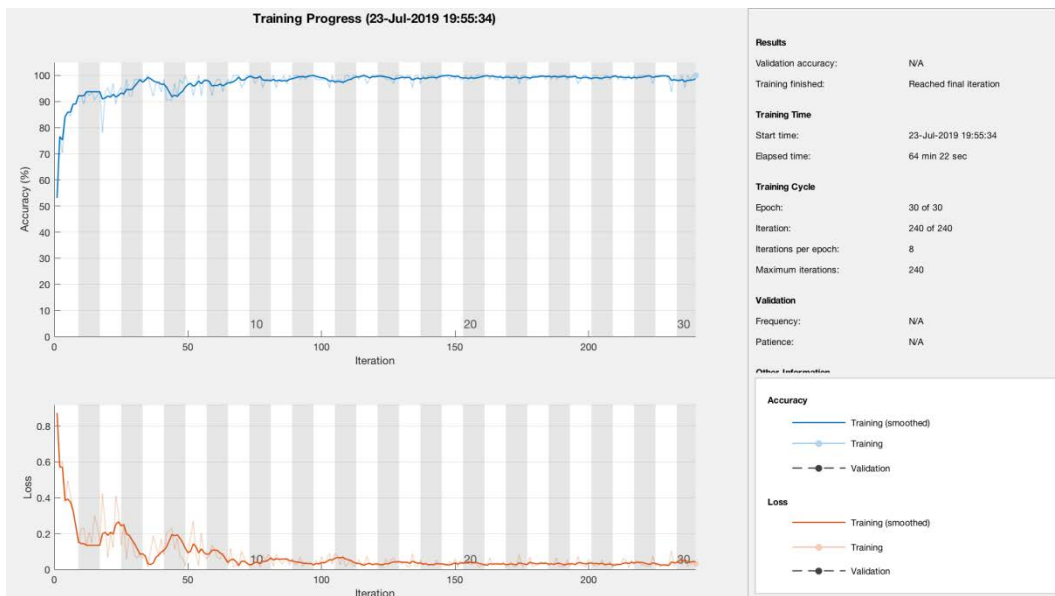


Figura 4.6 Monitoraggio progressi dell'addestramento della CNN.

4.2.4 Fase di testing della CNN

Dopo aver concluso la fase di training è importante avere una misura dell'efficacia della rete, ovvero avere informazioni su quanto bene ha imparato e con quale precisione essa riesca a svolgere il compito per cui è stata realizzata.

Distinguiamo due tipi di accuratezza:

- *accuratezza sul training set*: precisione con la quale la rete riesce ad effettuare predizioni sui dati che ha analizzato in fase di training;
- *accuratezza sul test set*: la precisione con la quale la rete riesce ad effettuare predizioni su dati mai visti prima, utilizzati come insieme di test;

Il *test set* è stato usato per avere una valutazione obiettiva del modello finale della CNN addestrata con il *training set*. Al sistema è stato fornito un insieme di immagini ad esso sconosciute (di cui si conosce l'etichetta) e si è valutata l'accuratezza in termini di percentuale di risposte corrette.

Per verificare la bontà effettiva del modello in termini di generalizzazione le predizioni prodotte utilizzando i dati del training set sono state confrontate con quelle fornite utilizzando i dati del test set. Questo perché, in generale, se le

performance su training set e test set sono molto diverse, potrebbe significare che nel training c'è overfitting.

4.3 U-NET – segmentazione disco e cup

L'analisi proposta, completamente automatizzata, si articola in tre fasi:

1. In un primo momento viene realizzata la segmentazione del disco ottico (se presente) in tutti i frame di un video classificati come 'buoni';
2. Successivamente, dopo aver individuato i frame in cui il disco ottico è interamente visibile, sono selezionati tra questi i tre che presentano il miglior contrasto;
3. Infine viene segmentata la cup.

Gli algoritmi di segmentazione semantica del disco ottico e della cup si basano su una rete convoluzionale con architettura a forma di U (U-Net [39-45]), un modello convenzionale per questo tipo di attività. Più precisamente sono state implementate in maniera distinta due reti convoluzionali, una per la segmentazione del disco ottico e una per la segmentazione della cup, che differiscono per dimensione dell'input e settaggio degli iperparametri.

4.3.1 Segmentazione del disco ottico

L'algoritmo proposto analizza ogni fotogramma 'buono' di un video. Per il training set sono state utilizzate immagini campionate casualmente tra le 1000 selezionate manualmente dal *subset 2*, tralasciando il 20% di questo insieme come set di validazione.

In fase di pre-processing i fotogrammi vengono ridimensionati a 512×1024 pixel tramite l'applicazione di una funzione di trasformazione bi-cubica.

Le CNN sono note per la necessità di un set di dati di grandi dimensioni: per questo motivo, con lo scopo di aumentare la numerosità campionaria a disposizione e incrementare la significatività statistica dei dati, è stata utilizzata una tecnica di data augmentation, che consiste nel capovolgere orizzontalmente ciascuna immagine.

L'architettura della *U-net* [39-45] addestrata per la segmentazione del disco ottico è strutturata in due parti:

4.3 U-NET-SEGMENTAZIONE DISCO E CUP

- *Encoder*: un percorso di contrazione lungo il quale la dimensione spaziale dell'immagine si riduce gradualmente, mentre la profondità aumenta; viene utilizzato per l'estrazione di feature utili all'apprendimento del contesto dell'immagine;
- *Decoder*: un percorso di espansione simmetrica che consente di ottenere una classificazione di etichette su ogni singolo pixel (in particolare, le nostre etichette sono "disco" e "sfondo"). Lungo questo percorso la dimensione spaziale dell'immagine aumenta gradualmente e la profondità diminuisce.

La parte contraente prende in input un'immagine 512×1024 (ovvero il frame pre-elaborato), applica una prima convoluzione 3×3 con passo 2 (che sotto-campiona l'immagine) e, alla fine, presenta cinque blocchi.

Ogni blocco è costituito da due unità convoluzionali (*Convolutional Layer 3×3 , Batch Normalization Layer, ReLU Layer*) seguite da un *Max Pooling Layer 2×2* (per sottocampionare l'immagine).

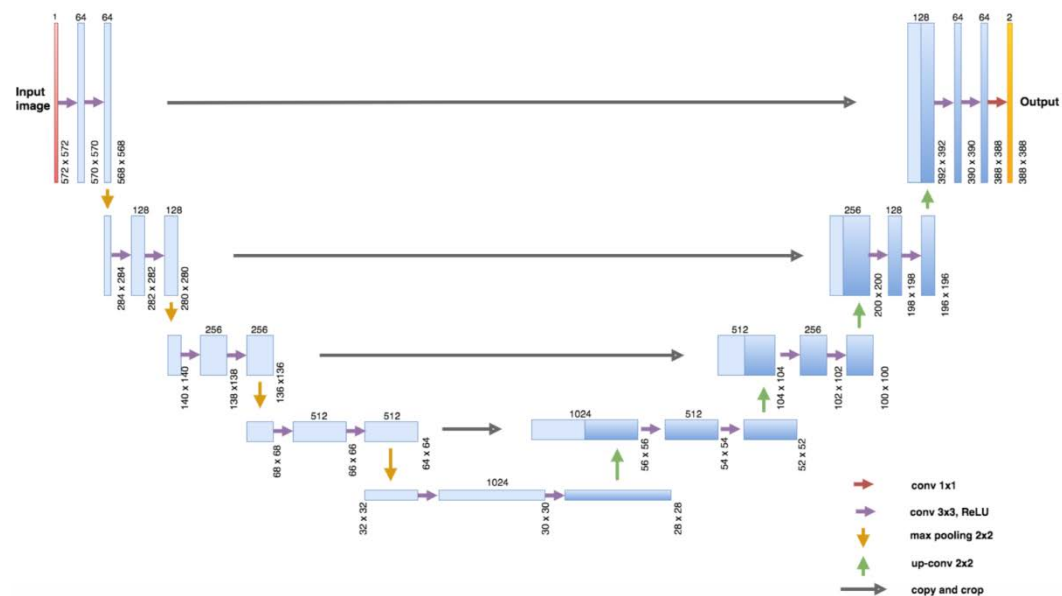


Figura 4.7 Esempio di Architettura U-Net.

La parte di espansione comprende anch'essa cinque blocchi, ciascuno dei quali concatena l'immagine sovra-campionata con quella corrispondente nella parte

dell'encoder e poi applica due unità convoluzionali (*Convolutional Layer 3x3*, *Batch Normalization Layer*, *ReLU Layer*).

Dopo i cinque blocchi, è stata implementata un'altra convoluzione trasposta, al fine di tornare alle dimensioni iniziali, e una convoluzione 1×1 .

Per quanto riguarda gli iperparametri, la rete è stata addestrata con un learning rate di 0,01, mini-batch di 20 immagini per 5 epoche e i dati del training set sono stati mescolati prima di ogni epoca di allenamento.

Poiché le classi sono sbilanciate (i pixel del disco ottico sono sottorappresentati rispetto ai pixel dello sfondo), sono state bilanciate usando come peso l'inverso della frequenza della classe.

4.3.2 Selezione dei frame

Per la selezione dei frame migliori sono stati utilizzati i seguenti criteri:

- dimensione dell'area del disco ottico segmentato;
- contrasto della regione corrispondente nell'immagine originale.

Dopo aver calcolato l'area più grande (*maxArea*) si procede ad una ulteriore analisi di tutti quei frame in cui la dimensione del disco ottico segmentato è maggiore di $0,9 * maxArea$.

Per ciascuno di questi frame, viene calcolata poi l'energia del gradiente (GRAE) [46], definita come la grandezza media delle derivate parziali dell'intensità dei pixel lungo le direzioni x e y .

Infine il GRAE è utilizzato come indice per la selezione dei 3 frame migliori (ovvero i 3 fotogrammi con area del disco ottico segmentato più grande e con GRAE maggiore).

Si è deciso di selezionare i 3 frame migliori anziché un singolo frame per rendere l'algoritmo robusto ad eventuali errori nella selezione di un singolo frame.

4.3.3 Segmentazione della cup

La segmentazione automatica della cup è eseguita nei tre migliori frame precedentemente selezionati.

Per ogni fotogramma, viene considerata la regione dell'immagine originale corrispondente al disco ottico segmentato, ridimensionata a 256×256 pixel.

Per l'addestramento, non avendo a disposizione abbastanza immagini acquisite con D-EYE, sono state utilizzate le immagini del *dataset 2*, un dataset pubblico di immagini retiniche acquisite con normali fundus camera ed analizzate manualmente da esperti oftalmologi. Per ogni immagine è stata considerata solo la parte relativa al disco ottico: la regione che comprende il disco è stata ritagliata, lasciando un piccolo bordo. Inoltre, al fine di sfocare le immagini e renderle simili (stesso GRAE, intervallo [2.8-5.2]) a quelle del *dataset 1*, acquisite con l'oftalmoscopia dello smartphone, è stato applicato un filtro gaussiano.

Delle 6 segmentazioni manuali, è stata considerata quella effettuata dal primo esperto di glaucoma (risultati analoghi nel training sono stati ottenuti considerando le segmentazioni effettuate da ciascuno degli altri esperti di glaucoma).

La U-Net utilizzata per la segmentazione della cup ha una architettura simile a quella descritta precedentemente per il disco ottico.

In questo caso, la struttura consiste in una parte di contrazione di cinque blocchi. Ogni blocco è costituito, come la rete precedente, da due unità convoluzionali (*Convolutional Layer 3x3*, *Batch Normalization Layer*, *ReLU Layer*), seguite da uno strato di *Max Pooling 2x2* (per sottocampionare l'immagine).

La parte di espansione è composta anch'essa da cinque blocchi, ognuno dei quali concatena l'immagine sovra-campionata con quella corrispondente nella parte dell'encoder e poi applica due unità convoluzionali (*Convolutional Layer 3x3*, *Batch Normalization Layer*, *ReLU Layer*).

I parametri degli strati convoluzionali e completamente connessi sono stati inizializzati in modo casuale con una distribuzione gaussiana a media zero.

Per quanto riguarda gli iperparametri, sono state provate diverse combinazioni di valori: alla fine è stato adottato come ottimizzatore la discesa stocastica del gradiente con l'aggiunta di un momento pari a 0,9 e in fase di training abbiamo eseguito 100 epoche. La dimensione del batch è stata impostata a 20 e il learning rate è stato inizialmente impostato a 0,01 e poi diviso per 2 ogni 75 epoche.

Capitolo 5

Risultati

In questo capitolo verranno riportati i principali risultati conseguiti nelle varie analisi. Verrà presentata in un primo momento una misura, in termini di accuratezza, delle prestazioni dell'algoritmo di *classificazione*; in particolare verranno messi a confronto i risultati ottenuti nel training set e nel test set.

Successivamente verranno ripercorsi i vari step previsti dall'algoritmo di *segmentazione semantica*; è bene ribadire che in questo caso la metrica di valutazione delle performance è basata sulla classificazione dei singoli pixel.

Per verificare l'efficienza dell'addestramento delle U-net, in aggiunta all'accuratezza, sono stati presi in considerazione anche gli indici statistici di sensibilità e specificità.

In un generico problema di classificazione binaria, l'insieme dei dati da classificare è suddiviso in due categorie, che possiamo indicare in modo convenzionale come *positivi (P)* e *negativi (N)*.

Le informazioni della classificazione, ovvero i valori predetti dal modello, possono essere riassunti in una tabella a doppia entrata, come quella in *Figura 5.1*, che prende il nome di *matrice di confusione*.

		CLASSE PREDETTA	
		+	-
CLASSE REALE	+	<i>TruePositive (TP)</i>	<i>FalseNegative (FN)</i>
	-	<i>FalsePositive (FP)</i>	<i>TrueNegative (TN)</i>

Figura 5.1 Matrice di confusione per classificatore binario.

Questa tabella raccoglie il numero di:

- *Veri Positivi (TP)*: predizioni positive corrette (istanze positive classificate correttamente come positive);
- *Falsi Positivi (FP)*: predizioni positive errate (istanze negative classificate dal modello erroneamente come positive);
- *Veri Negativi (TN)*: predizioni negative corrette (istanze negative classificate correttamente come negative);
- *Falsi Negativi (FN)*: predizioni negative errate (istanze positive classificate erroneamente dal modello come negative).

A partire da questi valori, si possono ricavare i tre indici di affidabilità della classificazione sopra citati: *accuratezza*, *sensibilità*, *specificità*.

L'*accuratezza* della classificazione è definita come il numero di istanze classificate correttamente (somma di veri positivi e veri negativi) rispetto al numero totale di istanze classificate:

$$Accuratezza = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.1)$$

L'*accuratezza* è determinata, quindi dalla proporzione di falsi positivi e negativi; tanto più basse sono le loro quote, tanto più il classificatore è accurato.

La *sensibilità* misura la percentuale di veri positivi rispetto al numero totale di positivi effettivi (somma di veri positivi e falsi negativi):

$$Sensibilità = \frac{TP}{TP + FN} \quad (5.2)$$

La *specificità* esprime la frazione di veri negativi rispetto al numero totale di negativi effettivi (somma di veri negativi e falsi positivi):

$$Specificità = \frac{TN}{TN + FP} \quad (5.3)$$

La sensibilità è influenzata dalla quota di falsi negativi; un classificatore è tanto più sensibile quanto minore è la quota di falsi negativi. Specularmente la specificità è condizionata dal numero di falsi positivi; un modello è tanto più specifico quanto più bassa risulta la frazione di falsi positivi.

5.1 Selezione frame

I risultati ottenuti dall’algoritmo di classificazione sono stati valutati quantitativamente sia sui dati del training set che su quelli del test set, con l’obiettivo di capire se la rete addestrata sia in grado di generalizzare bene producendo risultati simili in entrambi i set di dati.

In una breve sintesi, per questa analisi è stato utilizzato il *subset 1* suddiviso nel seguente modo:

- *Training set*: la rete è stata allenata su 1751 immagini. Di queste, 804 rappresentano esempi di frame “buoni” e 947 costituiscono campioni di frame ‘cattivi’;
- *Test set*: le prestazioni del modello prodotto in fase di training sono state valutate su 122 frame, dei quali 25 contengono la retina, per cui sono “buoni” ed i rimanenti 97 sono frame di scarsa utilità connotati come “cattivi”.

Per valutare l’effettiva capacità della rete di generalizzare, la porzione di immagini che compone il test set non comprende esempi utilizzati in fase di training. Si tratta di insiemi di dati entrambi rappresentativi ai fini della classificazione, ma del tutto disgiunti. Nella *Tabella 5.1* vengono riassunti i valori di accuratezza ottenuti in fase di training e implementando sul test set l’algoritmo addestrato.

	Accuratezza
Training Set	97 %
Test Set	83 %

Tabella 5.1 Risultati ottenuti nel training e nel test della CNN implementata per la classificazione di frame.

La rete riesce a riconoscere correttamente la quasi totalità dei dati di addestramento, raggiungendo un'accuratezza del 97%. In fase di test le prestazioni dell'algoritmo confermano l'efficacia del training, con una percentuale di frame correttamente classificati pari all'83%. Sebbene la rete sia in grado di generalizzare bene ed i risultati siano molto buoni, c'è ancora margine di miglioramento. La disponibilità di un dataset più ampio, caratterizzato da una maggiore variabilità di dati, con molta probabilità porterebbe beneficio all'addestramento e di conseguenza anche le prestazioni potrebbero migliorare.

I risultati del test finale sono riportati in extenso nell'*appendice A*. È stata allegata una tabella, nella quale in corrispondenza di ciascun frame (definito nella prima colonna) vengono presentati, rispettivamente nella seconda e nella terza colonna, il valore atteso ed il valore predetto dal modello. L'ultima colonna si riferisce al massimo tra i due valori forniti in output dal *Softmax Layer*, vale a dire la probabilità con cui l'immagine è stata associata ad una delle due categorie 'buoni' o 'cattivi'.

5.2 Segmentazione disco e cup

5.2.1 Segmentazione del disco ottico

L'algoritmo proposto ha raggiunto un'elevata precisione nella segmentazione del disco ottico. L'accuratezza della validazione è stata del 98%, ottenuta confrontando l'analisi automatizzata e manuale su 200 fotogrammi estrapolati dai video del *subset 2 (validation set)*. È stata, invece, ottenuta un'accuratezza pari al 99% confrontando l'analisi automatizzata e manuale sul miglior frame di ciascuno dei 10 video del *subset 3 (test set)*. Nella **Tabella 5.2** sono stati riportati anche i valori di sensibilità e specificità.

	200 frame del <i>subset 2</i>	10 frame del <i>subset 3</i>
Accuratezza	98 %	99 %
Sensibilità	95 %	95 %
Specificità	98 %	99 %

Tabella 5.2 Risultati ottenuti sul validation set (subset 2) e test set (subset 3).

5.2.2 Selezione dei frame

Le ispezioni visive hanno rivelato che tutti i frame selezionati automaticamente, dai video del *subset 2* e del *subset 3*, sono caratterizzati dalla presenza dell'intero disco ottico e da un elevato contrasto. Inoltre, il miglior frame selezionato manualmente da ciascun video del *subset 3* rientrava sempre nei 3 fotogrammi selezionati automaticamente.

5.2.3 Segmentazione della cup

L'accuratezza della segmentazione della cup è del 96 %, ottenuta confrontando la segmentazione manuale e automatizzata nel miglior fotogramma su ciascun video del *subset 3* (*Tabella 5.3*).

	10 frame migliori del <i>subset 3</i>
Accuratezza	96 %
Sensibilità	78 %
Specificità	98 %

Tabella 5.3 Risultati ottenuti sul test set (subset 3).

5.3 Glaucoma

A partire da disco e cup segmentati può essere facilmente derivato il *Vertical Cup-to-Disc Ratio (VCDR)*, un importante indice per lo screening ed il follow-up dei pazienti affetti da glaucoma.

Abbiamo già discusso nel capitolo 1 di come, in un paziente glaucomatoso, si assista ad una progressiva perdita delle fibre nervose retiniche con conseguente aumento dell'escavazione papillare. Uno degli indici utili a quantificare l'entità di questo danno è appunto il rapporto cup-to-disc, cioè il rapporto tra il diametro dell'escavazione e diametro totale del disco ottico.

La papilla danneggiata dal glaucoma soprattutto nelle fasi precoci della malattia presenta un'escavazione di dimensioni maggiori in direzione verticale.

La **Figura 5.2** mostra esempi rappresentativi della segmentazione automatica del disco ottico e della cup in un soggetto sano e in un soggetto con glaucoma.

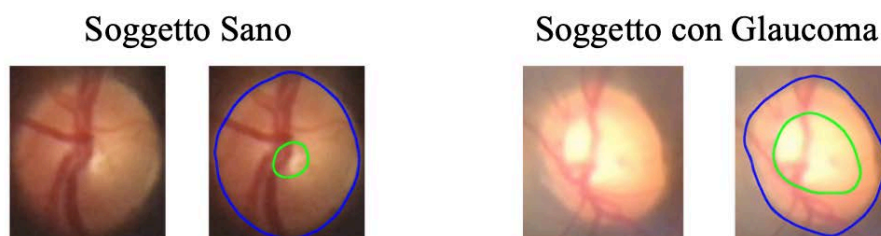


Figura 5.2 Esempi rappresentativi di segmentazione automatica di disco ottico e cup.

L'escavazione della papilla ottica è ben visualizzabile all'esame oftalmoscopico del fundus oculare.

Nonostante la qualità delle immagini acquisite con l'oftalmoscopia smartphone sia inferiore a quella delle immagini acquisite con le tradizionali fundus camera, uno studio precedente ha dimostrato come oftalmologi esperti possano calcolare il rapporto cup-to-disc anche dalle immagini catturate da oftalmoscopi basati su smartphone [50].

L'algoritmo implementato in questo progetto prevede una valutazione finale per ogni soggetto, che considera i VCDR medi ottenuti dai 3 frame selezionati automaticamente.

I VCDR rilevati sono ben correlati con quelli ottenuti manualmente (differenza media dell'8%). I VCDR automatizzati sono ≤ 0.37 per soggetti sani e ≥ 0.44 per soggetti con glaucoma (*Tabella 5.4*).

Risultati analoghi sono stati ottenuti considerando il rapporto orizzontale cup-to-disc (non riportati).

Questi dati indicano una differenza consistente tra gli indici ottenuti da soggetti sani e patologici. Questa differenza consente di distinguere chiaramente i 5 soggetti con glaucoma.

	Soggetti sani					Soggetti con glaucoma				
ID soggetto	1	2	3	4	5	6	7	8	9	10
VCDR Automatico	0.21	0.20	0.30	0.23	0.37	0.46	0.44	0.57	0.58	0.51
VCDR Manuale	0.21	0.22	0.30	0.24	0.33	0.57	0.49	0.51	0.58	0.60
% differenza	0.0	9.1	0.0	4.2	12.1	19.3	10.2	11.8	0.0	15.0

Tabella 5.4 Calcolo automatico e manuale del VCDR, ottenuto su ciascun soggetto del subset 3

Capitolo 6

Conclusioni

L'esame del fondo oculare tramite oftalmoscopio consente di visualizzare la retina, membrana che riveste l'interno del bulbo oculare, ed in particolare disco ottico, macula e fovea. Grazie a questo tipo di esame, semplice e non invasivo, è possibile diagnosticare patologie che riguardano la retina stessa o sistemiche, che riguardano quindi l'intero organismo. Tra le patologie che possono affliggere la retina, il glaucoma è la terza causa principale di cecità irreversibile e una delle principali cause di disabilità visiva in tutto il mondo; colpisce nella globalità circa *80 milioni* di persone. La malattia nelle fasi iniziali è asintomatica, pertanto più del *50 %* degli individui non entra in un iter diagnostico fino a quando il glaucoma raggiunge gli stadi avanzati [51]. In quest'ottica, perseguire uno screening diffuso è fondamentale per la diagnosi precoce e il trattamento di tale patologia, riducendo l'incidenza della cecità ad essa associata. L'oftalmoscopia smartphone rappresenta una risorsa cruciale per l'individuazione precoce di varie patologie, ed in particolare delle forme di glaucoma asintomatico all'esordio, per le quali una terapia tempestiva e mirata consentirebbe di limitare in modo significativo il numero di casi di malattia agli stadi avanzati.

In questa tesi è stata sviluppata un'analisi completamente automatizzata di video del fundus retinico acquisiti mediante l'oftalmoscopio digitale D-EYE, con focus incentrato sulla patologia glaucomatosa. L'algoritmo proposto, basato sull'utilizzo di reti neurali convoluzionali, seleziona i frame di interesse clinico ed esegue la segmentazione del disco ottico. Uno studio preliminare è stato fatto per la segmentazione della cup ed il conseguente calcolo del rapporto cup-to-disc verticale (VCDR), fornendo informazioni utili sul glaucoma.

Sono stati ottenuti buoni risultati ad ogni fase dell'analisi, raggiungendo una buona accuratezza sia nella classificazione dei frame di interesse clinico che nella

CAPITOLO 6. CONCLUSIONI

segmentazione del disco ottico e della cup; inoltre i VCDR calcolati automaticamente hanno indicato una differenza consistente tra soggetti sani e soggetti con glaucoma.

Al termine della sperimentazione è emerso come i rapporti cup-to-disc derivati sia manualmente che automaticamente dai video del dataset a disposizione fossero sovrastimati. Questo effetto, osservato anche in uno studio precedente [50] è probabilmente conseguenza di una saturazione eccessiva dell'immagine acquisita intorno alla cup (molto più luminosa rispetto al bordo), che la fa apparire più grande.

I risultati ottenuti dall'algoritmo proposto dimostrano che le reti neurali convoluzionali possono fornire un'analisi automatizzata affidabile e incoraggiano un ulteriore sviluppo del metodo proposto, con ottimizzazione delle reti neurali e con relativa sperimentazione su un set di dati più ampio.

Un'analisi automatizzata affidabile del video del fundus oculi può supportare l'uso dell'oftalmoscopia smartphone anche da parte di oftalmologi non esperti, come medici generici o paramedici, in un contesto di telemedicina [52-53]. Inoltre, un algoritmo del tutto automatico potrebbe fornire informazioni real time sulla qualità del video acquisito, dati clinici preliminari e consentire l'analisi remota di pochi frame anziché dell'intero video. Questo offrirebbe un notevole supporto durante l'esame oftalmico ed una tangibile possibilità di estendere servizi assistenziali di alta specializzazione ad una popolazione più ampia, in particolare nei paesi in via di sviluppo che presentano spesso condizioni economiche non sufficienti a garantire equità di accesso alle prestazioni sanitarie.

Bibliografia

1. https://www.d-eyecare.com/it_IT.
2. Arnalich-Montiel F, Muñoz-Negrete FJ, Rebolleda G, et al. *Cup-to-disc ratio: agreement between slit-lamp indirect ophthalmoscopic estimation and stratus optical coherence tomography measurement*. *Eye Lond Engl*. 2007; 21:1041–1049.
3. Biswas M, Kuppili V, Saba L, et al. *State-of-the-art review on deep learning in medical imaging*. *Frontiers In Bioscience*. 24:392-426, 2019.
4. Grewal PS, Oloumi F, Rubin U, Tennant MTS. *Deep learning in ophthalmology: a review*. *Canadian Journal of Ophthalmology*. 2018; 53(4):309-313.
5. Schmidt-Erfurth U, Sadeghipour A, Gerendas B, et al. *Artificial intelligence in retina*. *Progress in Retinal and Eye Research*. 2019; 67:1-29.
6. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, Tan GSW, Schmetterer L, Keane PA, Wong TY. *Artificial intelligence and deep learning in ophthalmology*. *British Journal of Ophthalmology*. 2019; 103(2):167-175.
7. Young B, Lowe JS, Stevens A, Heath JH. *Wheater's Functional Histology: A Text and Colour Atlas*. 5th edition. 2006.
8. Eagle RC. *Eye Pathology: An Atlas and Text*. 3rd edition. 2017.
9. American Academy of Ophthalmology. *Fundamentals and Principles of Ophthalmology*. Section 2. 2015-2016.
10. Sihota R, Tandon R. *Parsons' Diseases of the Eye*. 22nd Edition. Elsevier, 2015.
11. Netter FH. *Atlas of Human Anatomy*. 7th edition. Elsevier, 2018.
12. Mehta S. *Age-Related Macular Degeneration*. Primary Care: Clinics in Office Practice. 2015.
13. Jameson JL, Fauci AS, Kasper DL, Hauser SL, Longo DL, Loscalzo J. *Harrison's Principles of Internal Medicine*. 20th edition. McGraw-Hill Higher Education, 2018.

BIBLIOGRAFIA

14. Harjasouliha A, Raiji V, Garcia Gonzalez JM. *Review of hypertensive retinopathy*. Disease-a-Month. 2017; 63(3):63-69.
15. Midena E. *Malattie dell'Apparato Visivo*. CEDAM, Padova, 2006.
16. Cheung N., Mitchell P, Wong TY, *Diabetic retinopathy*. Lancet. 2010; 10(376):124-36.
17. Yau JW, Rogers SL, Lamoureux EL, et al. *Global prevalence and major risk factors of diabetic retinopathy*. Diabetes Care. 2012; 35(3):556-64.
18. Pearce JM. *The Ophthalmoscope: Helmholtz's Augenspiegel*. Eur Neurol. 2009, 61:244–249.
19. Nguyen CL, Wayenborgh JL. *Hermann von Helmholtz: The ophthalmoscope and some of his other contributions to ophthalmology*, Hist Ophthal Intern, 2015; 1: 165-177.
20. McMullen WH. *The evolution of the ophthalmoscope*. British Journal of Ophthalmology. 1917; 1(10):593-599.
21. Heitz RF. *The Earliest Visualizations of the Living Eye's Fundus by Immersion in Water*. Strasbourg, France, Archiwum Historii I Filozofii Medycyny. 2012; 75: 11-15.
22. Prévost B. 1810. *Considerations sur le brillant des yeux du chat et de quelques autres animaux*. Bibliothèque britannique. 11:196-211.
23. Hormière. *Instruments d'optique ophthalmique*. Editions TEC&DOC, Lavoisier. 2010.
24. Ravin JG. *Sesquicentennial of the ophthalmoscope*. Arch Ophthalmol. 1999; 117: 1634-1638.
25. Keeler CR. *The Ophthalmoscope in the Lifetime of Hermann von Helmholtz*, Arch Phenolphthalein. 2002, 120:194-201.
26. Jain AB, Prakash VJ, Bhende M. *Techniques of Fundus Imaging*. Scientific Journal of Medical & Vision Research Foundations. 2015; XXXIII:100–107.
27. Lucente A. *Evoluzione della fotografia retinica e Imaging Widefield*, oftalmologiadomani, Anno IX – 2018.
28. MathWorks. *Introducing Deep Learning with MATLAB*. Ebook: https://www.mathworks.com/content/dam/mathworks/tagteam/Objects/d/80879v00_Deep_Learning_ebook.pdf.
29. Pagnotta S. *Applicazioni di Reti Neurali a casi reali*. Eart Sciences Department, Pisa University. 2019.
30. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press. 2016. Ebook: www.deeplearningbook.org.
31. LeCun Y, Bengio Y, Hinton G. *Deep learning*. Nature, 2015; vol 521: 436-444.

32. Hinton GE. *L'apprendimento delle reti artificiali di neuroni*. Le Scienze, 1992.
33. Itamar A, Rose CR, Karnowski TP. *Deep Machine learning-A New Frontier in Artificial Intelligence Research*. The University of Tennessee, USA. IEE Computational Intelligence Magazine, 2010.
34. *Convolutional Neural Networks* (CNNs/ConvNets) <http://cs231n.github.io/convolutional-networks>.
35. Ivars Namatevs, *Deep Convolution Neural Network: Structure, Feature Extraction and Training Information Technology and management Science*. Vol. 20, pp. 40-77, Dicembre 2017.
36. LeCun Y, Bengio Y. *Convolutional Neural Networks for Image, Speech, and Time-Series*. The Handbook of Brain Theory and Neural Networks, M. A. Arbib. Ed. Cambridge, MA: MIT Press, 1995, 255-258.
37. Bengio Y. *Deep Learning Architectures for AI*. Vol.2, No. 1 (2009) 1-127.
38. Singh S, Kaushik S, Vats R, et al. *Semantic Segmentation Using Deep Learning*. IRJET. Volume: 06 Issue: 04| Apr 2019.
39. Ronneberger O, Fischer P, Brox T. *U-Net: convolutional networks for biomedical image segmentation*. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, 9351:234-241. Springer, Cham, 2015.
40. Yamashita R, Nishio M, Do RKG, et al. *Convolutional neural networks: an overview and application in radiology*. Insights Imaging (2018).
41. Litjens G, Kooi T, Bejnordi, BE, et al. *A survey on deep Learning in Medical Image Analysis*. Medical Image Analysis. 2017. Vol. 42: 60-88.
42. Chartran G, Cheng PM, Vorontsov E, et al. *Deep Learning: A Primer for Radiologists*. RadioGraphics. 2017; 37(7): 2013-213.
43. Grewal SG, Oloumi F, Rubin U, et al. *Deep Learning in Ophthalmology: a review*. Canadian Journal of Ophthalmology 2018.
44. Almazroa A, Alodhayb S, Osman E, et al. *Retinal fundus images for glaucoma analysis: the RIGA dataset*. Proc. SPIE 10579, Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications, 105790B, 2018.
45. Colonna A, Scarpa F, Ruggeri A. *Segmentation of Corneal Nerves Using a U-Net-Based Convolutional Neural Network*, In: Stoyanov D. et al. (eds) OMIA 2018, COMPAY 2018. LNCS, 11039. Springer, Cham, 2018.
46. Subbarao M, Choi TS, Nikzad A. *Focusing techniques*. Optical Engineering 1993; 32(11).
47. Bastawrous A, Rono HK, Livingstone IT, et al. *Development and validation of a smartphone-based visual acuity test (peek acuity) for clinical practice and community-based fieldwork*. JAMA Ophthalmol. 2015; 133:930-937.

BIBLIOGRAFIA

48. Russo A, Morescalchi F, Costagliola C, et al. *Comparison of smartphone ophthalmoscopy with slit-lamp biomicroscopy for grading diabetic retinopathy.* *Am J Ophthalmol.* 2015; 159: 360–364.
49. Russo A, Morescalchi F, Costagliola C, et al. *A novel device to exploit the smartphone camera for fundus photography.* *J Ophthalmol.* 2015:823139.
50. Russo A, Mapham W, Turano R, et al. *Comparison of Smartphone Ophthalmoscopy With Slit-Lamp Biomicroscopy for Grading Vertical Cup-to-Disc-Ratio.* *J Glaucoma.* 25(9):777-81, 2016.
51. Quigley HA, Broman AT. *The number of people with glaucoma worldwide in 2010 and 2020.* *British Journal of Ophthalmology.* 2006; 90:262-267.
52. Adam MK, Brady CJ, Flowers AM, et al. *Quality and diagnostic utility of mydriatic smartphone photography: the Smartphone Ophthalmoscopy Reliability Trial.* *Ophthalmic Surg Lasers Imaging Retina.* 46:631–637, 2015.
53. Bifulck E, Fink A, Pedersen D et al. *Smartphone imaging for the ophthalmic examination in primary care.* *Journal of the American Academy of PAs* 2018; 31(8): 34.

Ringraziamenti

Ora finalmente sento di poter mettere la parola 'fine', per farla felicemente seguire dalla parola 'inizio'. Si è concluso questo lungo e faticoso percorso, costellato di scelte e sacrifici, emozioni e soddisfazioni. È stata dura, ma ho coronato il mio sogno. Posso finalmente dire di avercela fatta!

Se ho raggiunto questo traguardo non lo devo solo a me stessa e alla mia tenacia, ma anche a tutte le persone che amo e stimo.

Il ringraziamento più importante e più sentito va alla mia grande e meravigliosa famiglia. Grazie mamma e papà. I miei successi li devo soprattutto a voi.

A te papà perché sei la mia roccia. Grazie per essere sempre dalla mia parte, per tutti i momenti in cui mi hai tirato su il morale. Grazie per avermi insegnato a non arrendermi mai e ad inseguire con determinazione i miei sogni.

A te mamma, perché sei una forza della natura, capace di affrontare ogni ostacolo che la vita ti pone davanti. Grazie per avere una soluzione ad ogni problema e un regalino pronto per ogni momento di sconforto. Per me sei davvero un esempio.

Vi ringrazio per avermi supportata moralmente, ma anche economicamente, senza farmi pesare un centesimo di quello che avete speso per me.

Siete la mia certezza. A voi devo tutto e senza di voi nulla sarebbe stato possibile.

Grazie alle mie sorelle, Marika e Francesca, che mi hanno sempre spronato ad andare avanti ed hanno sopportato tutte le mie ansie per ogni esame della mia carriera universitaria. "Stai tranquilla, che tanto sai tutto". Un po' per paura di deludervi, un po' per convincimento fosse vero, con queste parole mi avete sempre stimolata a dare il meglio di me stessa, grazie!

Un ringraziamento speciale va alla persona che mi completa, Mirko. Grazie per aver sempre creduto in me e per aver trovato il modo di farmi sorridere anche nei momenti più bui di questo percorso. A volta mi basta guardare i tuoi occhi e sentire il tuo forte abbraccio per capire che tutto si sistema. Sei il mio punto di riferimento

e ti sono immensamente grata per tutto l'aiuto che mi hai dato. In fondo questa laurea è anche un po' tua. Questa è la nostra laurea, la nostra conquista.

Grazie ai miei super nipotini, Tobia e Kilian. I vostri disegni, i vostri scherzetti, i bacini, gli abbracci, i fiorellini che mi avete regalato li porterò nel cuore per tutta la vita.

Un grazie anche alla mia cagnolina Pulce per aver studiato insieme a me dal primo all'ultimo esame. Molte volte ho invidiato la sua "vita da cani"!

Ringrazio di cuore la mia più fidata consigliera e compagna di chiacchierate interminabili, Anna. Grazie per essere un'amica leale e sincera, per essere sempre presente a tutte le ore del giorno e della notte rispondendo ad ogni mia richiesta di disperato aiuto. Grazie per tutti i momenti divertenti e per l'energia che insieme sprigioniamo. Senza di te le mie giornate non avrebbero lo stesso valore.

Un grande ringraziamento va alle mie amiche Sara, Valentina e Michela, che occupano un posto speciale nel mio cuore. Siete da sempre le mie più grandi sostenitrici. Grazie perché tutte le volte che ho pensato di mollare mi avete dato la forza ed il coraggio di andare avanti. Siete davvero speciali.

Grazie a Vanessa, perché sei la mia ricarica di energia. I tuoi "stay focused, be positive" sono la mia fonte di ispirazione ogni giorno. Mi trasmetti una forza incredibile.

Ringrazio profondamente Loris. Grazie per tutto quello che mi hai insegnato e per come me lo hai insegnato, per la capacità che hai di mettermi il buonumore e di regalarmi spensieratezza. So che su di te posso e potrò sempre contare, te ne sarò sempre grata. Ti stimo e ti ammiro e... sì, va bene, anche se ormai è un ritornello ricorrente nella tua vita, te lo ripeto: sei il numero uno!

Un immenso grazie alla mia numerosa compagnia di amici: siete le sfumature di colori che dipingono l'arcobaleno della mia vita. Vi voglio davvero un gran bene.

Un ringraziamento particolare al mio nonnino per essere così fiero di me, per capirmi e aspettarmi sempre, nonostante le mie inevitabili, lunghe assenze. Grazie alla mia "ninni", il mio angelo custode che mi ha accompagnata ad ogni esame ed ha vegliato su di me. La tua maglia è il mio inseparabile porta fortuna. Questo grande traguardo lo dedico tutto a te!