

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

Studio di applicazioni di *deep learning* alla
spettroscopia di RM cerebrale *in vivo*

Relatore:
Prof.ssa Claudia Testa

Presentata da:
Francesco Braglia

Anno Accademico 2022/2023

Abstract

La spettroscopia di risonanza magnetica (MRS) è una tecnica di indagine che possiede numerose applicazioni in chimica, biologia e medicina. In ambito clinico, è impiegata come metodo non invasivo per la quantificazione dei metaboliti cerebrali *in vivo*; evidenzia situazioni patologiche che comportano un'alterazione del normale rapporto delle concentrazioni dei metaboliti. Purtroppo, esistono numerose limitazioni che ne ostacolano un utilizzo più diffuso e sistematico.

Fortunatamente, negli ultimi tempi, l'integrazione di tecniche di intelligenza artificiale in supporto alle tradizionali metodologie di analisi della fisica biomedica ha mostrato risultati incoraggianti in termini di qualità dell'*imaging*, minori tempi di acquisizione e maggiore automazione delle procedure.

L'obiettivo di questa tesi è quello di presentare, attraverso l'analisi di tre studi svolti dal 2019 ad oggi, lo stato dell'arte delle applicazioni del *deep learning* alla MRS cerebrale *in vivo*. Si è osservato come, nonostante i risultati soddisfacenti ottenuti, l'impiego delle reti neurali nelle diverse fasi di un'indagine di MRS rimanga oggetto di ricerca scientifica e non sia ancora stato assimilato dalla pratica clinica.

Indice

1	Spettroscopia di risonanza magnetica	6
1.1	<i>Chemical shift</i>	8
1.2	Accoppiamento scalare	9
1.3	Sequenze per MRS	11
1.4	Soppressione dell'acqua e saturazione	12
1.5	Quantificazione	14
1.5.1	Quantificazione relativa	16
1.5.2	Stima della quantificazione assoluta	16
2	<i>Deep learning</i>	17
2.1	<i>Machine learning</i>	17
2.1.1	Dataset di <i>training</i> , validazione e <i>testing</i>	18
2.2	Percettroni e reti neurali	19
2.2.1	<i>Backpropagation</i>	21
2.3	Reti neurali convoluzionali	22
2.3.1	UNet	26
2.4	Reti neurali residuali	27
2.5	<i>Autoencoder</i>	28
3	Applicazioni	30
3.1	Selezione automatizzata del voxel	30
3.1.1	ResNet-50	31
3.1.2	Modellizzazione e implementazione	31
3.1.3	Dataset e procedure	32
3.1.4	Risultati	33
3.1.5	Discussione	34
3.2	<i>Imaging</i> spettroscopico con super-risoluzione	34
3.2.1	D-UNet	35
3.2.2	Dataset, generatore SI e <i>training</i>	36
3.2.3	Procedure	37
3.2.4	Risultati	38

3.2.5	Discussione	40
3.3	Quantificazione relativa dei metaboliti	42
3.3.1	Modellizzazione	42
3.3.2	Dataset	44
3.3.3	Procedure	45
3.3.4	Risultati	45
3.3.5	Discussione	48
4	Conclusioni	50
A	Metodi di <i>upscaling</i>	52
A.1	<i>Zero-filling</i>	52
A.2	Interpolazione bicubica	52
B	Metodi statistici	55
B.1	Coefficiente di determinazione	55
B.2	Errore medio assoluto percentuale	56
	Bibliografia	56

Introduzione

La spettroscopia di risonanza magnetica (MRS) è una tecnica di indagine spettroscopica che si basa sul fenomeno fisico della risonanza magnetica nucleare (NMR).

Nelle neuroscienze trova impiego come metodo analitico non invasivo per la quantificazione dei principali metaboliti cerebrali *in vivo*: creatina (Cr), glutammato (Glu), inositolo (Ins), *myo*-inositolo (mIns), N-acetil-aspartato (NAA), lattato (Lac), glutammina (Gln), fosfocolina (PCho), fosfocreatina (PCr), N-acetil-aspartil-glutammato (NAAG) e acido γ -amminobutirrico (GABA). Essa permette di effettuare indagini *in vivo* sui tessuti del cervello e solitamente viene associata all'*imaging* di risonanza magnetica per la diagnosi e la localizzazione spaziale di patologie cerebrali.

Queste ultime, infatti, sono spesso associate ad alterazioni nel normale rapporto tra le concentrazioni dei metaboliti. Ad esempio, le ricerche sui soggetti affetti da sclerosi multipla hanno evidenziato una diminuzione del rapporto NAA/Cr non solo nelle zone delle lesioni acute e croniche, ma anche nelle porzioni di sostanza bianca apparentemente sana [1].

La MRS, così come le altre metodologie di *imaging*, è stata affiancata negli ultimi tempi dallo sviluppo dell'intelligenza artificiale e in particolare di quella sua branca nota come *deep learning* (DL).

L'interesse per questo settore è testimoniato dalle numerose pubblicazioni, conferenze e review che presentano i risultati delle ricerche più recenti in materia; il motivo è rappresentato dal grande numero di compiti che i modelli DL sono in grado di svolgere. In ambito sanitario, le prime applicazioni per l'*imaging* medico sono relativamente recenti e risalgono al 2012, quando le performance delle reti neurali si sono mostrate migliori di quelle dei metodi tradizionali di computer vision [2]. Oggi, il DL è utilizzato comunemente per la segmentazione automatizzata delle immagini anatomiche, per la diagnosi di patologie, per la classificazione di lesioni (benigne o maligne).

Se MRI, PET e TC hanno potuto godere di significativi vantaggi da questa integrazione, alla MRS, in quanto tecnica utilizzata in maniera meno sistematica e con diverse limitazioni, è stato rivolto meno interesse, nonostante le potenzialità che questa possiede. Alla base di questo lavoro di tesi c'è stata la volontà di esporre lo stato attuale delle

applicazioni del DL alla MRS *in vivo* e valutare quali saranno i possibili sviluppi futuri. La struttura della tesi è riassunta nei punti seguenti:

- Il primo capitolo illustra i parametri e le tecniche che caratterizzano un esperimento di MRS richiamando i principi fisici della risonanza magnetica nucleare. Si elencano le sequenze per l'acquisizione e la soppressione dei segnali e si espone il problema della quantificazione.
- Il secondo capitolo fornisce un quadro generale sul *deep learning*. Dopo una parentesi sul *machine learning*, si affrontano i principali concetti alla base delle reti neurali e si illustrano quelle che compariranno negli esperimenti del capitolo successivo.
- Il terzo capitolo presenta tre studi, condotti rispettivamente negli anni 2020, 2019, 2023, che presentano potenziali applicazioni del DL alle diverse fasi di un esperimento MRS: acquisizione, processing, post-processing. Per ognuno di essi, sono stati riportati i procedimenti e i principali risultati, che includono un confronto con analoghi metodi tradizionali.

Capitolo 1

Spettroscopia di risonanza magnetica

La spettroscopia di risonanza magnetica (MRS) è un metodo analitico che viene impiegato in chimica, biologia e medicina. Nell'ambito delle neuroscienze, sul quale è focalizzato questo lavoro di tesi, costituisce la seconda applicazione più usata della NMR, dopo l'*imaging*.

Il risultato di un esperimento MRS è uno spettro, cioè un grafico che riporta l'intensità del segnale di NMR in funzione della frequenza; in corrispondenza delle frequenze di risonanza, il segnale presenta dei picchi più o meno accentuati. L'ampiezza dell'area sottesa da un picco è correlata alla concentrazione di un particolare metabolita.

Sia per MRI sia per MRS il principio fisico di base è il seguente: applicando un impulso a radiofrequenza e dei gradienti all'interno di un campo magnetico statico, si selezionano i nuclei dotati di una specifica frequenza di risonanza nella regione di interesse e si rileva il segnale emesso durante il rilassamento di questi; attraverso una trasformata di Fourier, si può ricostruire un'immagine spaziale (MRI) oppure una distribuzione dei segnali nel dominio delle frequenze (MRS).

Tuttavia, se nell'*imaging* si riceve il segnale proveniente dai nuclei ^1H delle molecole d'acqua, l'obiettivo della MRS è l'acquisizione del segnale dei nuclei appartenenti alle molecole metaboliche di interesse, che differiscono dall'acqua per il numero e il tipo di atomi presenti e per la geometria. In **Tabella 1.1** si riporta un elenco di nuclei che possono essere utilizzati per esperimenti di NMR in quanto dotati di spin non nullo; il nucleo di prozio ^1H , che ha un'abbondanza relativa e un rapporto giromagnetico elevati, è responsabile del segnale ricevuto in MRI.

In MRS, oltre a ^1H , risultano di interesse anche ^{13}C , ^{19}F , ^{23}Na e, soprattutto, ^{31}P . La ^{31}P MRS consente di reperire informazioni importanti sui processi energetici che si verificano nel cervello, che non sarebbero disponibili con la ^1H MRS; ciò deriva dal fatto che i metaboliti che forniscono energia contengono atomi di fosforo: adenosin-trifosfato (ATP), fosfocreatina (PCr) e varie forme di fosfato inorganico (Pi) [5]. La **Figura 1.1** mostra due spettri cerebrali caratteristici ottenuti con ^1H e ^{31}P MRS. Quest'ultima

Nucleo	Spin	γ (10^7 Hz/T)	Abbondanza isotopica (% in massa)	Abbondanza biologica (% in massa)
^1H	1/2	26.7520	99.985	10
^{13}C	1/2	6.7283	1.10	0.198
^{17}O	5/2	-3.6279	0.048	0.025
^{19}F	1/2	25.181	100	$7 \cdot 10^{-3}$
^{23}Na	3/2	7.08013	100	0.15
^{31}P	1/2	10.841	100	1

Tabella 1.1: Principali nuclei utilizzabili per NMR con relativi spin, rapporto giromagnetico, abbondanza naturale percentuale e abbondanza biologica percentuale [3, 4].

permette di distinguere PCr da Cr e GPC da PCho; con la spettroscopia protonica (utilizzando campi di intensità standard) è invece solamente possibile riconoscere i picchi della creatina totale (tCr) e della colina totale (tCho).

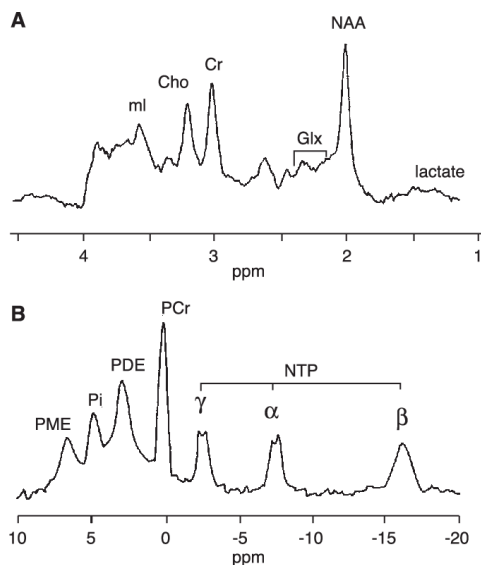


Figura 1.1: Spettro ^1H MRS (in alto) e spettro ^{31}P MRS (in basso). I picchi del secondo sono costituiti da fosfomonoesteri (PME), fosfodiesteri (PDE), fosfocreatina (PCr), fosfati inorganici (Pi) e dai fosfati α -, β - e γ - dei nucleotidi (tra cui l'ATP) [1].

In generale, la MRS è meno sensibile del MRI poiché la concentrazione dei nuclei dei metaboliti (^1H , ^{31}P , ecc.) è diversi ordini di grandezza inferiore rispetto a quella dei protoni dell'acqua.

Come noto, l'intensità del segnale di NMR dipende da diversi fattori:

- Tempi di rilassamento T_1 e T_2 : per un dato nucleo, dipendono dal tessuto considerato, in quanto vengono modificati in presenza di disomogeneità del campo

(fenomeni paramagnetici);

- Numero di nuclei: quanti più nuclei identici sono presenti in una molecola, tanto più intenso è il segnale corrispondente;
- Concentrazione dei nuclei: quanti più nuclei sono presenti nell'unità di volume analizzata (voxel), tanto più intenso è il segnale.

Per interpretare correttamente la natura degli spettri di NMR e trarre informazioni rilevanti per MRS, occorre tenere in considerazione altri due fenomeni che intervengono nel determinare la forma caratteristica del segnale proveniente dalle diverse molecole.

1.1 *Chemical shift*

Il *chemical shift* è il parametro più importante che caratterizza gli spettri di risonanza magnetica. Esso permette infatti di distinguere non solo molecole differenti, ma anche atomi differenti all'interno di una stessa molecola.

Come noto, la frequenza di risonanza di un nucleo è proporzionale all'intensità del campo magnetostatico B_0 percepito dal nucleo. Se questo assumesse lo stesso valore per tutti i nuclei, l'informazione fornita dallo spettro sarebbe poco rilevante, in quanto tutti i nuclei acquisirebbero energia alla stessa frequenza. Tuttavia, proprio il campo magnetico esterno induce delle correnti elettroniche che generano a loro volta un campo magnetico $B_{0\sigma}$ proporzionale a B_0 [6].

Si introduce un campo efficace:

$$B_{eff} = B_0(1 - \sigma) \quad (1.1)$$

dove σ è un parametro di schermatura; la frequenza di Larmor risulta pertanto:

$$\nu = \frac{\gamma B_0(1 - \sigma)}{2\pi} \quad (1.2)$$

Spostamento chimico è il nome del fenomeno di separazione della frequenza di risonanza di Larmor da un'arbitraria frequenza scelta come riferimento. Siccome la schermatura è molto piccola, anche la separazione lo è.

Per evitare la dipendenza dal valore di intensità del campo magnetico utilizzato, lo spostamento chimico è espresso in maniera adimensionale in parti per milione (ppm):

$$\delta = \frac{\nu_s - \nu_{rif}}{\nu_{rif}} \cdot 10^6 \quad (1.3)$$

dove ν_s è la frequenza di risonanza del campione e ν_{rif} è una frequenza di riferimento. Tipicamente, la frequenza di riferimento utilizzata è quella dei protoni del tetrametilsilano (**Figura 1.2**); questi, per via della simmetria della molecola e della bassa elettronegatività del silicio, sono sottoposti a schermature della stessa intensità. Ad esempio, per un esperimento di ^1H MRS a 1.5 T la frequenza di riferimento del TMS è 63.86 MHz.

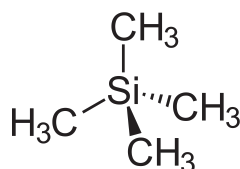


Figura 1.2: Formula di struttura del tetrametilsilano (TMS) [7].

1.2 Accoppiamento scalare

I singoli picchi di risonanza attesi per un atomo si dividono spesso in due o più componenti. Lo splitting dei picchi è conseguenza di un'interazione tra gli spin nucleari mediata dagli elettroni di legame condivisi negli orbitali molecolari. Una trattazione approfondita richiederebbe l'uso della meccanica quantistica, ma è comunque possibile comprendere i concetti fondamentali nel caso più semplice, quello dell'accoppiamento di due nuclei (**Figura 1.3**).

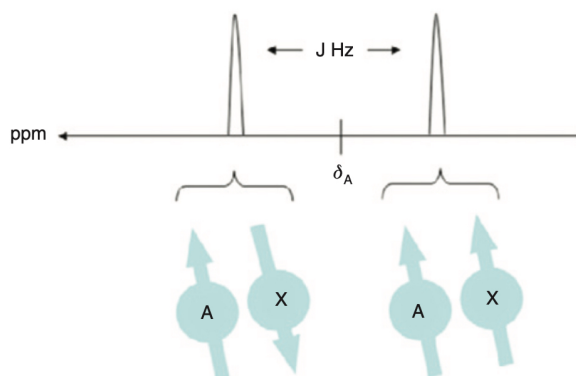


Figura 1.3: Effetto dell'interazione scalare tra due nuclei A e X. Lo stato di spin di X determina l'energia di risonanza di A, pertanto nello spettro si producono due distinti picchi corrispondenti ai due casi possibili. Il parametro J (con le dimensioni di una frequenza) misura la distanza tra i due massimi [8].

Si hanno due nuclei, indicati con A e X. Se non si verificasse alcuna interazione, il picco di risonanza del nucleo A sarebbe uno solo e si troverebbe in una posizione indicata

con δ_A nella scala dello spostamento chimico. Tuttavia, lo spin (up o down) del nucleo X influenza A, in quanto modifica la quantità di energia (di risonanza) che il nucleo deve assorbire per potersi eccitare e dunque anche la sua frequenza di risonanza. Ne derivano così due picchi, che corrispondono alle interazioni di A con i due possibili spin di X; lo scostamento rispetto a δ_A è lo stesso per i due massimi. Siccome le probabilità che lo spin di X sia up oppure down sono uguali, il segnale avrà la stessa intensità in entrambi i casi. Ovviamente, l'interazione è mutua, dunque per il picco di X avviene lo stesso.

Quando si ha a che fare con una molecola organica complessa con più atomi di idrogeno legati ad atomi diversi, è possibile utilizzare delle regole empiriche per stabilire il numero di multipletti attesi per ogni picco. L'accoppiamento si verifica solo per nuclei non equivalenti, vale a dire protoni di gruppi funzionali diversi; il numero di picchi che si presentano è pari a uno più il numero di nuclei (protoni) accoppiati adiacenti. Le intensità relative sono date da una distribuzione binomiale, mentre la distanza tra i picchi è fissata per un multipletto ed è pari a J , detta *costante di accoppiamento* [8, 9]. La **Figura 1.4** riporta un esempio di interazione scalare osservata nella molecola dell'etanolo.

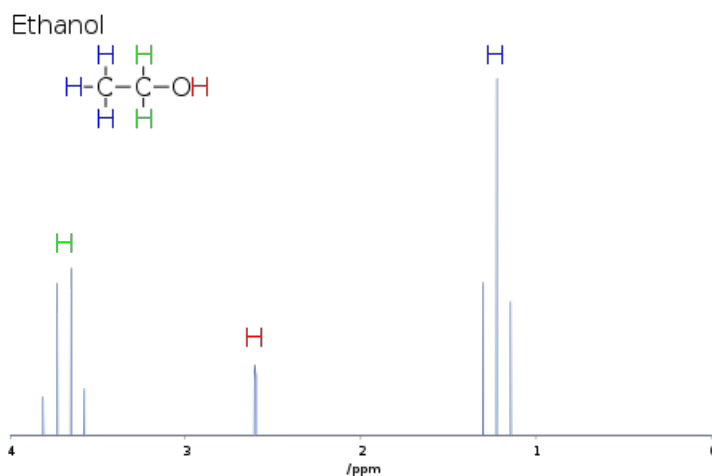


Figura 1.4: Formula di struttura e spettro NMR dell'etanolo (CH₃CH₂OH). I due idrogeni del metilene (-CH₂) interagiscono con i tre del metile (-CH₃), pertanto si genera un quadrupletto di picchi; i tre idrogeni del metile, al contrario, interagiscono con i tre del metilene, dunque si genera un tripletto di picchi. L'idrogeno del gruppo idrossile (-OH) non si accoppia e dunque il suo segnale risulta un singoletto; l'assenza di interazione è dovuta al fatto che il nucleo ¹⁶O possiede spin nullo [9].

1.3 Sequenze per MRS

Come nel caso del MRI, anche per la MRS esistono specifiche sequenze di impulsi e gradienti di campo magnetico per la selezione del voxel di interesse. Le più utilizzate sono la PRESS e la STEAM (**Figura 1.5**).

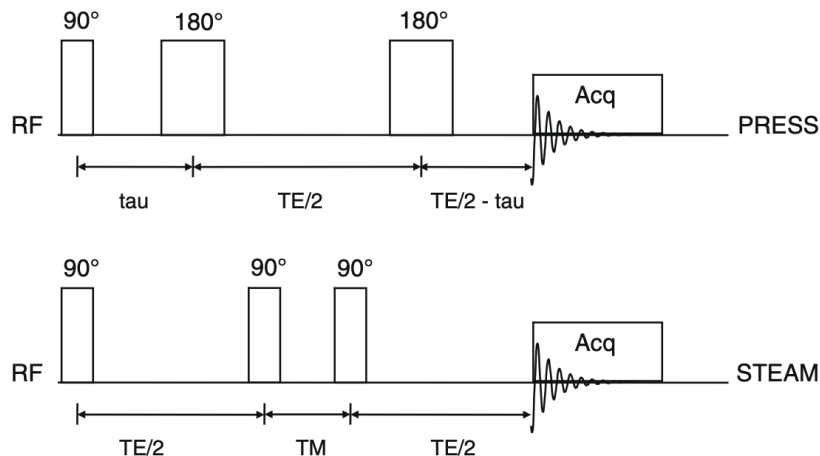


Figura 1.5: Sequenze PRESS (in alto) e STEAM (in basso) [10].

La sequenza PRESS (*point resolved spectroscopy*) consta di tre impulsi a radiofrequenza applicati in successione; contemporaneamente ad ognuno di questi, è applicato un gradiente di selezione spaziale. Il primo impulso, a 90° , è associato ad un gradiente lungo la direzione x , pertanto seleziona una sezione (*slice*) del campione sul piano yz . Il secondo impulso, a 180° , e il gradiente lungo y eccitano una *slab* nel piano xz ; l'intersezione di questa con quella selezionata in precedenza definisce la porzione di interesse, allineata con la direzione z . Proprio lungo la direzione z agisce l'ultimo gradiente, che, insieme al terzo impulso (180°), determina il voxel di interesse (VOI).

La sequenza STEAM (*stimulated echo acquisition mode*) usa tre impulsi selettivi a 90° , ognuno dei quali è applicato contestualmente ad un gradiente. Anche in questo caso, la selezione del VOI avviene per intersezione delle tre *slice* corrispondenti ai tre gradienti. La STEAM è caratterizzata da un rapporto segnale-rumore (SNR) più basso rispetto alla PRESS e oggi è sempre meno utilizzata.

È importante tenere presente che le sequenze per MRS, al contrario di quelle per MRI, non fanno uso di gradienti di codifica in frequenza, in quanto i nuclei precedono già con frequenze diverse per effetto dello spostamento chimico ed è proprio questo fenomeno che permette di discriminare i segnali di molecole diverse [5].

Per ottenere informazioni utili ai fini diagnostici, a seconda dei casi può essere necessario analizzare una piccola regione spaziale oppure mappare un'intera superficie (o

volume). Si delineano quindi due approcci: la spettroscopia *single-voxel* (SVS) e il *chemical shift imaging* (CSI).

La SVS utilizza una delle due sequenze presentate sopra per isolare il segnale proveniente da un solo voxel. Affinché gli spettri siano caratterizzati da SNR accettabili, i voxel devono avere dimensioni sufficientemente grandi, in modo da avere una concentrazione di metaboliti (o, più precisamente, di protoni che contribuiscono ai picchi) apprezzabile [8].

Il CSI, invece, acquisisce simultaneamente il segnale di più voxel differenti. Il nome deriva dal fatto che per ogni voxel viene ricostruito l'intero spettro, che contiene nuclei con diversi *chemical shift*.

Il primo step del CSI consiste nella selezione di una regione di interesse (voxel di eccitazione); successivamente, l'applicazione di opportuni gradienti mappa questa regione, suddividendola in più voxel di risoluzione da ognuno dei quali, in successione, si riceve il segnale. Solitamente, dunque, il CSI è preceduto da una sequenza PRESS o STEAM. Come detto in precedenza, in MRS non è possibile utilizzare gradienti di codifica in frequenza; è necessario perciò impiegare dei gradienti di codifica in fase nelle direzioni spaziali lungo le quali si desidera creare una risoluzione spaziale (2D o 3D).

L'informazione che viene acquisita in un CSI può essere presentata graficamente sovrapponendo gli spettri acquisiti nei voxel con le corrispondenti regioni cerebrali di un'immagine MR (**Figura 1.6**). Per questo motivo, il CSI è noto anche come *magnetic resonance spectroscopic imaging* (MRSI) o più semplicemente come *spectroscopic imaging* (SI).

1.4 Soppressione dell'acqua e saturazione

L'acqua è presente nei tessuti cerebrali con una concentrazione che è all'incirca 10^4 volte più grande di quella dei metaboliti. Per evitare che il segnale di questi sia totalmente coperto da quello dell'acqua, occorre applicare opportune sequenze di soppressione. Comunemente si utilizza la *chemical shift selective suppression* (CHESS, **Figura 1.7**). Consiste di tre impulsi selettivi a 4.7 ppm, seguiti da tre gradienti nelle tre direzioni spaziali; l'effetto è lo sfasamento degli spin e la soppressione del segnale dei protoni nel picco del segnale dell'acqua. L'ampiezza spettrale della soppressione CHESS è circa 75 Hz, che a 1.5 T corrisponde a 1.2 ppm, pertanto la sequenza non influenza il segnale di metaboliti che hanno picchi per frequenze minori di 4.1 ppm o maggiori di 5.3 ppm [8].

Oltre a quello dell'acqua, possono arrecare disturbo anche i segnali provenienti dall'esterno del voxel, i quali non sono completamente eliminati dalla procedura di selezione del VOI. Si rende quindi spesso necessaria l'applicazione di impulsi di *saturazione*, che hanno lo scopo di eliminare il segnale di specifiche bande nelle zone limitrofe al voxel. Nel momento in cui l'impulso termina e viene creata la magnetizzazione trasversale, vengono applicati dei gradienti molto intensi che defasano gli spin e riducono a zero il

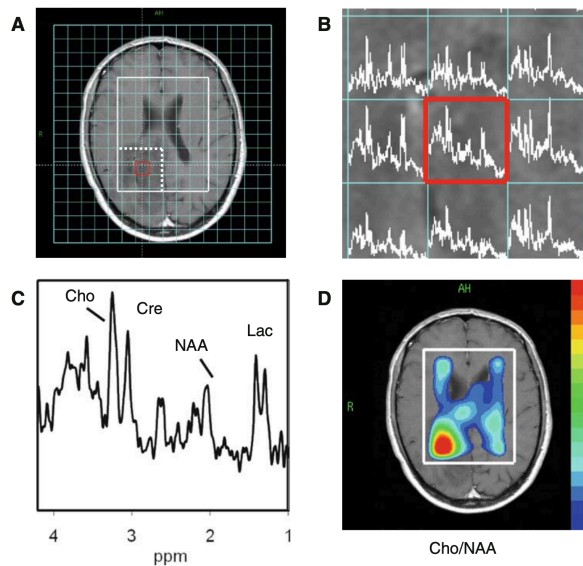


Figura 1.6: Immagini anatomiche e spettroscopiche ottenute su un paziente affetto da oligodendroglioma [10]. **A**: immagine MR in cui è selezionata la regione di interesse. **B**: immagine spettroscopica dei voxel. **C**: spettro del voxel contornato in rosso. **D**: ricostruzione in scala a falsi colori del rapporto delle concentrazioni di colina e NAA; il rosso indica il valore massimo ed è in corrispondenza della regione tumorale.

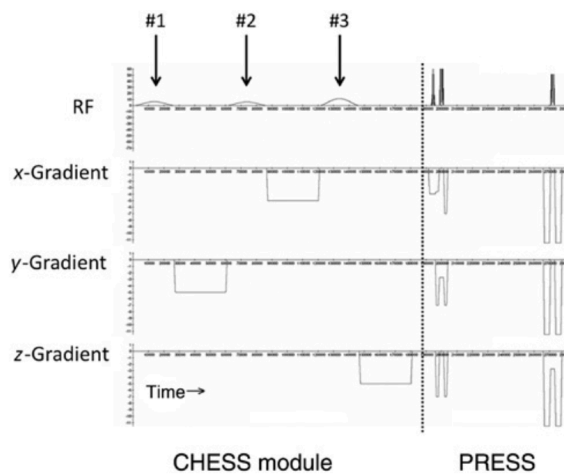


Figura 1.7: Sequenza CHESSE per la soppressione del segnale dell'acqua, seguita da una sequenza PRESS per la selezione del VOI [8].

segnale. Le bande di saturazione (**Figura 1.8**) sono quindi *slice* adiacenti al VOI e vengono determinate sulla base dell'immagine scout di MR acquisita come riferimento [8].

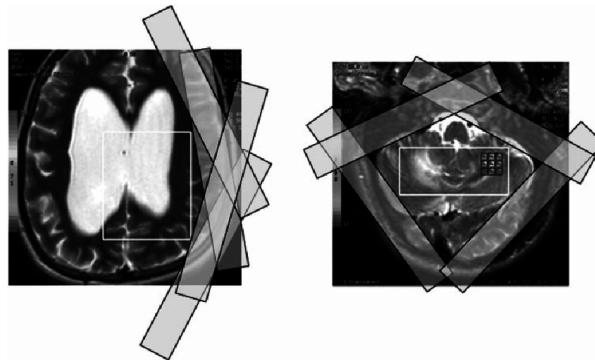


Figura 1.8: Esempi di posizionamento di bande di saturazione. A sinistra sono collocate lungo il lato sinistro del cranio, a destra lungo le ossa della fossa posteriore del cervello, in entrambi i casi per evitare una contaminazione dovuta ai lipidi [11].

1.5 Quantificazione

Esistono diversi approcci alla quantificazione dei metaboliti.

Il primo, più tradizionale, è l'integrazione dei picchi: si può infatti dimostrare che la concentrazione di un metabolita è proporzionale all'area sottesa dallo spettro di quel metabolita. Operativamente, si scelgono intervalli di frequenze che contengono i picchi del metabolita di interesse e integrando numericamente si calcola l'area delimitata. Occorre comunque sottrarre il contributo della *baseline*, cioè un segnale di fondo dovuto a protoni non considerati nell'analisi (ad esempio, quelli delle macromolecole).

Il secondo metodo è il *peaking*. Consiste nell'adattare i picchi del segnale a funzioni modello definite matematicamente, di cui si conoscono gli integrali. In generale risulta più accurato del precedente.

Attualmente, si è consolidato l'uso di un software (LCModel) che sfrutta set di segnali di base già noti. Il LCModel (*Linear Combination Model*), come suggerito dal nome, interpreta il segnale ottenuto *in vivo* come una combinazione lineare dei segnali dei diversi metaboliti presenti, rilevati da opportuni campioni *in vitro* ottenuti nelle stesse condizioni sperimentali. Questi sono soluzioni di metaboliti con concentrazioni note, i cui spettri hanno un rapporto segnale-rumore alto.

Sovrapponendo il segnale ottenuto con gli spettri dei metaboliti (**Figura 1.8**), il programma adibito esegue un fit e sulla base dei valori dei parametri della regressione risale alle concentrazioni.

Il vantaggio fondamentale di questa procedura è che due metaboliti che si sovrappongono ad una specifica frequenza di risonanza possono comunque essere discriminati se i loro spettri caratteristici sono diversi negli altri punti. In generale, il LCModel si è dimostrato più efficiente del *peaking*, siccome utilizza tutta l'informazione disponibile nello spettro.

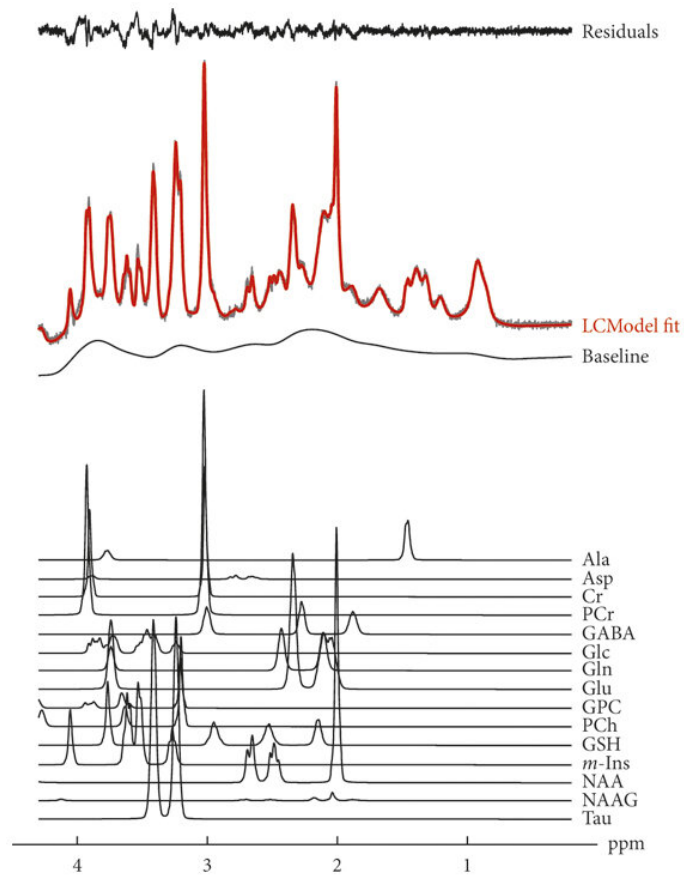


Figura 1.9: Combinazione lineare degli spettri dei metaboliti (in basso) adattata al segnale rilevato (in alto). In generale, occorre considerare la presenza di una *baseline*, cioè un segnale di base dovuto ai protoni che non vengono inclusi nell'analisi ma che danno comunque un contributo (ad esempio, quelli delle macromolecole) [12].

Dal punto di vista operativo, la difficoltà principale è che non c'è possibilità di convertire direttamente l'intensità del segnale rilevato in un valore di concentrazione assoluta: il motivo è che il fattore di scala varia a seconda della regione cerebrale e del soggetto considerati.

Ciononostante, esistono due approcci per trarre informazioni significative dagli spettri MRS: la quantificazione relativa di un metabolita rispetto ad un altro e la stima della quantificazione assoluta [8, 13].

1.5.1 Quantificazione relativa

La quantificazione relativa è una modalità di analisi che ricostruisce il rapporto tra la concentrazione del metabolita in esame e quella di un altro metabolita endogeno assunto come riferimento. Tipicamente, questo ruolo è assegnato alla creatina, per via del suo spettro con un SNR alto e della stabilità della sua concentrazione nei tessuti. L'utilizzo di questa normalizzazione ha l'effetto di ridurre la varianza dei dati dovuti alla variabilità del fattore di scala tra segnale e concentrazione.

Questa modalità possiede comunque delle limitazioni, dovute al fatto che la concentrazione di creatina può variare anche notevolmente tra soggetti diversi. Un problema significativo si verifica quando esiste una differenza sistematica in questo parametro tra due gruppi di soggetti: ciò comporta la possibilità che tra i risultati si ottengano falsi negativi o falsi positivi [8].

1.5.2 Stima della quantificazione assoluta

La stima della quantificazione assoluta sfrutta il riferimento del segnale dell'acqua nei tessuti cerebrali.

In questo approccio si rende quindi necessaria la conoscenza delle frazioni di sostanza grigia (GM), sostanza bianca (WM) e liquido cerebro-spinale (CSF) nel voxel; inoltre, occorrono due acquisizioni distinte con gli stessi parametri: una per il metabolita e una per l'acqua.

La concentrazione dell'acqua nel voxel è ricavata dall'analisi della segmentazione del voxel:

$$C_{acqua} = F_{GM} \cdot C_{GM} + F_{WM} \cdot C_{WM} + F_{CSF} \cdot C_{CSF} \quad (1.4)$$

dove F_{GM} , F_{WM} e F_{CSF} sono le frazioni del voxel costituite rispettivamente da GM, WM e CSF, mentre C_{GM} , C_{WM} e C_{CSF} sono le concentrazioni dell'acqua rispettivamente nella GM (43.30 M), nella WM (36.08 M) e nel CSF (55.51 M). Una volta ottenuta la concentrazione dell'acqua, si risale al fattore di scala utilizzando la relazione con il segnale dell'acqua.

Infine, si calcola la concentrazione del metabolita nella maniera seguente:

$$C_{metabolita} = \frac{S_{metabolita}(0)}{S_{acqua}(0)} \cdot \frac{N_{acqua}}{N_{metabolita}} \quad (1.5)$$

dove $S_{metabolita}(0)$ e $S_{acqua}(0)$ sono i segnali estrapolati a $TE = 0$ rispettivamente del metabolita e dell'acqua, mentre N_{acqua} e $N_{metabolita}$ sono il numero di protoni che producono rispettivamente il segnale dell'acqua e quello del metabolita. Ovviamente, $N_{acqua} = 2$, in quanto il contributo viene dai nuclei dei due atomi di idrogeno della molecola [8].

Capitolo 2

Deep learning

Il *deep learning* (DL, *apprendimento profondo*) è un'area del *machine learning*, a sua volta branca dell'intelligenza artificiale (AI). Questa è una scienza che mira a ricreare nelle macchine una forma di intelligenza simile a quella umana, rendendole in grado di svolgere compiti complessi senza la necessità di una programmazione esplicita da parte dell'uomo.

L'architettura di base di un modello di apprendimento profondo è rappresentata dalle *reti neurali artificiali* (*artificial neural networks*, ANN), che sono ispirate a quelle biologiche. La caratteristica essenziale è la presenza di più livelli di apprendimento (*layer*) organizzati in scala gerarchica, in modo che l'input di quelli più interni sia costituito dall'output dei precedenti. In particolare, gli algoritmi di DL imparano a riconoscere strutture gerarchiche e livelli di rappresentazione e astrazione per comprendere e analizzare autonomamente i pattern dei dati che vengono loro forniti in input. L'aggettivo profondo (*deep*) deriva da questa impostazione strutturale.

Nei paragrafi successivi, dopo una breve esposizione dei principi fondamentali del *machine learning*, si presentano alcuni modelli di reti neurali artificiali di interesse per le applicazioni in campo biomedico oggetto di questa tesi.

2.1 *Machine learning*

Con il termine *machine learning* (*apprendimento automatico*) si intende una branca dell'intelligenza artificiale che studia e sviluppa tecniche per predisporre le macchine alla comprensione autonoma dei problemi e allo sviluppo di algoritmi per risolverli, in maniera appunto "automatica".

L'apprendimento di una macchina passa attraverso un algoritmo in grado di imparare dai dati. Si riporta la definizione più citata di algoritmo di apprendimento (Tom M. Mitchell, 1997): "Si dice che un programma apprende dall'esperienza E con riferimento a alcune classi di compiti T e con misurazione della performance P se le sue performance

nel compito T , come misurato da P , migliorano con l'esperienza E " [14]. Aniché definire formalmente queste entità, è utile fornire alcuni esempi intuitivi [15].

I compiti T (*task*) sono solitamente determinati in relazione a come la macchina dovrebbe gestire un esempio. Un esempio è un insieme di caratteristiche (*features*) misurate quantitativamente a partire da un oggetto o da un fenomeno; viene rappresentato come un vettore $\mathbf{x} \in \mathbb{R}^n$, dove ogni componente x_i è una caratteristica (ad esempio, i valori dei pixel di un'immagine). Tra i compiti più comuni:

- **Classificazione:** al computer è chiesto di dire a quale delle k categorie disponibili appartiene un oggetto, dunque l'algoritmo deve produrre una funzione $f : \mathbb{R}^n \rightarrow \{1, \dots, k\}$;
- **Regressione:** date delle coppie di variabili in input, l'algoritmo deve produrre in output una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ che rappresenti le relazioni tra queste variabili.

La misurazione della performance P valuta le abilità dell'algoritmo. Tipicamente si misurano l'*accuratezza* (frazione di esempi per cui il modello produce l'output corretto) o equivalentemente il *tasso di errore* (frazione di esempi per cui il modello produce un output non corretto). Siccome l'interesse è rivolto alle capacità della macchina di gestire dati nuovi, si utilizzano in queste prove dei *test set* differenti da quelli usati per addestrare (istruire) il sistema (*training set*).

La modalità di esperienza E dell'algoritmo determina una categorizzazione degli algoritmi di *machine learning*:

- **apprendimento non supervisionato:** alla macchina viene fornito un set di dati e il modello deve trovare una struttura nell'input fornito, ad esempio una distribuzione di probabilità;
- **apprendimento supervisionato:** alla macchina viene fornito un set di dati a ognuno dei quali è associato l'output desiderato (*target*) e sulla base di questa guida il modello deve trovare una regola in accordo con questi dati;
- **apprendimento per rinforzo:** l'algoritmo interagisce con un ambiente che gli fornisce un feedback.

2.1.1 Dataset di *training*, validazione e *testing*

Nell'ambito del *machine learning*, le macchine prendono decisioni sulla base di algoritmi che rappresentano, solitamente, modelli matematici. Questi modelli sono creati sulla base di opportuni set di dati che vengono sottoposti agli algoritmi durante il processo di apprendimento. Questo insieme di dati viene suddiviso in tre parti con finalità differenti [16]:

1. **training dataset**: è un set di dati che viene passato alla macchina, la quale è guidata in un processo di apprendimento supervisionato in cui modifica i suoi parametri; l'obiettivo è adattare i suoi output ai target forniti. In generale, la scelta della modalità di valutazione della performance influisce sul risultato a cui giunge la macchina;
2. **validation dataset**: è un insieme di dati utilizzato per selezionare i valori degli iperparametri del modello (l'architettura; ad esempio, uno degli iperparametri delle reti neurali è il numero di strati che le compongono). Può essere omesso;
3. **testing dataset**: è un set indipendente da quello di *training*, ma che segue la stessa distribuzione di probabilità. È utilizzato per misurare le capacità dell'algoritmo alla fine del processo di apprendimento (*training* e validazione).

2.2 Percettroni e reti neurali

Il perceptrone è l'esempio più semplice di rete neurale artificiale. La sua unità costitutiva fondamentale è rappresentata dal neurone artificiale (**Figura 2.1**), un modello matematico elaborato da W. McCulloch e W. Pitts nel 1943 per rappresentare il comportamento del neurone biologico (**Figura 2.2**).

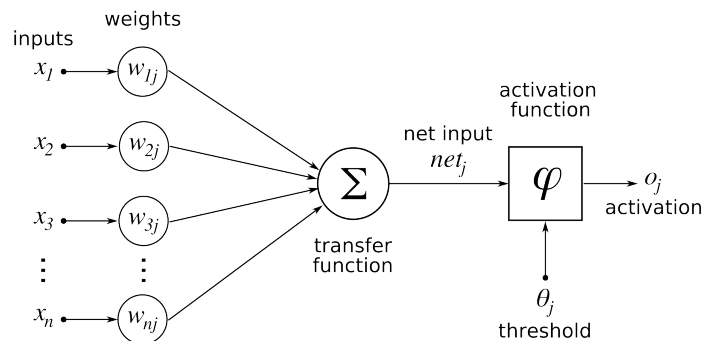


Figura 2.1: Schema logico del neurone di McCulloch-Pitts [17].

Così come quest'ultimo riceve degli stimoli (segnali elettrici) dall'esterno, anche il neurone di McCulloch-Pitts riceve in ingresso degli input.

In particolare, esso mappa i suoi ingressi \mathbf{x} in un valore di uscita $f(\mathbf{x})$, nel seguente modo: dato il vettore di input \mathbf{x} , questo viene moltiplicato scalarmente per il vettore dei pesi \mathbf{w} ; alla somma viene aggiunto un *bias* b e il risultato ottenuto φ diventa l'argomento della *funzione di attivazione* θ , che determina se il neurone produce o meno un output o . Questo output costituisce l'input di altri neuroni, nello stesso modo in cui nelle reti neurali biologiche i neuroni sono connessi tra loro in più livelli gerarchici.

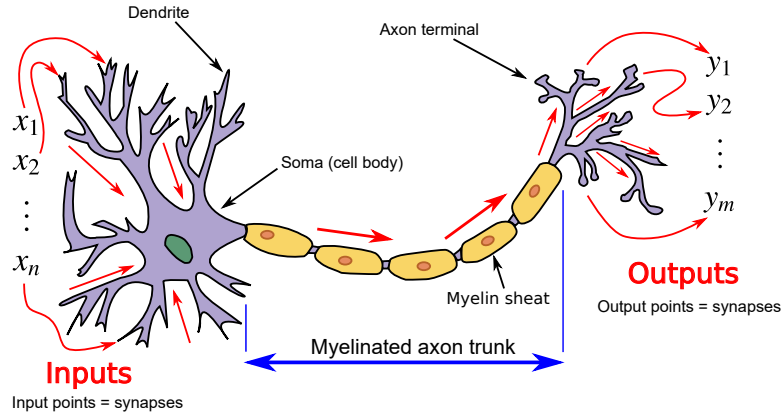


Figura 2.2: Schema del neurone biologico. I dendriti ricevono gli stimoli esterni, i quali vengono elaborati e trasmessi come differenze di potenziale attraverso l'assone fino alle sinapsi terminali [18].

Tra le funzioni di attivazione più utilizzate [19] ci sono la *Rectified Linear Unit* (ReLU):

$$\sigma = \begin{cases} \max(0, x), & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases} \quad (2.1)$$

e la *Gaussian Error Linear Unit* (GELU):

$$g(x) = \frac{1}{2}x \left(1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right) \quad (2.2)$$

L'algoritmo di apprendimento di questa semplice architettura è supervisionato e consta dei seguenti passi [20]:

1. Inizializzazione dei pesi;
2. Per ogni esempio j del *training set*
 - (a) calcolo dell'output: $y_j(t) = f(\mathbf{x}_j)$;
 - (b) aggiornamento dei pesi: $w_i(t+1) = w_i(t) + r \cdot (d_j - y_j(t)) x_{j,i}$, dove r e d_j sono, rispettivamente, un parametro detto tasso di apprendimento e l'output desiderato.

Il perceptrone è costituito da uno o più neuroni artificiali, organizzati su più livelli e connessi in modo che l'output di un livello costituisca l'input del successivo. A seconda

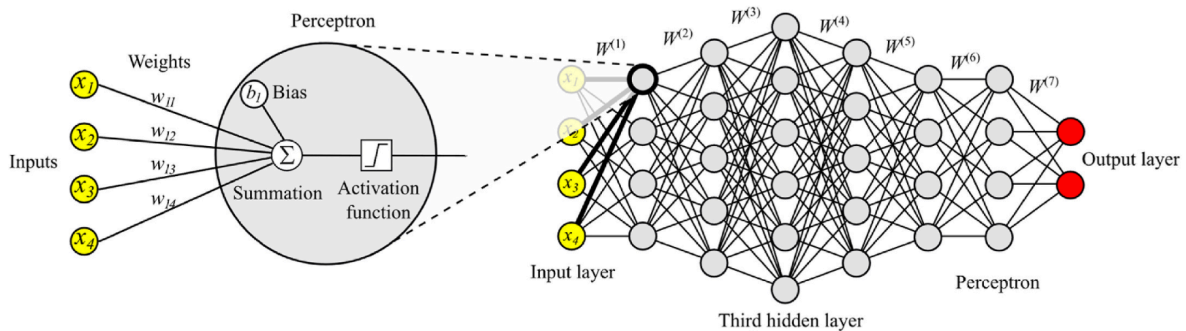


Figura 2.3: Schema di un percettrone *multi-layer*. A sinistra è visibile in dettaglio la struttura dei neuroni artificiali, che corrispondono ai nodi connessi dalle linee. Per ogni coppia di strati consecutivi i e $i + 1$, è definita una matrice $W^{(i)}$, i cui elementi sono i pesi delle connessioni tra i neuroni di un *layer* e quelli del *layer* successivo [2].

che siano o meno presenti strati nascosti (*hidden layers*), distinguiamo percettroni *multi-layer* e *single-layer* (**Figura 2.3**). Estendendo

Il primi sono un esempio di rete neurale *feed-forward* (FNN): il flusso dell'informazione è unidirezionale, siccome non sono presenti cicli o loop.

L'architettura di una FNN è definita dal numero di *layer*, dal numero di neuroni in ogni strato e dalle connessioni tra i *layer*. Lo scopo di una FNN è quello di approssimare, tramite l'output dell'ultimo strato, una data funzione f ; questo obiettivo passa attraverso il *training* della rete, svolto tipicamente con il metodo di *backpropagation* [21].

L'importanza delle funzioni di attivazione si manifesta nel *teorema di approssimazione universale*: se la funzione di attivazione è non lineare, allora con una semplice FNN a due *layer* è possibile realizzare un approssimatore universale di funzioni [19].

2.2.1 *Backpropagation*

La retropropagazione dell'errore è il più comune algoritmo di apprendimento per una FNN; la sua struttura ricalca quella del modello di apprendimento di un semplice neurone, presentato sopra.

Il processo con cui una rete impara a riconoscere le caratteristiche e i pattern di un input deve essere guidato da un indicatore, che valuti le performance nel corso dell'addestramento. Si definisce perciò una *funzione costo* o *obiettivo* (*loss function* o *objective function*), che deve essere minimizzata attraverso l'assegnazione di opportuni valori ai pesi dei nodi. Una scelta ricorrente per la *loss function* è l'errore quadratico medio (*mean*

squared error):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.3)$$

dove y_i è il valore di output della rete mentre \hat{y}_i è quello fornito per l'addestramento. Richiamando quanto già esposto, in un modello di apprendimento supervisionato per ogni input x_i viene calcolato l'output $f(x_i)$, che deve essere confrontato con il target y_i . A questo scopo, fissata la coppia (x_i, y_i) , l'algoritmo calcola i gradienti della funzione di costo:

$$\frac{\partial C}{\partial w_{jk}^l}(y_i, g(x_i)) \quad (2.4)$$

dove w_{jk}^l è il peso tra il k -esimo nodo nello strato $l - 1$ e il j -esimo nodo nello strato l . Il calcolo del gradiente fa parte della tecnica della *discesa del gradiente*, un metodo di ottimizzazione che consente di determinare il minimo (o il massimo) di una funzione di più variabili. Intuitivamente, si tratta di far variare (aggiornare) i pesi w_{jk}^l in modo da muoversi in direzione opposta al gradiente e giungere ad un punto di minimo (locale) della funzione [22]. Ad ogni iterazione (cioè per ogni coppia di elementi del dataset di *training*), ogni peso riceve un aggiornamento, in quanto al suo valore attuale viene sottratto un termine proporzionale al gradiente della funzione costo:

$$w \rightarrow w - \eta \nabla C(x_i) \quad (2.5)$$

dove w è il vettore di pesi di un *layer* ed η è il tasso di apprendimento.

Si parla di retropropagazione in quanto il calcolo inizia dai layer vicini all'output e procede all'indietro.

Un problema ricorrente nell'applicazione di questo algoritmo è quello della *scomparsa del gradiente*. Nell'algoritmo di *backpropagation*, i gradienti nei vari livelli vengono calcolati secondo la regola della catena delle derivate: poiché le funzioni di attivazione di molti *layer* hanno gradienti con valori nell'intervallo $[0, 1]$, il prodotto decresce esponenzialmente con la profondità della rete. Questo implica che i pesi degli strati vicini all'output si aggiornano, mentre quelli vicino all'input non lo fanno. Una soluzione è rappresentata da opportune scelte dei pesi e inizializzazioni dei parametri [23, 24].

2.3 Reti neurali convoluzionali

Le reti neurali convoluzionali (CNN) costituiscono una classe di reti neurali ispirate alla struttura della corteccia visiva animale; le loro applicazioni principali si trovano nel campo del riconoscimento delle immagini. In particolare, le CNN vengono addestrate per

imparare ad estrarre caratteristiche ricorrenti (pattern) all'interno di una certa categoria di oggetti, in modo da poterle utilizzare come elemento per distinguere e riconoscere quella specifica categoria. La **Figura 2.4** mostra la struttura di una CNN.

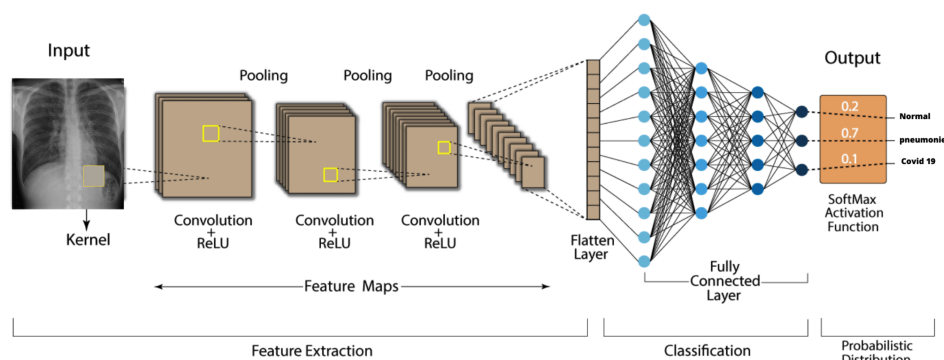


Figura 2.4: Schema di una CNN impiegata per il riconoscimento di una porzione di un'immagine radiologica. L'unità fondamentale è il *layer* di convoluzione, che estrae le caratteristiche di un'immagine secondo livelli di astrazione via via crescenti; il risultato finale dell'analisi è costituito dalle probabilità delle opzioni diagnostiche prese in considerazione [25].

L'elemento costitutivo di una CNN è il *layer di convoluzione*. Questo strato è caratterizzato da tre parametri: altezza, larghezza e profondità (H, W, D).

L'altezza e la larghezza definiscono la dimensione del filtro di convoluzione, mentre la profondità è il numero di filtri in quel layer [21].

Per comprendere il funzionamento del filtro, occorre ricordare come un'immagine digitale sia costituita da pixel, ognuno dei quali è definito da tre o un valore numerico (colori o livello di grigio). L'operazione di convoluzione consiste nello svolgimento di un prodotto di Hadamard del filtro (anche detto *kernel*) con l'immagine (matrice o tensore) di input. Posizionando il *kernel* sull'immagine, i pixel di una porzione vengono ad essere in corrispondenza con quelli del filtro; sommando tra di loro i valori dei pixel moltiplicati per i pesi definiti dagli elementi del filtro, si ottiene il risultato. In seguito, traslando il kernel fino a coprire tutti i pixel dell'immagine di partenza, si ottiene la matrice di convoluzione (**Figura 2.5**). Il parametro che definisce questo processo è il *passo* (*stride*): è il numero di pixel che definiscono la traslazione del kernel tra due applicazioni successive [26].

Il risultato dell'applicazione di una convoluzione ad un'immagine è una *feature map* o *feature channel*.

Il *pooling* è un'operazione che viene eseguita nel passaggio tra layer di convoluzione; diminuisce la dimensione della rappresentazione per ridurre il numero di parametri nel calcolo e facilitare l'operazione di estrazione delle *feature* (da qui la denominazione alternativa *down-sampling*). Esistono vari tipi di pooling, ma quello più comune è il *max*

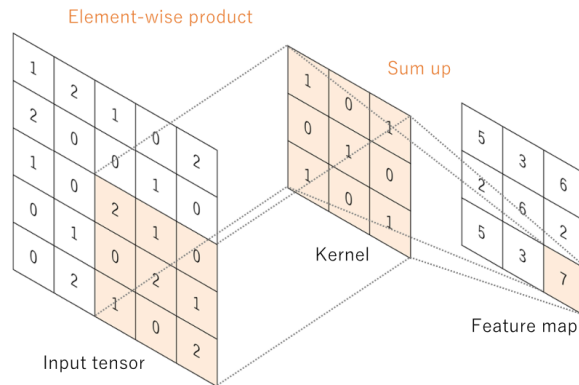


Figura 2.5: Convoluzione dell'immagine con il *kernel*; il pixel evidenziato nella *feature map* si ottiene dal prodotto di Hadamard: $2 \cdot 1 + 1 \cdot 0 + 0 \cdot 1 + 0 \cdot 0 + 2 \cdot 1 + 1 \cdot 0 + 1 \cdot 1 + 0 \cdot 0 + 2 \cdot 1 = 7$ [26].

pooling (Figura 2.6). Esso sostituisce un cluster di neuroni in input con un unico pixel definito dal maggiore dei valori dei singoli pixel considerati.

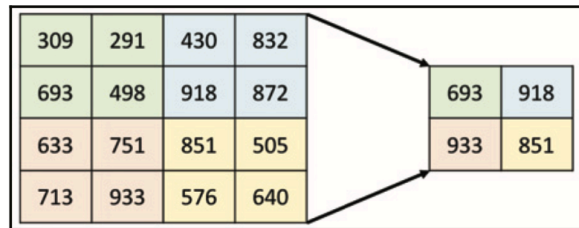


Figura 2.6: Esempio di *max pooling* 2 x 2 [17].

Dopo il percorso di estrazione delle *feature* si trova un *flatten layer*, cioè uno strato che trasforma l'output delle convoluzioni in un vettore; questo costituisce l'input di un *layer fully connected* (FC), in cui vengono sviluppati i calcoli numerici che forniranno i risultati del riconoscimento. Due strati consecutivi sono detti FC se esistono connessioni tra ogni nodo di un *layer* e tutti i nodi dell'altro.

L'ultimo strato della CNN consiste di una funzione detta Softmax, che fornisce in output un vettore di probabilità dei possibili outcome. Matematicamente, è definita come [19]:

$$S(y)_i = \frac{\exp(y_i)}{\sum_{j=1}^n \exp(y_j)} \quad (2.6)$$

dove y è il vettore di input che consiste di n elementi, n è il numero di categorie della classificazione e y_i è la componente i -esima del vettore.

Spesso si rende necessario avere un'immagine di output con le stesse dimensioni di quella in input. In tal caso, si fa uso di un *up-sampling*, di cui esistono diversi esempi:

- *nearest neighbor*: è un'operazione che ripete righe e colonne di pixel secondo un opportuno fattore per ottenere un'immagine delle dimensioni desiderate (**Figura 2.7**);

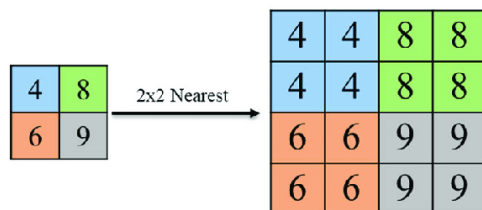


Figura 2.7: Esempio di *up-sampling nearest neighbor*. Il fattore di *up-sampling* è (2,2), perciò ogni pixel nell'input produce una matrice 2x2 di output [27].

- *un-pooling*: in seguito ad un'operazione di *max pooling*, la matrice di pixel viene ricostruita nelle sue dimensioni originali sostituendo con degli zero i valori diversi dai massimi (**Figura 2.8**).

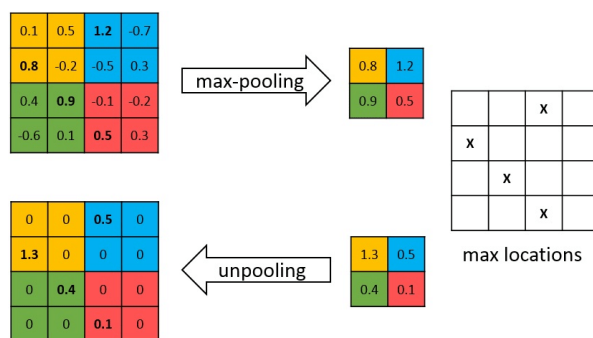


Figura 2.8: Esempio di *un-pooling*. Il sistema tiene memoria di dove erano collocati i massimi precedenti nella matrice di partenza e li colloca quelli nuovi durante l'up-sampling [28].

Simile all'*up-sampling* è la convoluzione trasposta (o *up-convolution*, **Figura 2.9**): anch'essa produce in output un'immagine di dimensioni maggiori rispetto a quella di input. Di fatto, la convoluzione trasposta effettua l'operazione inversa della convoluzione: anziché far scorrere il *kernel* sull'input, il *layer* fa scorrere l'input sul *kernel* e per ogni step viene effettuata la somma dei pixel pesata con i coefficienti del *kernel*.

In maniera intuitiva, il concetto di convoluzione può essere trasportato anche al caso 1D (**Figura 2.10**); questo tipo di reti convoluzionali viene solitamente impiegato nel processing di oggetti unidimensionali quali segnali.

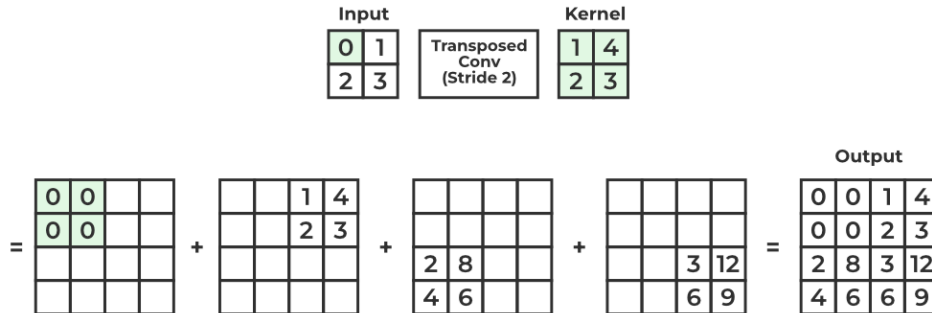


Figura 2.9: Esempio di convoluzione trasposta [29].

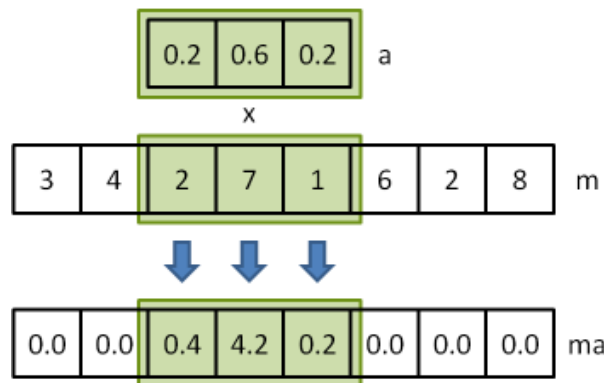


Figura 2.10: Esempio di convoluzione 1D; il kernel viene fatto scorrere in una sola dimensione e per ogni spostamento del filtro si effettua un prodotto scalare [30].

2.3.1 UNet

L'architettura UNet (**Figura 2.11**) è una CNN che è stata sviluppata nel 2015 dall'Università di Friburgo per la segmentazione delle immagini nel campo biomedico. Con segmentazione si intende il processo di suddivisione di un'immagine in gruppi di pixel che hanno caratteristiche comuni; ad esempio, in ambito biomedico, la segmentazione di un'immagine (di TC, PET o MR) consiste nell'individuazione e nel riconoscimento dei diversi tessuti visibili e delle loro separazioni.

La rete consiste di un percorso di contrazione (parte sinistra) e un percorso di espansione (parte destra), che comportano la caratteristica forma a "U" dello schema.

Il percorso di contrazione funge da encoder ed è costituito dalla ripetizione, su ogni livello, di due convoluzioni 3×3 , ognuna seguita da una ReLU e un *max pooling* 2×2 per

il *down-sampling*. Ad ogni pooling si raddoppia il numero di *feature channel*, ma allo stesso tempo si dimezza la dimensione dell'immagine.

Nel percorso di espansione, che funge da decoder, si applica un *up-sampling* alla *feature map*, seguito da una convoluzione trasposta 2×2 che dimezza il numero di *feature channel*; successivamente, una concatenazione con la *feature map* corrispondente proveniente dal percorso di contrazione e due convoluzioni 3×3 , ognuna delle quali seguita da una ReLU [31].

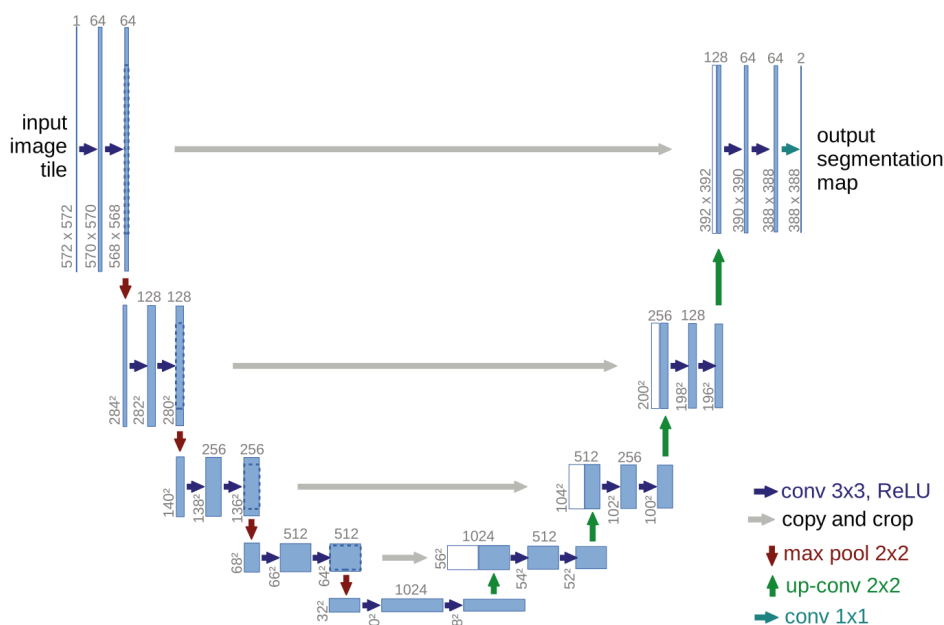


Figura 2.11: Schema di un'architettura UNet. Nel percorso di contrazione, l'immagine riduce progressivamente le sue dimensioni, ma aumenta il numero di canali; in quello di espansione, essa torna alle sue dimensioni originali e riduce il numero di canali [31].

Le connessioni tra il percorso di contrazione e quello di espansione (*skip connection*) servono per facilitare il riconoscimento e la collocazione sull'immagine delle *feature* estratte.

2.4 Reti neurali residuali

Una rete neurale residuale (ResNet) è formata dalla ripetizione di un'unità fondamentale, il *blocco residuale* (Figura 2.12). Questo è tipicamente costituito da due o tre *layer* in sequenza che ricevono in input x e producono in output $F(x)$; l'output complessivo del blocco è rappresentato da $H(x) = F(x) + x$, siccome esiste una *shortcut* o *skip connection* che somma direttamente l'input x all'output dei *layer*. L'ipotesi alla base è che sia più

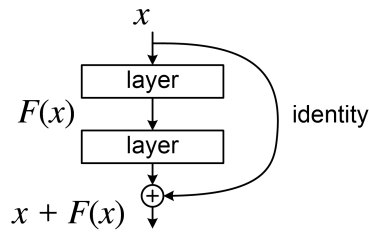


Figura 2.12: Blocco residuale [32].

semplice per la rete apprendere il residuo $F(x) = H(x) - x$ piuttosto che la relazione complessiva $H(x)$.

Il vantaggio fondamentale delle *skip connection* è che rappresentano una soluzione al problema della scomparsa del gradiente: creano percorsi per la minimizzazione del gradiente alternativi a quelli non lineari e ciò permette di aggiornare i pesi anche dei primi *layer* [24, 32].

2.5 *Autoencoder*

Un *autoencoder* (Figura 2.13) è una rete neurale artificiale che ha lo scopo di apprendere delle codifiche per dati non etichettati. Svolge due compiti: codifica i dati in input (*encoder*) e decodifica per riprodurre i dati originali (*decoder*) secondo una rappresentazione efficiente. Il significato di efficienza dipende generalmente dal contesto: ad esempio, può indicare una riduzione delle dimensioni di un'immagine, oppure una rimozione del rumore.

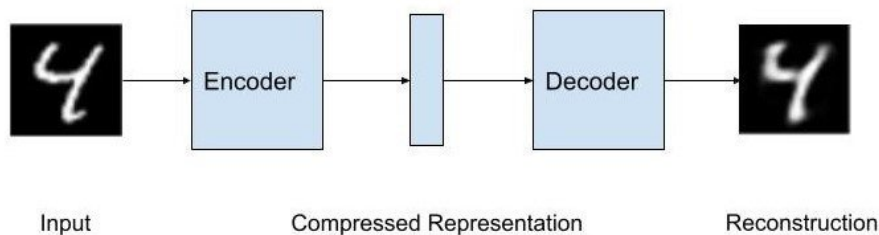


Figura 2.13: Schema di funzionamento di un *autoencoder*. L'*encoder* trasforma l'input in una rappresentazione più compatta, mentre il *decoder* ricostruisce dalla codifica l'oggetto iniziale [33].

Matematicamente, un *autoencoder* è definito da:

- lo spazio dei "messaggi" decodificati $\mathcal{X} = \mathbb{R}^n$ e quello delle "parole" della codifica (detto *latent space*) $\mathcal{Z} = \mathbb{R}^m$;

- la famiglia delle funzioni di codifica $E_\phi : \mathcal{X} \rightarrow \mathcal{Z}$ e quella delle funzioni di decodifica $D_\theta : \mathcal{Z} \rightarrow \mathcal{X}$ (parametrizzate rispettivamente da ϕ e θ).

Qualora siano presenti più textitlayer nascosti sia per l'*encoder* che per il *decoder*, si parla di *deep autoencoder*. Tra i vantaggi della profondità ci sono la diminuzione della quantità di dati necessaria per il *training* e una maggiore capacità di compressione (che si traduce in una rappresentazione più efficiente delle funzioni) [34].

Capitolo 3

Applicazioni

La MRS è l'unica tecnica non invasiva in grado di quantificare il metabolismo cerebrale *in vivo*. Purtroppo, però, numerose complicazioni e limitazioni compromettono le possibilità di utilizzare questa metodologia di indagine.

Il primo ostacolo è costituito dalla scarsa concentrazione dei metaboliti nei tessuti (1-10 mM). Ciò comporta la necessità di acquisire il segnale da voxel di grandi dimensioni, affinché gli spettri che si ottengono siano caratterizzati da un SNR accettabile; ne derivano immagini spettroscopiche (SI) a bassa risoluzione. Ad esempio, i segnali di creatina, colina e N-acetil-aspartato (i quali sono i metaboliti con la concentrazione più elevata, 10-30 mM) vengono acquisiti in voxel di 8 cm³; quello del lattato in voxel di dimensioni ancora maggiori, 15-30 cm³ [8].

In generale, la bassa sensibilità della MRS implica anche tempi di acquisizione più lunghi rispetto a quelli caratteristici del MRI. Inoltre, a causa della bassa intensità dei campi magnetici usati in ambito clinico, negli spettri i segnali dei diversi metaboliti si sovrappongono, problema esacerbato anche dalla disomogeneità del campo all'interno del voxel; ciò rende difficoltosa la quantificazione.

Infine, in MRS è molto difficile costruire protocolli di acquisizione standardizzati, anche a causa della bassa specificità [35].

Nei paragrafi che seguono, vengono presentati tre studi che forniscono una panoramica dello stato dell'arte delle applicazioni di DL alla MRS *in vivo*; ognuno di questi è relativo ad una delle tre fasi principali di un esperimento MRS: acquisizione, processing e analisi.

3.1 Selezione automatizzata del voxel

Nella pianificazione dell'acquisizione, uno dei principali ostacoli è rappresentato dalla definizione del voxel di interesse all'interno di una regione cerebrale in cui si è osservata la presenza di una lesione. L'obiettivo della MRS è acquisire informazioni utili ai fini

diagnostici in corrispondenza di quella lesione, che può corrispondere ad una patologia tumorale. La selezione del VOI, ad oggi, è svolta per via manuale da un esperto.

Lo studio di Bolan et al. [36] ha proposto l'utilizzo di una CNN per la segmentazione di una lesione tumorale e di un metodo di ottimizzazione per massimizzare la regione tumorale contenuta nel voxel.

In particolare, sono stati confrontati i risultati della selezione manuale eseguita da un esperto con quelli della rete.

3.1.1 ResNet-50

Per la segmentazione automatizzata è stata impiegata una CNN, in particolare una U-Net 2D modificata: il percorso di espansione, cioè l'encoder, è sostituito da un modello ResNet-50.

La ResNet-50 (Figura 3.1) è un'architettura costituita da 5 blocchi convoluzionali, che riducono progressivamente le dimensioni dell'immagine; in particolare si tratta di blocchi convoluzionali residuali. La U-Net utilizzata nello studio era stata già precedentemente

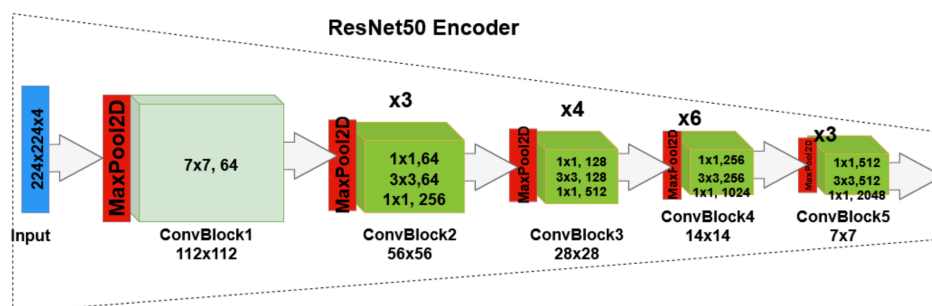


Figura 3.1: Architettura ResNet50 [37]. In ogni blocco convoluzionale, la notazione $k \times k$, n indica la dimensione del filtro (k) e la profondità del filtro (n).

addestrata sul dataset ImageNet [38], dunque è stato effettuato un *fine tuning*: si tratta di un ri-addestramento della rete su nuovi dati. Una scelta di questo tipo è vantaggiosa quando si ha a che fare con dataset limitati, che da soli non garantirebbero risultati ottimali.

3.1.2 Modellizzazione e implementazione

La sperimentazione sulla rete è consistita di due step: segmentazione della lesione e ottimizzazione geometrica. Per valutare la prima, è stato usato il *coefficiente di Sørensen-Dice*, che è una misura della sovrapposizione tra la segmentazione effettuata e quella

assunta come *gold standard*; dati due insiemi X e Y è definito come:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (3.1)$$

dove $|X|$, $|Y|$ e $|X \cap Y|$ sono le cardinalità dei due insiemi e della loro intersezione, rispettivamente.

Invece, l'indicatore della performance della rete per l'ottimizzazione geometrica è costituito da una funzione obiettivo, che tiene in considerazione la dimensione del voxel e la frazione di voxel che contiene la lesione:

$$F_{obiettivo}(\theta) = \exp\left(-\frac{1}{2}\left(\frac{V_{target} - \mu_V}{\sigma_V}\right)^2\right) \cdot \exp\left(-\frac{1}{2}\left(\frac{f_{target} - \mu_f}{\sigma_f}\right)^2\right) \quad (3.2)$$

dove V_{target} è il volume dell'intersezione tra il voxel e la regione che contiene la lesione (*gold-standard*) e f_{target} è la frazione del voxel che contiene la lesione, mentre (μ_V, σ_V) e (μ_f, σ_f) sono la media e la deviazione standard delle distribuzioni di V_{target} e f_{target} , rispettivamente; θ è un vettore di nove parametri geometrici che definiscono la localizzazione del voxel (posizione, dimensioni, angoli di rotazione del parallelepipedo rispetto al riferimento).

La scelta di costruire la funzione obiettivo come prodotto di due gaussiane è legata alla dipendenza della performance dai parametri: il SNR è basso se V_{target} è troppo piccolo, mentre il segnale è disomogeneo se V_{target} è troppo grande; la performance è ottimale se $f_{target} \sim 1$.

3.1.3 Dataset e procedure

Per prima cosa, la rete CNN è stata ulteriormente addestrata, facendo uso dei dati della repository BraTS del 2018 [39]. Di questi, solo le immagini T₂w-FLAIR e le segmentazioni manuali sono state usate. Di 285 casi, 200 sono stati riservati al *training*, 43 alla validazione, 42 al *testing*.

Invece, Il dataset utilizzato come banco di prova per il confronto tra selezione manuale e automatica è formato da 60 immagini anatomiche assiali T₂w acquisite su pazienti con gliomi di basso grado, con una sequenza FLAIR a 3 T. Come *gold-standard* sono state utilizzate le immagini prodotte da un neuro-oncologo.

Nella performance manuale, l'esperto spettroscopista ha selezionato manualmente il voxel e acquisito i dati con una sequenza MEGA-PRESS. Questa, come suggerito dal nome, è data dall'integrazione della PRESS con la MEGA (dal nome dei ricercatori che l'hanno proposta, *Meshcher-Garwood*); si tratta di una sequenza ideata per isolare il segnale del GABA, i cui nuclei sono sottoposti ad accoppiamento scalare [8].

Per quel che riguarda la rete, invece, 36 dei 60 casi disponibili sono stati dedicati al

training, 8 alla validazione, 8 al *testing*; 8 casi non sono stati utilizzati perché riservati ad un eventuale futuro *testing set*. Il *fine-tuning* (*training* e validazione) della rete si è reso necessario per adattarla alle differenze del dataset in esame con il BraTS.

3.1.4 Risultati

La segmentazione automatizzata ha prodotto i seguenti coefficienti di Sørensen-Dice: 0.81 sul dataset di validazione del BraTS, 0.88 e 0.87 sui dataset di validazione e *testing* dei gliomi.

Per definire i parametri della funzione obiettivo, si sono osservati i risultati della selezione manuale da parte dell'esperto (**Figura 3.2**): $\mu_V = 8.5 \text{ mL}$, $\mu_f = 1$; le deviazioni standard (SD) sono state scelte in modo da produrre gradienti elevati attorno alla soluzione: $\sigma_V = 2 \text{ mL}$, $\sigma_f = 0.25 \text{ mL}$. La **Figura 3.3** riporta il risultato di *training*/validazione e

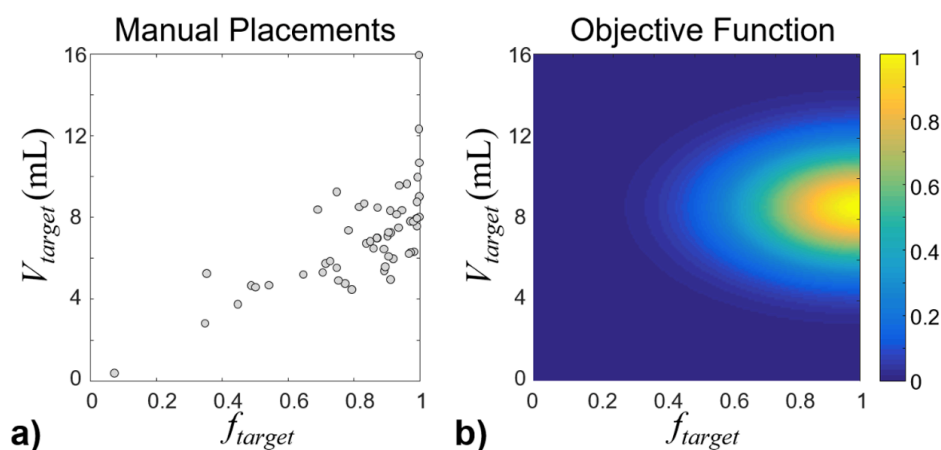


Figura 3.2: Collocazioni manuali (a) e funzione obiettivo proposta (b) in termini di V_{target} e f_{target} .

di *testing* della rete.

La **Tabella 3.1** riporta medie e SD di f_{target} , V_{target} e V_{voxel} sia per la selezione manuale sia per quella automatizzata. Si osserva come, nel set di *training* e validazione, f_{target} della modalità automatizzata sia maggiore di quella della modalità manuale: ciò indica che i voxel sono posizionati in maniera più precisa.

Inoltre, le SD di tutti i parametri sono più piccole nel metodo automatizzato: si può desumere quindi che le collocazioni del voxel eseguite dalla rete siano caratterizzate da una maggiore coerenza rispetto a quelle effettuate manualmente.

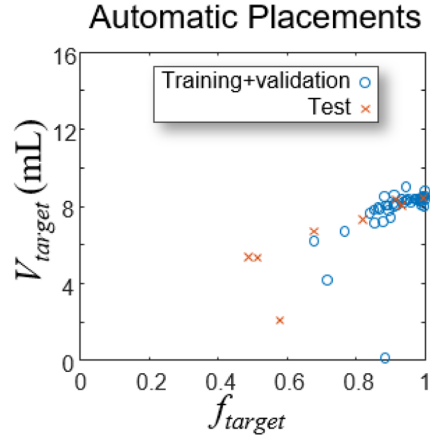


Figura 3.3: Parametri della collocazione del voxel per le fasi *training*/validazione e *testing* della rete.

	f_{target}		V_{target} (mL)		V_{voxel} (mL)	
	manuale	autom.	manuale	autom.	manuale	autom.
<i>training</i> e validazione	0.81, 0.23	0.926, 0.075	7.47, 4.25	7.85, 1.43	9.14, 4.81	8.47, 1.40
test	0.69, 0.23	0.74, 0.20	7.30, 4.09	6.45, 2.15	11.32, 8.42	8.75, 2.26

Tabella 3.1: Medie e SD relative ai parametri della selezione, nel caso manuale e in quello automatizzato [36].

3.1.5 Discussione

Il metodo di selezione del voxel con l'uso di una rete CNN si è mostrato più efficiente di quello manuale in termini di risultati e di tempo impiegato.

Ciononostante, ci sono alcune limitazioni da tenere in conto. Lo studio ha di fatto valutato solo l'individuazione della regione da parte della rete e non la qualità dell'indagine MRS nel voxel; inoltre, i casi di gliomi a disposizione nell'esperimento erano molto pochi.

3.2 *Imaging* spettroscopico con super-risoluzione

Lo studio di Iqbal et al. [40] ha messo in evidenza la capacità di una particolare architettura UNet (D-UNet) di ricostruire immagini spettroscopiche ad alta risoluzione (HRSI, *high resolution spectroscopic images*) a partire da immagini di risonanza magnetica pesate in T_1 (T_1w) e dalle corrispondenti immagini spettroscopiche a bassa risoluzione (LRSI). L'immagine spettroscopica è una sezione assiale del cervello su cui, in scala di falsi colori, viene riportata la concentrazione di un metabolita per ogni pixel.

3.2.1 D-UNet

La D-UNet (*densely connected UNet*) è un modello UNet in cui le *feature map* vengono trasportate a tutti i *layer* successivi, con l'obiettivo di rendere la rete più intelligente strato dopo strato e velocizzare il processo di *training*; la differenza con una UNet è dunque la presenza di concatenazioni. In **Figura 3.4** è riportato lo schema della D-UNet utilizzata nello studio, composta da tre *max pooling layer*.

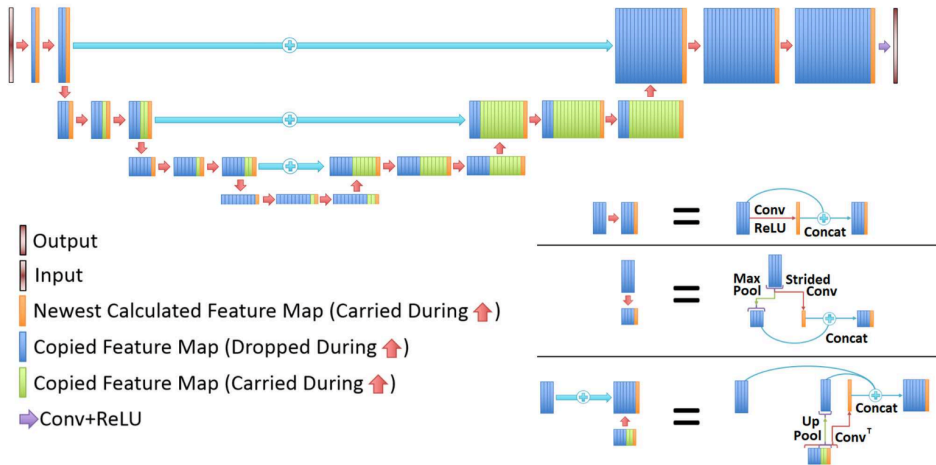


Figura 3.4: Schema dell'architettura D-UNet usata nello studio. Nella parte destra dell'immagine, è riportata una legenda relativa alle trasformazioni tra i *layer*: si noti come ad ogni convoluzione segua una concatenazione, in particolare ogni *feature map* prodotta da una convoluzione è trasportata fino alla fine del *layer* (su ognuno dei livelli). Le *feature map* che sono andate incontro a *max pooling* non vengono trasportate in quanto esistono già versioni di queste caratteristiche a risoluzione migliore.

Questa rete richiede due oggetti in input: un'immagine T_{1w} (128×128) e la corrispondente immagine LRSI (16×16 , 24×24 o 32×32).

L'output è costituito da un'immagine HRSI (128×128). Per il *training* della rete, come descritto nel paragrafo seguente, si utilizzano immagini T_{1w} , HRSI e LRSI generate appositamente. La funzione di costo è rappresentata dall'errore quadratico medio:

$$MSE = \sum \sum \frac{(O - HRSI)^2}{m^2} \quad (3.3)$$

dove O è l'output della rete, $HRSI$ è la vera immagine ad alta risoluzione prodotta dal generatore (detta *ground truth*) e associata alla LRSI di input e m è la dimensione (in pixel) dell'output (128 in questo caso). La somma è eseguita su entrambe le dimensioni spaziali dell'immagine.

Essendo le immagini LRSI caratterizzate da risoluzioni diverse, si sono in realtà rese

necessarie nell'esperimento tre diverse reti D-UNet, da usare a seconda della risoluzione dell'input.

3.2.2 Dataset, generatore SI e *training*

Per sopperire alla mancanza di grandi quantità di dati e immagini spettroscopiche, necessari per il *training* della D-UNet, si è impiegato un generatore di immagini spettroscopiche.

Per prima cosa si è proceduto con una segmentazione di immagini assiali T₁w (27 provenienti da dataset MATLAB e 416 dal progetto OASIS [41]) in base all'intensità.

Una volta determinate le intensità massima della sostanza bianca (WM_{max}) e minima della sostanza grigia (GM_{min}), si sono prodotte immagini WM e GM nel modo seguente:

$$WM = \frac{S - GM_{min}}{WM_{max}} \cdot M \quad (3.4)$$

$$GM = (1 - WM) \cdot M \quad (3.5)$$

dove S è l'intensità del segnale originale e M è una maschera per selezionare la regione del cervello. In seguito, il generatore ha prodotto un'immagine T₁w *aumentata* (*augmented*, aT₁w):

$$aT_1w = R(S_n^{r_1} + L) \quad (3.6)$$

dove S_n è il segnale dell'immagine T₁w normalizzato e r_1 è un numero casuale compreso tra 0.25 e 0.5; $R()$ è una trasformazione che ruota l'immagine con un angolo estratto casualmente da una distribuzione uniforme tra -15° e 15° e la tronca casualmente in ogni direzione di un numero di pixel compreso tra 0 e 40; infine, L è una matrice di dati che rappresentano delle lesioni (fino a 6), di intensità, posizione e dimensioni diverse. La stessa trasformazione è stata applicata anche alle corrispondenti immagini WM e GM. L'operazione di *augmentation* serve per generare dati aggiuntivi a partire dalle immagini del dataset, che in questo modo viene ampliato. Assumendo poi che il metabolismo (e quindi i segnali) delle regioni WM e GM fossero correlati alla concentrazione, un'immagine ad alta risoluzione è stata ottenuta secondo la seguente espressione:

$$HRSI = r_2 \cdot WM + (1 - r_2) \cdot GM + B + r_3 \cdot L \quad (3.7)$$

dove r_2 e r_3 sono due numeri casuali compresi rispettivamente tra 0 e 1 e tra -1 e 1, mentre B è una matrice che aggiunge un segnale di bias in maniera casuale.

Infine, la HRSI è stata sottoposta ad un troncamento nello spazio k per produrre la corrispondente LRSI.

È utile notare che, in virtù della variabilità di r_1 , r_2 e r_3 , è stato possibile ottenere diverse terne di immagini aT_{1w}, HRSI e LRSI corrispondenti da una stessa immagine T_{1w} di input e, inoltre, diverse coppie di HRSI e LRSI fissata l'immagine aumentata aT_{1w}. Ciò ha permesso di avere molti dataset prodotti da una singola immagine T_{1w} disponibili per il *training* della D-UNet (si veda **Figura 3.5**).

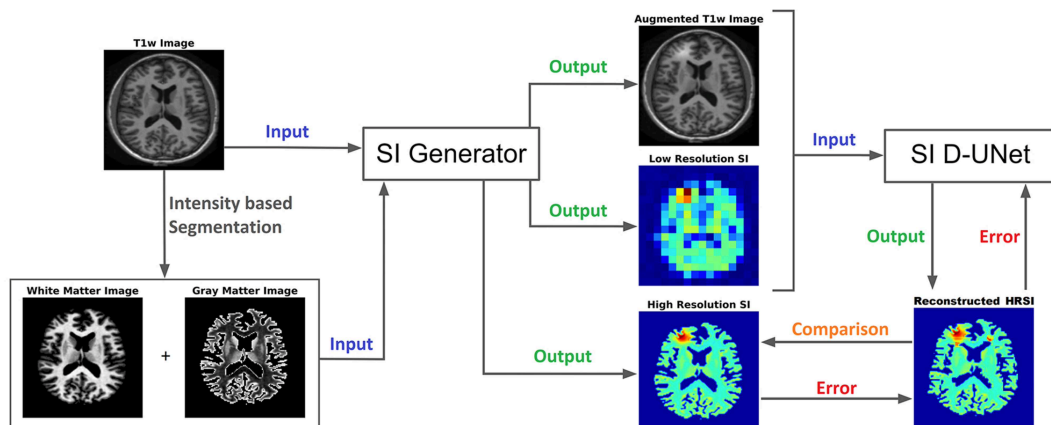


Figura 3.5: Schema della fase di *training*. Il generatore SI riceve in input le immagini T_{1w}, WM e GM e produce in output una HRSI, una LRSI e una aT_{1w}. Queste ultime due vengono trasformate dalla D-UNet nell'immagine HRSI ricostruita. Il confronto di questa con il *ground truth* produce un errore che viene passato alla rete e permette l'aggiornamento dei parametri. Alla fine del processo di *training*, i parametri per la ricostruzione saranno ottimizzati e gli errori minimizzati.

3.2.3 Procedure

Per il processo di *training* si sono utilizzate 102000 coppie di HRSI-LRSI prodotte dal generatore a partire da 135 immagini assiali: ognuna delle tre D-UNet è stata addestrata per 102 epoche (un'epoca è costituita, in questo esperimento, da un set di 1000 coppie HRSI-LRSI diverse).

Per quel che riguarda la fase di *testing*, invece, si sono prodotte 169 terne di immagini aT_{1w}, HRSI, LRSI a partire da 169 immagini assiali disponibili dal progetto OASIS; la bontà della ricostruzione delle immagini è stata valutata con il calcolo dell'errore quadratico medio (MSE), analogamente a quanto esposto per il *training* nel paragrafo precedente. Il confronto è stato ripetuto più volte con diversi livelli di rumore (frazioni dell'intensità massima del segnale) inseriti nella LRSI, con l'obiettivo di determinare l'effetto sulla ricostruzione. Questa valutazione assume particolare importanza nell'ottica di una possibile applicazione della procedura in ambito clinico, dove i dati di MRS sono

tipicamente affetti da alto rumore.

Lo studio ha incluso anche una valutazione della performance della rete nel caso di dati ottenuti *in vivo*. Pertanto, con uno scanner 7 T sono stati acquisiti i segnali provenienti da 10 volontari sani e, in seguito alla soppressione e alla quantificazione con LCM, sono state prodotte mappe spettroscopiche ad alta risoluzione (128×128 pixel, 1.7×1.7 mm²). Queste HRSI sono state in seguito sottoposte a *down-sampling* così da ottenere LRSI 32×32 , che, insieme alle corrispondenti immagini T_{1w}, hanno costituito l’input della D-UNet.

Infine, è stata studiata anche l’abilità della rete nella ricostruzione spettrale dei dati. A partire da immagini T_{1w} del *testing set* sono stati prodotti, grazie a una simulazione GAMMA [42], gli spettri dei principali metaboliti (NAA, Glu, Gln, Cr, Cho, mI); da qui, sovrapponendo i dati sulle ampiezze con le immagini anatomiche si sono ottenute le HRSI e le LRSI corrispondenti. In maniera analoga a quanto esposto prima, le D-UNet hanno ricostruito le immagini ad alta risoluzione e in particolare gli spettri, oggetto di confronto con i *ground truth*.

3.2.4 Risultati

Nella fase di *testing*, le ricostruzioni ad opera della D-UNet sono state confrontate con quelle effettuate da due metodi di *upscaling* convenzionali, l’interpolazione *zero-filling* e l’interpolazione bicubica (**Appendice A**). Si è visto che, mentre il metodo *zero-filling* superava le performance della D-UNet e dell’interpolazione bicubica nel caso 16×16 , la D-UNet risultava migliore nei casi 24×24 e 32×32 . La **Figura 3.6** riporta un esempio di ricostruzione a partire da un’immagine T_{1w}.

Per la valutazione quantitativa si rimanda alla **Tabella 3.2**, dove sono presentati i valori degli errori quadratici medi (MSE) per i diversi metodi impiegati e per i diversi livelli di rumore: si osserva come la performance della D-UNet superi quelle dei metodi tradizionali per immagini alla stessa risoluzione (32×32).

Metodo	Rumore (frazione dell’intensità massima)		
	Basso (2-5 %)	Medio (15-20 %)	Alto (30-40 %)
16×16 (Zero-fill)	1.109	1.652	4.505
16×16 (Bicubica)	2.794	3.129	3.820
16×16 (D-UNet)	1.863	2.420	2.761
24×24 (D-UNet)	1.139	1.316	1.745
32×32 (D-UNet)	0.7460	0.9722	1.599

Tabella 3.2: Valori di MSE nella fase di *testing*. Si nota come D-UNet 32×32 superi nettamente gli altri metodi; inoltre, aumentando il livello di rumore a medio-alto, anche le D-UNet 16×16 e 24×24 funzionano meglio dei metodi tradizionali.

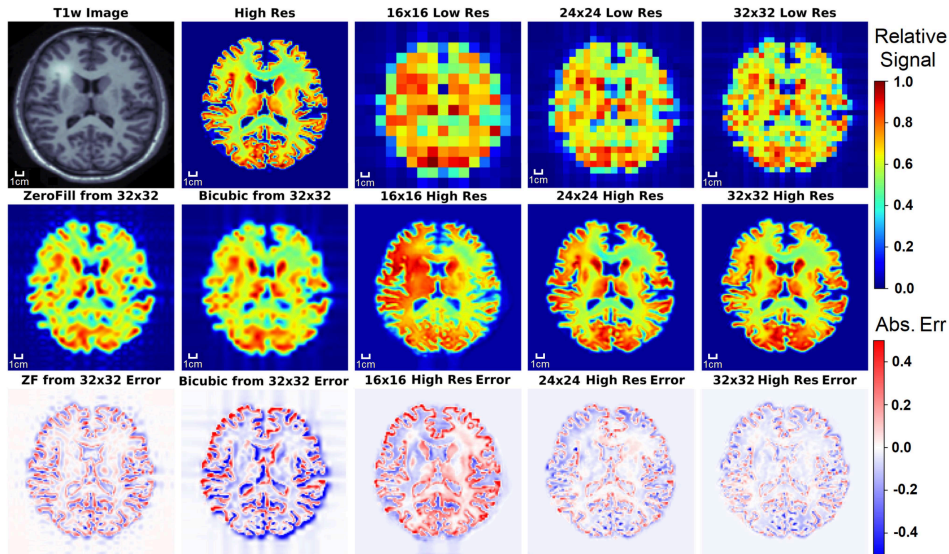


Figura 3.6: Risultato della D-UNet. Sono riportate anche le ricostruzioni con le interpolazioni *zero-filling* e bicubica per l'immagine 32×32 . Osservando l'ultima riga, che raffigura gli errori di ricostruzione sulla sezione assiale del cervello in scala di colori, si nota come la D-UNet comporti un errore minore rispetto ai metodi *zero-filling* e bicubico.

Anche i risultati dell'utilizzo della D-UNet sui dati acquisiti *in vivo* (**Figura 3.7**) hanno dimostrato che il metodo di ricostruzione è accurato. Infatti, le immagini ricostruite presentano variazioni regionali di concentrazione dei metaboliti simili a quelle delle HRSI. Anche i risultati quantitativi confermano le osservazioni. Ad esempio, è noto come la concentrazione di Glu sia maggiore nella sostanza grigia, mentre quella di Cho sia maggiore in quella bianca. In **Figura 3.8** sono stati riportati i rapporti Glu/Cr e Ch/Cr mediati sui dieci soggetti per valori differenti di intensità dell'immagine T_1w , che è proporzionale alla quantità di sostanza bianca in base alle relazioni stabilite nell'elaborazione del modello dell'esperimento. Si nota come all'aumentare della frazione di sostanza bianca, il rapporto Glu/Cr diminuisca mentre il rapporto Ch/Cr cresca.

Inoltre, per entrambi i rapporti il coefficiente di determinazione R^2 tra HRSI e ricostruzioni è risultato essere maggiore di 0.99.

Infine, in **Figura 3.9** è riportato l'esempio di una ricostruzione spettrale in due punti, corrispondenti ad una regione sana e ad una lesione. Si osserva che, nel pixel corrispondente a sostanza bianca sana, la D-UNet 16×16 tende a sottostimare le ampiezze (fino a 20% in meno per la colina). Nel pixel sulla lesione, invece, le reti 24×24 e 32×32 sovrastimano le concentrazioni; per la colina (che è presente in quantità anormale), questo errore coinvolge tutta l'area della lesione e produce uno scostamento notevole tra le ricostruzioni e il *ground truth*.

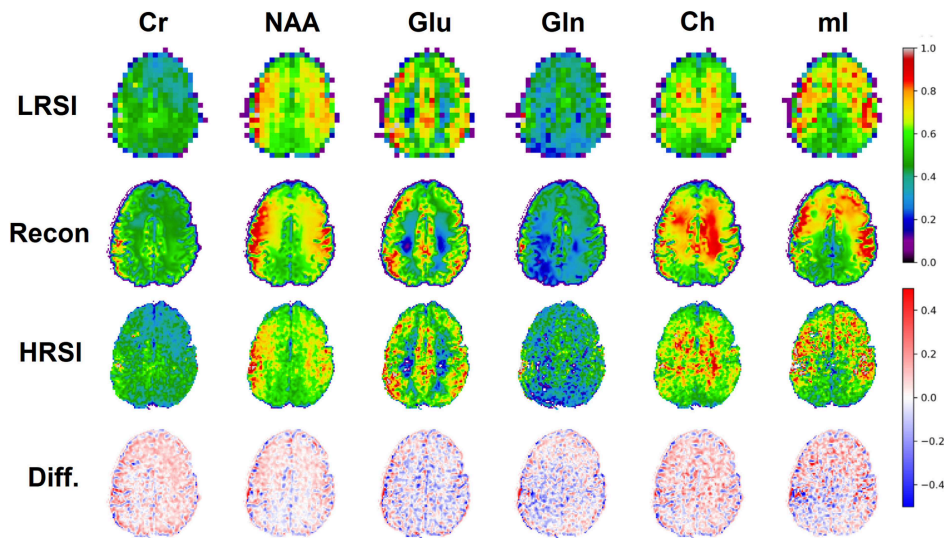


Figura 3.7: Esempio di applicazione della rete D-UNet sulle SI *in vivo* di un volontario sano per alcuni metaboliti. Le HRSI sono state acquisite con un protocollo accelerato e in seguito sono state sottoposte a *down-sampling* per ottenere le LRSI 32x32. Insieme alle ricostruzioni, compaiono anche le differenze di queste rispetto alle HRSI.

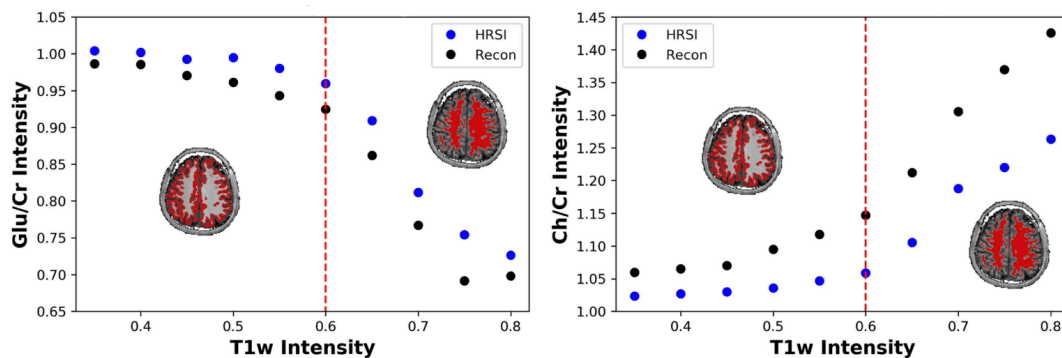


Figura 3.8: Rapporti medi Glu/Cr (sinistra) e Ch/Cr (destra) delle immagini HRSI e delle ricostruzioni in funzione dell'intensità dell'immagine T_1w ; la linea rossa verticale indica il punto in cui le frazioni di sostanza grigia e sostanza bianca coincidono.

3.2.5 Discussione

Il lavoro presentato ha dimostrato la capacità di questa tecnica di superare uno dei limiti fondamentali dell'*imaging* spettroscopico, vale a dire la scarsa risoluzione spaziale. A partire da immagini LRSI e T_1w , si sono ottenute ricostruzioni ad alta risoluzione molto simili al *ground truth* per quel che riguarda i risultati quantitativi. Il vantaggio fondamentale apportato da questa tecnica di ricostruzione è il risparmio di tempo: teori-

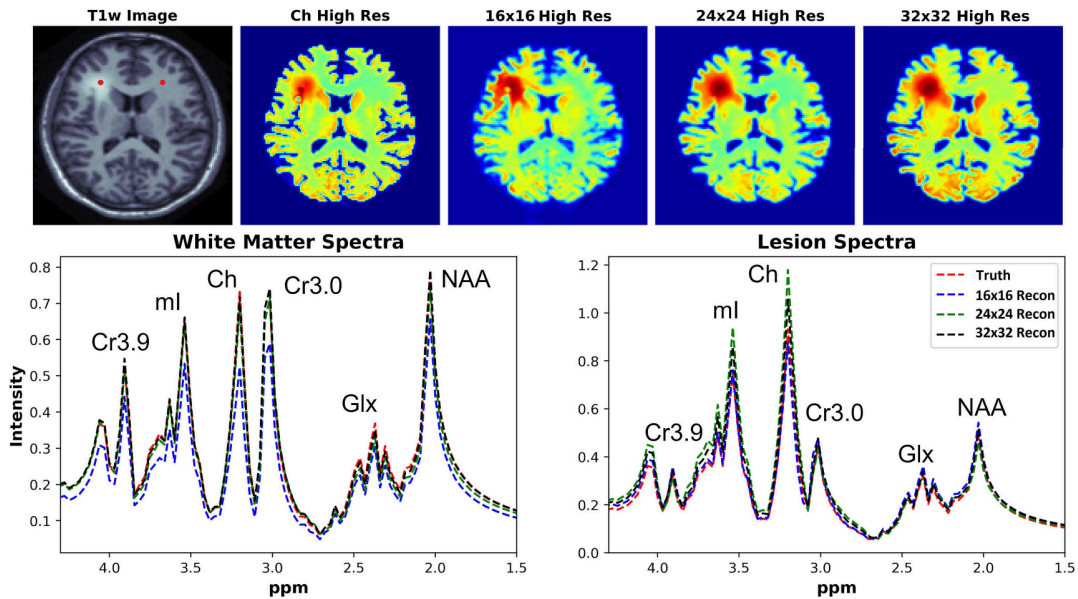


Figura 3.9: Ricostruzioni delle ampiezze (intensità) spettrali in due pixel di immagini ricostruite (in basso) e immagini spettroscopiche della colina (in alto). Nella lesione, le D-UNet 24×24 e 32×32 sovrastimano la concentrazione di Cho, come ben visibile sia nelle SI che negli spettri; nella regione sana, invece, la 16×16 sottostima l'intensità del segnale.

camente, l'acquisizione di immagini $Recon_{32 \times 32}$ dura tra il 75 e il 94% in meno di quella delle HRSI. La D-UNet potrebbe essere inoltre utilizzata per accelerare i protocolli di acquisizione.

Allo stesso tempo, però, la ricostruzione degli spettri si è mostrata poco soddisfacente, in quanto la presenza di alcuni metaboliti risultava alterata (sia sottostimata che sovrastimata), sia nelle regioni sane che in quelle corrispondenti a lesioni tumorali.

Occorre fare alcune precisazioni volte a circoscrivere l'ambito di validità dell'esperimento: se i risultati *in vivo* hanno mostrato valori bassi di MSE, questo potrebbe dovuto al fatto che l'immagine LRSI è stata ottenuta dalla HRSI tramite un *down-sampling*; ciò riduce la possibilità di interferenze, artefatti e contaminazioni del segnale, che comportano errori di ricostruzione. Nell'ottica di immagini SI a bassa risoluzione acquisite sperimentalmente, si prospettano naturalmente errori più grandi.

Per concludere, il test sulla ricostruzione spettrale ha mostrato che la rete è in grado di riprodurre le alterazioni nella concentrazioni dei metaboliti, ma anche che queste non sono molto accurate. Questo ovviamente costituisce un limite significativo per l'applicabilità.

3.3 Quantificazione relativa dei metaboliti

Attualmente, i due metodi più utilizzati per la quantificazione delle concentrazioni dei metaboliti sono il *Linear Combination Model* (LCM) e il *peaking*; tra i due, il LCM mostra una maggiore efficacia, facilità di utilizzo e adattabilità.

Shamaei et al. [43] hanno sviluppato e testato un algoritmo di LCM, implementato tramite DL, per la quantificazione di dati di MRS. In particolare, hanno sviluppato una rete DAE (*deep autoencoder*) che codifica l'input all'interno di un *latent space* e in seguito lo ricostruisce in maniera auto-supervisionata.

3.3.1 Modellizzazione

Il segnale di MRS (nel dominio temporale) può essere descritto come una combinazione lineare dei segnali dei metaboliti, di un segnale di base (che include il contributo delle macromolecole) e del rumore:

$$S(t) = \left(\sum_{m=1}^M A_m X_m(t) + A_b B(t) \right) e^{(\Delta\alpha + 2i\pi\Delta f)t} e^{i\Delta\theta} + \xi(t) \quad (3.8)$$

dove A_m e $X_m(t)$ sono rispettivamente il fattore di scala dell'ampiezza e la funzione modello di base per il m -esimo metabolita; A_b e $B(t)$ sono l'ampiezza e il segnale delle macromolecole (MM); $\Delta\alpha$, Δf , $\Delta\theta$ sono rispettivamente il fattore globale di smorzamento, lo shift di frequenza e lo shift di fase; $\xi(t)$ è il rumore.

In questo lavoro, il metodo di quantificazione LCM è stato assimilato ad un *deep autoencoder*: i parametri del modello costituiscono i parametri del *latent space*, mentre gli spettri di base e le conoscenze a priori rappresentano il risultato della decodifica.

La struttura del DAE utilizzato in questo studio è presentata in **Figura 3.10**. L'*encoder* è costituito da una successione di otto *layer* di convoluzione 1D seguiti da funzioni di attivazione GELU e da *layer* FC; il *decoder* consta di una funzione modello che ricostruisce il segnale di input usando i parametri del *latent space* e gli spettri di base. È da notare la presenza di una funzione Softplus [19], applicata alle ampiezze per evitare che queste risultino negative:

$$\text{Softplus}(x) = \ln(1 + e^x) \quad (3.9)$$

La funzione di costo che deve essere minimizzata è l'MSE.

Il vantaggio fondamentale di questo approccio consiste nel fatto che non si rende necessario conoscere i valori *ground truth* per il *training*; la rete minimizza le differenze tra l'input originale e la ricostruzione. Questo è il motivo per cui si parla di apprendimento auto-supervisionato: la rete comprende autonomamente l'output di riferimento, senza la necessità di *labels* sui segnali di ingresso.

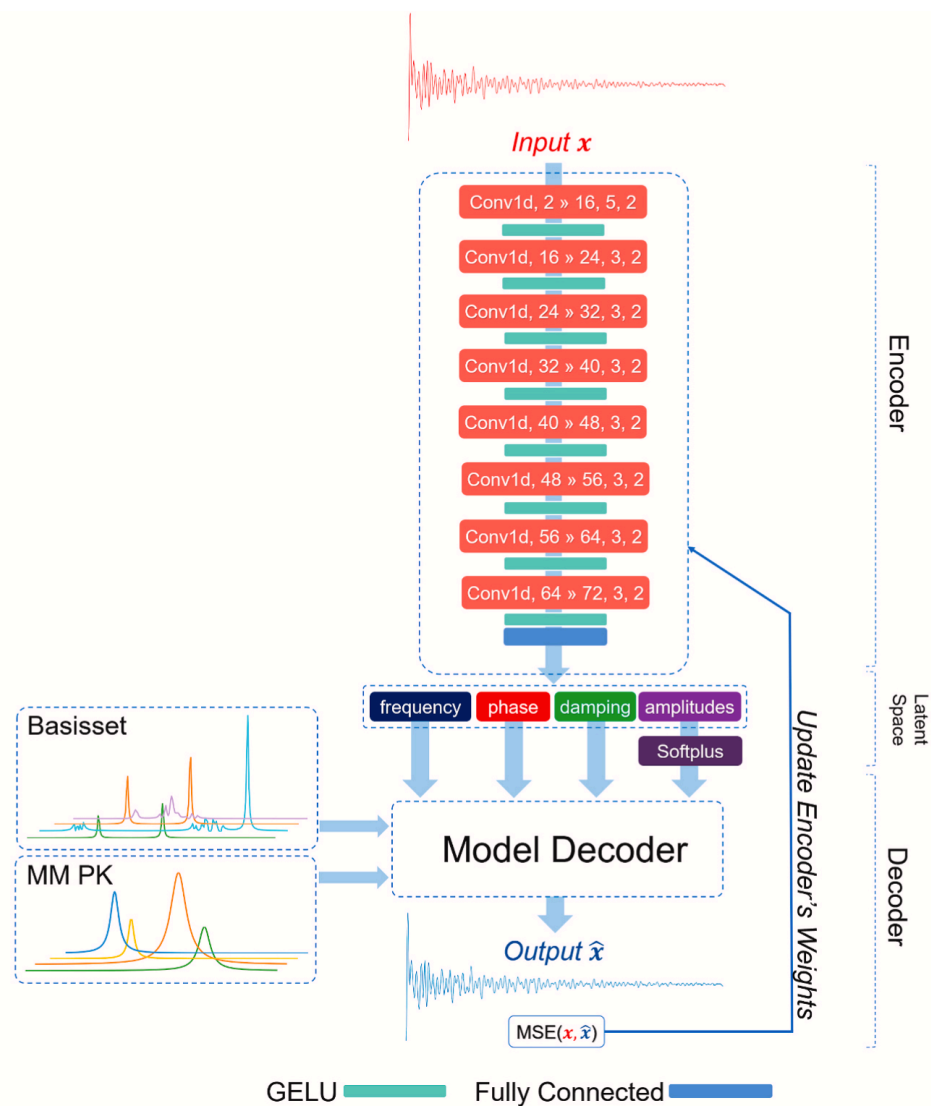


Figura 3.10: Rappresentazione della rete DAE. L'input è un segnale complesso in funzione del tempo (FID), che viene ricostruito sfruttando l'**Equazione 3.7**, gli spettri di base e le conoscenze a priori sulle macromolecole (*macromolecule prior knowledge*, MM PK). La notazione Conv1d $i \gg o, k, s$ sta ad indicare la dimensione dell'input (i), quella dell'output (o), quella del *kernel* (k) e il passo (s) del *layer* convoluzionale.

In questo studio sono stati presi in considerazione tre approcci diversi per la gestione del segnale proveniente dalle MM:

- quantificazione tramite pattern numerici (DQ-nMM);

- quantificazione tramite una combinazione di componenti parametriche relative alle diverse MM (DQ-pMM);
- quantificazione tramite componenti parametriche MM (incluse nell'**Equazione 3.7**) e un termine di regolarizzazione che viene aggiunto alla funzione di costo:

$$\left\| \left(x_{input}(t) - \hat{x}_{metaboliti}(t) \right) \cdot \left(\frac{1 + \text{sgn}(t - t_r)}{2} \right) \right\|_2^2 + \lambda \|t_r\|_1 \quad (3.10)$$

dove sgn è la funzione segno, t_r è un parametro temporale determinato dalla rete, x_{input} e $\hat{x}_{metaboliti}(t)$ sono i segnali di input e ricostruito, λ è un parametro che caratterizza la rilevanza del termine di regolarizzazione. Il primo termine dà importanza alla coerenza tra l'input e la ricostruzione; t_r deve essere grande abbastanza da far sì che il contributo delle MM sia trascurabile. Al tempo stesso, però, la rete deve evitare di tagliare la regione informativa del FID e considerare le parti costituite prevalentemente da rumore; questo è il motivo della presenza del secondo termine, che fa aumentare la funzione di costo per istanti di tempo troppo lontano dall'inizio del FID.

3.3.2 Dataset

Per le fasi di *training* e *testing*, è stato simulato un dataset a partire dagli spettri prodotti dal gruppo di ricerca dell'ISMRM (*International Society for Magnetic Resonance in Medicine*) sul fitting dei dati di MRS [44]; si è ottenuto un numero di segnali pari a 96000. I valori dei parametri che compaiono nell'**Equazione 3.7** sono stati estratti da una distribuzione uniforme, all'interno di range definiti da studi precedenti. Per dimostrare l'applicabilità del modello alla quantificazione *in vivo*, sono stati utilizzati anche i dati della repository pubblica Big GABA [45, 46]; in particolare, sono stati scelti 48 dataset provenienti da 48 volontari sani (acquisizioni a 3 T, *single-voxel*, sequenza PRESS).

Per ampliare il numero di segnali disponibili, si è praticata una *data augmentation*: come già anticipato nella sezione precedente, si tratta di una tecnica che permette di simulare dati realistici effettuando piccole modifiche ad un dataset già esistente e di dimensioni ridotte. Rimandando allo studio per la descrizione dettagliata del procedimento, per il caso *in vivo* si sono ottenuti un *training dataset* e un *testing dataset* costituiti rispettivamente da 28864 e 320 segnali. Questi ultimi derivano dai dati di soli cinque soggetti ($320 = 5 \times 64$).

3.3.3 Procedure

Tutte e tre le reti (DQ-nMM, DQ-pMM, DQ-rpMM) sono state addestrate e testate sul dataset simulato; solo la DQ-rpMM è stata testata anche sui dati *in vivo* (320 segnali estratti casualmente dai 28864 disponibili).

Sono stati poi eseguiti dei test statistici per valutare le performance delle DAE sia in termini assoluti sia rispetto a quelle di due algoritmi tradizionali di tipo LCM, QUEST (inclusa l'opzione Subtract) e FiTAID (nei domini temporale e delle frequenze). Entrambi appartengono al pacchetto software jMRUI [47].

QUEST utilizza un set di base di segnali di metaboliti e MM e ha un approccio semi-parametrico; QUEST-Subtract, che rappresenta un'alternativa alla modalità standard di utilizzo, agisce secondo questa sequenza: separa il segnale del metabolita dal background, elabora un modello e lo adatta tramite un fit non lineare con il metodo dei minimi quadrati [48].

FiTAID (*Fitting Tool for Arrays of Interrelated Datasets*) è uno strumento che offre la possibilità di eseguire fit sia nel dominio dei tempi che in quello delle frequenze, di imporre restrizioni sui parametri e di scegliere la strategia di fit preferita [49].

Per tutti e tre i metodi (DAE, QUEST e FiTAID) si è voluto valutare la capacità di stima delle ampiezze, tramite il confronto con i valori veri (noti in quanto provenienti da un dataset simulato): sono stati calcolati il coefficiente di determinazione R^2 e l'errore assoluto medio in percentuale (MAPE).

Sono stati effettuati inoltre ulteriori test sui singoli modelli di DAE, tra cui un confronto fra le tre architetture (DQ-nMM, DQ-pMM, DQ-rpMM) e un'analisi di tipo Monte Carlo sulla DQ-nMM.

La DQ-nMM è stata anche sottoposta ad uno studio sull'effetto delle dimensioni del dataset. Ne sono stati simulati cinque di dimensioni differenti (1000, 12000, 24000, 96000, e 384000 elementi); per ognuno di essi il 90% è stato allocato per l'addestramento e il 10% per la validazione.

Infine, è stato preso in considerazione il termine di regolarizzazione dell'**Equazione 3.9**: è stata studiata la performance della DQ-rpMM per diversi valori di λ ($\{0, 1, 2, 3, 4, 6, 8\}$), su un dataset di 24000 elementi (anche in questo caso il 90% dedicato al *training* e il 10% alla validazione).

3.3.4 Risultati

La **Tabella 3.3** riporta i risultati dei test condotti sulle performance delle reti e dei metodi nella determinazione delle ampiezze, per alcuni dei metaboliti (dataset simulati).

Si osserva che la DQ-nMM ha prodotto risultati migliori di QUEST e FiTAID, per quel che riguarda R^2 e MAPE.

Anche i tempi di processing mostrano valori a favore delle reti DAE: 0.031, 0.507 e

		Cr	GABA	Gln	Glu	GPC
DQ-nMM	R^2	0.88	-3.89	-0.53	0.68	0.67
	MAPE	6.02	33.04	19.11	8.88	16.27
DQ-pMM	R^2	0.73	-71.35	-2.32	-2.01	0.23
	MAPE	9.07	101.61	31.10	29.88	24.01
DQ-rpMM	R^2	0.76	-57.36	-3.18	-1.56	0.40
	MAPE	8.55	101.28	33.49	27.31	21.97
FiTAID (freq.)	R^2	0.80	-26.13	-2.02	0.38	0.03
	MAPE	7.85	99.65	25.95	13.01	27.32
FiTAID (FID)	R^2	0.69	-26.27	-1.94	0.44	0.03
	MAPE	8.43	100.00	25.51	12.13	27.39
QUEST	R^2	0.83	-10.07	-1.84	0.55	0.10
	MAPE	7.12	52.73	24.54	11.05	26.06
QUEST (Sub)	R^2	0.35	-72.01	-6.71	-1.27	-0.22
	MAPE	15.45	146.23	47.05	24.54	29.80
		Ins	NAA	NAAG	PCho	PCr
DQ-nMM	R^2	0.71	0.98	0.46	0.62	0.54
	MAPE	8.30	2.96	19.15	18.22	9.12
DQ-pMM	R^2	0.21	0.56	-5.31	0.34	-0.70
	MAPE	14.89	12.17	78.63	25.56	18.07
DQ-rpMM	R^2	0.45	0.57	-4.00	0.50	-0.39
	MAPE	12.06	11.55	68.75	21.05	16.63
FiTAID (freq.)	R^2	0.34	0.96	0.24	-0.06	0.13
	MAPE	12.80	3.50	25.84	27.02	12.29
FiTAID (FID)	R^2	0.60	0.95	0.11	-0.01	-0.47
	MAPE	10.57	3.92	27.52	27.15	13.55
QUEST	R^2	0.35	0.97	0.30	0.10	0.19
	MAPE	12.88	3.36	24.84	25.75	11.79
QUEST (Sub)	R^2	-0.26	0.51	-2.19	-0.30	0.00
	MAPE	17.84	14.38	56.88	31.95	13.40

Tabella 3.3: Risultati del *testing* dei metodi di quantificazione: R^2 e MAPE (percentuale).

0.517 *ms* per segnale per DQ-nMM, DQ-pMM, DQ-rpMM rispettivamente; 2500, 2500, 560, 1518 *ms* per segnale per FiTAID (tempi), FiTAID (frequenze), QUEST, QUEST-Subtract rispettivamente.

La **Figura 3.11** illustra i plot delle concentrazioni stimate dai diversi metodi, in funzione dei *ground truth*, per alcuni metaboliti; è da leggersi con riferimento alla legenda della **Figura 3.12**.

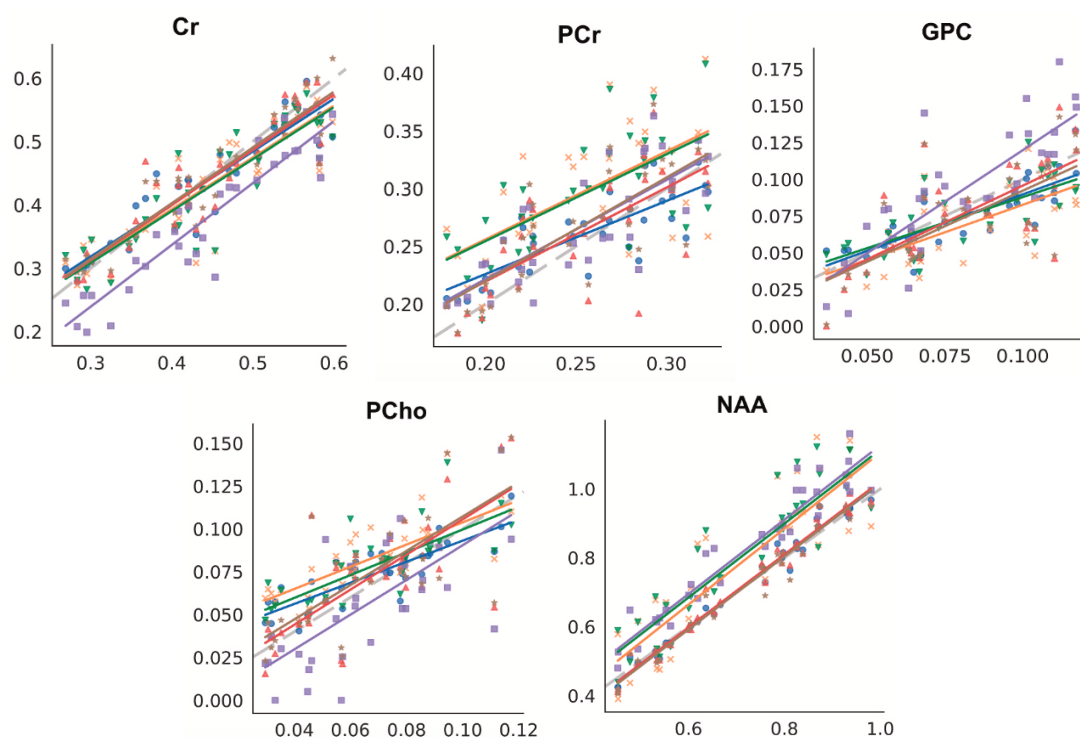


Figura 3.11: Plot delle concentrazioni di Cr, PCr, GPC, PCho e NAA. In ascissa stanno i valori *ground truth*, mentre in ordinata quelli stimati dai diversi metodi utilizzati. Questi ultimi si differenziano per il simbolo e il colore della resa grafica; per distinguerli si faccia riferimento alla legenda della Figura successiva. La rete DQ-nMM ha mostrato i risultati migliori nella quantificazione di PCr e PCho, i cui picchi si sovrappongono con quelli di Cr e GPC e che pertanto non sono distinguibili *in vivo* con sequenze standard.

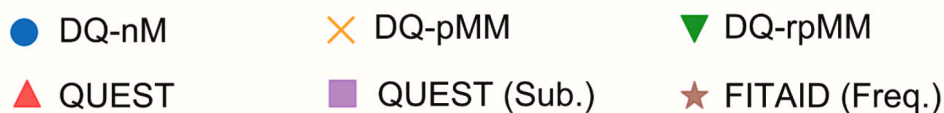


Figura 3.12: Legenda per il grafico di Figura.

	tNAA		tCho		Glx	
	Media	SD	Media	SD	Media	SD
1° individuo	1.69	0.025	0.25	0.005	1.28	0.016
2° individuo	1.69	0.017	0.259	0.005	1.402	0.033
3° individuo	1.729	0.021	0.205	0.022	1.424	0.013
4° individuo	1.616	0.065	0.233	0.018	1.288	0.073
5° individuo	1.656	0.043	0.237	0.005	1.351	0.053

Tabella 3.4: Medie e SD dei rapporti tra le concentrazioni dei metaboliti e quelle della creatina (totale) per i cinque individui nel *testing dataset*; ad ogni individuo corrispondono 64 segnali. tNAA: NAA + NAAG, tCho: Cho + PCho, Glx: Glu + Gln.

In generale, per i metaboliti con concentrazioni elevate (Cr, NAA, Ins, Glu) le tecniche che usavano metodi numerici (DQ-nMM, QUEST, FiTAID) hanno mostrato buoni risultati; DQ-nMM ha ottenuto un buon risultato anche nella quantificazione di NAAG, che è resa complicata dalla bassa concentrazione e dalla sovrapposizione con NAA. DQ-pMM ha ottenuto le performance peggiori; DQ-rpMM, che utilizza come DQ-pMM un approccio parametrico, si è dimostrata migliore grazie ai termini di regolarizzazione dell’**Equazione 3.9**.

Per quel che riguarda i dati *in vivo*, sono stati calcolati media, deviazione standard e coefficiente di variazione per il rapporto tra la concentrazione di ogni metabolita e quella della creatina (**Tabella 3.4**). Per il tNAA in particolare, le medie risultano coerenti tra i diversi soggetti e le deviazioni standard sono le minime. Ovviamente, i dati fanno riferimento a Cr, NAA, Cho e Glx totali in quanto con sequenze di acquisizione standard a 3 T non è possibile separare distinguere gli spettri di Cr e PCr, di NAA e NAAG, di Cho e PCho e di Glu e Gln.

Infine, in **Figura 3.13** sono presentati i risultati dello studio sugli effetti della dimensione del *training dataset* (DQ-nMM) e del valore del parametro λ (DQ-rpMM). In generale, si osserva come la scelta di λ influenzi la performance dell’algoritmo, in maniera diversa (positivamente o negativamente) a seconda del metabolita. Per quel che riguarda la dimensione del *training dataset*, l’esperimento ha evidenziato come la DQ-rpMM necessiti di almeno 12000 elementi di *training* per poter eseguire quantificazioni ottimali; al tempo stesso, però, si nota una leggera diminuzione del valore di R^2 per dataset molto grandi.

3.3.5 Discussione

Lo studio ha presentato un approccio alla quantificazione dei metaboliti tramite un modello DL di apprendimento auto-supervisionato e ha dimostrato che è possibile integrare

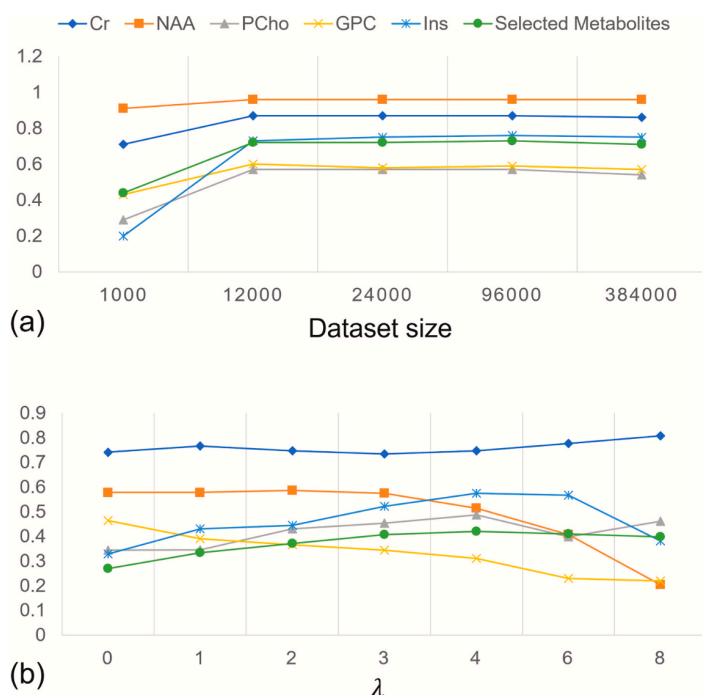


Figura 3.13: Andamento di R^2 in funzione del numero di elementi del *training dataset* per la DQ-nMM (a) e del valore numerico di λ per la DQ-rpMM (b), per ognuno dei cinque metaboliti in legenda e per l'insieme di questi. Il coefficiente è stato calcolato sul set di validazione.

questo algoritmo nelle tecniche LCM per quantificare dati sia simulati sia acquisiti *in vivo*.

Tuttavia, questi esperimenti presentano delle limitazioni che occorre tenere presente. Innanzitutto, le buone performance delle reti sono in parte riconducibili al fatto che il dataset di *training* e quello di *testing* sono stati simulati secondo gli stessi parametri. Pertanto, una quantificazione eseguita su spettri simulati secondo parametri differenti richiederebbe un *training* precedente specifico.

In secondo luogo, la simulazione potrebbe generare spettri che non sono rappresentativi di tutti i dati clinici, perché, come noto, esiste una grande variabilità di dati relativi alle concentrazioni dei metaboliti cerebrali; per di più questo sistema utilizza risposte dei metaboliti che sono state simulate attraverso calcoli quanto-meccanici *in silico*, dunque potrebbero non esaurire la complessità dei segnali *in vivo*.

Infine, il modello necessita di grandi quantità di dati per il *training*; ciò potrebbe compromettere la sua funzionalità in alcuni contesti in cui ottenere dataset estesi è complesso.

Capitolo 4

Conclusioni

La ricerca svolta in questo lavoro di tesi ha permesso di osservare in quali ambiti il *deep learning* possa intervenire per rendere la spettroscopia di risonanza magnetica *in vivo* una tecnica di più ampio utilizzo, ai fini diagnostici.

Lo studio di Bolan et al. [36] ha proposto un metodo computazionale che fa uso di una UNet per la segmentazione e collocazione automatizzata del voxel, all'interno di una lesione tumorale. È stato effettuato un confronto tra la performance della rete e quella manuale: il modello di DL si è dimostrato più abile nell'includere le zone delle lesioni nel VOI, ha selezionato voxel di dimensioni minori (migliori in termini di risoluzione spaziale) e ha ottenuto risultati più coerenti (deviazioni standard minori). Anche la segmentazione, che è un compito ormai svolto con regolarità dalle CNN, è risultata efficiente. Tuttavia, non è stata testata la qualità degli spettri acquisiti sul voxel così selezionato.

Per quel che riguarda la ricostruzione di immagini ad alta risoluzione, il lavoro di Iqbal et al. [40] ha utilizzato una D-UNet per la ricostruzione di immagini spettroscopiche ad alta risoluzione a partire da corrispondenti dati a bassa risoluzione. Introducendo livelli di rumore medio-alti, la ricostruzione della D-UNet presentava errori nettamente minori di quelli delle immagini a cui erano state applicate le interpolazioni bicubica e *zero-filling*. Ciononostante, non sono stati valutati tutti i metaboliti e la ricostruzione degli spettri non si è dimostrata particolarmente accurata.

Infine, il modello LCM implementato tramite autoencoder da Shamaei et al. [43] ha superato le performance degli algoritmi LCM tradizionali nella quantificazione dei metaboliti, sia in termini di tempi sia in termini di qualità del fit.

Nonostante questi risultati promettenti, si è osservato come queste applicazioni siano stadi preliminari nell'ambito della ricerca scientifica e siano ancora piuttosto lontane da un uso clinico effettivo. Ciò è principalmente dovuto al fatto che gli esperimenti considerati si sono svolti sotto alcune condizioni specifiche, che generalmente non rispecchiano la realtà della pratica clinica.

Innanzitutto, occorre sottolineare che una limitazione delle reti neurali è rappresentata dalla grande quantità di dati che servono per il *training*; grandi dataset non sono sempre disponibili, soprattutto nel caso di protocolli di acquisizione non standard. Per questo motivo, si fa uso di dataset simulati e sottoposti a *data augmentation*; tuttavia, questi non possono rappresentare tutta la grande variabilità che si può riscontrare nella realtà degli spettri. Come noto, infatti, le concentrazioni dei metaboliti possono variare significativamente tra individui diversi, senza che queste differenze siano riconducibili a situazioni patologiche.

Inoltre, i dataset sintetici sono prodotti tramite simulazioni quanto-meccaniche del comportamento dei metaboliti e, in generale, queste faticano a riprodurre la complessità dei segnali acquisiti *in vivo*; questi ultimi, infatti, seppur in minima parte, includono anche i contributi di macromolecole e lipidi.

Infine, la dipendenza dei modelli di DL dai particolari dataset usati per il *training* produce dei *bias*. Nell'ottica di applicazioni in ambito clinico, risulta di fondamentale importanza sviluppare dei metodi per individuare e correggere questi *bias*; una soluzione potrebbe essere affiancare modelli di DL a modelli basati sui dati.

Appendice A

Metodi di *upsampling*

A.1 *Zero-filling*

Il metodo *zero-filling*, conosciuto anche come *textitzero padding*, è utilizzato per aumentare il numero di voxel di griglie o immagini spettroscopiche ed è basato sulle proprietà matematiche della trasformata di Fourier (FT).

Come noto, durante l'acquisizione del segnale di NMR si mappa con un numero finito di step lo spazio ausiliario k . Il metodo di *zero-filling* consiste nell'aggiungere alla matrice dei dati punti dello spazio k (che corrispondono a punti di intensità nulla nel FID, da qui il nome). Con il calcolo della FT, questo intervento ha l'effetto di aumentare il numero di pixel mantenendo invariato il *field of view*.

Di fatto, questa operazione non aumenta realmente la risoluzione dell'immagine, bensì la rende più informativa e diminuisce la percezione della suddivisione in pixel.

In **Figura A.1** si riporta un esempio di applicazione del metodo *zero-filling* ad un'immagine del miocardio.

A.2 Interpolazione bicubica

L'interpolazione è una tecnica di analisi numerica che consente di determinare, sulla base del trend descritto da un insieme di punti del piano, nuovi punti che appartengano alla stessa funzione. Un esempio di interpolazione è quello della regressione lineare, vale a dire la determinazione di un polinomio di primo grado (una retta, nel piano cartesiano) che rappresenti al meglio un set finito di punti.

L'interpolazione bicubica è un'estensione del caso 1D, in quanto la funzione rappresentata è di terzo grado in due variabili (2D). Nell'ambito dell'*image processing* è una delle tecniche utilizzate per effettuare l'*upsampling* delle immagini, cioè l'aumento della risoluzione.

Si consideri una funzione f e le sue derivate parziali f_x , f_y , f_{xy} , che assumono valori noti

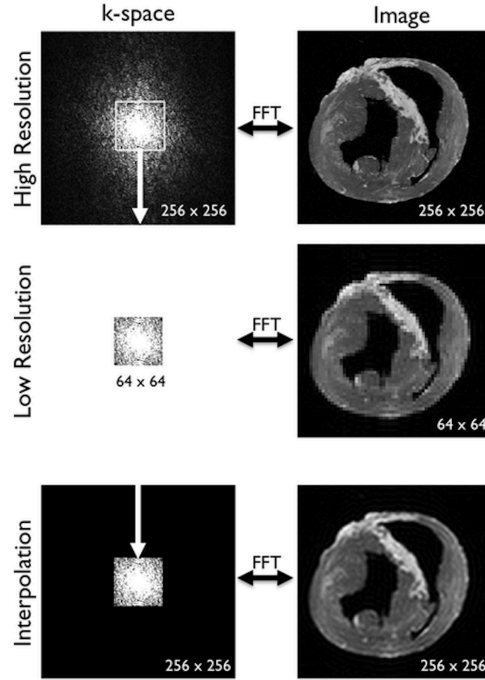


Figura A.1: Spazio k (sinistra) e corrispondenti immagini del miocardio (destra) [50]. In alto, è stato effettuato un campionamento ad alta risoluzione da cui è derivata tramite una trasformata di Fourier un'immagine 256×256 ; per produrre un'immagine a bassa risoluzione (64×64) sono stati selezionati solo i punti centrali dello spazio k . Infine, in basso, per riottenere un'alta risoluzione è stato applicato il metodo zero-filling all'immagine 64×64 , includendo nel campionamento dello spazio k punti aggiuntivi di intensità nulla (neri nell'immagine).

nei punti del piano $(0, 0)$, $(0, 1)$, $(1, 0)$ e $(1, 1)$. La superficie interpolante è una funzione cubica nelle due variabili x, y :

$$p(x, y) = f(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} x^i y^j \quad (\text{A.1})$$

L'interpolazione consiste nella determinazione dei 16 coefficienti a_{ij} , che si traduce in un sistema di 16 equazioni in 16 incognite:

$$\begin{aligned}
 f(0, 0) &= p(0, 0) = a_{00} \\
 f(1, 0) &= p(1, 0) = a_{00} + a_{10} + a_{20} + a_{30} \\
 &\dots \\
 f_x(0, 0) &= p_x(0, 0) = a_{10} \\
 &\dots \\
 f_y(0, 0) &= p_y(0, 0) = a_{01} \\
 &\dots \\
 f_{xy}(0, 0) &= p_{xy}(0, 0) = a_{11}
 \end{aligned}$$

Il risultato dell'interpolazione è una funzione con parametri noti; è possibile dunque valutare il valore delle variabili che identificano l'immagine in ulteriori punti dello spazio. Ciò permette di suddividere questa in un numero più grande di pixel e quindi avere una risoluzione migliore. In **Figura A.2** si riporta un confronto tra il metodo *nearest neighbor* e l'interpolazione bicubica; la seconda è caratterizzata da variazioni graduali delle tonalità, mentre la prima mantiene un effetto pixel [51].

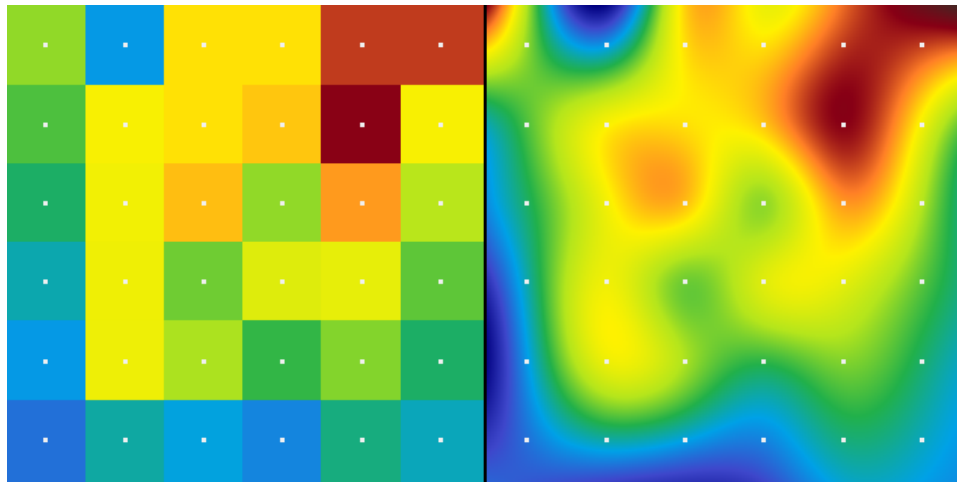


Figura A.2: Interpolazione *nearest neighbor* (sinistra) e interpolazione bicubica (destra) della stessa immagine [52].

Appendice B

Metodi statistici

B.1 Coefficiente di determinazione

Il coefficiente di determinazione, meglio noto come R^2 , misura la capacità di un modello di spiegare la variabilità dei dati in termini del modello statistico scelto. Più precisamente, è una stima della frazione di varianza della variabile dipendente che è prevista dal modello. Esistono diverse definizioni di R^2 ; la più generale è riportata a seguire.

Sia dato un insieme di valori della variabile dipendente: y_1, \dots, y_n ; ognuno di questi è associato ad un valore della funzione risultante dal fit: f_1, \dots, f_n . Si definiscono la *somma dei quadrati dei residui* (**Equazione B.1**) e la *devianza* (**Equazione B.2**):

$$SS_{res} = \sum_i (y_i - f_i)^2 = \sum_i e_i^2 \quad (\text{B.1})$$

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \quad (\text{B.2})$$

dove si introducono i residui $e_i = y_i - f_i$ e la media $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. Si definisce quindi il coefficiente di determinazione come:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (\text{B.3})$$

R^2 è un indicatore della bontà del fit. Assume, di norma, valori tra 0 e 1: il primo indica che il modello è totalmente incapace di prevedere la variabilità dei dati; il secondo stabilisce una perfetta aderenza dei dati al modello. La **Figura B.1** mostra tre esempi di regressione lineare con diversi valori di R^2 . Esistono casi in cui R^2 è minore di zero: questo accade quando la funzione modello è stata costruita secondo una conoscenza a priori e non sulla base dei dati [53].

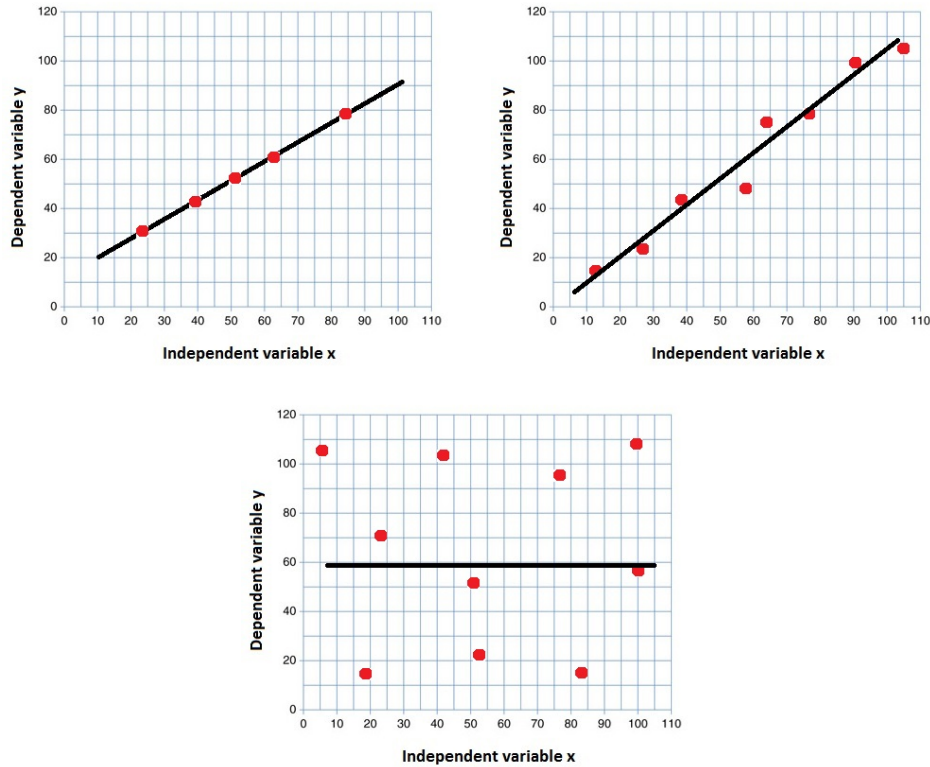


Figura B.1: Esempi di regressione lineare con: $R^2 = 1$ (a sinistra in alto), $R^2 = 0.83$ (a destra in alto) e $R^2 = 0$ (in basso) [54].

B.2 Errore medio assoluto percentuale

L'errore medio assoluto percentuale (*mean absolute percentage error*, MAPE) è un indicatore spesso usato come funzione di costo nei problemi di regressione e nella valutazione della bontà dei modelli. È definito come:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (\text{B.4})$$

dove A_t è il valore effettivo e F_t è il valore previsto.

In generale, il MAPE risulta meno indicativo di R^2 per alcuni motivi. Innanzitutto non può essere usato per valori nulli o molto vicino a zero, in quanto il denominatore risulterebbe molto piccolo e la frazione tenderebbe a valori molto grandi. In secondo luogo, il MAPE penalizza maggiormente gli errori negativi ($A_t < F_t$), quindi presenta un *bias* [55].

Bibliografia

- [1] A. J. Ross e P. S. Sachdev, «Magnetic resonance spectroscopy in cognitive research», *Brain Research Reviews* **44**, 83-102 (2004).
- [2] A. Anaya-Isaza, L. Mera-Jiménez e M. Zequera-Diaz, «An overview of deep learning in medical imaging», *Informatics in Medicine Unlocked* **26** (2021).
- [3] I. P. Gerothanassis, A. Troganis, V. Exarchou e K. Barbarossou, «Nuclear magnetic resonance (NMR) spectroscopy: Basic principles and phenomena, and their applications to chemistry, biology and medicine», *Chemistry Education Research and Practice* **3**, 229-252 (2002).
- [4] *Composition of the human body*, https://en.wikipedia.org/wiki/Composition_of_the_human_body.
- [5] P. K. Mandal, «Magnetic Resonance Spectroscopy (MRS) and its Application in Alzheimer's Disease», *Concepts in Magnetic Resonance Part A* **30A**, 40-64 (2007).
- [6] D. G. Gadian, *Nuclear magnetic resonance and its application to living systems*, 1^a ed. (Oxford University Press, New York, 1982).
- [7] *Tetramethylsilan* — *Wikimedia Commons, the free media repository* (2008), https://commons.wikimedia.org/wiki/File:Tetramethylsilan_stereo.svg.
- [8] M. H. Buonocore e R. J. Maddock, «Magnetic resonance spectroscopy of the brain: a review of physical principles and technical methods», *Reviews in the Neurosciences* **26**, 609-632 (2015).
- [9] *J-coupling*, <https://en.wikipedia.org/wiki/J-coupling>.
- [10] M. van der Graaf, «In vivo magnetic resonance spectroscopy: basic methodology and clinical applications», *European Biophysics Journal* **39**, 527-540 (2010).
- [11] A. Lin, T. Tran, S. Blumi, S. Merugumala, H. Liao e B. D. Ross, «Guidelines for Acquiring and Reporting Clinical Neurospectroscopy», *Seminars in Neurology* **32**, 432-453 (2012).

- [12] Y. Hwang, M. Lee, C. Yun, Y. Kim, H. Baek, B. S. Han e D. Y. Kim, «Dynamic Variation in Hippocampal Metabolism after Acute Stress Exposure: An In Vivo Proton Magnetic Resonance Spectroscopy Study at 9.4 T», *Journal of Spectroscopy* (2021).
- [13] S. W. Provencher, «Estimation of metabolite concentrations from localized in vivo proton NMR spectra», *Magnetic Resonance in Medicine* **30**, 672-679 (1993).
- [14] *Machine learning*, https://en.wikipedia.org/wiki/Machine_learning.
- [15] I. Goodfellow, Y. Bengio e A. Courville, *Deep Learning*, 1^a ed. (MIT Press, Cambridge MA, 2016).
- [16] *Training, validation, and test data sets*, https://en.wikipedia.org/wiki/Training_validation_and_test_data_sets.
- [17] W. Di, A. Bhardwaj e J. Wei, *Deep Learning Essentials*, 1^a ed. (Packt Publishing, Birmingham, 2018).
- [18] *Neuron and myelinated axon* — *Wikimedia Commons, the free media repository* (2019), <https://commons.wikimedia.org/wiki/File:Neuron3.svg>.
- [19] *Activation function*, https://en.wikipedia.org/wiki/Activation_function.
- [20] *Perceptron*, <https://en.wikipedia.org/wiki/Perceptron>.
- [21] L. Valente, «A Variational Autoencoder Application for real-time Anomaly Detection at CMS», tesi di laurea magistrale (Università di Bologna, 2022).
- [22] *Backpropagation*, <https://en.wikipedia.org/wiki/Backpropagation>.
- [23] *Vanishing gradient problem*, https://en.wikipedia.org/wiki/Vanishing_gradient_problem.
- [24] T. Angeli, «Applicazione delle reti neurali residuali ai mercati energetici e loro interpretazione come problema di controllo ottimo stocastico», tesi di laurea magistrale (Università di Bologna, 2021).
- [25] Z. Rguibi, A. Hajami, D. Zitouni, A. Elqaraoui e A. Bedraoui, «CXAI: Explaining Convolutional Neural Networks for Medical Imaging Diagnostic», *Electronics* **11** (2022).
- [26] R. Yamashita, M. Nishio, R. K. G. Do e K. Togashi, «Convolutional neural networks: an overview and application in radiology», *Insights into Imaging* **9**, 611-629 (2018).

- [27] M. Sheu, S. M. Salahuddin Morsalin, S. Wang, L. Wei, S. Hsia e C. Chang, «FHI-UNet: Faster Heterogeneous Images Semantic Segmentation Design and Edge AI Implementation for Visible and Thermal Images Processing», *IEEE Access* **10**, 18596-18607 (2022).
- [28] O. E. David e N. S. Netanyahu, «DeepPainter: Painter Classification Using Deep Convolutional Autoencoders», *Artificial Neural Networks and Machine Learning – ICANN 2016*, 20-28 (2016).
- [29] *What is Transposed Convolutional Layer?*, <https://www.geeksforgeeks.org/what-is-transposed-convolutional-layer/>.
- [30] A. Cole, A. McEwan e S. Singh, «An Analysis of Programmer Productivity versus Performance for High Level Data Parallel Programming», *Concurrent Systems Engineering Series* **68**, 111-130 (2011).
- [31] O. Ronneberger, P. Fischer e T. Brox, «U-Net: Convolutional Networks for Biomedical Image Segmentation», *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234-241 (2015).
- [32] *Residual neural network*, https://en.wikipedia.org/wiki/Residual_neural_network.
- [33] A. Singh e T. Ogunfunmi, «An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications», *Entropy* **24** (2022).
- [34] *Autoencoder*, <https://en.wikipedia.org/wiki/Autoencoder>.
- [35] J. Hwang e C. S. Choi, «Use of *in vivo* magnetic resonance spectroscopy for studying metabolic diseases», *Experimental & Molecular Medicine* **47** (2015).
- [36] P. J. Bolan, F. Branzoli, A. L. Di Stefano, L. Nichelli, R. Valabregue, S. L. Saunders, M. Akçakaya, M. Sanson, S. Lehéricy e M. Marjańska, «Automated Acquisition Planning for Magnetic Resonance Spectroscopy in Brain Cancer», *Medical Image Computing and Computer Assisted Intervention*, 730-739 (2020).
- [37] S. Lefkovitz e L. Lefkovitz, «U-Net architecture variants for brain tumor segmentation of histogram corrected images», *Acta Universitatis Sapientiae Informatica* **14**, 49-74 (2022).
- [38] *ImageNet*, <https://www.image-net.org/index.php>.
- [39] *Multimodal Brain Tumor Segmentation Challenge 2018*, <https://www.med.upenn.edu/sbia/brats2018/data.html>.

- [40] Z. Iqbal, D. Nguyen, G. Hangel, S. Motyka, W. Bogner e S. Jiang, «Super-Resolution ^1H Magnetic Resonance Spectroscopic Imaging Utilizing Deep Learning», *Frontiers in Oncology* **9** (2019).
- [41] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, R. L. Buckner, «Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults», *Journal of Cognitive neuroscience* **19**, 1498-1507 (2007).
- [42] S. A. Smith, T. O. Levante, B. H. Meier e R. R. Ernst, «Computer Simulations in Magnetic Resonance. An Object-Oriented Programming Approach», *Journal of Magnetic Resonance, series A* **106**, 75-105 (1994).
- [43] A. Shamaei, J. Starcukova e Z. Starcuk, «Physics-informed deep learning approach to quantification of human brain metabolites from magnetic resonance spectroscopy data», *Computers in Biology and Medicine* **158** (2023).
- [44] M. Marjańska, D. K. Deelchand, R. Kreis e 2016 ISMRM MRS Study Group Fitting Challenge Team, «Results and interpretation of a fitting challenge for magnetic resonance spectroscopy set up by the MRS study group of ISMRM», *Magnetic Resonance in Medicine* **87**, 11-32 (2022).
- [45] M. Mikkelsen, P. B. Barker, P. K. Bhattacharyya et al., «Big GABA: Edited MR spectroscopy at 24 research sites», *NeuroImage* **159**, 32-45 (2017).
- [46] M. Mikkelsen, D. L. Rimbault, P. B. Barker et al., «Big GABA II: Water-referenced edited MR spectroscopy at 25 research sites», *NeuroImage* **191**, 537-548 (2019).
- [47] *jMRUI* — *Software for the clinical and biomedical MRS*, <http://www.jmrui.eu/>.
- [48] D. Stefan, F. Di Cesare, A. Andrasescu, E. Popa, A. Lazariiev, E. Vescovo, O. Strbak, S. Williams, Z. Starcuk e M. Cabanas, «Quantitation of magnetic resonance spectroscopy signals: the jMRUI software package», *Measurement Science and Technology* **20** (2009).
- [49] D. G. Q. Chong, R. Kreis, C. S. Bolliger, C. Boesch e J. Slotboom, «Two-dimensional linear-combination model fitting of magnetic resonance spectra to define the macromolecule baseline using FiTAID, a Fitting Tool for Arrays of Interrelated Datasets», *Magnetic Resonance Materials in Physics, Biology and Medicine* **24**, 147-164 (2011).
- [50] H. Ashikaga, H. L. Estner, D. A. Herzka, E. R. Mcveigh e H. R. Halperin, «Quantitative Assessment of Single-Image Super-Resolution in Myocardial Scar Imaging», *IEEE Journal of Translational Engineering in Health and Medicine* **2** (2014).

- [51] *Bicubic interpolation*, https://en.wikipedia.org/wiki/Bicubic_interpolation.
- [52] *Beispiel Bikubische C2 Interpolation* — *Wikimedia Commons, the free media repository* (2018), https://commons.wikimedia.org/wiki/File:Beispiel_Bikubische_C2_Interpolation.png#mw-jump-to-license.
- [53] *Coefficient of determination*, https://en.wikipedia.org/wiki/Coefficient_of_determination.
- [54] *Coefficient of Determination, R-squared*, <https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/coefficient-of-determination-r-squared.html>.
- [55] *Mean absolute percentage error*, https://en.wikipedia.org/wiki/Mean_absolute_percentage_error.