

Data Science Lab in Medicine: Medical Imaging & Big Data

Federico Signoretta

May 10, 2020

Contents

1 Imaging Techniques	1
1.1 Imaging Process	1
1.2 Principali tecnologie di imaging	2
1.2.1 TAC	2
1.2.2 PET	3
1.2.3 PET/TAC (sistema ibrido)	4
1.2.4 Risonanza Magnetica (MRI)	5
2 Obiettivi del Medical Imaging	9
2.1 Passato	9
2.2 Presente	10
2.3 Futuro	11
3 Medical-image Quantification	11
3.1 Analisi qualitativa	12
3.2 Analisi quantitativa	13
3.3 Calcolo del volume della lesione: problematiche	14
3.3.1 Movimento	15
4 Medical-image Segmentation	17
4.1 Image thresholding methods	17
5 Machine Learning nel Medical-imaging	20
5.1 Support-Vector Machine	20
5.2 Applicazioni del ML nell'imaging medico	22
6 Pre-processing	24
7 Features extraction	24
7.1 Voxel-level approach	24
7.2 Anatomical measurements	24
7.3 Data-driven techniques	24
8 Features selection	25
8.1 A-priori (ROI)	25
8.2 Univariate feature selection	25
8.3 A-posteriori (Wrapped)	25

9 Training-Validation-Test-Bias	26
9.1 Training	26
9.2 Validation	26
9.3 Metriche	26
9.3.1 Accuracy	27
9.3.2 Sensitivity and specificity	27
9.3.3 ROC	27
9.4 Bias	27

1 Imaging Techniques

In questa sezione verranno presentate le principali tecniche di acquisizioni di immagini mediche (passato, presente e futuro). In particolare, verrà affrontata l'integrazione delle immagini con altre fonti nell'ambito medico, la quale risulta essere fondamentale per questo tipo di analisi.

Le *imaging techniques* nella medicina si basano sull'interazione delle radiazioni elettromagnetiche con la materia ed i tessuti organici. Nelle *basse frequenze* delle radiazioni elettromagnetiche, si ha l'**imaging ad ultrasuoni** (tra i 10^6 ai 10^8 Hz) le quali vengono sfruttate dalle apparecchiature ecografiche. La risonanza magnetica sfrutta le frequenze tra i 10^7 ai 10^8 Hz, i sistemi di termografia tra i 10^{12} ai 10^{14} Hz, i sistemi di imaging che si basano sulla fluorescenza stimolata (10^{14} Hz), la radiografia e la TAC (CT) tra i 10^{16} ai 10^{19} Hz ed, infine, la PET (tomografia ad emissione di positroni) tra i 10^{19} ai 10^{20} Hz.

Le immagini mediche derivanti da questi strumenti chiaramente conterranno informazioni molto diverse tra loro, poiché l'interazione tra le onde elettromagnetiche e la materia è differente: in particolare, è possibile notare dalla figura 1 che cambia sia la scala di colori che la risoluzione.

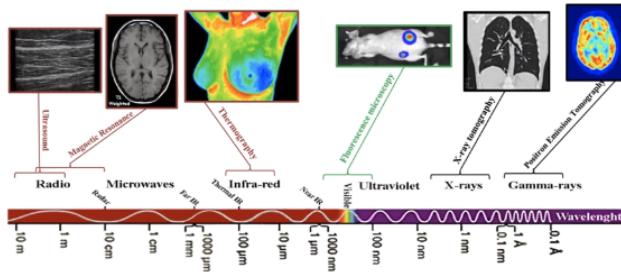


Figure 1:

1.1 Imaging Process

Il processo di ricostruzione dell'immagine è un processo molto complesso: si basa su sistemi di acquisizione dei segnali generati dall'interazione tra la radiazione elettromagnetica e materia. Sfortunatamente, non viene raccolta solo l'informazione utile, ma vengono raccolte altre informazioni non rilevanti che prendono il nome di disturbo. Sia il segnale che il rumore vengono elaborate nel tempo, nell'energia e nello spazio in profili di distribuzione del segnale a diverse proiezioni. Tali profili vengono mappati su una griglia in uno spazio bidimensionale in modo da ricostruire l'immagine.

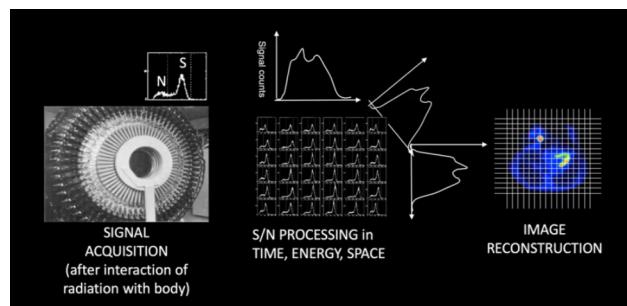


Figure 2: Acquisizione e ricostruzione dell'immagine

Nel caso della tomografia il discorso è equivalente, ma viene complicato ulteriormente, siccome si

avranno più strati di profili (più piani) per ricostruire l'immagine attraverso più strati di griglie bidimensionali: l'integrazione di queste informazioni crea il "tomo", ossia il volume che rappresenta la sezione corporea interessata.

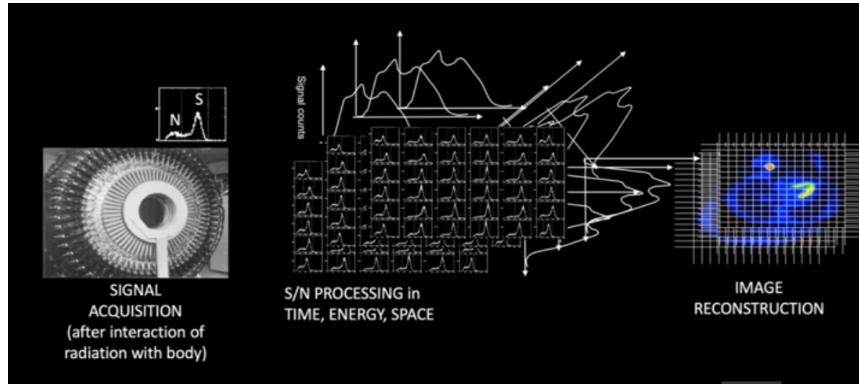


Figure 3: Acquisizione e ricostruzione dell'immagine nella tomografia

1.2 Principali tecnologie di imaging

1.2.1 TAC

Una delle tecniche più utilizzate negli ospedali è la tomografia a raggi-X. È composta da una sorgente di raggi-X (tubo), che ruota attorno al corpo che deve essere analizzato e emette delle radiazioni che vengono catturate da un multiple-detector, il quale riorganizza e trasmette le informazioni ad un calcolatore: attraverso questo processo vengono generate delle immagini bidimensionali della sezione di interesse (figura 4).

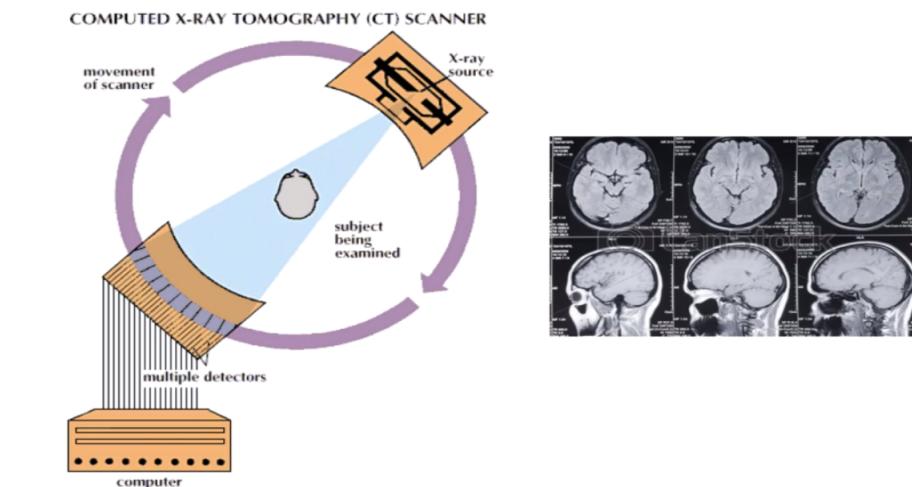


Figure 4: TAC (CT)

Per calcolare il **profilo di attenuazione** della tac vengono calcolati i raggi-X che attraversano ogni voxel dell'immagine: dato I_0 l'intensità iniziale nota, d la distanza dalla sorgente e μ_i il coefficiente di attenuazione di ciascun voxel, è possibile definire l'intensità in uscita (radiazione emessa) come:

$$I = I_0 \cdot e^{-d \sum_i^n \mu_i}$$

I dati relativi all'attenuazione della radiazione di ogni singolo voxel rappresenta la proiezione di un dato oggetto/corpo interposto tra la sorgente e il rilevatore di raggi-X.

Per la ricostruzione dell'immagine viene utilizzato un metodo chiamato **retro-proiezione** (Back-projection): vengono utilizzati degli algoritmi che riposizionano tutte le immagini registrate per poi tornare a quelle originali andando per esclusione.

Le immagini retro-proiettate presentano delle code lungo i profili dovuti all'attenuazione parziale della radiazione. Questa è responsabile di una sorta di sfocatura: per mitigare questo effetto viene filtrato mediante degli algoritmi (ad esempio filtri di convoluzione) prima di essere retro-proiettato. I **filtri** utilizzati dipendono dai tessuti che vengono utilizzati (tessuti ossei, etc.).

Dal punto di vista numerico l'immagine viene creata in base al profilo di attenuazione corrispondente, assegnando un valore a ciascun voxel (attenuazione media del volume del tessuto esaminato). Ogni valore del voxel viene assegnato ad un valore della matrice dell'immagine, generalmente formata da 512x512 pixel. Alla fine della ricostruzione dell'immagine ogni pixel corrisponde, nel caso della TAC, al numero di Hounsfield (la scala Hounsfield anche chiamata numero CT, è una scala di unità di misura usata per descrivere quantitativamente la radiodensità). La scala di grigi associata (8 bit - 256 valori di grigio) viene interpretata nel seguente modo: maggiore è l'attenuazione della radiazione più chiara è la scala verso il bianco; viceversa, minore è l'attenuazione, più scura è la scala verso il nero.

L'unità di Hunsfield (o il numero TAC) è un valore adimensionale proporzionale alla densità dei tessuti. Come unità principale ci si riferisce alla densità dell'acqua. I valori associati alle diverse densità dei tessuti della materia si trovano sopra o sotto a questa unità. La maggior parte dei tessuti liquidi e molli si trova in un range di valori compresi tra -100 e +100 (figura 5). Dunque, in questo tipo di analisi (TAC), l'informazione che si ottiene è la densità dei tessuti: gli organi e i tessuti possono essere iperdensi, ipodensi o isodensi in relazione ad un altro organo o tessuto o in relazione alla densità dell'acqua.

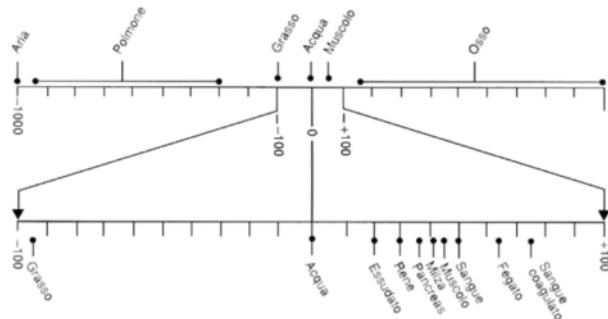


Figure 5: Scala Hunsfield

1.2.2 PET

La Tomografia ad Emissione di Positroni è uno degli strumenti più importanti nell'ambito medico (PET), ma poco utilizzata negli ospedali poiché molto complessa: viene somministrato al paziente un radio-farmaco, ossia una molecola che viene trattata sinteticamente in cui un elemento neutro viene sostituito con un elemento che emette particelle (che emette il positrone). In particolare, tale sistema raccoglie la distribuzione di radioattività tracciante nell'organo preso in esame.

I radio-farmaci somministrati al paziente si basano su diversi radio-isotopi, che differiscono per tempi di dimezzamento. Sono elementi presenti nel nostro organismo, per questo sono definiti naturali o pseudo-naturali (l'impatto sul nostro organismo è molto basso, interferiscono poco). Le produzioni di questi isotopi si basano su reazioni chimiche e nucleari.

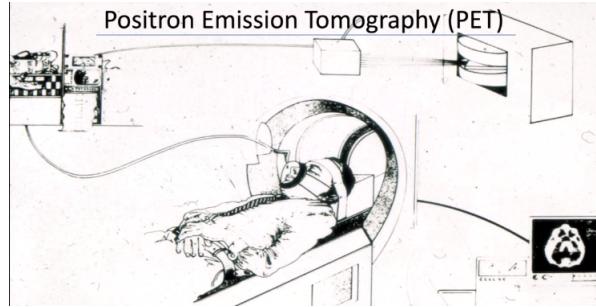


Figure 6: PET

Il macchinario in grado di produrre isotopi per la PET si chiama ciclotrone: è un sistema di accelerazione delle particelle che deve essere presente nelle vicinanze (o all'interno dell'ospedale stesso), a seconda dell'isotopo prodotto, in base ai tempi di dimezzamento. Questo spiega la minore diffusione di questo sistema negli ospedali.

ISOTOPES	$T_{1/2}$	
^{11}C	20.4 min	"natural"
^{13}N	10.0 min	"natural"
^{15}O	2.0 min	"natural"
^{18}F	109.8 min	"pseudo-natural"

Figure 7: Isotopi e tempi di dimezzamento

La PET si basa sul seguente principio fisico: è legato all'interazione del positrone emesso da un nucleo del tessuto del paziente (al quale è stato somministrato il radio-farmaco) con le particelle del tessuto che circonda il nucleo. Quindi, il positrone emesso può interagire con gli elettroni della materia circostante (materia dei tessuti molli, ossa, etc.) con delle interazioni diverse, generando due raggi gamma γ con direzione antiparallela e collineare, la cui energia è pari a 511 keV.

Come nella TAC, anche per la ricostruzione dell'immagine nella PET viene utilizzato il metodo di retro-proiezione (o combinazioni con essa, come i filtri) in modo da migliorare la combinazione tra rumore e segnale. L'intensità dell'immagine dipende dal numero di fotoni raccolti dopo l'emissione. La qualità dell'immagine dipende dai filtri, dagli algoritmi di ricostruzione: possono essere iterativi (OSEM) o mediante algoritmi di correzione (VUE POINT). La durata di acquisizione influenza sulla qualità dell'immagine, ma bisogna ovviamente acquisire una buona qualità in tempi ragionevoli, per non sovraccaricare il paziente.

1.2.3 PET/TAC (sistema ibrido)

Un altro sistema è il sistema ibrido, in cui vengono accoppiato il sistema TAC e il sistema PET. È costituito da uno scanner in cui sono co-presenti (adiacenti) sia TAC che PET. La differenza è che la PET si basa sull'emissione di radiazioni, mentre la TAC si basa sulla trasmissione. Per la PET, è il paziente (a cui è stato somministrato il radio-farmaco) ad essere sorgente ed emissione di radiazione.

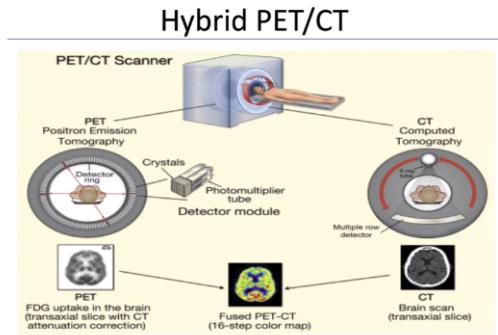


Figure 8: Sistema ibrido PET/TAC

Il sistema PET, è un sistema di rilevatori disposti attorno al paziente (a forma circolare) che raccolgono la radiazione emessa dal paziente attraverso un sistema elettronico basato su cristalli e fotomoltiplicatori (che gestiscono il segnale e il rumore) e che ricostruiscono le immagini in sezioni bidimensionali registrate nello stesso spazio di riferimento della TAC. In questi sistemi vengono appunto combinate le immagini derivanti dal sistema TAC e PET, le quali forniscono informazioni diverse sulla sezione presa in esame.

1.2.4 Risonanza Magnetica (MRI)

La risonanza magnetica è una tecnica di indagine sulla materia basata sulla misura della precessione dello spin di protoni o di altri nuclei dotati di momento magnetico quando sono sottoposti a un campo magnetico.

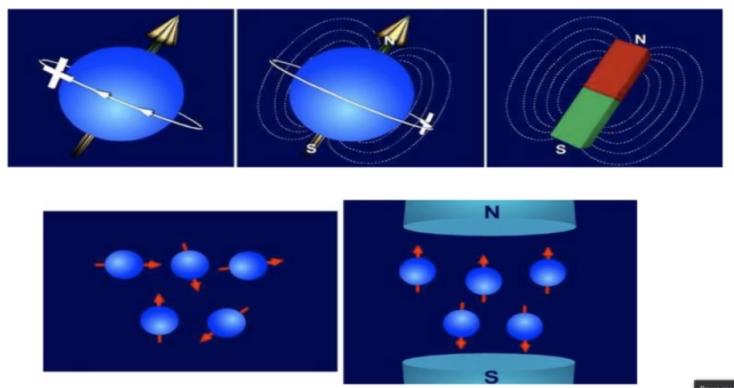


Figure 9: MRI

[Da Wikipedia]

Nel modello classico lo spin del nucleo atomico viene descritto come un momento magnetico, rappresentato da un vettore $\vec{\mu}$.

Quando un nucleo atomico interagisce con un campo magnetico uniforme e costante, rappresentato dal vettore \vec{B}_0 , $\vec{\mu}$ tende a orientarsi nella direzione del campo magnetico, risentendo di una coppia, data da $\vec{\tau} = \vec{\mu} \times \vec{B}_0$, che provoca la rotazione (precessione) di $\vec{\mu}$ attorno alla direzione delle linee di forza del campo magnetico \vec{B}_0 con una precisa frequenza angolare ν_0 , detta "frequenza di Larmor", che dipende solo dal tipo di nucleo e dalla intensità del campo magnetico (\vec{B}_0).

La tecnica NMR non osserva un singolo nucleo, invece misura statisticamente l'effetto di una popolazione di nuclei presenti entro il campione di materiale sotto esame. Viene quindi definito un vettore di magnetizzazione \vec{M} come risultante della somma vettoriale dei momenti magnetici $\vec{\mu}$ dei nuclei, ciascuno dei quali potrà avere la componente parallela a \vec{B}_0 , con lo stesso verso di \vec{B}_0 (allineamento parallelo), oppure con verso opposto (allineamento antiparallelo). In base alle leggi della meccanica quantistica si ha un leggero eccesso di momenti magnetici allineati in un verso rispetto all'altro, e quindi il vettore \vec{M}_0 risultante non è nullo ed è misurabile.

Il vettore \vec{M} , risultante dall'azione del campo magnetico sugli spin nucleari, è proporzionale a quella dell'eccesso di spin (popolazione N_1) definito dal modello quantistico e orientazione uguale a quella del campo esterno B_0 applicato. In definitiva, è possibile ottenere da un piccolo volume di materia un vettore magnetizzazione misurabile dato dalla composizione dei vettori (non misurabili) $\vec{\mu}$ dovuti agli spin dei nuclei contenuti nella materia. Per rilevare questo vettore \vec{M} occorre perturbare il sistema dal suo stato di equilibrio, ad esempio applicando un secondo campo magnetico \vec{B}_1 perpendicolare a \vec{B}_0 e variabile nel tempo (\vec{B}_1 può essere indotto per mezzo di un segnale a radiofrequenze). Se \vec{B}_1 ruota intorno a \vec{B}_0 con frequenza uguale a $\vec{\nu}_0$ cambiando il sistema di riferimento e ponendosi solidali con \vec{B}_1 , si osserva che il vettore \vec{M} precede attorno a \vec{B}_1 con frequenza angolare $\vec{\nu}_1$, variando in questo sistema di riferimento la propria orientazione rispetto alla direzione del campo \vec{B}_0 .

I campi \vec{B}_0 e \vec{B}_1 sono detti rispettivamente campo di polarizzazione e campo di eccitazione. La rotazione che il vettore \vec{M} subisce rispetto al campo principale per effetto del campo \vec{B}_1 dipende dall'energia assorbita dai nuclei e quindi anche dal "tempo di applicazione" T di \vec{B}_1 stesso.

L'angolo è detto angolo di flip. Tramite opportuni valori del campo \vec{B}_1 applicato e del tempo T è possibile ruotare il vettore \vec{M} di 90° e in questo caso si parla di impulso di 90° ; è anche possibile capovolgere la direzione del vettore \vec{M} con un impulso di 180° detto impulso di inversione o "impulso pigreco". In questo caso una parte dei nuclei della popolazione N_1 ha acquistato energia tale da far cambiare direzione ai loro momenti magnetici $\vec{\mu}$, tanto che si dovrebbe raggiungere la situazione ideale in cui il numero dei nuclei con $\vec{\mu}$ antiparalleli egualia il numero di nuclei con $\vec{\mu}$ paralleli ($N_1 = N_2$ questa situazione si raggiungerebbe, in base alla legge di Boltzmann, solo quando la temperatura del sistema di spin è infinita), per cui il sistema di spin non è più in equilibrio termodinamico con l'ambiente (reticolo). Perciò, una volta spento \vec{B}_1 , il sistema di spin dovrà cedere l'eccesso di energia al reticolo.

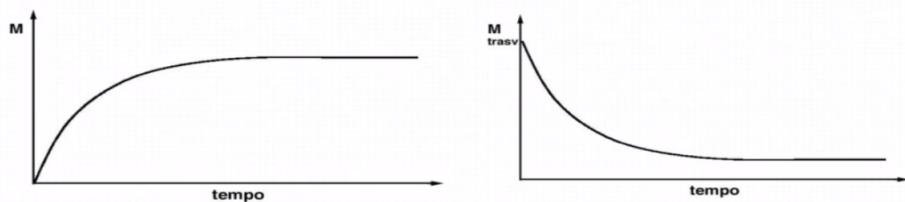


Figure 10: Processo di rilassamento longitudinale e trasversale

Terminata la perturbazione dovuta al campo \vec{B}_1 si ristabilisce l'equilibrio di partenza tra spin degli atomi del campione e campo \vec{B}_0 con determinate modalità temporali. L'ampiezza del vettore \vec{M} non è conservata durante quello che viene definito "processo di rilassamento". Tale processo coinvolge il rilassamento trasversale ed il rilassamento longitudinale.

La costante di tempo T_1 , che governa il ritorno all'equilibrio della componente longitudinale del vettore \vec{M} , è definita tempo di rilassamento spin-reticolato, in quanto coinvolge i trasferimenti di energia che avvengono tra il sistema di spin e il resto dell'ambiente.

T1-weighted images

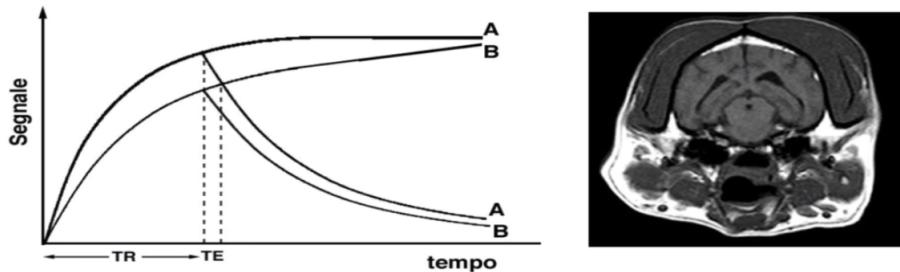


Figure 11: Rilassamento spin-reticolato

La costante di tempo T_2 , che governa l'annullamento della componente trasversale del vettore \vec{M} , è definita tempo di rilassamento spin-spin in quanto coinvolge le interazioni tra i momenti magnetici dei singoli nuclei, cioè è legata alla dinamica temporale che porta gli spin atomici a perdere di coerenza e quindi a sfasarsi.

T2-weighted images

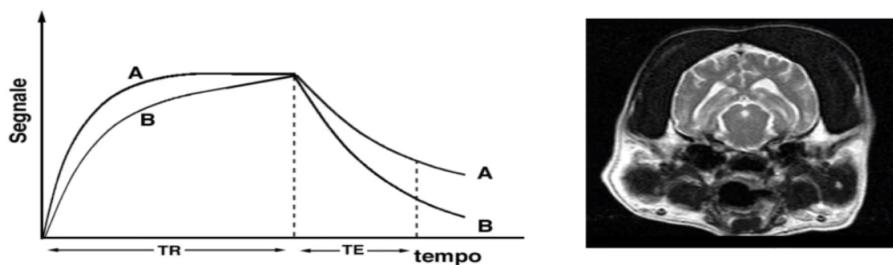


Figure 12: Rilassamento spin-spin

Nel caso che il campo \vec{B}_0 non sia perfettamente omogeneo a livello locale a causa di disomogeneità del campo esterno applicato, o a causa di differenze puntuali di

suscettività magnetica del sistema, oppure per l'applicazione di un campo magnetico caratterizzato da un preciso gradiente, la frequenza di precessione dei nuclei viene a dipendere anche dalla posizione che essi occupano rispetto a tali disomogeneità locali. Ogni pacchetto di spin può precedere allora a una sua velocità nei diversi punti del campione, sfasandosi di fatto rispetto agli altri. Si osserva dunque un decadimento del vettore \vec{M} più rapido di quello che ci si aspetterebbe dal solo rilassamento spin-spin dovuto al minore ordine del sistema. Questo fenomeno è considerato nella costante di tempo T_2^* . Il tempo di rilassamento T_2 è sempre minore o uguale a T_1 .

Una volta terminata l'azione perturbante del campo \vec{B}_1 , dopo un tempo di applicazione T_p , si segue l'andamento del ritorno all'equilibrio della magnetizzazione macroscopica \vec{M} che tende a riallinearsi al campo \vec{B}_0 . Il segnale prodotto dalla variazione nel tempo del vettore \vec{M} viene misurato in laboratorio usando una bobina a induzione elettromagnetica posta attorno al campione in direzione ortogonale al campo esterno, che si comporta come una antenna: le variazioni della componente trasversale di \vec{M} si vanno a concatenare alla bobina, inducendo in essa una piccola forza elettromotrice (misurabile tramite un ricevitore a radiofrequenza) che oscilla alla frequenza di Larmor.

Il segnale NMR, detto FID (Free Induction Decay, decadimento libero dell'induzione) è approssimativamente monocromatico e oscilla alla frequenza di Larmor, attenuandosi in maniera esponenziale con il tempo in funzione della costante di tempo T_2^* .

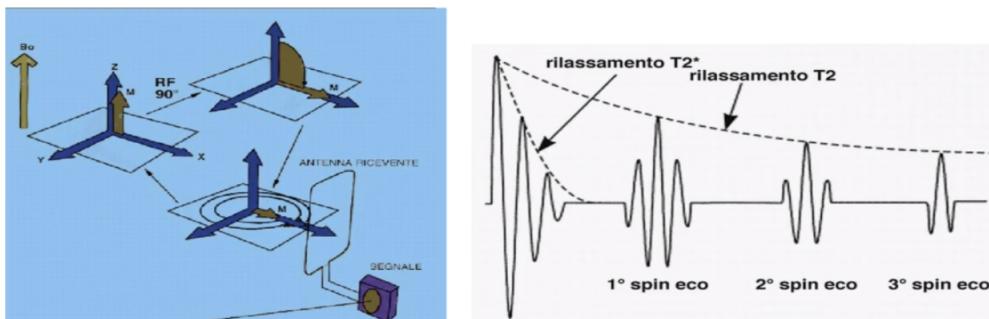


Figure 13: Processo di rilassamento

Per la formazione di immagini, si utilizzano sequenze di eccitazione opportunamente progettate che consentono di enfatizzare la dipendenza del FID dai tre parametri: densità protonica ρ , T_1 , T_2 .

Un parametro caratteristico di tutte le sequenze è il tempo di ripetizione T_R , ossia l'intervallo di tempo fra l'inizio di una sequenza di eccitazione e l'inizio della successiva. Inoltre non si misura direttamente il FID ma un segnale "eco", composto da 2 FID speculari.

Il passo fondamentale che ha permesso di avere dal segnale NMR anche una codifica spaziale sul volume del campione è stata l'introduzione dei gradienti di campo magnetico: questo ha permesso di ottenere immagini spaziali del "campione" esaminato. Se oltre al campo magnetico principale \vec{B}_0 e a quello rotante \vec{B}_1 viene applicato a una piccola zona del materiale in esame un campo magnetico variabile linearmente nel volume del campione, ma di intensità molto minore di quello polarizzante, la frequenza di risonanza di Larmor in quella zona cambia in funzione della somma tra il campo

principale e il valore in quel punto dell'intensità del campo secondario.

Risulta quindi possibile (sapendo in quale area viene applicato il campo secondario) legare il segnale di ritorno a coordinate spaziali, e di conseguenza avere anche una misura della densità protonica in un ben preciso punto del materiale, in particolare il segnale raccolto dall'antenna sarà una somma di termini oscillanti avente la forma di una trasformata di Fourier della distribuzione delle densità dei protoni del volume acquisito.

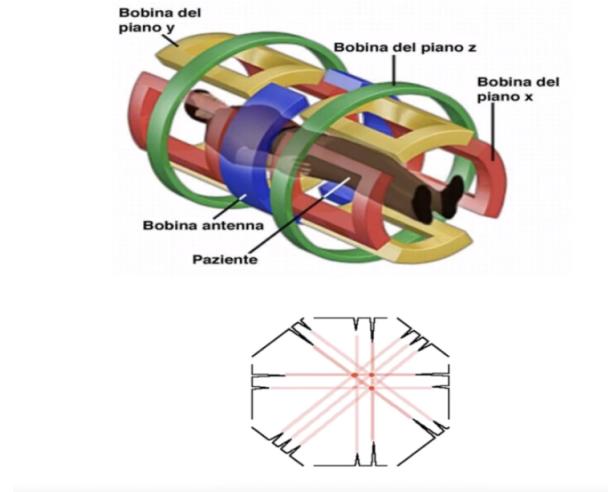


Figure 14: Ricostruzione dell'immagine da MR

Antitrasformando il segnale $S(t)$ si ottiene la distribuzione della densità di protoni $\rho(r)$. La parte centrale dello spazio k (basse frequenze) determina il contrasto dell'immagine, mentre le parti esterne (alte frequenze) influiscono sui dettagli dell'immagine. Impostando particolari sequenze di segnali di eccitazione che pesino selettivamente i contributi al segnale dati dai diversi parametri (ρ, T_1, T_2, T_2^*), si possono ottenere informazioni sul tipo di tessuto in esame.

La rappresentazione in termini di Fourier del segnale $S(t)$ permette una facile visualizzazione grafica delle varie strategie che possono essere usate per l'acquisizione delle immagini. Al crescere del tempo, l'evoluzione del segnale descrive una traiettoria nello spazio k , e in ogni suo punto il valore della trasformata k è dato dal valore di $S(t)$. In linea di principio si può percorrere una qualunque traiettoria nello spazio k impostando opportunamente l'andamento temporale dei gradienti.

2 Obiettivi del Medical Imaging

In questa sezione verranno presentati i principali obiettivi del passato, presente e futuro (medicina personalizzata) nell'imaging medico. Quindi, verranno presentate le principali sfide e le relative metodologie atte ad affrontarle.

2.1 Passato

L'imaging medico nel passato aveva come principale obiettivo quello di identificare la malattia in modo non invasivo (prelievi, etc.) e in vivo (quando il paziente è ancora in vita), attraverso il fenotipo (*[da Wikipedia] il fenotipo è l'insieme di tutte le caratteristiche manifestate da un*

organismo vivente, quindi la sua morfologia, il suo sviluppo, le sue proprietà biochimiche e fisiologiche comprensive del comportamento. Questo termine viene utilizzato in associazione al termine genotipo, dove per questo si intende la completa costituzione genetica di un individuo o di un organismo vivente, che è solo in parte espressa nel corpo del vivente.

Quindi il fenotipo è ciò che è evidentemente il vivente, il suo genotipo invece è la informazione genetica che è in lui contenuta, che ha generato il fenotipo; solo una parte esigua del genotipo si esprime nel fenotipo.). In sintesi, le sfide del passato nella medicina sono state:

- comprendere le malattie (insorgenza e progressione della malattia)
- effettuare una diagnosi (capire se un soggetto è affetto o non affetto da una malattia). Rimane ancora oggi l'obiettivo principale
- monitorare la terapia nel tempo

Dal punto di vista fisico-informatico, le sfide del passato nell'imaging medico riguardano il miglioramento della qualità dell'immagine (sia trasmissive che emissive) attraverso:

- il miglioramento del rapporto Signal-to-Noise: il segnale che viene immagazzinato, passando attraverso il corpo umano, risulta essere attenuato. Pertanto, sono stati sviluppati degli algoritmi in grado di correggere le informazioni mancanti e quindi di correggere l'attenuazione del segnale. Inoltre, sono stati sviluppati dei metodi (come il metodo di Monte Carlo) in grado di ridurre il disturbo e quindi correggerlo. Infine, sono state proposte delle metodologie in grado di ottimizzare i sistemi di acquisizione delle immagini basate sull'estrazione e fitting sulle curve di performance
- il miglioramento dei sistemi di "rivelazione" delle immagini: gran parte di questo lavoro viene svolto dai bio-ingegneri/informatici. Si cerca di migliorare la rivelabilità, ossia di ottimizzare la risoluzione delle immagini derivanti da apparecchiature mediche: spesso le immagini "grezze" non riescono a fornire delle informazioni utili a causa della bassa qualità dell'immagine e, ad esempio, delle lesioni molto piccole o poco contrastate potrebbero non essere visibili. Anche in questo caso vengono utilizzate dei metodi di simulazione di Monte Carlo. Questa ottimizzazione ha lo scopo di limitare l'analisi superflua sui pazienti.
Una tecnica molto importante è la **SPM** (Statistical Parametric Mapping). È una tecnica di elaborazione delle immagini radiologiche basata sulla seguente pipeline: l'immagine viene normalizzata (attraverso atlanti standard) e filtrata per ridurre il rapporto segnale-rumore. Dopodiché, vengono effettuati dei test statistici (tra i gruppi e tra un singolo soggetto e un gruppo) basate sui Voxel-wise. I risultati vengono mappati su uno spazio conosciuto (atlanti standard) coi corrispondenti valori di probabilità.

2.2 Presente

La sfida dell'imaging medico nel presente riguarda la misurazione della malattia attraverso biomarcatori quantitativi. Come nel "passato", con processi non invasivi ed in vivo attraverso il fenotipo. Oggi si sa che anche se vi sono soggetti con diagnosi simili, ma che in realtà sono differenti per quanto riguarda la risposta alla terapia e la prognosi (evoluzione della malattia). Da alcuni studi, è stato possibile osservare che pazienti con diagnosi simili, dopo un periodo di tempo, hanno ottenuto degli outcome clinici differenti tra loro: la spiegazione è che ci sono dei genotipi e fenotipi che ne caratterizzano un diverso comportamento. I biomarcatori genomici, ad oggi, sono gli unici ad aver identificato il comportamento differente tra i soggetti. Le sfide di "oggi" sono:

- identificare delle misure come candidato marcitore
- garantire che l'accuratezza sia elevata
- fornire tecniche per rendere riproducibili/ripetibili i metodi di misurazione dei marcatori

I problemi sono legati alla fisica delle immagini, che vengono risolti con metodi computazionali. I problemi sono:

1. Segmentazione del volume di interesse (VOI): è molto complesso per oggetti con forme piccole, eterogenee e irregolari
2. Risoluzioni spaziali limitate: dovute agli effetti di volume parziale che vanno ad impattare sull'informazione dell'immagine

Le soluzioni sono delle tecniche di correzione: queste non possono essere applicate sui pazienti, pertanto vengono testati su dei "fantocci". I fantocci sono delle ricostruzioni con materiali che dal punto di vista dell'imaging medico risultano essere equivalenti (e vale per tutte le tecniche viste precedentemente - TAC,PET,etc.). In particolare, sono tecniche di segmentazione automatica dei volumi delle lesioni e correzione automatica del volume parziale. Questi fantocci sono costruibili attraverso delle stampanti 3D, come gusci riempiti di materiali (equivalenti al caso reale). Risultano essere molto utili, poiché ci si trova in un ambiente totalmente controllato ed è, quindi, possibile valutarne l'accuratezza.

Tali tecniche risultano essere molto importanti, ad esempio, nella ricerca dei marcatori, poiché è possibile osservare e categorizzare i pazienti in sotto-tipi: ad esempio, un tipo di marcatore può essere il volume di una lesione che attraverso ai modelli di segmentazione automatica e alle opportune correzioni è possibile osservare se il paziente sta reagendo alla terapia ed in caso negativo proporne una differente.

2.3 Futuro

Sebbene è stato detto che al momento la genomica è l'unica branca della medicina in grado di distinguere pazienti diversi, un importante articolo scritto su *"The New England Journal of Medicine"* ha dimostrato che anche tale metodologia ha dei limiti: un esempio è il "fallimento" dei biomarcatori genomici per l'eterogeneità intra-tumorale. La speranza è che l'imaging 3D vada in supporto della genomica. Una possibile soluzione è la Voxel-based features from images (RADIONICS): dopo la parte di segmentazione automatica, si passa alla fase di features extraction. Queste caratteristiche estratte, risultano essere importanti per comprendere meglio il comportamento alla terapia del paziente, attraverso la correlazioni con altre features. Questa tecnica risulta essere un metodo *in vivo* e non invasivo.

Quindi, l'imaging medico del futuro ha come sfida quella di predirre il corso della malattia, la risposta alla terapia di un singolo soggetto in modo non invasivo e *in vivo*. In che modo? Accoppiando grosse moli di immagini mediche (Big Data) e strumenti di Intelligenza Artificiale. In particolare, vengono utilizzati modelli di Machine Learning di tipo predittivo (come ad esempio SVM). Degli esempi di applicazione degli algoritmi di ML riguardano l'Alzheimer e il Parkinson. I metodi che vedremo sono molto recenti e vi sono ancora pochi lavori in merito e bisogna cercare di raggiungere la ripetibilità, riproducibilità e la significatività necessaria.

3 Medical-image Quantification

Per quantificazione delle immagini s'intende una varietà di algoritmi che hanno l'obiettivo di estrarre dall'immagine medica un parametro (solitamente è un parametro fisico) che ha un significato clinico, al fine di fornire un supporto della decisione del medico. In particolare, la quantificazione delle immagini mediche è richiesta nel caso di patologie oncologiche: i tumori sono degli oggetti molto complessi dal punto di vista biologico. Dal punto di vista macroscopico sono costituiti da tessuti e masse che hanno delle forme e densità molto eterogenee ed hanno un metabolismi globali molto complessi. Dal punto di vista microscopico sono caratterizzati dalla proliferazione cellulare in aree diverse, ipossia (mancanza di ossigeno) e neo-angiogenesi (processi di morte cellulare causati dal tumore).

Una delle tecniche più utilizzate per studiare i tumori in modo non invasivo e *in-vivo* è il sistema ibrido PET/TAC. Le tecniche PET e TAC vengono eseguite, attraverso un macchinario in figura 15, in modo consecutivo, mantenendo gli assi delle immagini equivalente per le due tecniche (a

meno di una traslazione dell'asse z). Quindi, sono immagini co-registrate che hanno lo stesso spazio di riferimento.

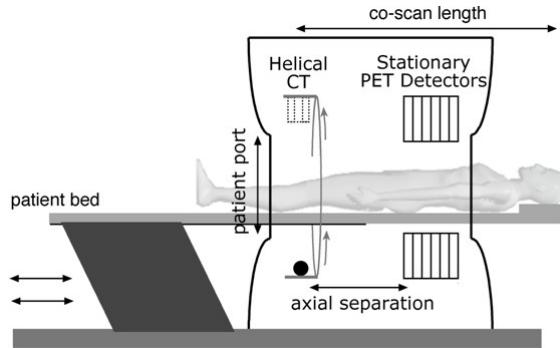


Figure 15: PET/CT process

Nella figura 16 è possibile osservare un esempio di immagini ottenute da tale tecnica: ovviamente, la lesione tumorale si trova nella medesima posizione in tutte le immagini. Le immagini riportate sono il risultato della TAC (immagine di tipo anatomico), della PEC (immagine di tipo molecolare) ed, infine, una loro combinazione (immagine "fusa"), sia nella visione coronale che trasversale.

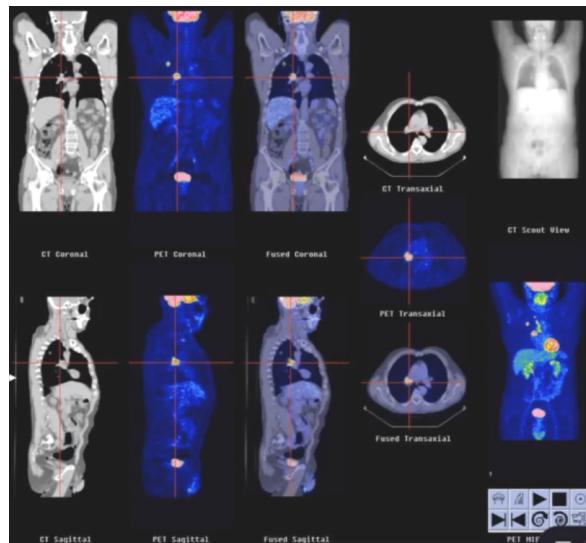


Figure 16: PET/CT results

Gli obiettivi principali di queste tecniche sono:

1. localizzare da un punto di vista anatomico il tumore
2. caratterizzarlo da un punto di vista tessutale e molecolare
3. guidare il medico nella scelta della terapia
4. monitorare la terapia (durante e dopo)

3.1 Analisi qualitativa

Il primo modo di valutare le immagini PET/TAC è di tipo qualitativo. La fusione delle immagini PET e TAC forniscono un'immagine molto utile, sia per individuare le lesioni, sia per comprenderne meglio la loro localizzazione anatomica e sono utili per la scelta della terapia. In figura

17, è possibile osservare un esempio di immagine derivante da PET/TAC, in cui si evidenzia la differenza tra le immagini attraverso PET, TAC e PET/TAC. Quest'ultimo consente di individuare le lesioni al 96,6%, rispetto all'utilizzo della sola PET, la quale individua l'84,3% delle lesioni (altri esempi sulle slides - se verranno pubblicate).

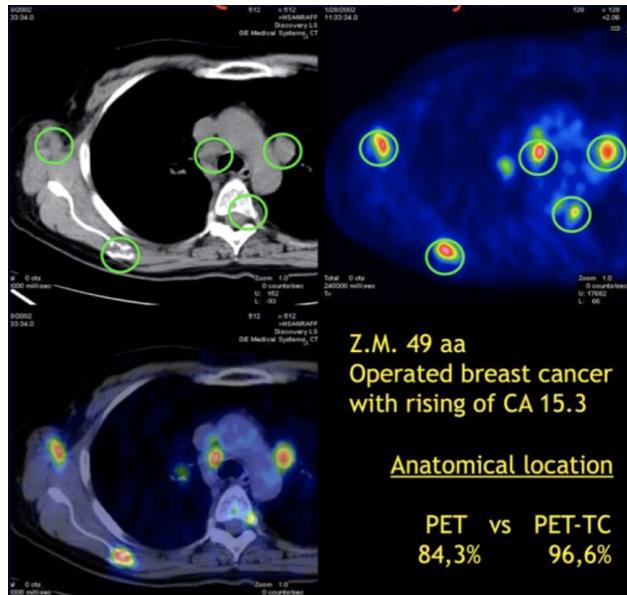


Figure 17: PET/CT results

Lo studio qualitativo è utile anche durante il monitoraggio (durante e dopo il trattamento). Infatti, è possibile ottenere delle informazioni attraverso il confronto tra immagini catturate (ad esempio, prima e dopo la terapia). In particolare, viene valutata:

- la massa residua
- le ricorrenze
- il funzionamento della terapia, ossia la risposta al trattamento

3.2 Analisi quantitativa

Oltre all'analisi qualitativa, ricopre un ruolo molto importante quella quantitativa. Infatti, ad esempio, essa guida la pianificazione del trattamento della radio-terapia, attraverso il PTV (*Planning Tumor Volume*). Tale pianificazione ha bisogno di:

- un particolare lettino molto sottile, in modo che il paziente sia "poco deformato"
- uno specifico posizionamento molto accurato, segnato anche da segnalatori laser
- gating respiratorio (ossia viene prestata attenzione al respiro del paziente, soprattutto per gli organi che hanno un impatto dalla respirazione, come quelli nella zona del torace)
- software dedicati per la simulazione, visualizzazione e segmentazione
- tools che utilizzano le immagini con diverse tecniche (come l'object detection)

Questi punti definiscono il workflow per la pianificazione del trattamento della terapia, in cui è fondamentale l'interazione tra medici e data scientist.

Vi sono principalmente due problematiche che posso presentarsi durante il trattamento alla radio-terapia: lo scorretto posizionamento ed i cambiamenti (dal punto di vista biologico) del target, ossia il Tumor Volume (TV). I tipi di volume sono, secondo gli standard dettati dal ICRU:

- GTV: Gross Tumor Volume
- CTV: Clinical Tumor Volume
- BTV: Biological Tumor Volume

Per arrivare a queste misure è necessario ottenere dei biomarcatori estratti dalle immagini, che hanno informazioni inerenti alle tre metriche sopra-menionate.

Il primo parametro è il **SUV** (Standardized Uptake Value): viene utilizzata per le analisi PET e cerca di standardizzare la dose che è stata captata all'interno del tumore. La formula è la seguente:

$$SUV = \frac{\text{Decay corrected dose / ml of tumor}}{\text{Injected dose / patient weight in gr}}$$

Dunque, è un parametro che tiene in considerazione la dose iniettata e il peso del paziente. Altre misure analoghe, più sofisticate, sono: il SUV_{BW} (body-surface-area corrected), il SUV_{BSA} (body-weighted corrected) ed il SUV_{LBM} (lean body-mass corrected). Infine, il SUV è molto importante perché fornisce informazioni sulla proliferazione cellulare del tumore: tanto più il tumore consuma lo zucchero iniettato, tanto più sarà aggressivo.

Il SUV è molto importante, poiché è in grado di fornirci informazioni circa l'effetto del trattamento. In particolare, l'*European Journal of Cancer* (1999), ha definito degli intervalli di valori che corrispondono alla risposta del trattamento:

- se il SUV aumenta per valori maggiori del 25% (assieme alla presenza di nuove lesioni), allora la malattia progredisce (PD - Progressive metabolic disease)
- se il SUV aumenta per valori minori del 25%, o diminuisce per valori maggiori del 15%, allora la malattia è stabile (SD - Stable metabolic disease)
- se il SUV diminuisce per valori maggiori del 15% dopo un ciclo di chemioterapia, o diminuisce per valori maggiori del 25% dopo diversi cicli di chemioterapia, allora vi è una risposta parziale del metabolismo (PR - Partial metabolic response)
- se il SUV diminuisce per valori maggiori del precedente, allora vi è una risposta completa del metabolismo (CR - Complete metabolic response)

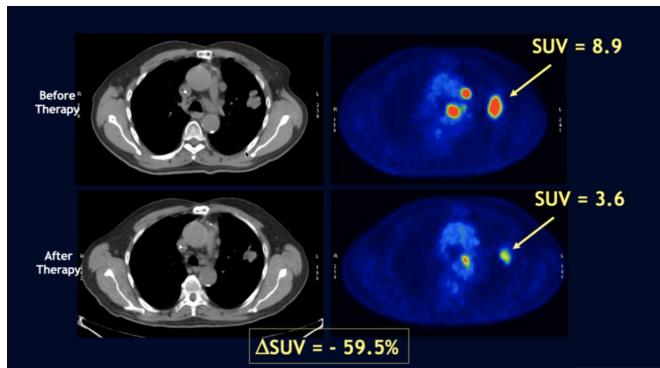


Figure 18: Esempio di applicazione del SUV su un paziente con carcinoma ai polmoni: in questo caso il $\Delta SUV = -59.5\%$, quindi il trattamento ha fatto effetto ed è possibile proseguire con la terapia

3.3 Calcolo del volume della lesione: problematiche

Calcolare il volume di una lesione non è un compito facile. È possibile arrivare alla segmentazione della lesione attraverso le immagini PET/MRI-DWI) e CT/MRI.

Le immagini PET/MRI-DWI sono affette da diverse problematiche. Infatti, vi è una scarsa risoluzione spaziale, il rapporto segnale-rumore è basso (conseguentemente, lo è anche il rapporto lesione-background) ed, infine, sono affette da problemi di movimento. Come aspetti positivi si ottiene l'eterogeneità nelle lesioni e nei tessuti (background). Le immagini CT/MRI hanno le seguenti caratteristiche: la risoluzione spaziale è alta, il segnale è alto, il rumore è medio, il rapporto lesione-background è medio/basso ed, infine, sono anch'esse affette da problemi di movimento. Come per le immagini precedenti, si ottiene l'eterogeneità nelle lesioni e nei tessuti (background).

Ovviamente, sia la modalità anatomica che quelle funzionale sono indispensabili, poiché forniscono informazioni diverse: quella anatomica localizza la lesione, mentre quella funzionale ne determina la proliferazione funzionale.

3.3.1 Movimento

Il problema nel **movimento**, risiede nel fatto che spesso è molto difficile catturare l'immagine per target in movimento. Per esempio, una lesione polmonare, può apparire allungata se durante l'acquisizione il paziente respira, creando così un'immagine non realistica, distorta. Per ridurre tale problema è stato introdotto un sistema del monitoraggio del respiro del paziente chiamato **respiratory gating**. In particolare, la respirazione può essere facilmente descritta da una funzione sinusoidale e verrà registrata attraverso un marker posto sul paziente, il quale ne traccerà l'andamento. Con l'ingresso di questa nuova componente, si parlerà di 4D PET/CT, poiché oltre alle coordinate spaziali viene inserita una nuova dimensione: il tempo. In questo modo, dalla somma delle immagini ottenute, si otterrà un'immagine molto definita, in cui la lesione risulta essere molto concentrata.

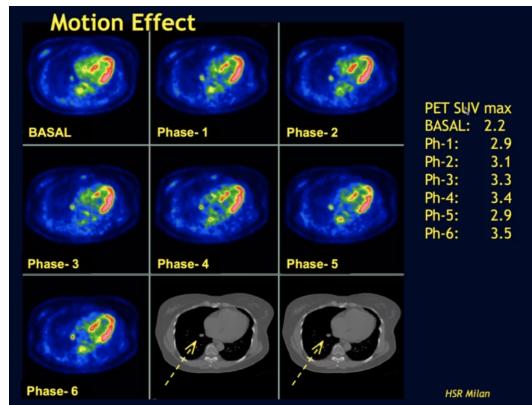


Figure 19: esempio di cambiamento del valore SUV nelle varie fasi, rispetto al valore base, grazie all'inserimento del respiratory gating

Un'altra strategia per stimare il SUV, è la stima mediante dei "fantocci": viene simulato il paziente, nella sua forma e composizione. Per esempio, vengono utilizzate delle sfere, con un particolare diametro, per simulare le lesioni oncologiche: vengono riempite con delle concentrazioni di zucchero radioattivo e ne vengono studiate le relative immagini ottenute. È possibile, come è possibile vedere in figura 20, che l'errore è inversamente proporzionale al diametro della sfera.

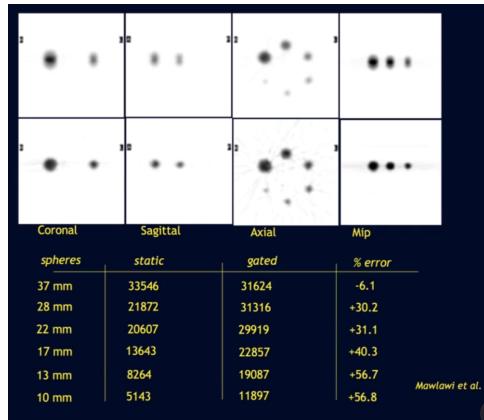


Figure 20: motion effect

Per questa ragione viene studiato il **Partial Volume Effect**: questo fenomeno fisico causa la variazione dell'oggetto nella sua dimensione e nel suo contenuto informativo, attraverso l'effetto di image blurring e tissue fraction effect.

Gli oggetti più piccoli appaiono più grandi, causando una diminuzione del segnale. Questo effetto prende il nome **image blurring** (sparpagliamento dell'immagine), ed è causato dalla risoluzione finita dei sistemi di misurazione medici. Per cui, se si vanno a confrontare degli oggetti molto piccoli, si potrebbero ottenere degli errori molto grandi.

Il secondo effetto è dovuto al campionamento dell'immagine: la discretizzazione dell'immagine provoca delle perdite di informazione lungo i contorni. Altri due effetti che concorrono al PVE sono lo spill-in (i conteggi dell'esterno vengono spinti all'interno dei contorni) e lo spill-out (i conteggi dell'interno vengono spinti all'esterno), i quali causano un ingrandimento apparente della lesione.

Conosciuto il problema del PVE, vengono eseguite delle correzioni sul SUV. Data un'immagine di una lesione ideale, viene confrontata la distribuzione di densità di una lesione reale. In particolare, si cerca di stimare di quanto è sceso il massimo (il picco) del reale rispetto a quello ideale: tale fattore prende il nome di Recovery Coefficient (RC). Attraverso l'RC, è possibile ottenere un'immagine della lesione corretta (figura 21)

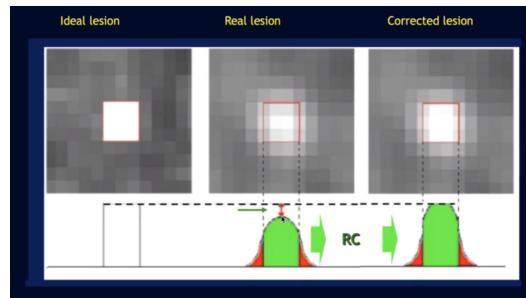


Figure 21: motion effect

I RC vengono stimati attraverso dei fantocci: viene stimato il rapporto tra i conteggi degli errori e i conteggi ideali, ottenendo delle curve (figura 22): tali funzioni vengono costruite in funzione delle attività e dei diametri, da cui è possibile recuperare il coefficiente di recupero (RC). I RC dipendono sia dal volume che dalla concentrazione/background della lesione.

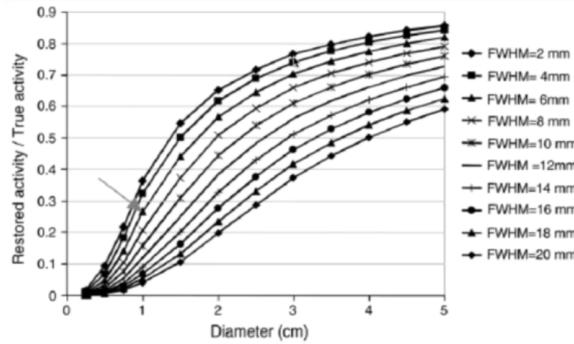


Figure 22: RC curve

4 Medical-image Segmentation

La segmentazione delle immagini è il processo di partizione di un'immagine in regioni significative.

Definizione. Per segmentazione s'intende il partizionamento di un'immagine I , in n subregioni differenti S_i , con $i = 1, \dots, n$:

$$I = \sum_{i=1}^n S_i, \quad S_i \subset I, \quad S_i \cap S_j = \emptyset \quad \forall i \neq j$$

Come già descritto nella sezione 3, i tumori sono degli oggetti molto complessi dal punto di vista biologico, con diverse proprietà a livello microscopico e macroscopico, che vanno a complicare le tecniche di segmentazione. Infatti, vi possono essere delle lesioni eterogenee differenti, ossia il tumore appare molto diverso, se vengono confrontate le immagini PET e TAC. Analogamente, il background del tessuto molto eterogeneo può apparire differente delle immagini PET e TAC. Pertanto, vengono seguiti due diversi approcci per la segmentazione delle lesioni:

- metodi qualitativi: basati sull'ispezione visiva e del contornamento manuale da parte del medico. Tali metodi hanno il grosso difetto di dipendere dalla modalità d'immagine (PET-TAC), dal medico e dal livello della colorbar utilizzata.
- metodi quantitativi: basate sull'elaborazione delle immagini e del semi-auto/autocontornamento da parte degli algoritmi

Di seguito verranno presentati dei metodi quantitativi molto utili al supporto dei medici.

4.1 Image thresholding methods

La tecnica del **image thresholding** è basata sugli algoritmi cosiddetti "a soglia", ed è una tecnica relativamente semplice.

Metodo. Durante un processo di sogliatura, i singoli pixel dell'immagine sono catalogati come "pixel oggetto" se il loro valore è maggiore di una certa soglia e come "pixel di sfondo" se il valore è sotto la soglia. Solitamente l'immagine binaria in uscita ha valore pari a "1" dove è presente l'oggetto (dentro la soglia) e pari a "0" per lo sfondo (fuori dalla soglia). Formalmente, si ha:

$$\text{lesion} = T[I(x)] = \begin{cases} 1, & I(x) \geq T \\ 0, & I(x) < T \end{cases}$$

dove x rappresenta il digital number (DN) ovvero il valore del pixel nell'immagine (figura 23).

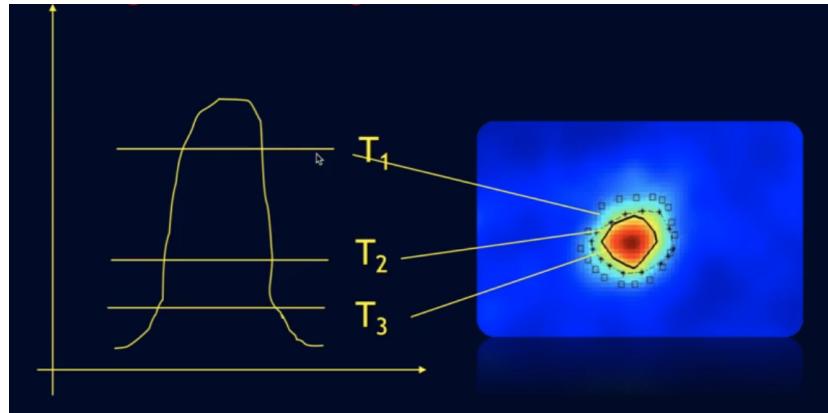


Figure 23: image thresholding methods

Il parametro chiave in una sogliatura è la scelta del valore di soglia. Esistono diverse metodologie per la scelta automatica di questo valore, oppure è possibile sceglierlo a mano.

1. Un semplice metodo potrebbe essere quello di creare un'istogramma e osservare dove vi è il punto più basso in mezzo all'istogramma; quindi si sceglie quel valore come soglia

Quindi, si cerca di ottimizzare questi algoritmi (computazionalmente poco dispendiosi) tenendo in considerazione alcuni parametri dell'immagine. I metodi per la scelta del valore T sono i seguenti:

- (a) si cerca di utilizzare delle T fisse: 40-50-60% del massimo valore. Questo metodo funziona bene per gruppi di lesioni molto simili tra loro
- (b) si estraggono delle funzioni che portano a delle soglie ottimali in funzione del volume della lesione. Per fare un esempio:

$$\%T = a + b \log_{10} BTV + q$$

Questo tipo di soglia viene definita *adattativa*

- (c) si utilizza un valore T assoluto per tutte le lesioni ($T = 2.5$). Se il SUV supera tale soglia, siamo in presenza di un linfonodo tumorale
- (d) si stima la soglia attraverso il rapporto lesione-background (LB) nel seguente modo:

$$T = a + b \frac{1}{LB}$$

In questo caso è necessaria la pre-calibrazione attraverso fantocci. Vi sono diverse funzioni che possono essere utilizzate per ottenere il valore soglia, ciascuna applicabile in specifici contesti.

- (e) si utilizzano delle funzioni iterative, le quali cercano di minimizzare l'errore.
2. Tecniche si basano sulle differenze dei gradienti all'interno e all'esterno (background) della lesione: tali tecniche si basano sulle pendenze delle curve dei profili dell'immagine che quindi mirano ad identificare i contorni delle lesioni. Possono essere edge detector (Sobel operator) o ridge detector (Watershed Transform).
 3. Tecniche si basano sui contorni attivi deformabili: sono algoritmi basati su funzioni geometriche o di deformazioni di superfici sotto forze interne ed esterne, dove quelle esterne lavorano come gradienti direzionali. I contorni vengono definiti quando avvengono delle variazioni molto ripide ("snakes"), che appunto identificano superfici diverse. Vi sono diverse metodi basati su questa tecnica: Gradient Vector Flow (GVF), equazioni differenziali di Poisson

con evoluzione iterativa. Un esempio di quest'ultimo metodo è il level-set: si parte dal pre-processing dell'immagine attraverso un filtraggio per passare al processo iterativo per l'individuazione dei contorni. Funziona molto bene per lesioni bifocali (esigenza di definire un contorno non regolare, con superfici vuote)

4. Tecniche di addestramento (learning methods). Hanno come obiettivo la discriminizzazione del segnale dentro la lesione dal tessuto circostante, basandosi su delle features estratte. Può essere supervisionato o non-supervisionato. Gli algoritmi di classificazione più utilizzati sono KNN, SVM e ANN, mentre gli algoritmi di clustering più utilizzati sono k-means, fuzzy C-means e Expectation maximization.

Le metriche più utilizzate per comprendere le performance dei classificatori sono la *Dice Score Coefficient*:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

e la *recall*:

$$recall = \frac{|X \cap Y|}{|Y|}$$

dove X è l'area predetta e Y quella vera. Le reti neurali maggiormente utilizzate sono riportante in figura 24

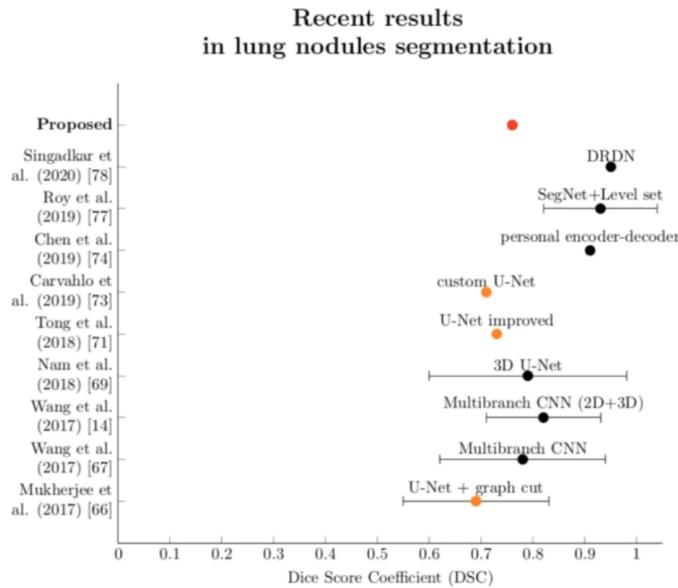


Figure 24: Comparazione di diverse architetture

5. Tecniche basate su modelli stocastici. L'ipotesi di partenza di questi metodi è la distribuzione di intensità statisticamente indipendente tra la lesione ed il background. Uno dei principali modelli è il Gaussian Mixture Model (GMM). Tale metodo è in grado di fornire 3 classi: la regione del tumore, il background e la classe "incerta", ossia la regione tra la lesione e il background. Viene stimata la probabilità di appartenenza di ciascun voxel alle classi, col metodo detta massima verosimiglianza. Per definire le tre aree, si utilizzano modelli di Markov, wavelet-based multiscale decomposition e Fuzzy locally adaptive Bayesian (FLAB).

Ovviamente, non vi è una tecnica migliore in modo assoluto, infatti esse vanno confrontate tra di loro: grazie alle stampanti 3D si è in grado di creare dei fantocci in modo da avere un termine di

paragone e quindi per testare la bontà dei modelli.

In conclusione, la segmentazione delle lesioni è un problema molto complesso nell'imaging medico: si è potuto osservare che è molto importante l'utilizzo del respiratory gating per la corretta rilevazione delle immagini soggette a distorsioni dovute alla respirazione, che è importante migliorare la definizione delle immagini attraverso la correzione del volume parziale (soprattutto nel caso di lesioni molto piccole) ed, infine, che i metodi di segmentazione possono essere usati per estrarre i volumi delle lesioni, tenendo in considerazione la complessità della lesione e le diverse proprietà delle immagini.

5 Machine Learning nel Medical-imaging

L'obiettivo dell'utilizzo del Machine Learning nell'imaging medico è di fornire maggiori informazioni ai medici. In particolare, sono utili per aumentare l'efficacia dell'analisi medica e soprattutto di fornire informazioni utili al fine di decisioni strategiche in problemi che normalmente sono molto difficili da studiare. Un altro obiettivo importante è l'estrazione di dati significativi su una mole di dati molto grande, in breve tempo e in modo versatile. Quindi, in sintesi, l'utilizzo del ML offre una diagnosi assistita, una valutazione clinica oggettiva e fornisce diagnosi con alti livelli di accuratezza.

L'imaging medico **in vivo** svolge un ruolo cruciale per la diagnosi di molte malattie neurologiche, incluse quelle neurodegenerative e psichiatriche. Dunque, è possibile prevenire tali malattie (diagnosi precoce) e di differenziare la tipologia di malati. In questo caso, vengono estratte delle informazioni dalle immagini, le quali potrebbero non essere facilmente osservabili (nella diagnosi precoce) o facilmente confondibili all'occhio umano (del medico).

I sistemi di supporto alle decisioni nel settore medico si dividono principalmente in:

1. MultiVariate Analysis: MultiVariate (or MultiVoxel) Pattern Analysis (MVPA)
2. Machine Learning

Le principali differenze tra i due sistemi sono: che il ML si focalizza sulla classificazione, mentre la statistica sull'inferenza; nel ML si trattano dati che non hanno distribuzioni note a priori e vengono trattati su problemi con elevata dimensionalità; infine, l'aspetto computazionale del ML rispetto alla statistica gioca un ruolo centrale.

Nel Machine Learning è possibile categorizzare gli algoritmi in base al tipo di dati/problema e/o alla presenza o meno delle labels.

	Supervised Learning	Unsupervised Learning
Discrete	Classification	Clustering
Continuous	Regression	Dimensionality Reduction

5.1 Support-Vector Machine

Gli algoritmi di SVM sono tra gli algoritmi più utilizzati in medicina, specialmente nell'imaging medico. In particolare, SVM è un classificatore binario che ha come obiettivo la generazione di un modello predittivo per la discriminazione di un nuovo campione, date le caratteristiche apprese. Quindi è in grado di separare l'iper-spazio composto dalle features, attraverso degli iper-piani. L'iper-piano scelto sarà quello che massimizza la distanza di discriminazione tra le classi. In formule si ha:

$$y(x) = \sum_{n=1}^N w_n t_n k(x, x_n) + b$$

dove N è il numero di campioni del training set, w_n sono i pesi assegnati dal SVM, t_n è la label della classe associata, $k(x, x_n)$ è la funzione kernel, b è il bias ed, infine, $y(x)$ è la class label predetta (output) per il campione x .

I campioni più importanti, risultano essere quelli più vicini all'iper-piano, poiché sono coloro che permettono all'algoritmo di discriminare nel modo più opportuno. In altre parole, viene minimizzato l'errore di classificazione empirica, massimizzando la distanza di separazione tra le classi del training set. I campioni che giacciono sui margini degli iper-piani (figura 25), vengono chiamati *vettori supporto* e sono gli unici campioni in cui l'algoritmo assegna un peso diverso da zero.



Figure 25: linear kernel

Vi sono alcuni parametri geometrici da fissare, tra i quali vi è la *kernel function*: può essere lineare, polinomiale o Gaussian Radial Basis (RBF).

- **kernel linerari.** $k(x_i, x_j) = (x_i \cdot x_j)$

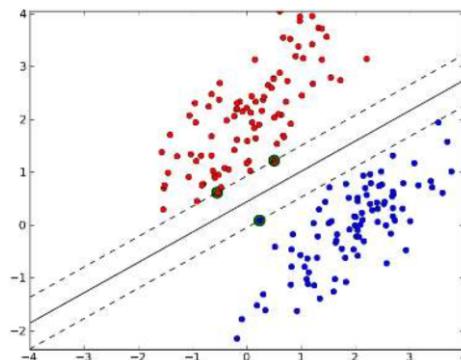


Figure 26: linear kernel

- **kernel polinomiali.** $k(x_i, x_j) = (x_i \cdot x_j)^d$ or $k(x_i, x_j) = (x_i \cdot x_j + 1)^d$

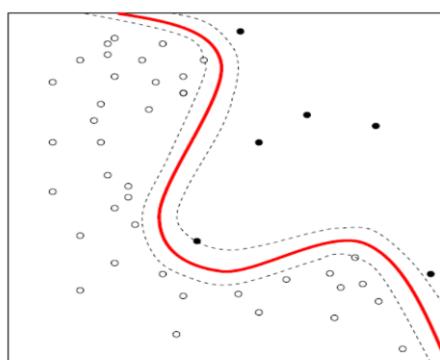


Figure 27: polynomial kernel

- **kernel RBF.** $k(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \quad \gamma > 0$

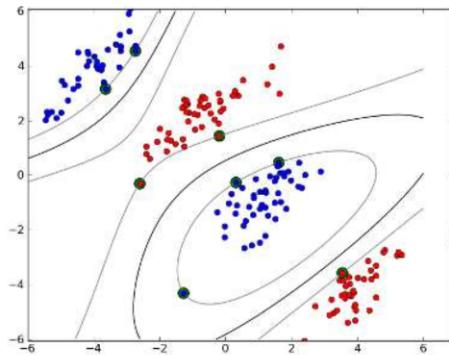


Figure 28: RBF kernel

I vantaggi delle SVM sono che sono strumenti molto potenti e che riescono a lavorare bene anche con pochi dati. Gli svantaggi sono legati all'alta onerosità computazionale, alla difficoltà di svolgere compiti non-binari ed all'assenza di una probabilità di appartenenza ad entrambe le classi classificate.

Infine, è possibile affermare che sebbene i kernel polinomiali riescano a trovare una superficie di separazione "migliore" di quella lineare, hanno il difetto di adattarsi troppo bene ai dati di training, causandone l'overfitting. Per questa ragione vengono preferiti i modelli lineari.

5.2 Applicazioni del ML nell'imaging medico

Un esempio di applicazione è la classificazione dell'Alzheimer: attraverso immagini celebrali ottenute da risonanza magnetica, si vuole classificare se il paziente è sano oppure no, attraverso features (segni di atrofia cerebrale) molto difficili da individuare. In figura 29, è possibile vedere la classificazione delle immagini in due classi, attraverso SVM.

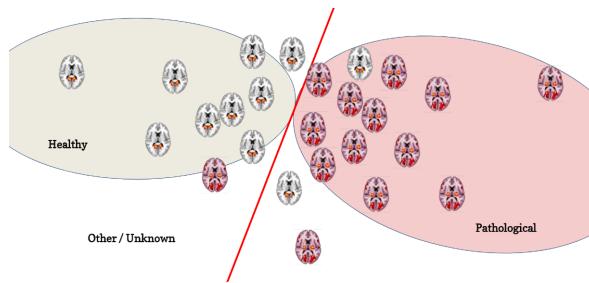


Figure 29: SVM-Alzheimer

In sintesi, il ML è utile per:

- diagnosi precoce/differenziale
- svolgere compiti di prognosi: quale sarà la corsa della malattia? Come evolverà il paziente in futuro?
- fare screening sui pazienti (*[da Wikipedia]* mira ad individuare il più precocemente possibile, nella popolazione i soggetti affetti da malattia. A differenza dei test medici eseguiti nella pratica diagnostico-clinica generale, le procedure dello screening prevedono che gli esami medici siano eseguiti a tappeto su tutta la popolazione in oggetto che, per definizione, non ha alcun sintomo né segno clinico di malattia)

- creare algoritmi utili per indirizzare la scelta del trattamento da parte del medico

Tornando all'esempio dell'Alzheimer, è possibile osservare l'andamento della malattia, stratificando la popolazione a rischio in aree, le quali descrivono la condizione clinica del paziente (figura 30) in tre differenti fasi. In questo caso il ML è utile nella fase di MCI, per creare diagnosi precoce della malattia.

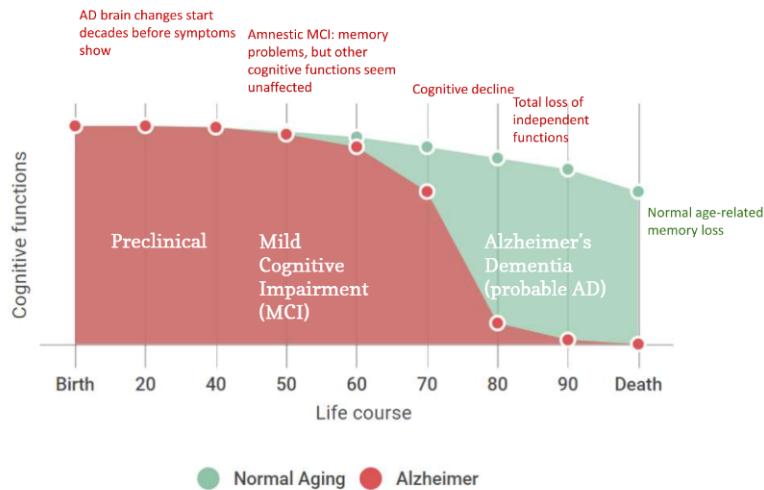


Figure 30: Corso di vita della malattia di Alzheimer

I vantaggi delle analisi multivariate (MVPA) rispetto a quelle univariate, risiedono nel fatto che durante la costruzione del classificatore vengono utilizzate un numero notevole di features che lo rendono maggiormente "sensibile". Infatti, le immagini di dimensione $n \times m \times k$ (che sono discretizzate in matrici/tensori), vengono "appiattite" in vettori di lunghezza $l = n \times m \times k$. Solitamente i vettori, in cui ogni elemento costituisce una feature, sono di lunghezza 10^6 .

Il workflow è schematizzato in figura 31.

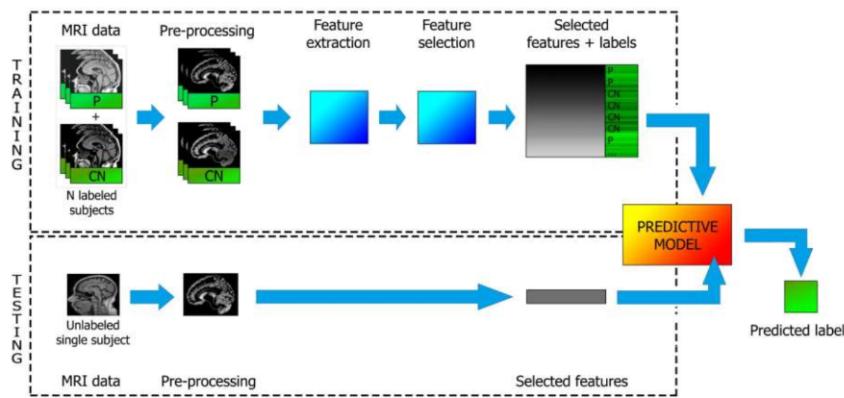


Figure 31: Workflow per la creazione dei modelli di ML

6 Pre-processing

Come descritto in figura 31, prima di passare alla fase di estrazione e selezione delle features, è necessario passare per la fase di pre-processing (ovviamente sia sui dati di training che di test). I due passaggi fondamentali durante questa fase sono la normalizzazione spaziale (ossia la co-registrazione a determinati standard) e lo smoothing (ossia il filtraggio).

Lo **smoothing**, è una tecnica largamente utilizzata per l'analisi delle immagini mediche e consente sia di distinguere le differenze tra i vari soggetti che diminuire il rumore nelle immagini.

Questo processo consiste nel mediare l'intensità del segnale per ogni voxel con le intensità dei suoi vicini; spesso vengono utilizzati filtri come quelli Gaussiani. Tale processo aiuta a normalizzare la distribuzione degli errori, migliorare il rapporto segnale-rumore ed, infine, tiene conto delle trasformazioni locali degli errori che si verificano in fase di co-registrazione (fasi di deformazione e traslazione)

7 Features extraction

Gli steps successivi al pre-processing, sono la features extraction e selection. La **features extraction** è un'operazione che è in grado di generare un nuovo set di features come funzione dei dati di input originali, che consentono una riduzione della dimensionalità del problema, mantenendo solo le informazioni rilevanti. Questo nuovo set dovrà avere le seguenti caratteristiche sulle features:

1. riduce la ridondanza
2. aumenta la rilevanza
3. riduce la dimensionalità, come conseguenza dei punti 1 e 2

I vantaggi sono l'aumento delle performance di classificazione e la diminuzione dei costi computazionali.

7.1 Voxel-level approach

Vi sono vari approcci per la features extraction, i quali usano le informazioni direttamente ad un livello voxel per la classificazione. Nel caso delle immagini neuronali si utilizzano: voxel intensities, probability values from segmented, etc.

Il vantaggio è quello di non aver escluso informazioni a priori. LO svantaggio è di aver incluso informazioni ridondanti ed irrilevanti per la classificazione, le quali potrebbero ridurre le performance in termini di accuracy.

7.2 Anatomical measurements

In medicina, una tecnica molto utilizzata e semplice per estrarre features dalle immagini è quella che si basa sul dominio di conoscenza del medico. Quindi, spesso, si tratta di estrarre delle caratteristiche anatomiche dalle immagini, basandosi su biomarcatori noti a priori, come per esempio lo spessore, il volume , la curvatura, la forma, etc. Tale approccio prende il nome di **radiomica** (estrazione delle infromazioni da immagini radiologiche).

Le features morfologiche più utilizzate in questo approccio sono: il Metabolic Target Volume (MTV), la superficie, la Spherical Disporportion (che è il rapporto tra la superficie della lesione e quella di una sfera-equivalente), la sfericità ed il rapporto superficie-volume.

7.3 Data-driven techniques

Tale approccio si basa sull'estrazione automatica delle features, senza avere una informazioni sul dominio di applicazione.

Hanno il vantaggio di identificare pattern molto complicati all'interno dei dati, i quali permettono

di identificare similarità e differenze anche su set di dati molto elevate. Gli svantaggi sono legati alla bassa interpretabilità delle features estratte e all'overfitting dei modelli. Quindi, nell'ambito medico le tecniche maggiormente preferite sono quelle in cui le features sono guidate dal medico avendo una maggiore interetabilità.

Degli esempi di tecniche di data-driven sono la Principal Components Analysis (PCA) e la Independent Components Analysis (ICA).

8 Features selection

La **features selection** ha lo stesso intento della features extraction, ossia ridurre la dimensionalità e la ridondanza dei dati e per questo vengono spesso confuse. La principale differenza è che la feature extraction genera un nuovo set di features partendo dal set di dati originale, mentre la features selection seleziona un sotto-insieme di features dal set di dati originale.

Dunque, la feature selection risulta essere uno step opzionale quando i dati contengono informazioni ridonanti o irrilevanti per il processo di apprendimento. La feature extraction è, invece, uno step obbligatorio per la ragioni sopracitate.

Il vantaggio è legato alla maggiore interpretabilità delle features, poiché restringe i dati ad un sotto-insieme di features discriminate senza alcuna trasformazione.

8.1 A-priori (ROI)

Partendo da tutti i voxel dell'immagine, i criteri per la selezione delle features più informative, possono basarsi sulla conoscenza **a-priori** di alcune caratteristiche. Per esempio, un modo utilizzato per discriminare le informazioni è la restrizione locale su determinate zone anatomiche (Region Of Interest - ROI). Quindi, sotto certe condizioni si possono selezionare delle ROI:

- basandosi su biomarcatori strutturali
- basandosi sui risultati presenti in letteratura
- con un approccio esclusivo

I vantaggi di questo approccio sono la ridotta irrilevanza dei dati e, quindi, una maggiore interpretabilità del modello. Come contro, è possibile che vengano escluse delle informazioni a priori potenzialmente utili. Per tale ragione, molti studi basati su questo approccio, consigliano di esplorare le regioni escluse da loro a priori.

8.2 Univariate feature selection

Un altro possibile approccio è assegnare un punteggio/rank alle features attraverso metodi statistici univariati, attraverso la loro rilevanza rispetto alla classificazione. Per esempio, è possibile effettuare un t-test per osservare le features con maggiore "potere statistico", applicando una soglia. Tale metodo può essere applicato sia a livello voxel che di features estratte. Gli svantaggi sono legati all'utilizzo di una tecnica univariata prima di quella multivariata, riducendone le potenzialità di quest'ultima e legati al rischiodi perdere generalizzazione del modello

8.3 A-posteriori (Wrapped)

Il metodo **a-posteriori**, seleziona le features in base alle loro performance rispetto al modello di classificazione. In particolare, si selezionano le features che hanno più potere discriminativo a livello di classificazione. Questo algoritmo prevede la selezione di sub-set diversi in modo iterativo, i quali otterranno verranno scartati o mantenuti in base alla loro performance sul modello di classificazione. Lo schema è rappresentato in figura 32.

Essendo un processo a-posteriori, questo metodo è parte del processo di valutazione delle performance della classificazione. Infatti, è molto importante compiere una buona valutazione al fine

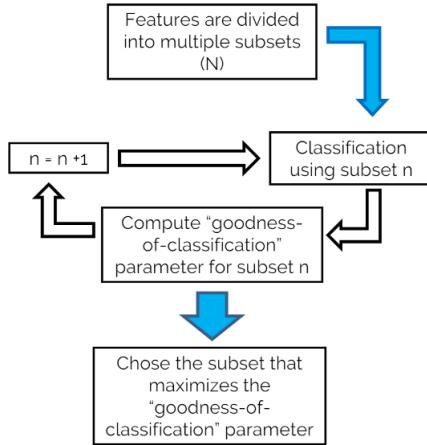


Figure 32: A-posteriori Method

di evitare l'overfitting. Le misure maggiormente utilizzate sono l'accuracy, l'area sotto la ROC curve, i pesi assegnati dal SVM, etc.

9 Training-Validation-Test-Bias

9.1 Training

Per descrivere la fase di training del modello, si consideri il workflow in figura 31, in cui vengono trattate le immagini relative all'Alzheimer. In particolare, è possibile notare che nel set di training compaiono sia persone affette da tale malattia che persone sane (con range di età simili). Le immagini, vengono scomposte in tre immagini, prima della fase di pre-processing: materia bianca, materia grigia e fluido cerebrospinale. Dopodiché, vengono estrarre e selezionate le features con le modalità sopra-descritte. In fase di training, è molto importante la scelta degli iper-parametri: possono essere scelti attraverso *grid search* o mantenuti quelli di default (SMBO?).

9.2 Validation

Un buon processo di validazione permette di ottenere il minimo bias possibile. Il validation set è molto importante per poter osservare quanto il modello è in grado di generalizzare su immagini nuove. Quindi, l'obiettivo è avere sia un basso errore nel training set, che sul validation set. Spesso, per questo tipo di analisi, vengono utilizzate le stesse immagini sia sul training che sul test, dal punto di vista della coorte considerata e provenienti dallo stesso centro clinico, cadendo nell'errore di ottenere modelli che generalizzano poco se utilizzati in ambienti e in coorti di pazienti diverse. Per esempio, lo strumento di acquisizione delle immagini incide sulle caratteristiche e sulla qualità dell'immagine stessa, che conseguentemente influirà sul modello implementato.

Per cercare di limitare il problema dell'overfitting, vengono utilizzate tecniche di cross validation, k-fold cross validation e leave-one-out cross validation (quest'ultima non raccomandabile per ricerche scientifiche).

9.3 Metriche

Le metriche maggiormente utilizzate sono: accuracy, sensitivity and specificity e AUC.

9.3.1 Accuracy

L'*accuracy* è senza dubbio la metrica più utilizzata per i problemi di classificazione. Essa è definita come:

$$\text{accuracy} = \frac{\# \text{ correctly classified samples (for all classes)}}{\# \text{ classified samples}}$$

Se l'*error rate* è definito come il numero di campioni mal classificati (per tutte le classi) diviso per il totale di campioni classificati, allora vale la seguente uguaglianza:

$$\text{error rate} = 1 - \text{accuracy}$$

9.3.2 Sensitivity and specificity

Le misure di *sensitivity* e *specificity* sono le misure maggiormente utilizzate in medicina, poichè vi è un significato clinico intrinseco ai gruppi. Esse misurano il tasso di campioni classificati correttamente sul numero dei positivi (pazienti patologici, ad esempio) e sul numero dei negativi (pazienti normali), rispettivamente nelle due classi.

La **sensitivity**, chiamata anche *recall*, è la capacità di classificare correttamente i pazienti patologici (classe positiva) e viene utilizzata per fare screening sui pazienti, ossia:

$$\text{sensitivity} = \frac{\# \text{ true positive}}{\# \text{ true positive} + \# \text{ false negative}}$$

La **specificity**, chiamata anche *True Negative Rate*, è la capacità di classificare correttamente i pazienti sani (ad esempio, nel caso del Covid-19, questa misura è molto importante), ossia:

$$\text{specificity} = \frac{\# \text{ true negative}}{\# \text{ true negative} + \# \text{ false positive}}$$

9.3.3 ROC

Nella teoria delle decisioni, le curve **ROC** (Receiver Operating Characteristic, anche note come Relative Operating Characteristic) sono degli schemi grafici per un classificatore binario. Lungo i due assi si possono rappresentare la sensitivity e la specificity. In altre parole, si studiano i rapporti fra allarmi veri (hit rate) e falsi allarmi. È un metodo molto utile per confrontare le curve ottenute dai medici. L'area sottesa alla curva è la quantificazione delle performance ottenute dal modello, statisticamente più consistenti rispetto all'accuracy

9.4 Bias

Il principale problema dei bias è legato all'uso del cosiddetto record-wise validation, anziché subject-wise. Il bias, consiste nel fatto che durante la fase di splitting del set di dati in train-validation, è necessario tenere in considerazione a quale soggetto si riferiscono determinate immagini: tale errore potrebbe causare risultati sulle performance del modello molto buone, ma distorte dal fatto che il soggetto potrebbe essere presente sia sul training set che sul validation set.

Un altro tipo di bias è il cosiddetto **circularity/double dipping**, ossia che i dati utilizzati per allenare il classificatore o per ottimizzare i parametri (inclusa la fase di feature extraction/selection o tuning degli iper-parametri) sono gli stessi per testare l'abilità di generalizzazione. Una possibile soluzione è quella di utilizzare dei test per i gold standard, ossia per le etichette da assegnare ai soggetti durante la fase di addestramento, indipendenti da quelli utilizzati dal classificatore. Un'altra possibile soluzione è l'utilizzo del follow-up clinico: addestrare i modelli su coorti cliniche in cui si ha la certezza di patologia ad un determinato tempo.