

UNIVERSITÀ DEGLI STUDI DI BARI

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

CORSO DI LAUREA IN FISICA

Tesi di Laurea

**TECNICHE DI ANALISI DI NEUROIMMAGINI
PER LA DIAGNOSI DELLA MALATTIA DI
ALZHEIMER**

Relatore:

Prof. Roberto Bellotti

Laureando:

Bruno Coladonato

ANNO ACCADEMICO 2010-2011

*Lo spirito può stare sulla punta di un dito,
eppure è grande quanto la totalità dell'universo.*

“I maestri di Huainan”

Indice

Introduzione	iv
1 Il problema sanitario	1
1.1 La Malattia di Alzheimer	1
1.1.1 Descrizione generale	2
1.1.2 Evidenze mediche	3
1.2 Il decadimento cognitivo lieve	4
1.3 Impatto sociale e costi	6
2 Diagnosi di AD con le neuroimmagini: principi e materiali per la ricerca	8
2.1 Le neuroimmagini e la diagnostica medica	9
2.1.1 Le neuroimmagini	9
2.2 La Risonanza Magnetica	11
2.3 Le neuroimmagini nella diagnosi di AD	13
2.4 Risorse, basi di dati e collaborazioni internazionali per lo studio dell'AD	16
2.4.1 ADNI	17
2.4.2 IBSR	20
2.4.3 OASIS	20
2.4.4 Altre risorse	21
2.5 Coordinate stereotassiche e atlanti	22
2.5.1 Talairach	22
2.5.2 LONI-ICBM	23

3	Neuroimaging diagnostico per l'AD: algoritmi base e metodi "classici"	24
3.1	Fase di <i>pre-processing</i>	26
3.1.1	Trasformazioni affini	29
3.1.2	Trasformazioni non lineari	31
3.1.3	Ricampionamento	32
3.1.4	Filtro di media	34
3.1.5	Filtro gaussiano	34
3.1.6	Registrazione	35
3.2	Segmentazione	37
3.2.1	Segmentazione mediante mappe statistiche parametriche	38
3.3	Classificazione	41
3.3.1	Algoritmi di decisione	42
3.3.2	Support Vector Machine	44
3.4	Tecniche di diagnosi dell'AD con le immagini di risonanza magnetica	54
3.4.1	Algoritmi basati sui voxel	56
4	Un nuovo algoritmo (pre-processing)	62
4.1	Riduzione del rumore	63
4.1.1	Steerable pyramid filter	64
4.1.2	Structural Similarity Index	66
4.1.3	Calibrazione automatica della soglia di rumore	68
4.2	Registrazione affine	70
4.3	Normalizzazione dei livelli di grigio	73
4.3.1	Algoritmo delle k-medie	75
4.4	Estrazione dei volumi di interesse	76
5	Un nuovo algoritmo (calcolo e scelta delle caratteristiche, classificazione)	80
5.1	Filtraggio	81
5.2	Scrematura dei dati	85
5.2.1	Alberi di decisione	85

INDICE	iii
5.2.2 Algoritmo Random Forest	91
5.3 Classificazione	94
6 Riepilogo e conclusioni	98
Bibliografia	102

Introduzione

L'enorme progresso tecnico-scientifico che ha caratterizzato l'ultimo secolo, ha visto lo svilupparsi di sempre più numerose specializzazioni che hanno raggiunto, nei rispettivi campi, livelli di profondità e dettaglio solo poco tempo fa impensabili. Le novità introdotte in ogni ambito della scienza e della tecnica hanno consentito un consistente aumento della conoscenza dei fenomeni e della possibilità del loro controllo, spesso però accompagnate da una crescente separazione tra discipline diverse, che ha finito così col condizionarne la direzione di sviluppo.

Negli ultimi decenni questa tendenza culturale è venuta però gradualmente cambiando: in parte spinta da esigenze tecnologiche, economiche e sociali, e in parte stimolata dalla ricerca di idee nuove e capaci di approcci innovativi, la visione interdisciplinare dell'avventura scientifica ha conquistato spazi sempre più ampi e complessi, rinnovando paradigmi e modi di operare. Da un lato questo ha portato alla creazione di gruppi di ricerca caratterizzati da competenze le più diverse, ma complementari rispetto ai problemi da affrontare; dall'altro, più semplicemente se vogliamo, ha ampliato gli obiettivi di indagine delle scienze applicate, ponendo quesiti che richiedono prospettive diverse e a volte inattese.

La diagnosi medica mediante tecniche di analisi automatica delle immagini può considerarsi appartenente ad entrambe le categorie menzionate. L'autentica rivoluzione apportata in Diagnostica Medica dall'avvento di strumenti di indagine come la Tomografia Assiale Computerizzata, la Risonanza Magnetica Nucleare, la Tomografia ad Emissione di Positroni, con tutte le loro varianti, è stata frutto dell'intensa collaborazione, tra gli altri, di fisici, ingegneri, medici radiologi, matematici, informatici; si è avval-

sa inoltre dell'intenso sviluppo delle tecnologie elettronica e dei calcolatori. Tuttavia, nell'uso consueto, è sempre l'occhio esperto del medico che effettua le valutazioni del caso. Il passo successivo, ossia quello della diagnosi con metodi automatici, assume invece aspetti che pongono problemi attinenti sia al calcolo automatico che alla matematica applicata, passando ora per la teoria delle decisioni, ora per tecniche di *data mining*, muovendosi comunque in quella vasta area spesso definita col termine generico e suggestivo di *intelligenza artificiale*. Tutto senza trascurare metodi più "classici" come il trattamento del segnale e l'analisi statistica dei dati.

Questa tesi si propone di fornire una rassegna delle principali tecniche di analisi delle neuroimmagini, dedicate al problema della diagnosi della malattia di Alzheimer, esaminando l'intera catena di elaborazione automatica che, partendo dall'immagine acquisita dalle macchine di Risonanza Magnetica, arriva a fornire un'ipotesi di diagnosi. Verranno considerate, oltre agli aspetti fondamentali e agli algoritmi di uso consolidato nella ricerca, le procedure più recenti presenti in letteratura.

Focalizzare la ricerca su un particolare problema diagnostico, come in questa tesi, oltre all'ovvio interesse che la soluzione del problema stesso comporta, consente di definire un preciso terreno di indagine, in grado di rivelare le diverse potenzialità e attitudini delle tecniche a disposizione; consente inoltre di indagare sulla natura stessa del problema attraverso le risposte che i vari approcci sono in grado di fornire.

La rilevanza umana e sociale di un problema che comporta la sofferenza di pazienti e loro familiari costituisce, infine, uno stimolo che, al di là di ogni altro, basterebbe da solo a giustificare ogni entusiasmo nella ricerca delle migliori soluzioni possibili.

Capitolo 1

Il problema sanitario

1.1 La Malattia di Alzheimer

Le demenze sono delle patologie cronico-degenerative destinate ad avere nel corso dei prossimi anni, come è opinione diffusa nella ricerca medica, un incremento esponenziale della prevalenza¹, e dell'incidenza², con conseguente aumento della spesa sociale e sanitaria. Esse hanno un impatto notevole in termini socio-sanitari sia perché un sempre maggior numero di famiglie sono interessate al problema, sia perché richiedono una qualificata rete integrata di servizi sanitari e socio-assistenziali.

Nel Rapporto Mondiale Alzheimer 2009 [1], Alzheimer's Disease International (ADI) ha calcolato in 35,6 milioni le persone affette da demenza nel mondo nel 2009 [1][2], che nel 2020 i malati saranno oltre 42 milioni, e che essi aumenteranno a 65,7 milioni nel 2030 ed a 115,4 milioni entro il 2050.

L'aumento del numero di casi rilevati negli ultimi anni è giustificato da un lato dalla cosiddetta *transizione demografica*, cioè la tendenza a un progressivo aumento della vita media e ad una riduzione delle nascite con conseguente invecchiamento della popolazione, in atto in Italia come in tut-

¹ Prevalenza = (numero di soggetti malati in un dato istante)/(numero di soggetti sani esposti alla malattia nello stesso istante). Esprime la probabilità che un soggetto appartenente alla popolazione *sia* malato.

² Incidenza = (nuovi casi di malattia in un fissato periodo di tempo)/(soggetti inizialmente sani che possono contrarre la malattia); il periodo di tempo è scelto dallo sperimentatore in base alle caratteristiche della patologia. Stima la probabilità che un soggetto *contragga* la malattia in tale periodo.

to il mondo; dall'altro dall'avanzamento delle tecniche diagnostiche che ne rende possibile oggi la diagnosi in sempre più casi.

Descritta per la prima volta dal neuropatologo tedesco Alois Alzheimer nel 1906, la *malattia di Alzheimer* (Alzheimer's Disease, AD) è considerata la più comune causa di demenza, rappresentandone dal 50-56% dei casi secondo alcune stime [3], fino al 75% secondo altre (in tabella le percentuali dei vari tipi di demenza secondo ADI).

Principali tipi di demenza (World Alzheimer Report 2009)

Tipo di demenza	Proporzione
Malattia di Alzheimer (AD)	50–75%
Demenza di tipo vascolare (VaD)	20–30%
Demenza a corpi di Lewy (DLB)	<5%
Demenza fronto-temporale (FTD)	5–10%

L'AD è caratterizzata da una progressiva alterazione della funzionalità neuronale, che porta a un deterioramento delle funzioni cognitive e delle capacità esecutive, accompagnate da modificazioni comportamentali.

Sebbene sia stata descritta in ogni periodo dell'età adulta, la maggior parte dei pazienti ha più di sessant'anni, per cui questo tipo di demenza può essere definito come un fenomeno età-correlato: esso colpisce più del 5% degli individui tra i 60 e i 65 anni di età, e questa percentuale aumenta rapidamente al 19% dopo i 75 anni e al 30% dopo gli 85 [4], evidenziando un drastico incremento di incidenza dopo i 65 anni [5]. Sempre dopo i 65 anni, la prevalenza raddoppia ad ogni incremento di età di cinque anni: il principale fattore di rischio per la demenza è perciò da considerarsi proprio l'età avanzata [1].

1.1.1 Descrizione generale

La malattia di Alzheimer è una patologia neurodegenerativa, irreversibile e progressiva, che distrugge lentamente la memoria e le abilità cognitive con alterazione della personalità e del comportamento [7].

La caratteristica peculiare di questa malattia è la presenza di una atrofia cerebrale diffusa, la cui progressione risulta particolarmente accelerata nelle fasi precoci di malattia e maggiormente evidente a livello dell'ippocampo e delle porzioni mediali dei lobi temporali.

La storia naturale della malattia di Alzheimer prevede una lunga fase iniziale, detta *prodromica*, che può essere considerata una fase intermedia tra l'invecchiamento cognitivo fisiologico e una sindrome dementigena vera e propria; in questo periodo iniziale la patologia si manifesta con alcuni sintomi premonitori tra i quali: perdita di memoria, difficoltà nelle attività quotidiane, problemi di linguaggio, disorientamento nel tempo e nello spazio, diminuzione della capacità di giudizio, difficoltà nel pensiero astratto, cambiamenti di umore, di comportamento, di personalità e mancanza di iniziativa.

In una seconda fase l'amnesia diviene più pronunciata e si accompagna a segni di agnosia, afasia, aprassia (difficoltà nel riconoscere le persone, deterioramento del linguaggio, difficoltà nel movimento), e a disturbi nelle funzioni esecutive [8].

La fase terminale sopraggiunge, in generale, dopo circa venti anni dall'esordio, ma, nei casi più gravi, anche dopo soli tre anni. In questo stadio della malattia si osserva una disintegrazione delle funzioni vitali, apatia ed inerzia; segue la morte causata da patologie collaterali, come quelle a carico dell'apparato respiratorio.

1.1.2 Evidenze mediche

L'esame del cervello di un tipico paziente affetto da AD permette di rilevare due tipi di formazioni caratteristiche: placche senili amiloidi (SP), che si depositano a livello extracellulare, e depositi di aggregati, o *grovigli*, neurofibrillari (NFT), visibili a livello intracellulare [9]. Le aree cerebrali dove si registra il maggior accumulo delle placche amiloidi sono la *corteccia temporale*, la *corteccia parietale* e *l'ippocampo*, che sono proprio aree associate con le funzioni della memoria e dell'apprendimento [10].

In effetti la semplice presenza o assenza delle placche senili e dei grovigli neurofibrillari non è determinante dal punto di vista diagnostico: placche e

grovigli neurofibrillari sono presenti anche nell'encefalo di individui anziani che non mostrano segni di demenza. È necessario quindi determinare la quantità di tali formazioni che, soprattutto, devono essere accompagnate da segni clinici di demenza [10].

Al momento dell'autopsia [11], negli stadi avanzati di malattia, il cervello presenta un aspetto diffusamente atrofico, con la conseguente dilatazione delle cavità ventricolari nonché, a livello della corteccia, l'allargamento dei solchi e l'assottigliamento delle circonvoluzioni; inoltre il suo peso è solitamente ridotto del 20% o più a causa di una significativa perdita di cellule nervose, sinapsi e dendriti [8]. Le aree più vulnerabili del cervello, dove si registra un pronunciato processo neurodegenerativo, sono i *lobi temporali mediali* (MTL), con particolare interessamento dell'ippocampo, della *corteccia entorinale* (ERC) e della *circonvoluzione paraippocampica*.

In realtà una diagnosi certa di AD può essere effettuata solo *postmortem* con l'esame autoptico; tuttavia, incrociando opportunamente test cognitivi ed esami clinici specifici, è possibile formulare con ragionevole certezza una diagnosi di AD, differenziandola dalle altre forme di demenza, anche in soggetti in vita: questo, però, solo quando la malattia ha già raggiunto una fase avanzata.

1.2 Il decadimento cognitivo lieve

Negli ultimi anni l'attenzione della ricerca scientifica e clinica nell'ambito della malattia di Alzheimer si è spostata sulla diagnosi precoce, focalizzandosi in particolare sulla fase di transizione tra il normale processo di invecchiamento e la demenza [12].

Delineare i confini di questa situazione intermedia tra invecchiamento cerebrale "fisiologico" e demenza, capire fin dove si è autorizzati a considerare "normale" il declino della memoria o delle altre funzioni cognitive e quando invece considerarli come fase iniziale di una malattia dementigena è un compito difficile, sul quale sono tuttora aperti ampi dibattiti.

Questo tentativo di individuare uno stato clinicamente ben definito che sia prodromico dell'AD, ha portato nel tempo alla definizione di numerosi quadri clinici adatti a distinguere i soggetti affetti da lievi disturbi cognitivi

età-correlati dai soggetti sani (Age-Associated Memory Impairment, Benign Senescent Forgetfulness, Age-Related Cognitive Decline, Age-Associated Cognitive Decline, ...).

Oggi la caratterizzazione prevalentemente utilizzata nella ricerca sull'AD per descrivere lo stadio intermedio tra l'invecchiamento cerebrale normale e la demenza è il cosiddetto *decadimento cognitivo lieve*, o *Mild Cognitive Impairment* (MCI) [12][13] con cui si fa riferimento a soggetti con età > 65 anni.

I criteri diagnostici per definire il MCI includevano inizialmente [12][17]:

- disturbo soggettivo di memoria, possibilmente confermato da un familiare;
- disturbo oggettivo di memoria, dimostrato con test neuropsicologici, di entità superiore a quello atteso relativamente all'età e al livello di educazione;
- integrità delle altre funzioni cognitive; conservata autonomia nella vita quotidiana;
- assenza di demenza.

In seguito il concetto di MCI è stato ampliato introducendo un "MCI non amnestico" (na-MCI), nel quale il paziente ha disturbi in domini cognitivi diversi dalla memoria, ad esempio nelle funzioni esecutive e nel linguaggio [14], distinto dal "MCI amnestico" (a-MCI), in cui il soggetto mostra i già citati disturbi nella memoria [12]. Per includere un più largo spettro di condizioni sono stati introdotti anche dei sottotipi diagnostici, che però in questo lavoro non interessano.

Similmente all'AD, uno dei principali fattori di rischio nell'insorgenza del decadimento cognitivo sembra essere l'età: si stima che la tendenza a sviluppare MCI vari dal 15% registrato ad un'età inferiore ai 75 anni, al 30% ad un'età superiore agli 85 [15].

Come già detto, l'interesse in questa nuova definizione diagnostica risiede nel fatto che si ritiene il MCI essere una patologia che *precede* l'insorgere della demenza, e in particolar modo l'AD [16]: è stato infatti evidenziato

che gli individui affetti da MCI mostrano un aumentato rischio, ben 6,7 volte superiore, di sviluppare demenza rispetto ad un soggetto sano [14].

Questo suggerisce di cercare un criterio di diagnosi precoce dell'AD focalizzando la ricerca proprio sui pazienti affetti da MCI: questi possono in breve tempo sviluppare demenza, secondo alcuni studi approssimativamente dal 10% al 15% in un anno, e addirittura dal 40% all'80% in cinque, sei anni [13][15].

Nonostante ciò, molte persone colpite da tale patologia possono rimanere stabili o mostrare solo un lento deterioramento cognitivo, inoltre un numero sorprendente di pazienti, dal 17% al 32%, mostra persino un miglioramento spontaneo delle abilità cognitive [12][15].

Tutto questo, unito alla notevole incertezza delle stime contenute nella letteratura medica, suggerisce prudenza nelle conclusioni, ma nel contempo indica che la strada intrapresa con la ricerca delle fasi prodromiche dell'AD nei soggetti affetti da MCI è molto promettente e mostra fondamenta particolarmente solide.

1.3 Impatto sociale e costi

Dare una valutazione dell'impatto sociale causato dalle demenze è sicuramente un compito impossibile: le persone affette da demenza, i loro familiari, gli amici, sono colpiti a livello personale, emozionale, finanziario e sociale. Nessuna cifra può esprimere il disagio che la disabilità causata dalla malattia può comportare.

Tuttavia le cifre sono sicuramente idonee a quantificare il costo economico che ad essa consegue, ed è interessante dare un'idea della possibile ricaduta sulla ricchezza che la ricerca può fornire col suo lavoro.

Saranno sufficienti pochi dati, tutti ricavati dal Rapporto Mondiale Alzheimer 2010 [6], che proprio a questi temi ha dedicato la sua attenzione:

“Nel 2010 i costi totali mondiali stimati ammontano a 604 Miliardi \$ US. Questi costi rappresentano circa l'1% del PIL (Prodotto Interno Lordo) mondiale . . . ”

“Se i costi globali della demenza fossero una nazione, rappresenterebbe la 18^a economia mondiale . . . ”

“La dimensione del costo globale della demenza è spiegabile se si considera che circa lo 0.5% della popolazione mondiale vive con la demenza.”

Si aggiunga a questo che: “I costi dell’assistenza informale (assistenza non rimborsata sostenuta dalle famiglie ed altri soggetti) ed i costi diretti dell’assistenza sociale (fornita dall’assistenza sociale professionale delle comunità ed in strutture residenziali) contribuiscono in proporzioni simili (42%) al totale mondiale dei costi, mentre i costi sanitari diretti restano a livello molto più basso (16%)”; e che: “ . . . nelle nazioni a basso reddito pro capite, l’assistenza informale incide per la maggior parte del totale dei costi sostenuti e quelli dell’assistenza sociale diretta sono di entità trascurabile” [ibid.], per capire quale impatto formidabile l’AD possa avere anche e soprattutto in quei paesi dove l’assistenza sanitaria è fortemente carente.

Queste malattie toccano quindi profondamente le economie di tutti i paesi, pur se in modo diverso. Anche in quest’ottica la ricerca può rivestire un’importanza particolarmente significativa.

Capitolo 2

Diagnosi di AD con le neuroimmagini: principi e materiali per la ricerca

L'uso di immagini per la diagnosi medica, le cui potenzialità furono evidenti sin dalla nascita stessa della radiologia, ha acquisito nel tempo un rilievo sempre maggiore, e conosce ancor oggi uno sviluppo rapido e di vaste proporzioni: gli strumenti tecnologici e le moderne acquisizioni della scienza medica non solo consolidano e perfezionano le tecniche già note, ma continuano ad aprire nuovi e a volte inaspettati scenari applicativi.

In questo contesto il ruolo della fisica applicata alla medicina si è andato ulteriormente ampliando, superando i tradizionali ambiti della pratica radiologica e di medicina nucleare, e inserendosi in quel nuovo quadro di interdisciplinarietà che oggi è ormai prassi in molta ricerca avanzata sui problemi della salute umana.

Il fisico, già valido ausilio nella comprensione di molti processi fisiologici, nonché coprogettista e consulente nell'uso di apparecchiature mediche radio-nucleari, oggi è presente, coi suoi metodi e le sue competenze, ovunque la ricerca richieda l'uso di strumentazioni e tecnologie biomediche avanzate.

2.1 Le neuroimmagini e la diagnostica medica

Negli ultimi venti-trenta anni, molti progressi della fisica applicata alla medicina sono stati effettuati nel campo delle immagini.

La digitalizzazione delle immagini, oltre a rappresentare una evoluzione tecnologica naturale nel settore dell'*imaging* medico, si è rivelata essere anche decisiva: ha consentito e stimolato lo sviluppo di tecniche avanzate come Tomografia Assiale, Risonanza Magnetica, Tomografia a Emissione di Positroni, ecc., oltre che fornito a medici e chirurghi strumenti sempre più accurati, che hanno permesso la nascita di nuovi metodi di diagnosi e nuove procedure chirurgiche.

Di più: trasformare una immagine in una sequenza numerica ha consentito l'ingresso di nuove sofisticate tecniche matematiche e informatiche, suggerendo la possibilità di ottenere anche informazioni *quantitative*, sia di tipo morfologico che funzionale, da strumenti usati fino ad allora solo come, anche se fondamentale, ausilio alla valutazione clinica umana.

In realtà, a tutt'oggi valutazione medica e diagnosi sono sempre appannaggio finale esclusivo dell'uomo, ma si scorge tuttavia la possibilità sia di fornire al medico dati più precisi di quanto possa rilevare un occhio anche allenato, sia di svincolare la ricerca dalla soggettività da cui nessuna persona può considerarsi esente. Questo può essere un elemento fondamentale nella ricerca delle cause e/o delle evidenze cliniche che accompagnano l'evoluzione delle malattie, oltre che contribuire a chiarire quali parametri accompagnino il corretto e naturale funzionamento del corpo umano.

È un problema di misura che comporta problematiche radicalmente nuove, con molti aspetti dalle caratteristiche originali, da affrontare con strumenti innovativi, e in grado di porre difficoltà a volte formidabili: un terreno dove la professionalità del fisico trova un ambito di azione naturale.

2.1.1 Le neuroimmagini

Nel campo dello studio del sistema nervoso centrale l'elaborazione e l'analisi delle immagini, sia quando l'attenzione è focalizzata sull'indagine *morfologica*, ossia sulla individuazione delle varie strutture anatomiche di cui valutare dimensioni e caratteristiche, sia quando si indirizza la ricerca

sugli aspetti *funzionali* che lo riguardano, come il metabolismo durante lo svolgimento di particolari attività o i flussi sanguigni in determinate aree di interesse, spinge la ricerca a concentrarsi, tra gli altri, su tre problemi specifici:

- l'identificazione di particolari regioni e strutture ritenute clinicamente rilevanti (*segmentazione, analisi di importanza*);
- la valutazione delle loro dimensioni e densità (*morfometria*);
- la determinazione della classe di appartenenza di un individuo tra persone affette o non affette da una particolare malattia (*classificazione*);

il tutto con la richiesta aggiuntiva, ed essenziale, di eseguire questi compiti in maniera completamente *automatica*.

La scienza delle neuroimmagini, o *neuroimaging*, è così di fatto diventata una disciplina a sé stante, ma profondamente interconnessa con informatica e matematica applicata: essa usa metodi sia squisitamente specialistici, come quelli legati a filtraggio e preelaborazione del segnale (costituito dall'immagine), sia di uso generalizzabile e comune a discipline affatto diverse come ad esempio l'economia, cosa che d'altronde non deve sorprendere quando si usano strumenti concettuali molto generali.

Per ultimo è importante sottolineare come la digitalizzazione delle neuroimmagini abbia consentito un ulteriore fondamentale ampliamento degli strumenti di ricerca: la condivisione attraverso *basi di dati internazionali*, che mette a disposizione dei ricercatori di ogni parte del mondo immagini in formato standardizzato, controllate e validate dal punto di vista medico al massimo livello disponibile. Questo permette ai vari gruppi di ricerca di utilizzare materiale comune e di caratteristiche note, facilitando confronti obiettivi fra i risultati ottenuti, e favorendo collaborazioni in rete tra gruppi diversi.

Quest'ultimo aspetto è in fase di evoluzione, ma sono già disponibili da alcuni anni banche di dati riconosciute e utilizzate a livello internazionale.

2.2 La Risonanza Magnetica

Nel campo delle neuroimmagini, quello della *Risonanza Magnetica* (MR) è il settore che forse più di tutti si è giovato dell'impressionante impulso innovativo offerto dai recenti sviluppi tecnici, tanto da diventare oggi il principale strumento della diagnostica neuroradiologica: basandosi sul fenomeno della risonanza magnetica nucleare dei nuclei di idrogeno, la MR è capace di produrre immagini tridimensionali con risoluzione dell'ordine del millimetro e oltre.

La sua nascita si può far risalire ai primi anni '70 con gli esperimenti compiuti da Lauterbur su campioni "in vitro" prima e su animali da esperimento poi [18]; i rapidi sviluppi di elettronica ed informatica, oltre che della tecnologia dei magneti, hanno quindi portato alla produzione di apparecchiature (*scanner*) evolute operanti a campi magnetici di intensità elevata, consentendo l'impiego di tecniche di acquisizione sempre più avanzate.

La generazione del segnale MR si basa essenzialmente sulla *precessione di Larmor* che il momento magnetico di un protone compie in presenza di un campo magnetico esterno. Si pone l'attenzione proprio sul protone sia perché dotato di un momento magnetico significativo, sia in quanto nucleo dell'atomo di idrogeno, elemento presente largamente e in proporzioni differenti nei diversi tessuti del corpo umano.

A causa del suo spin, quando è immerso in un campo magnetico B_0 sufficientemente intenso¹, il protone assume una di due possibili orientazioni: "parallela", ossia con componente L_z concorde con B_0 , o "antiparallela", con componente L_z discorde; quest'ultima rappresenta uno stato di energia maggiore rispetto alla prima, con differenza tra i due stati proporzionale a B_0 .

Quando i protoni sottoposti al campo sono molti, vi è prevalenza di orientazioni a bassa energia (parallele) rispetto a quelle ad energia più alta (antiparallele), disparità tanto maggiore quanto più intenso è il campo esterno: il risultato complessivo di questo "eccesso di spin" è una magnetiz-

¹ Gli apparecchi attualmente disponibili usano campi che vanno da 1 a più di 10 Tesla, con questi ultimi riservati comunque solo a scopo di ricerca.

zazione netta del tessuto sottoposto ad esame, diretta come B_0 e crescente con esso.

Come già detto, in queste condizioni il momento magnetico precessa intorno alla direzione del campo con frequenza di Larmor $\nu_L = \gamma \cdot B_0 / 2\pi$. In equilibrio termico ogni protone precessa indipendentemente dagli altri, per cui la precessione netta totale è nulla.

Per creare un segnale utile si procede quindi con l'applicazione di un campo elettromagnetico, detto B_1 , oscillante proprio alla frequenza di Larmor ν_L , e normale al campo statico B_0 . Si ottengono in questo modo due diversi effetti:

- si sincronizza la precessione degli spin dei singoli protoni, portandoli in fase tra loro;
- si provoca la transizione di alcuni protoni, mediante assorbimento di energia dal campo oscillante, dallo stato di spin parallelo a quello antiparallelo, provocando di fatto il “passaggio” del loro momento magnetico dal lato opposto del piano xy , con conseguente diminuzione della magnetizzazione lungo la direzione di B_0 .

Entrambi questi processi contribuiscono alla formazione di un segnale indotto in un'opportuna bobina di ricezione.

Per acquisire il segnale è però necessario interrompere il campo oscillante, quindi il sistema tende a tornare al precedente equilibrio: gli stati antiparalleli si diseccitano con un parametro caratteristico di rilassamento T_1 (o T_1), e la precessione perde coerenza di fase secondo un proprio parametro di rilassamento T_2 (o T_2 , o anche T_2^*). Il segnale MR (*eco*) è quindi costituito dall'energia in radiofrequenza rivelata dal sistema fisico durante i processi di rilassamento.

L'uso opportuno dei parametri di rilassamento da parte dell'operatore, secondo precisi e stabiliti protocolli radiologici, consente poi di ottenere immagini ottimizzate a seconda delle regioni e dei tessuti da esaminare².

² I suddetti processi avvengono secondo leggi complesse, determinate dall'interazione dei parametri di rilassamento con i parametri di *sequenza* scelti dall'operatore e denominati *tempo di ripetizione* (TR) e *tempo di eco* (TE), dell'ordine delle centinaia di ms il primo, delle decine di ms il secondo [18].

Il problema di determinare la collocazione nello spazio delle singole porzioni di tessuto viene infine risolto facendo ricorso ad un campo magnetico a gradiente, detto semplicemente *gradiente*, aggiunto ai precedenti³. Le frequenze in gioco rientrano infatti nel campo delle radiofrequenze (RF), rendendo impossibile la necessaria risoluzione spaziale con metodi basati su proiezioni come nella Tomografia Assiale (trasformata di Radon e simili): basti considerare che un protone libero in un tipico scanner MR da 1,5 Tesla ha una frequenza di Larmor dell'ordine dei 63 MHz, cui corrisponde una lunghezza d'onda di quasi cinque metri.

Un campo magnetico opportunamente variabile nelle tre direzioni dello spazio⁴ genera invece, grazie alla dipendenza di ν_L dall'intensità del campo, frequenze di Larmor diverse a seconda della posizione: quest'ultima è perciò determinabile senza difficoltà dall'analisi in frequenza del segnale, rendendo così possibile la costruzione finale dell'immagine.

2.3 Le neuroimmagini nella diagnosi di AD

L'idrogeno è presente in maniera consistente in larga parte del corpo umano: principalmente come componente dell'acqua, che da sola ne costituisce circa il 70% del peso, ma anche di molti altri tessuti, come in particolare quello adiposo. Le diverse quantità e la varietà delle strutture molecolari di cui è parte ne determinano, in campo magnetico esterno, un comportamento peculiare e riconoscibile, per esempio: magnetizzazione netta più intensa nei tessuti che ne hanno maggior concentrazione [19]; moto rotazionale più vicino alla frequenza di Larmor nelle molecole di medie dimensioni come quelle lipidiche (con rilassamento T1 particolarmente veloce); interazione energetica poco efficiente alla frequenza ν_L per le molecole piccole come quelle d'acqua a causa del loro veloce stato di rotazione, e per quelle grandi come le macromolecole proteiche per ragioni opposte; e via dicendo [18]. Si sono così individuati sperimentalmente, a seconda dei tessuti presenti nei vari distretti anatomici, i protocolli più opportuni per l'acquisizione di immagini.

³ In realtà il campo oscillante B_1 è applicato ad impulsi, *durante* i quali e *tra* i quali viene applicato ad intervalli il campo gradiente [18].

⁴ La variazione del campo gradiente è dell'ordine di alcune decine di mTesla/m [20].

Nel campo specifico delle neuroimmagini da Risonanza Magnetica (MRI⁵) del cervello, i tessuti coinvolti sono praticamente solo tre: sostanza grigia (*gray matter*, GM), sostanza bianca (*white matter*, WM), liquido cerebrospinale (*cerebrospinal fluid*, CSF), quest'ultimo con forte componente acquosa.

Questo ha consentito nel tempo di elaborare e fissare protocolli *standard* di acquisizione per tutte le neuroimmagini MR di comune uso diagnostico. Nelle ricerche che interessano la diagnosi di AD basata sulle normali scansioni di risonanza magnetica *strutturale*, ossia che mostrano la distribuzione spaziale dei diversi tessuti, il protocollo generale di riferimento è quello denominato semplicemente "T1", e le MRI che genera vengono indicate in letteratura come immagini "pesate in T1" (*T1-weighted MRI*); se non indicato diversamente, questo è dato per sottinteso.

La ricerca condotta su immagini prodotte partendo da una base comune assicura non solo un indispensabile fondamento di omogeneità alla conseguente analisi, ma mette a disposizione un vastissimo archivio potenziale di scansioni MR eseguite nel corso degli anni nelle strutture sanitarie di diverse parti del mondo. A ciò si aggiunge la possibilità di sviluppare in futuro nuovi ausili diagnostici disponibili nella pratica, localmente o mediante invio telematico delle immagini, senza dover rinnovare l'*hardware* già oggi operativo nei servizi sanitari.

Naturalmente sono presenti ancora molti fattori variabili nelle singole acquisizioni, responsabili di disomogeneità e non sempre controllabili. Alcuni di essi sono noti, come le esatte dimensioni e il numero dei singoli elementi dell'immagine (*voxel*), o, quello che è il più caratterizzante di tutti, l'intensità del campo B_0 , che varia a seconda del modello di apparecchiatura utilizzata, ma che fortunatamente ha da tempo trovato nel valore di 1,5 Tesla un compromesso tra precisione e costi che ne ha fatto lo *standard* commercialmente più diffuso. Altri risiedono invece in caratteristiche più sfuggenti, come quelle legate alle specifiche costruttive e di messa a punto di ogni singolo apparecchio MR, agli elementi contestuali presenti nella misura,

⁵ La sigla MRI può indicare sia le immagini di risonanza magnetica (*Magnetic Resonance Image*), che la tecnica di trattamento e analisi delle stesse (*M.R. Imaging*). Nel seguito verrà usata sempre col primo significato.

e ai margini di regolazione e di filtraggio nelle mani dell'operatore radiologo, questi ultimi squisitamente legati alle preferenze e all'abilità soggettiva.

Il problema di uniformare i dati per la ricerca da questo punto di vista viene affrontato nella fase di preelaborazione, trattata più avanti.

Ma non è questo che rende oggi la diagnosi delle demenze degenerative in generale, e della malattia di Alzheimer in particolare, una delle maggiori sfide dell'analisi automatica delle immagini applicata alla neurologia. L'uso delle neuroimmagini in tali condizioni patologiche, infatti, pone questioni molto più sottili.

In questo campo l'*imaging* convenzionale ha fino ad ora avuto tipicamente il compito di *escludere* la presenza di problemi che, pur diversi dalla demenza, fossero tuttavia in grado di *mimare* la sintomatologia; all'interno dell'ambito stesso delle demenze il suo ruolo è stato generalmente limitato al tentativo di *differenziarne* la natura, evidenziando ad esempio problemi vascolari piuttosto che un'atrofia diffusa.

Il problema che ora si pone, invece, è arrivare ad una possibile diagnosi *diretta* dell'AD, e non quando l'atrofia che interessa il cervello è talmente importante da essere facilmente accertabile anche con un semplice esame visivo della MRI, ossia in una fase avanzata, ma cercando di coglierne prove diagnostiche in uno stadio il più possibile precoce, ossia quando le modificazioni morfologiche delle aree interessate non sono ancora visibili all'occhio del radiologo.

La richiesta poi di automatizzare completamente questo processo risponde a precise necessità:

- fornire una base diagnostica il più possibile *oggettiva* e condivisa, per consentire il monitoraggio e lo studio dell'evoluzione della patologia a partire dalle sue fasi iniziali;
- realizzare protocolli di calcolo che forniscano risultati *confrontabili* in base alla loro effettiva capacità discriminatoria, e non ad abilità ad essi esterne;
- sfruttare la capacità dei calcolatori di elaborare massicce quantità di dati per estrarre informazioni utili in maniera *originale* rispetto alla

valutazione umana, evitando quindi qualunque intervento manuale sui dati stessi;

- analizzare il tipo di *caratteristiche* estratte automaticamente dalle immagini che si rivelano significative per la diagnosi, per meglio individuare le modificazioni morfologiche tipiche dell'esordio della malattia.

Come detto in precedenza, la diagnosi certa di malattia di Alzheimer è possibile solo quando vi è presenza nel cervello di particolari depositi proteici extracellulari e strutture (grovigli) intracellulari che si aggiungono alle altre evidenze cliniche: allo stato attuale però, nonostante l'intensa ricerca in atto, non esistono strumenti di imaging o altri test in grado di accertare tale presenza "in vivo".

Le immagini di risonanza magnetica sono invece uno strumento ideale per tentare una valutazione precisa dello stato di atrofia cerebrale, quantificare la distribuzione regionale del danno, cercarne segni morfologici premonitori.

Si tratta, in ultima analisi, di individuare in una persona riduzioni o alterazioni significative della materia grigia rispetto ai soggetti sani di pari età, compito ben più difficile che seguire nel tempo l'evoluzione delle MRI dello stesso paziente. Le differenze anatomiche *inter-individuali* (tra individui diversi) sono spesso ben più nette di quelle *intra-individuali* (nello stesso individuo), quindi l'analisi non può ridursi alla mera ricerca di differenze, ma deve piuttosto trovare un affidabile *riferimento di normalità*, e discriminare solo quelle caratteristiche che lo differenziano da un soggetto affetto da patologia.

Il primo di questi problemi, il riferimento di "normalità", viene considerato in un prossimo paragrafo; il secondo, la ricerca di caratteristiche discriminanti, sarà argomento del capitolo seguente.

2.4 Risorse, basi di dati e collaborazioni internazionali per lo studio dell'AD

Grazie alla qualità e alla risoluzione delle sue immagini, ma soprattutto grazie all'uso di radiazioni non ionizzanti, la tecnica della risonanza magnetica è diventata largamente popolare in campo diagnostico e nella ricerca:

la possibilità di ottenere immagini strutturali dettagliate su ogni piano geometrico l'ha resa una tecnologia di elezione nello studio delle malattie degenerative del cervello, portando anche allo sviluppo di nuove tecniche di imaging strutturale (*Diffusion Tensor Imaging*, DTI) e funzionale (*functional Magnetic Resonance Imaging*, fMRI; *functional Diffusion Tensor Imaging*, fDTI) basate su di essa [21].

Il progresso della ricerca sui metodi computazionali di analisi dell'immagine, ha presto fatto emergere la necessità di disporre di una base condivisa per il loro sviluppo, la sperimentazione e il confronto.

Una base di dati (*database*) internazionale, contenente immagini di qualità validate da esperti, e disponibile facilmente in rete informatica, si è rivelata lo strumento ideale allo scopo: diverse collaborazioni si sono per questo sviluppate negli anni, alcune più generali, altre con obiettivi più specifici, come ad esempio proprio la lotta alla malattia di Alzheimer.

L'incalzare dei risultati delle ricerche ha portato ad un continuo miglioramento dei protocolli e del materiale a disposizione, ma d'altro canto ha creato qualche difficoltà nella standardizzazione dei metodi, e soprattutto nel confronto delle diverse metodologie di analisi: i vari gruppi di ricerca fanno oggi riferimento a fonti diverse, rendendo non perfettamente uniforme il materiale sperimentale di base. L'alta qualità raggiunta dalle neuroimmagini rende però i dati più omogenei, compensando in parte il problema.

2.4.1 ADNI

Le note seguenti riguardano quella che oggi è forse la principale risorsa internazionale di MRI per lo studio della malattia di Alzheimer: le problematiche qui accennate forniscono un'idea comunque generale delle difficoltà che si incontrano nell'acquisizione di questo tipo di dati, e possono aiutare a meglio inquadrare i risultati delle ultime ricerche in corso.

L'*Alzheimer's Disease Neuroimaging Initiative* (ADNI) [22] è una collaborazione americana di: National Institute on Aging (NIA), National Institute of Biomedical Imaging and Bioengineering (NIBIB), Food and Drug Administration (FDA), con l'industria farmaceutica e alcune fondazioni private attraverso la *Foundation for the NIH* (National Institutes of Health).

Si tratta di uno studio osservativo longitudinale⁶, multi-sito, condotto sia su soggetti anziani sani (dal punto di vista della demenza), sia con sintomatologia MCI, sia affetti da AD. Attivo dal 2004, si propone di stabilire come si possa accertare, e valutare nel tempo, l'insorgenza e la progressione della malattia di Alzheimer attraverso l'uso, singolo o combinato, degli strumenti diagnostici oggi ritenuti più idonei dalla ricerca: le neuroimmagini e i marcatori biochimici. Unitamente si procede anche con analisi cliniche sui fluidi biologici e attente valutazioni neuropsichiatriche.

Oltre al suddetto obiettivo finale, ADNI ha lo scopo di fornire liberamente alla comunità scientifica dati controllati e conseguenti nel tempo, sviluppare standard tecnici per gli studi longitudinali con neuroimmagini, determinare la metodologia ottima per l'acquisizione e l'analisi preliminare delle MRI, validare immagini e biomarcatori anche con la fornitura di indagini cliniche e psicometriche a questi correlate, contribuire infine allo sviluppo dei metodi di indagine clinica nei casi di MCI e AD.

Lo studio è partito dal reclutamento, su base volontaria, di complessivamente 800 adulti, tra i 55 e i 90 anni, presso 55 strutture accuratamente selezionate di Stati Uniti e Canada: 200 soggetti cognitivamente sani (soggetti di controllo), 400 con decadimento cognitivo lieve (MCI), e 200 affetti da malattia di Alzheimer, sono stati ivi sottoposti, ad intervalli di tempo stabiliti⁷, a risonanza magnetica del cervello (MRI), tomografia ad emissione di positroni con opportuno marcatore (FDG-PET), analisi con bio-marcatori di liquido cerebrospinale (CSF), esami clinici e prelievo di campioni. Attualmente il database include oltre 32'000 scansioni MRI e PET [21].

La massima cura è stata posta nel mantenere la consistenza dei dati nel tempo e attraverso le diverse piattaforme di acquisizione, affinché tutte le immagini acquisite con tecnologia analoga abbiano qualità simile (risoluzione spaziale, rapporto contrasto/rumore, minimizzazione degli artefatti, ecc.).

⁶ Con *longitudinale* si indica uno studio condotto mediante ripetute osservazioni dello stesso (o di pochi) soggetto(i) in un lungo periodo di tempo; si contrappone allo studio *trasversale*, composto invece da diverse osservazioni in diversi (o numerosi) soggetti nello stesso momento.

⁷ I soggetti di controllo a 0, 6, 12, 24, 36 mesi; gli MCI anche a 18 mesi; gli AD solo a 0, 6, 12, 24 mesi.

Un'attenzione particolare è stata poi riservata all'analisi morfometrica del cervello, e di conseguenza proprio alle sequenze 3D di immagini da risonanza magnetica, pesate in T1, ad alta risoluzione e qualità, acquisite con opzioni accuratamente ottimizzate per ogni tipo di macchinario usato⁸; viene inoltre sottoposto a scansione anche un apposito fantoccio (*phantom*) progettato per il monitoraggio continuo e la calibrazione delle piattaforme MR.

Tutti i soggetti partecipanti sono stati sottoposti a scansioni MR a 1,5 Tesla⁽⁹⁾ con risoluzione di circa 1 mm³ per voxel, possibilmente isotropica, e al massimo di 1,5 mm in ogni direzione. Ogni acquisizione è stata ripetuta due volte, selezionando quindi quella di qualità migliore; quando sono di qualità equivalente, le MRI possono essere combinate per incrementare il rapporto S/N.

Oltre alle piccole correzioni nelle mani dell'operatore allo scanner, intrinseche al processo di acquisizione, lo studio ADNI prevede anche una successiva elaborazione per una ulteriore correzione degli artefatti generati dalle stesse macchine MR, ove possibile usando apposito software fornito dalle case costruttrici: si tratta soprattutto di non linearità nel campo gradiente, disuniformità causate dalle bobine di ricezione, imperfetto allineamento delle sezioni (*slices*) 2D che formano la sequenza 3D, e così via. Sono tutte correzioni specifiche per ogni sistema, valutate ed applicate da esperti del settore, necessarie a garantire a tutte le immagini la necessaria uniformità e indipendenza dai dettagli tecnici.

Sono comunque a disposizione della comunità scientifica sia le immagini corrette, sia quelle prive di correzione, sia anche immagini con solo alcune delle correzioni possibili.

⁸ Per questa ragione le piattaforme utilizzate sono state scelte esclusivamente tra i modelli non obsoleti dei tre principali costruttori mondiali: GE Healthcare, Philips Medical Systems, Siemens Medical Solutions.

⁹ Sono previste anche altre modalità di imaging: a 3 Tesla (su metà dei soggetti), con miglior rapporto segnale/rumore e risoluzione, ma con problemi di *shift* chimico e maggior velocità di variazione della suscettibilità; FDG-PET (sull'altra metà); MR a pesatura T2 per individuare altre patologie presenti (su tutti).

Tutto questo intenso lavoro di preparazione trova una spiegazione nella necessità di elaborare le MRI della base di dati in maniera del tutto automatica: là dove un radiologo esperto è capace di leggere le immagini riconoscendone senza difficoltà i piccoli artefatti, i programmi di analisi potrebbero incorrere in equivoci sostanziali. Anche un difetto apparentemente trascurabile può sviare un algoritmo, per quanto accurato questo sia: vale anzi la regola empirica secondo la quale più è automatica l'analisi delle immagini, più questa è sensibile verso le imperfezioni.

È un tipo di problema che si incontra spesso nel campo della intelligenza artificiale.

2.4.2 IBSR

L'*Internet Brain Segmentation Repository* (IBSR) [23] è una raccolta di immagini MR, creata nel 1996 a cura del Center for Morphometric Analysis (CMA) del Massachusetts General Hospital. Sono in maggioranza acquisizioni pesate a T1, disponibili in *slices* con matrice di 256×256 pixel, generalmente con scala di grigi a 16 bit, di soggetti sani o di pazienti con tumori al cervello. Le scansioni sono effettuate presso lo stesso ospedale su macchine dello stesso modello e costruttore.

La sua caratteristica più interessante è la disponibilità di molte di esse anche in versione segmentata, con indicate esplicitamente materia grigia (GM), bianca (WM), o altro; e/o in versione *contornata* con l'intervento manuale di un esperto, ricampionate a 512×512 pixel, con indicazioni molto più dettagliate di varie strutture anatomiche.

L'uso principale di questi dati è la verifica degli algoritmi automatici di segmentazione MRI sia supervisionati che non.

Sono disponibili anche immagini ottenute con tecnologie diverse, e sono in fase di sviluppo altri *database* dedicati a studi su diverse patologie.

2.4.3 OASIS

L'*Open Access Series of Imaging Studies* (OASIS) [24] è frutto di una collaborazione tra la Howard Hughes Medical Institute (HHMI) della Harvard University, il Neuroinformatics Research Group (NRG) presso la Washing-

ton University School of Medicine, e il Biomedical Informatics Research Network (BIRN).

È costituito da immagini cerebrali di oltre 400 soggetti ampiamente distribuiti in età (dai 18 ai 96 anni), dei quali fornisce più MRI a 1,5 Tesla pesate in T1 e acquisite con un unico modello, fissato per tutti, di scanner MR; inoltre sono disponibili le medie delle stesse elaborate (*registrate*, v. cap. seg.) per la correzione del movimento, le versioni ripulite dalle strutture non encefaliche, le versioni corrette e registrate mediante atlante [25] in uno spazio anatomico standard [26], infine le versioni già segmentate in materia grigia, bianca e liquido cerebrospinale [27].

La base di dati è pensata soprattutto per testare algoritmi diversi in ogni punto fondamentale della catena di elaborazione automatica, e dispone quindi, oltre ai dati già citati, anche di ulteriori informazioni anatomiche utili, e di un accurato sistema di ricerca su basi cliniche e di popolazione.

2.4.4 Altre risorse

BIRN (Biomedical Informatics Research Network) [28]: è un progetto nato nel 2001 che si propone di favorire collaborazioni interdisciplinari a larga scala nel campo delle scienze biomediche, mediante l'uso delle più recenti infrastrutture informatiche di rete. Si caratterizza per il lavoro congiunto di informatici e ricercatori biomedici nel progetto ed implementazione di un'architettura distribuita di risorse condivise per la diagnosi e cura di varie malattie.

NDG (Neuroscience Database Gateway) [29]: sviluppato dalla Society's Brain Information Group (BIG), e ora ospitato dalla Yale University sotto la supervisione della Society's Neuroinformatics Committee, è un progetto pilota che organizza e rende disponibili riferimenti (*links*) a basi di dati, tutti di ambito neuroscientifico, di cinque tipi diversi: dati sperimentali, conoscenze di base, *software* dedicato alle neuroscienze, risorse bioinformatiche, soggetti fornitori di materiale per la ricerca. È attivo dal 2004.

EADC (European Alzheimer's Disease Consortium) [30]: da segnalare infine questa rete, che gode del contributo economico della Commissione Europea, composta da oltre 50 centri europei di eccellenza nella lotta alla malattia di Alzheimer e alle demenze ad essa correlate. EADC ha di recente avviato un progetto finalizzato allo sviluppo di un protocollo ottimizzato per la stima del volume dell'ippocampo mediante tracciamento manuale su MRI [31].

2.5 Coordinate stereotassiche e atlanti

Molte procedure di registrazione e/o segmentazione delle immagini MR si basano sul confronto e sull'adattamento riferiti ad una immagine-modello: gli atlanti anatomici del cervello forniscono tali riferimenti in uno spazio associato tridimensionale di coordinate standard (spazio *stereotassico*).

Anche in questo campo la presenza di standard diversi, conseguenza quasi inevitabile del rapido procedere del neuroimaging, è fonte di qualche difficoltà nell'analisi e nel confronto dei lavori di ricerca.

2.5.1 Talairach

L'atlante di Talairach e Tournoux (1988) [26] nacque dalla necessità, imposta soprattutto dalla neurochirurgia, di costruire una dettagliata mappa tridimensionale del cervello, che fosse anche munita di precisi riferimenti anatomici.

È stato costruito esaminando sottili sezioni di un cervello umano reale, riportandone fedelmente l'aspetto su carta, tracciandone con cura le regioni anatomiche, e munendo il tutto di un sistema di coordinate ortogonali: in questo modo si è creata una corrispondenza che consente di indicare ogni dettaglio morfologico semplicemente fornendone le coordinate, e in maniera molto più precisa che con un tradizionale sistema di tipo descrittivo.

Mantenendo questa corrispondenza tra morfologia e coordinate, che definisce il cosiddetto *spazio stereotassico*, segue che un cervello di dimensioni e forma diverse potrà essere descritto con le stesse coordinate solo a prezzo di una deformazione continua che lo porti a combaciare con esattezza sul modello, in modo da sovrapporne le strutture morfologiche.

Questa procedura, indispensabile in ogni analisi automatica, ha una precisione necessariamente limitata, e garantisce una migliore qualità dei risultati quando lo spazio è definito su un modello che ben rappresenti il cervello normale medio.

Il campione usato nell'atlante di Talairach e Tournoux, purtroppo, non risponde a questo requisito: ha una forma piuttosto discosta dalla media umana, e inoltre, conservato per qualche tempo in formalina prima del sezionamento, è risultato aver subito anche ulteriori deformazioni. Se si aggiunge che le regioni anatomiche sono tracciate solo su uno degli emisferi, e che la sua risoluzione non è migliore di 4 mm [49], si comprende come l'uso di questo atlante nel neuroimaging possa oggi essere giustificato solo dalla vasta notorietà che si è conquistato negli anni, e come un suo definitivo abbandono, in questo ambito, sarebbe ormai opportuno.

2.5.2 LONI-ICBM

Il *Laboratory of neuro imaging* (LONI) dell'UCLA [33], ha sviluppato una quantità di risorse messe in rete a disposizione dei ricercatori di tutto il mondo. Tra queste vari atlanti neurologici, tra cui uno dedicato alla ricerca specifica sulla malattia di Alzheimer (*Alzheimer's Disease Template*).

È ivi presente anche un'intera struttura dedicata alla ricerca internazionale sul neuroimaging (*International Consortium for Brain Mapping, ICBM*) [34], che ha sviluppato un vasto progetto per la creazione di un nuovo atlante probabilistico con associato sistema di riferimento [35].

Capitolo 3

Neuroimaging diagnostico per l'AD: algoritmi base e metodi “classici”

I metodi basati sulle neuroimmagini per la diagnosi della malattia di Alzheimer sono costituiti da un complesso insieme di processi di calcolo che possono essere grossolanamente suddivisi in due fasi: la preelaborazione, o preprocessamento (*pre-processing*), e le seguenti analisi e classificazione diagnostica.

La prima comprende tutte quelle fasi preliminari che vanno dalla correzione delle immagini da difetti tecnici e rumore, alla loro normalizzazione geometrica in uno spazio comune, alla normalizzazione dei livelli di intensità, all'applicazione di filtri, il tutto per rendere il materiale omogeneo ed idoneo al trattamento successivo; la seconda consiste essenzialmente nell'estrazione dall'immagine delle caratteristiche capaci di distinguere lo stato patologico da quello normale, e nella loro elaborazione da parte di un programma in grado di decidere in base ad esse la valutazione diagnostica finale.

Entrambe le fasi, la prima più “tecnica”, la seconda più legata ad approcci innovativi, sono composte da diversi algoritmi di vario grado di complessità e sviluppo, che muniscono lo sperimentatore di un vasto repertorio su cui operare le proprie scelte: tali scelte vanno ben valutate e soprattutto ben coordinate tra loro, perché il loro grado di interdipendenza, anche

quando si usano strategie puramente “sequenziali”, può risultare spesso più complesso di quanto ipotizzabile ad un primo esame.

Nel tempo, alcuni di questi algoritmi si sono notevolmente consolidati, e oggi possono essere utilizzati con sicurezza anche con una conoscenza solo superficiale del loro funzionamento; altri, invece, anche se a volte proposti dagli sviluppatori come soluzioni di riferimento per la loro robustezza, non possono prescindere dall’approfondimento delle loro caratteristiche intrinseche per farne un uso realmentre efficace.

D’altronde è ormai divenuto impossibile, per un singolo gruppo di ricerca, lo sviluppo autonomo di tutta la catena di calcolo che porta dalla singola neuroimmagine, prelevata da un archivio certificato, alla diagnosi finale. Lo sforzo è inevitabilmente concentrato su ambiti più ristretti, e l’uso combinato con codice prodotto, a scopo di ricerca o commerciale, da altri soggetti è pressoché obbligato.

L’organizzazione generale di un programma di analisi di MRI per la diagnosi dell’AD è quindi normalmente realizzata come sequenza di programmi specializzati separati, che operano in successione ognuno sui dati elaborati dal precedente. Il problema dell’interfacciamento tra programmi, collegato al moltiplicarsi degli standard in gioco, e del controllo dell’intera sequenza, trova una soluzione ideale nella definizione di una cosiddetta *pipeline* che gestisca in automatico l’attivazione dei programmi interessati ed il flusso di dati tra di essi, nonché gli eventuali controlli da parte dell’operatore. L’adesione ad un protocollo comune, come la pipeline LONI, fornisce una solida e comoda soluzione tecnica, facilitando così anche l’interazione tra gruppi diversi di ricercatori.

Non mancano anche approcci con un certo grado di parallelismo, nei quali la ricerca di prestazioni migliori passa attraverso la sinergia di algoritmi che interagiscono nella ottimizzazione dei parametri di lavoro: la prestazione globale è così generalmente aumentata rispetto alla corrispondente tecnica in sequenza, ma si perde fatalmente il controllo delle singole fasi del processo, rinunciando a testare, ed eventualmente sostituire, i passi che si rivelano, volta per volta, più critici.

Uno degli obiettivi su cui si concentra oggi la ricerca è quindi proprio la creazione, sperimentazione e verifica, di metodologie *complessive* di analisi

delle neuroimmagini: gli strumenti a disposizione sono, oltre alle tecniche di elaborazione tipiche delle immagini digitali e dei segnali, anche quelle sviluppate negli ambiti, fittamente intrecciati, dell'*intelligenza artificiale* e del *data mining*.

L'altro obiettivo fondamentale, e zona di frontiera della ricerca, è, naturalmente, la creazione, specializzazione e sviluppo di algoritmi *dedicati*, in particolare per ciò che riguarda la fase finale e di più alto livello della catena: estrazione dalle immagini di caratteristiche indicatrici della patologia, e *decisione* sullo stato di malattia.

Alla base di tutti i sistemi di diagnosi automatica per mezzo di neuroimmagini vi sono alcuni procedimenti fondamentali e comuni a tutti, che verranno ora esaminati. Si approfondiranno inoltre gli algoritmi che interessano la linea di ricerca seguita in questo lavoro.

3.1 Fase di *pre-processing*

Determinare con una neuroimmagine un possibile stato di malattia significa, in pratica, confrontare la scansione del soggetto con l'immagine tipo di una persona sana, e individuare le differenze che caratterizzano la malattia in questione, senza confonderle con la normale variabilità anatomica tra individui.

Chiaramente, per poter operare qualsiasi confronto è necessario dapprima “inquadrare” tutte le immagini nello stesso spazio geometrico, ad esempio fissando per tutte una comune origine e comuni assi di orientazione. A questo punto, però, basteranno le comuni differenze di forma e dimensioni del cervello tra soggetti diversi a rendere la corrispondenza tra le singole strutture anatomiche scarsa, se non addirittura nulla: è quindi sempre necessario introdurre qualche tipo di *deformazione* per ottenere almeno un certo grado di sovrapposizione, senza la quale l'analisi successiva sarebbe impensabile. Il carattere e l'entità di tale deformazione definisce e condiziona tutta la successiva elaborazione.

La prima di queste scelte, ossia il *tipo* di trasformazione che è opportuno adottare, è parte integrante e fondamentale del progetto dell'algoritmo,

come si vedrà nel seguito; la seconda, cioè la determinazione della particolare trasformazione tra le infinite del suo tipo, è specificata mediante un algoritmo. Il processo che determina la trasformazione, dallo spazio originale ad uno spazio di riferimento comune, di dati riferiti a diversi sistemi di coordinate, è detto *registrazione*¹ (*registration*).

Va chiarito che il termine si riferisce, a rigore, alla sola *determinazione* dei parametri di trasformazione, e non all’effettuazione della stessa, che può avvenire anche in un secondo tempo, e con diverse modalità: ad esempio, per individuare su una immagine “bersaglio” una specifica porzione anatomica delineata in un atlante (che è anch’esso un’immagine tridimensionale), si può registrare il bersaglio sull’atlante, e poi trasformare l’atlante sul bersaglio con i parametri appena determinati, trasferendo così su di esso la delineatura interessata. Tuttavia, quando le due fasi si susseguono immediatamente, si trova spesso indicato con “registrazione” l’intero processo così composto, affidando al contesto la distinzione tra i due significati.

Gli algoritmi di registrazione basano il loro funzionamento sul confronto, tra le due immagini da far corrispondere, di precise caratteristiche da usare per guidare l’allineamento; queste possono essere: punti fissati manualmente da un operatore su entrambe; dettagli facilmente individuabili automaticamente, come contorni o linee particolari; regolarità nel valore di intensità dei punti. In una procedura completamente automatica la prima soluzione è da escludersi; la seconda richiede la presenza di dettagli molto chiari che possano essere individuati con sicurezza da un algoritmo, cosa oggi possibile solo con i principali *solchi* presenti al livello più esterno della corteccia cerebrale; la terza richiede un approccio globale, o comunque non puntuale, e una valutazione della “distanza” tra immagini, quest’ultima da minimizzare.

Il tipo, o modello, di trasformazione da adottare, dipende dal grado di corrispondenza, o sovrapponibilità, che si vuole raggiungere in funzione delle fasi successive: una semplice trasformazione rigida può essere sufficiente solo per compensare piccoli spostamento di un singolo soggetto durante la fase di acquisizione (che coinvolge sempre più scansioni consecutive), mentre

¹ Si trova anche il termine *registratura*.

non è capace di allineare a sufficienza neanche il contorno esterno di crani diversi; all’opposto una trasformazione diffeomorfica, capace di deformazione continua e definita puntualmente, ha bisogno di tener conto con precisione delle nuove densità introdotte, per non vanificare l’idea stessa di misura del grado di atrofia.

Invertendo il punto di vista, la questione può anche essere posta come segue: rinunciando ad ogni tipo di vincolo, è sempre possibile trasformare una immagine affinché corrisponda esattamente ad un’altra; il problema è perciò non tanto la definizione del tipo di trasformazione, quanto piuttosto dei vincoli a cui sottoporla per ottenere un risultato efficace. La validità di una trasformazione può così essere ricondotta alla validità dei vincoli imposti.

Per fornire risultati corretti, ed evitare clamorosi disallineamenti, la registrazione viene in generale eseguita su immagini rese più omogenee da opportuni filtri (di *smooth*), tutti riconducibili al tipo passa-basso, che producono in definitiva quello che definiremmo come un effetto di “sfocatura”. Questo può sembrare sorprendente, soprattutto perché un operatore umano compirebbe questa operazione proprio con la ricerca di dettagli e riferimenti facilmente riconoscibili, cercando poi di sovrapporli al meglio nelle due immagini.

Una tipica registrazione automatica basata sulle intensità, invece, *precede* l’eventuale fase di riconoscimento delle strutture anatomiche interessate, e anzi è proprio un corretto allineamento delle immagini, tra di loro o in uno spazio anatomico prefissato, una condizione essenziale per il prosieguo del lavoro.

Il processo di registrazione avviene, come detto, tipicamente mediante un confronto *globale* delle immagini oggetto e riferimento, non basato sui dettagli: la corrispondenza non può essere esatta, e quindi la sfocatura dei contorni delle regioni anatomiche, riducendo le discrepanze non essenziali, favorisce la ricerca di una corrispondenza generale. Inoltre i parametri che determinano la successiva trasformazione dell’oggetto sul riferimento vengono scelti attraverso la minimizzazione di una qualche funzione che ne misuri la diversità, e questo tipo di processo può essere ingannato da eventuali

minimi *locali*, tra cui quelli generati da rumore di tipo impulsivo.

Va aggiunto infatti che gli scanner MR producono un segnale ottimizzato per l’analisi del radiologo, per cui nel processo di acquisizione il rumore viene filtrato per generare una immagine della massima qualità per l’esame *visivo*: un filtraggio eccessivo renderebbe solo inutilmente più confuso il risultato. Questo significa che nei tipici dati MRI è presente ancora del rumore, indifferente per l’occhio umano, ma in grado di condizionare il calcolo automatico: si tratta essenzialmente appunto di rumore di tipo impulsivo, e quindi ad alta frequenza, mentre gli effetti del rumore a bassa frequenza sono già filtrati al meglio nello scanner stesso, o, al più, in fase di post-acquisizione, e in maniera specifica per ogni singolo modello di macchinario.

Un ulteriore importante passaggio della preelaborazione è la cosiddetta *normalizzazione² dei toni di grigio*, che consiste nell’uniformare la distribuzione delle intensità nei punti corrispondenti delle diverse immagini: come la registrazione, anche questa è necessaria per operare correttamente confronti in tutti gli algoritmi basati sulle intensità per le misure morfometriche. Si ottiene con tecniche analoghe a quelle di registrazione, dalle quali non è concettualmente diversa, e anzi viene a volte effettuata contestualmente con questa, mediante un unico algoritmo.

In definitiva, la fase di pre-processing consiste essenzialmente in una registrazione delle immagini in uno spazio comune, e un riallineamento dei toni di grigio, accompagnati da uno o più processi di filtraggio che hanno lo scopo di esaltare le corrispondenze globali, favorendo la registrazione stessa, e attenuare i residui di rumore.

3.1.1 Trasformazioni affini

Sebbene le registrazioni più accurate devono includere qualche tipo di deformazione dell’immagine, in alcune fasi della catena di algoritmi di analisi

² Questo termine viene usato con diversi significati nel contesto dell’imaging, per esempio con “normalizzazione spaziale” si può intendere un particolare metodo di registrazione (che verrà esaminato in seguito), o la registrazione stessa delle immagini.

delle MRI può essere preferibile affidarsi a *trasformazioni rigide*, p. es. per non compromettere successive valutazioni volumetriche.

Le trasformazioni rigide coinvolgono solo traslazioni e rotazioni, e nel caso tridimensionale più generale coinvolgono fino a sei parametri: le tre traslazioni lungo gli assi (x_0, y_0, z_0) , e i tre angoli di rotazione (α, β, γ) intorno agli assi del riferimento. Le relative matrici si scrivono:

$$\begin{pmatrix} 1 & 0 & 0 & x_0 \\ 0 & 1 & 0 & y_0 \\ 0 & 0 & 1 & z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\begin{pmatrix} \cos \beta & 0 & \sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} \cos \gamma & \sin \gamma & 0 & 0 \\ -\sin \gamma & \cos \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Si può aggiungere anche una semplice deformazione generale, utile per compensare le diversità della forma del cranio, ma pure le deformazioni apparenti dovute a diverse posizioni della testa nello scanner: basta introdurre opportuni fattori di scala (k_x, k_y, k_z) lungo i tre assi coordinati. L'insieme di tutte queste operazioni porta alla costruzione di una matrice con nove parametri liberi totali, e definisce la cosiddetta *trasformazione affine a 9-parametri*.

La più generale delle trasformazioni lineari nello spazio tridimensionale si scrive in forma algebrica:

$$\begin{aligned} y_1 &= m_{11}x_1 + m_{12}x_2 + m_{13}x_3 + m_{14} \\ y_2 &= m_{21}x_1 + m_{22}x_2 + m_{23}x_3 + m_{24} \\ y_3 &= m_{31}x_1 + m_{32}x_2 + m_{33}x_3 + m_{34} \end{aligned}$$

equivalente alla forma matriciale:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix}$$

che contiene un totale di dodici parametri.

I tre parametri restanti aggiungono ulteriore libertà nelle deformazioni, e possono essere introdotti con una opportuna matrice di “deformazione di taglio” (*shear*). La matrice coi fattori di scala, e la deformazione appena definita, si scrivono rispettivamente:

$$\begin{pmatrix} k_x & 0 & 0 & 0 \\ 0 & k_y & 0 & 0 \\ 0 & 0 & k_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & s_1 & s_2 & 0 \\ 0 & 1 & s_3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

La trasformazione complessiva risultante è a volte detta semplicemente “affine”, ma in letteratura si preferisce precisare chiamandola *affine a 12-parametri*, considerato l’uso diffuso anche di quella a nove.

La corrispondenza tra i parametri di trasformazione rigida ($x_0, y_0, z_0, \alpha, \beta, \gamma$) e quelli di deformazione ($k_x, k_y, k_z, s_1, s_2, s_3$), con quelli affini ($m_{11}, m_{12}, \dots, m_{34}$), si definisce stabilendo la sequenza delle applicazioni (il prodotto non è commutativo) ed effettuando il prodotto matriciale.

3.1.2 Trasformazioni non lineari

Registrazioni molto più raffinate si possono ottenere solo con trasformazioni elastiche a numero di parametri molto più elevato, e vengono usate in approcci morfometrici non adatti all’analisi delle strutture interessate precocemente dalla AD, per cui non verranno trattate in questo lavoro. Per darne comunque un’idea, mi limiterò a qualche cenno su di esse.

L’approccio più immediato consiste nel suddividere l’immagine in volumi semplici e sufficientemente piccoli da raggiungere il dettaglio prefissato, quindi procedere con trasformazioni affini indipendenti su ognuno di essi, infine raccordare le trasformazioni tra loro. È evidente che, con questo metodo, non è in alcun modo garantita la corrispondenza uno a uno tra voxel, d’altronde indispensabile per poter effettuare qualunque trasformazione inversa: il problema può essere ragionevolmente aggirato solo con trasformazioni di piccola entità, e non senza problemi.

L’approccio matematicamente più corretto resta quindi la definizione di un autentico diffeomorfismo, computisticamente pesante e dal numero di

parametri virtualmente anche molto elevato. La costruzione di una tale trasformazione deve ricorrere a modelli complessi sviluppati in altre discipline: dalle prime implementazioni che modellavano un fluido viscoso nel suo “flusso” tra una immagine e l’altra, si arriva oggi a costruire campi di velocità che si evolvono nel tempo. Ognuno di questi modelli richiede un uso intenso di equazioni differenziali risolte come equazioni alle differenze finite.

3.1.3 Ricampionamento

Stabilita una mappa tra coordinate originali e trasformate, è necessario ricampionare l’immagine per applicare la trasformazione: i voxel trasformati giaceranno in posizioni diverse da quelle precedenti, per cui è necessario stabilirne le corrette intensità mediante un qualche tipo di interpolazione dalla immagine originale a quella trasformata.

L’approccio più semplice, detto di *ordine zero*, è quello che assegna al nuovo punto il valore del vicino non trasformato più prossimo: il valore del voxel originale non viene alterato, ma la qualità generale è molto degradata.

Il ricampionamento del *primo ordine* usa invece la *interpolazione trilineare*: il risultato è meno “squadrettato”, ma si perdono alcune informazioni di alta frequenza. La sua versione 2D, detta bi-lineare, calcola il valore da assegnare al punto r genericamente contenuto all’interno del quadrato con ai vertici i punti originali a, b, c, d interpolando prima il punto p tra a e b e collineare con essi, e analogamente il punto q tra c e d , mediante le:

$$v_p = \frac{(x_b - x_p)v_a + (x_p - x_a)v_b}{x_b - x_a}, \quad v_q = \frac{(x_d - x_q)v_c + (x_q - x_c)v_d}{x_d - x_c},$$

per poi calcolare finalmente il valore di r interpolando tra p e q :

$$v_r = \frac{(y_q - y_r)v_p + (y_r - y_p)v_q}{y_q - y_p}.$$

L’estensione a 3D è immediata, e interpola tra gli 8 punti più vicini.

Una migliore interpolazione si può ottenere coinvolgendo un più ampio numero di punti vicini, ad esempio con una *interpolazione polinomiale*. Se ad esempio il punto q giace sulla retta x tra b e c , essendo a, b, c, d i quattro pixel collineari più prossimi (due per lato), si costruisce per questi punti il

polinomio di coefficienti:

$$\mathbf{q} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & (x_b - x_a) & (x_b - x_a)^2 & (x_b - x_a)^3 \\ 1 & (x_c - x_a) & (x_c - x_a)^2 & (x_c - x_a)^3 \\ 1 & (x_d - x_a) & (x_d - x_a)^2 & (x_d - x_a)^3 \end{pmatrix}^{-1} \begin{pmatrix} v_a \\ v_b \\ v_c \\ v_d \end{pmatrix},$$

quindi il valore v_q può essere calcolato con

$$v_q = \mathbf{q}^T \begin{pmatrix} 1 \\ (x_q - x_a) \\ (x_q - x_a)^2 \\ (x_q - x_a)^3 \end{pmatrix}.$$

Le estensioni sul piano e nello spazio sono di nuovo immediate.

In pratica si sono sviluppati diversi algoritmi di interpolazione polinomiale (v. p. es. [36]), ma una certa presenza di artefatti da interpolazione è inevitabile.

Un metodo ottimo per evitare tali artefatti sarebbe trasformare le immagini nello spazio di Fourier [37], ma è preferibile evitare l'introduzione, per ragioni computazionali, di grandezze complesse.

Un metodo che fornisce risultati molto vicini ad una interpolazione di Fourier, e che si sviluppa nel campo reale, è l'interpolazione *sinc*, che viene implementata con la:

$$v = \sum_{i=1}^I v_i \cdot \frac{\frac{\text{sen}(\pi d_i)}{\pi d_i} \cdot \frac{1}{2} \left(1 + \cos \left(\frac{2\pi d_i}{I} \right) \right)}{\sum_{j=1}^I \frac{\text{sen}(\pi d_j)}{\pi d_j} \cdot \frac{1}{2} \left(1 + \cos \left(\frac{2\pi d_j}{I} \right) \right)},$$

dove v è il valore incognito del generico punto dello spazio, d_i la sua distanza dall' i -esimo voxel, v_i il valore di quest'ultimo. La funzione $\text{sinc } x = \text{sen } x / x$ si estende all'infinito, inoltre ogni valore va teoricamente calcolato usando tutti i voxel dell'immagine: nella pratica computazionale, invece, la $\text{sinc } x$ viene troncata con una funzione finestra, mentre si scelgono per il calcolo solo gli I voxel più vicini. Il livello qualitativo è molto elevato, inoltre, come per le interpolazioni polinomiali, il calcolo può essere eseguito semplicemente in sequenza nelle tre dimensioni: ciò rende questo metodo una delle interpolazioni di *ordine superiore* più usate.

Allo stato attuale, grazie agli algoritmi e ai calcolatori oggi a disposizione, il problema del ricampionamento non è da considerarsi critico nell’analisi delle neuroimmagini.

3.1.4 Filtro di media

È il filtro di smoothing concettualmente più semplice: esegue la media dei valori dei voxel in un definito intorno, assegnando il risultato al voxel centrale. La facile implementazione e il bassissimo costo computazionale ne hanno fatto in passato una scelta privilegiata, e tuttora viene usato affiancandolo a filtri più sofisticati.

La mancanza di pesatura ne rende l’intervento piuttosto “forte” anche su intervalli minimi, e il suo effetto sulle frequenze spaziali non è di facile controllo: le varianti pesate hanno caratteristiche nettamente migliori, e vengono costruite praticamente partendo da funzioni di proprietà note e desiderabili.

3.1.5 Filtro gaussiano

Usa una convoluzione discreta con un *kernel* gaussiano, in cui l’ampiezza a j unità di distanza dal centro è definita da:

$$g_j = \frac{e^{-j^2/2s^2}}{s\sqrt{2\pi}} .$$

In una dimensione la convoluzione w tra il kernel g e una funzione con valori v_i si scrive:

$$w_i = \sum_{j=-d}^d v_{i-j} g_j ,$$

mentre in due (tre) dimensioni basta costruire il relativo kernel sostituendo j con la distanza $D_{i,j}$ ($D_{i,j,k}$) del pixel (voxel), con la convoluzione che può essere effettuata sui due (tre) indici semplicemente in sequenza.

Il filtro gaussiano è esente da artefatti tipici di altri filtri [38], la qual cosa lo rende preferibile nell’imaging medico.

Nella pratica del calcolo, anche in questo caso, la funzione gaussiana non viene estesa fino all’infinito, bensì limitata al massimo a valori di circa sei FWHM (larghezza a metà altezza) della curva³.

3.1.6 Registrazione

Si è visto che l’operazione di registrazione tra due immagini consiste nel determinare i parametri, natura e numero dei quali sono definiti dal tipo di trasformazione scelta, che meglio trasformano una immagine sull’altra.

Deve quindi venir definita una grandezza che misuri la “diversità”, o distanza, tra le due immagini in funzione dei parametri di trasformazione, per poter poi cercare i valori che la minimizzano.

Per esempio, sia $d_i(\mathbf{p})$ una misura della differenza tra le immagini al voxel i quando il *vettore dei parametri* vale \mathbf{p} : si può usare come distanza globale la somma dei quadrati $\sum_i (d_i(\mathbf{p}))^2$, e cercare il vettore di incrementi \mathbf{l} da aggiungere ai parametri per minimizzare tale distanza. Una stima approssimata, basata su metodo del tipo di Gauss-Newton [39], si ottiene con la:

$$\begin{pmatrix} -\frac{\partial d_1(\mathbf{p})}{\partial p_1} & -\frac{\partial d_1(\mathbf{p})}{\partial p_2} & \cdots \\ -\frac{\partial d_2(\mathbf{p})}{\partial p_1} & -\frac{\partial d_2(\mathbf{p})}{\partial p_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \ell_1 \\ \ell_2 \\ \vdots \end{pmatrix} \simeq \begin{pmatrix} d_1(\mathbf{p}) \\ d_2(\mathbf{p}) \\ \vdots \end{pmatrix},$$

che si può scrivere semplicemente come $\mathbf{A}\mathbf{l} \simeq \mathbf{d}$. Stimato \mathbf{l} , si aggiorna il vettore dei parametri \mathbf{p} a $\mathbf{p} + \mathbf{l}$, e si prosegue iterativamente per stime successive

$$\mathbf{p}^{(n+1)} = \mathbf{p}^{(n)} + (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{d}$$

fino a soddisfare qualche criterio di convergenza, o fino ad un numero massimo di iterazioni.

La scelta dei valori iniziali viene a volte effettuata manualmente per partire da una condizione già prossima alla soluzione, ma la necessità di un calcolo completamente automatico richiede un approccio diverso: si possono allora stabilire delle stime basate su precise regole a priori, o partire da una selezione casuale; questo però può portare l’algoritmo lontano dalla condizione cercata, e farlo cadere in un minimo locale. Si preferisce quindi lanciare

³ FWHM = $s \cdot \sqrt{8 \ln(2)}$

il programma più volte con diversi insiemi di parametri iniziali, e stabilire dalla convergenza dei calcoli il risultato corretto: si possono costruire così algoritmi particolarmente robusti, e capaci di individuare automaticamente gli eventuali casi problematici.

Una soluzione diversa al problema dei valori iniziali può essere ottenuta grazie all’impiego di filtri di smoothing di ampiezza decrescente: si parte da immagini fortemente sfumate e se ne fa una prima registrazione, facendo così una prima stima dei parametri; senza effettuare alcuna trasformazione, si utilizza poi questa stima come valore iniziale per una seconda registrazione, questa volta con una versione meno sfumata delle stesse immagini iniziali; la nuova stima ottenuta darà i valori iniziali per una terza registrazione, a sfumatura ancora inferiore, tra le immagini non trasformate. Poche di queste iterazioni sono sufficienti ad ottenere un risultato stabile, soprattutto combinando la sfumatura con filtri sensibili alle transizioni di intensità tra regioni diverse [40].

La definizione dei parametri di registrazione può estendersi al di là della pura trasformazione spaziale, e includere anche altri fattori. Ad esempio, poiché basta una differenza nella scala dei grigi tra due immagini, altrimenti identiche, per aumentare una distanza d tra voxel corrispondenti basata proprio sulla differenza della loro intensità v , può essere opportuno inserire nella definizione di distanza da minimizzare una dipendenza anche da parametri di questo tipo: questo può garantire una corrispondenza spaziale migliore, ma soprattutto restituisce trasformazioni che includono anche un certo grado di normalizzazione dei livelli di grigio, facilitando il compito, ad esempio, della eventuale successiva *segmentazione*.

Una distanza così definita avrà la forma generale:

$$d_i(\mathbf{p}) = f(\mathbf{x}_i, \mathbf{p}_s, \mathbf{p}_v) ,$$

dove la dipendenza funzionale f dai vettori delle coordinate dell’ i -esimo voxel \mathbf{x}_i , dei parametri di trasformazione spaziale \mathbf{p}_s , e dei parametri di trasformazione dell’intensità \mathbf{p}_v , può assumere forme più o meno complesse a seconda del modello usato.

È questo un esempio di approccio combinato alla registrazione spaziale e alla normalizzazione dei livelli di grigio.

3.2 Segmentazione

Lo scopo finale dell'intero procedimento, ossia la diagnosi di AD, viene perseguito valutando lo stato di atrofia della sostanza grigia del cervello: la fase di preelaborazione, appena vista, ha lo scopo di rendere i dati abbastanza omogenei da essere confrontati con un'immagine di riferimento che descrive un soggetto medio “normale”; si tratta ora di trovare il modo di effettuare tale confronto.

Tutti gli algoritmi della classe di cui si occupa questo lavoro fondano le loro valutazioni sui valori di intensità dei singoli voxel, basandosi sull'assunto che ogni tessuto, dopo un'opportuno pre-processing, sia caratterizzato da valori tipici. In linea di principio questo è vero, ma il segnale che costituisce le immagini MR è in realtà distribuito statisticamente intorno a tali valori, con ampie sovrapposizioni lungo la gamma dei toni di grigio.

Un modo per affrontare il problema, e che costituisce l'approccio più frequente in letteratura, passa attraverso la cosiddetta *segmentazione*, ossia il partizionamento dell'immagine in regioni disgiunte e ben definite che possano poi essere confrontate a fini diagnostici. La strada qui intrapresa, però, non consente questa suddivisione con la precisione sufficiente per effettuare valutazioni di volume o densità di parti anatomiche, per cui si cerca piuttosto una suddivisione nei tre tessuti principali del cervello, ossia materia bianca (WM), grigia (GM) e liquido cerebrospinale (CSF): l'idea è di valutare semplicemente la quantità di materia grigia, che è quella che subisce l'atrofia.

La distribuzione effettiva dei livelli di grigio nelle immagini non consente però un vero partizionamento, per cui questo viene sostituito da un insieme di immagini di probabilità, una per tessuto, su cui valutare le quantità totali: si tratta quindi, se così si può dire, di un partizionamento “sfumato”, e, per estensione, si parla ancora di segmentazione.

Questa strada è quella più frequentemente seguita in letteratura, ampiamente sperimentata, nonché distribuita all'interno di un pacchetto software integrato. Sarà descritta nel prossimo paragrafo.

Una diversa soluzione, su cui ferisce oggi una intensa ricerca, si affida invece all'analisi diretta delle immagini, senza passare da fasi di segmentazione

se non, al più, per selezionare particolari zone di interesse sulle quali fissare l’attenzione. Un approccio di questo tipo, però, non dispone di strumenti sufficienti per valutare l’appartenenza ai diversi tessuti dei voxel “ambigui”: occorrono perciò caratteristiche aggiuntive da *estrarre* dalle immagini, e strumenti software particolarmente sofisticati per elaborare l’imponente massa di dati che così si genera.

È una strada aperta che dispone di vaste possibilità ancora inesplorate, e che verrà esaminata, in un prossimo capitolo, in un suo recentissimo sviluppo.

3.2.1 Segmentazione mediante mappe statistiche parametriche

Il cervello può essere suddiviso in tre principali tessuti, che ne caratterizzano quasi totalmente la massa: materia grigia (GM), bianca (WM), liquido cerebrospinale (CSF).

Una misura del grado di atrofia, più o meno localizzata, passa attraverso l’identificazione della WM, che rappresenta il tessuto interessato da degenerazione nella AD.

Tradizionalmente questa suddivisione viene operata a mano, quindi subisce una valutazione soggettiva che è causa di incertezze e non ripetibilità.

Un possibile approccio per la segmentazione automatica è basato su un semplice modello statistico, e sul confronto con una distribuzione di riferimento dei tessuti.

Si assume che la MRI consista di un certo numero di tipi di tessuto, che determinano l’intensità di ogni voxel. L’appartenenza ad un dato tipo di tessuto determina una intensità del voxel conforme ad una distribuzione normale multivariata, che può essere descritta da un vettore di media, una matrice di covarianza, e dal numero di voxel appartenenti alla distribuzione.

Queste informazioni si aggiungono alla conoscenza a priori fornita da *immagini di probabilità*, una per ognuno dei tre tessuti principali, costruite su un largo numero di MRI, e fornite da una base di dati specializzata.

Il Montreal Neurological Institute (MNI), forse il principale riferimento mondiale per questo tipo e altri tipi di risorse di neuroimaging, costruisce tali

immagini di probabilità partendo da un ampio numero di scansioni di risonanza magnetica: queste vengono segmentate in GM, WM, e CSF da esperti, quindi trasformate in immagini binarie, dove cioè ogni voxel prende uno tra due valori a seconda dell'appartenenza o meno al tessuto, infine registrate in uno spazio comune mediante una registrazione affine a 9-parametri⁴. Ogni MRI produce quindi tre immagini, una per tessuto, i cui voxel valgono 0 o 1; una media tra tutte le immagini prodotte, calcolata su ogni voxel, fornisce quindi una stima della probabilità di corrispondenza di quel voxel al tessuto in questione. Tali probabilità costituiscono infine proprio il valore attribuito ai voxel nelle tre immagini di probabilità che costituiscono il prodotto finale di questa procedura.

(Alcune considerazioni vanno fatte sul significato delle immagini di probabilità: le MRI di base su cui sono costruite costituiscono un campione rappresentativo scelto accuratamente da esperti, che viene continuamente ampliato e perfezionato; la variabilità anatomica coperta è molto estesa, poiché i cervelli umani si diversificano tra loro sia per forma, sia per dimensione, sia per orientamento, e quindi la corrispondenza del singolo soggetto alla media va sempre considerata con adeguati strumenti statistici. La considerazione delle probabilità andrebbe pertanto sempre congiunta all'informazione sull'esatta procedura con cui è stata ottenuta: nel nostro caso (MNI) si può dire che il valore v_i del voxel i rappresenta la probabilità che un elemento di immagine appartenente ad un dato tessuto venga trasformato nella posizione i con una registrazione affine a 9-parametri (tre traslazioni, tre rotazioni, tre fattori di scala ortogonali), in generale diversa da quella ottenuta con una trasformazione rigida o da una a 12-parametri.)

Preso una immagine MR da segmentare, una prima stima della distribuzione dei tessuti si ottiene registrandola sulle immagini di probabilità con una trasformazione affine a 12-parametri (la trasformazione spaziale *lineare* che più ne assicura la corrispondenza): a questo punto ad ogni voxel vengono

⁴ Una registrazione rigida non va bene perché, oltre alle già ampie differenze geometriche individuali, la scala di una MRI dipende fortemente dal processo e dal macchinario di acquisizione, e va quindi obbligatoriamente corretta; una registrazione affine a 12-parametri, invece, può introdurre una deformazione delle strutture anatomiche non accettabile in una immagine di riferimento.

assegnate come probabilità di appartenenza ad ogni tessuto semplicemente i valori corrispondenti delle relative immagini di probabilità.

Questi valori iniziali vengono quindi corretti con un processo iterativo basato su un semplice modello bayesiano basato sull'ipotesi che ad ogni tessuto corrisponda un tipico valore del voxel, diverso per ogni immagine ma in essa costante⁵, e da determinare: la probabilità $p[c_i|v]$ che un valore v indichi appartenenza al raggruppamento (*cluster*) c_i viene dato dalla nota formula di Bayes:

$$p[c_i|v] = \frac{p[v|c_i] p[c_i]}{\sum_k p[v|c_k] p[c_k]},$$

dove $p[v|c_i]$ è la probabilità che l'appartenenza a c_i produca un voxel di valore v , qui valutata secondo un modello statistico gaussiano, e $p[c_i]$ è la probabilità di appartenenza del voxel a c_i , valutata questa in base all'immagine di probabilità come $p[c_i] = n_i b_i / \sum_j b_j$, con n_i = numero di voxel appartenenti a c_i e b_i = valore dello stesso voxel nell'immagine di probabilità che comprende c_i .

Ad ogni iterazione vengono ricalcolati i valori medi per ogni cluster, le matrici di covarianza, e i numeri di voxel per ogni cluster, questi ultimi come somma normalizzata ad uno di probabilità. Questi parametri vengono infine utilizzati per il calcolo, dalla formula di Bayes, di nuove immagini di probabilità, a partire dalle quali si calcolerà il ciclo seguente.

La convergenza non è garantita, quindi occorre fissare anche un opportuno numero massimo di iterazioni raggiunto il quale si ferma la procedura.

Va aggiunto che i cluster in realtà possono essere sei o sette, poiché devono comprendere anche tessuti non appartenenti al cervello e lo sfondo dell'immagine; e che i modelli della distribuzione delle intensità sono multivariati perché la variazione del livello di grigio a parità di tessuto dipende, oltre che dalla risposta intrinseca alla scansione, anche da fattori come la sovrapposizione di più tessuti nello stesso voxel, la struttura delle zone di transizione, e così via.

⁵ Questo rende l'algoritmo sensibile alle disuniformità, quando non corrette, prodotte da alcuni scanner.

3.3 Classificazione

Il passo finale di tutte queste procedure di calcolo, e al quale tutte tendono, è la suddivisione, o *classificazione*, delle immagini MR tridimensionali sottoposte all’analisi automatica in due categorie diagnostiche: appartenenti a soggetti sani, o a soggetti malati.

Differentemente dalle comuni analisi statistiche, che forniscono risultati descritti da probabilità e livelli di significatività, si chiede a questi algoritmi di decidere una diagnosi: parliamo quindi di algoritmi di *decisione*, e l’ambito di lavoro in cui essi operano si inquadra nella cosiddetta *intelligenza artificiale*.

Nel nostro caso, la decisione affidata all’algoritmo consiste nella scelta di una tra due classi, per cui parleremo di classificatori. Precisamente: chiameremo *classificatore* un procedimento che elabora oggetti muniti di un determinato insieme di caratteristiche, ed in base ad esse assegna all’oggetto una classe di appartenenza.

Questi algoritmi costituiscono oggi un campo di analisi e di ricerca estremamente vivace ed in continua evoluzione, sia per il loro interesse intrinseco e teorico, ma anche e soprattutto per le enormi prospettive ed applicazioni pratiche che spaziano tra i più svariati campi applicativi, tra cui appunto la diagnostica medica mediante elaborazione delle immagini digitali.

Di algoritmi classificatori ne sono stati proposti vari, ognuno dotato di peculiarità che possano renderlo efficace in particolari ambiti, ma che spesso si rivelano una scelta inadeguata in altri. Il “classificatore perfetto” allo stato attuale appare ancora lontano: tuttavia alcuni algoritmi hanno già dimostrato la loro validità fornendo ottime prestazioni in svariate ricerche e, pur soggetti a nuovi sviluppi e perfezionamenti, hanno raggiunto uno “stato dell’arte” riconosciuto universalmente come riferimento.

Nel nostro caso, il classificatore che si è rivelato più efficace, al punto da diventare *il* classificatore di riferimento nella diagnosi con neuroimmagini, è quello chiamato “macchina a vettori di supporto”: è quasi il metodo di classificazione standard, e si ritrova nella maggioranza dei lavori di diagnosi basata su MRI.

Verrà trattato dopo aver inquadrato brevemente il problema della de-

cisione automatica con algoritmi ad apprendimento, alla cui classe questo classificatore appartiene.

3.3.1 Algoritmi di decisione

In generale un *algoritmo di decisione* fornisce in uscita un valore di predizione, ossia una *decisione*, basandosi su un insieme di valori forniti in ingresso, detti *attributi* o *caratteristiche*, che descrivono un generico oggetto da analizzare: i valori degli attributi in ingresso possono essere sia discreti, come nel nostro caso, che continui, e altrettanto vale per i valori in uscita. Riferendosi a questi ultimi, si parla di *classificazione* in presenza di valori in uscita discreti, o di *regressione* nel caso di quelli continui.

Da quanto detto sembra si parli sostanzialmente di una funzione, e tale sarebbe infatti nel caso ideale: nulla però può assicurarci il determinismo, inoltre non è in alcun modo nota la relazione funzionale che lega le variabili. È proprio la ricerca di questa, in ultima analisi, l'obiettivo finale dell'algoritmo.

Mancando informazioni funzionali tra variabili e output, tutta la conoscenza che abbiamo del problema risiede solo nei dati, e da essa va estratta con un opportuno processo di *apprendimento*.

Volendo formalizzare, la questione si può esprimere in questo modo: data una funzione ignota f , anche di tipo non deterministico (in sostanza con input effettivo incompleto o in parte casuale), con valore di uscita $f(\mathbf{x})$, dove il vettore \mathbf{x} indica i valori degli attributi in ingresso, e presa una collezione (nota) di coppie del tipo $(f(\mathbf{x}), \mathbf{x})$, dette *esempi*, si vuole ricavare da essi una funzione h , di valore $h(\mathbf{x})$, e detta *ipotesi*, che *approssimi* f .

La procedura utilizzata nel nostro caso utilizza il cosiddetto metodo di *apprendimento supervisionato*, basato sulla seguente idea: si prende un insieme di oggetti di cui sono già noti sia il valore degli attributi che la classificazione corretta, e con questi si addestra l'algoritmo, regolandone i parametri in modo da ottenere la giusta classificazione per tutti i dati di addestramento; l'algoritmo così addestrato è ora pronto per dati di classificazione ignota.

Naturalmente, per quanto le caratteristiche possano essere scelte con cura, i dati di addestramento non possono in generale rappresentare tutti gli ingressi possibili, in altre parole non contengono tutta l'informazione necessaria per costruire una classificazione perfetta; inoltre anche la stessa classificazione esatta di tutti i dati di addestramento in pratica non solo spesso non avviene, ma in molti algoritmi sarebbe addirittura indice di un possibile *sovradattamento* ai dati, con relativo aumento degli errori.

I dati reali, infatti, contengono sempre una certa quantità di rumore (informazione estranea al problema), e un algoritmo che classificasse con esattezza tutto l'insieme di addestramento, potrebbe risultare in realtà “troppo” calibrato su quei valori, rumore compreso: gli stessi oggetti, acquisiti in condizioni di rumore diverse, potrebbero generare così classificazioni discordanti.

Questo può essere espresso dicendo che il modello che si cerca non deve essere complesso quanto, o più, dei dati che l'hanno generato, ma piuttosto il giusto compromesso tra semplicità e consistenza con i dati stessi.

In pratica, avendo a disposizione un certo numero (congruo) di esempi, si procede suddividendolo in due parti: una la si usa per realizzare l'addestramento dell'algoritmo, *insieme di addestramento* o di *training*; l'altra, *insieme di test*, viene utilizzata per verificare su dati “freschi” le sue effettive prestazioni. Solo a questo punto, se il controllo ha raggiunto risultati soddisfacenti, il classificatore può considerarsi pronto.

Questa suddivisione è fondamentale nelle tecniche di apprendimento supervisionato, e le procedure con cui viene effettuata, per sfruttare al massimo il contenuto informativo dei dati a disposizione, possono risultare alquanto complesse. Il principio generale comunque non cambia.

Si aggiunge, per completezza, che formalmente lo spazio delle funzioni h nel quale viene cercato il classificatore è detto *spazio delle ipotesi*, e il *problema dell'apprendimento* viene detto *realizzabile* se la funzione cercata f , funzione *obiettivo*, è contenuta in esso, ovvero è una delle h generabili dai dati, *irrealizzabile* nel caso opposto.

3.3.2 Support Vector Machine

Le *macchine a vettori di supporto*, o *Support Vector Machines* (SVM), o ancora, più in generale, *macchine kernel*, rappresentano un recente e promettente metodo per la classificazione di dati sia lineari, sia non lineari (ossia separabili linearmente o no, v. seguito): sono caratterizzate da un algoritmo di addestramento che conduce a risultati molto efficienti, ai quali si aggiunge la capacità di rappresentare funzioni anche molto complesse. Le prestazioni, nel tipo di classificazione che ci serve, sono considerate da molti le migliori sia per velocità che per accuratezza; l'apprendimento è però piuttosto lento, e quando la mole di attributi diventa particolarmente grande, come è il caso di quelli estratti dalle MRI, il carico computazionale diventa facilmente eccessivo: oltre alla crescita del tempo di calcolo, si può arrivare addirittura a peggiorarne le capacità classificatorie.

L'idea base di questo classificatore, ideato per la classificazione booleana, si basa sulla ricerca di una superficie $n-1$ dimensionale che separi in maniera ottimale i dati positivi da quelli negativi nello spazio n dimensionale degli attributi, partizionando lo spazio stesso in regioni di diversa classificazione.

Trovata tale superficie, dato che ogni nuovo vettore di caratteristiche sarà rappresentato da un punto nell' n -spazio, sarà sufficiente vedere in quale delle due regioni esso vada a cadere, per decidere della sua classificazione.

Cercando la *superficie di decisione* tra le funzioni polinomiali, è sempre possibile costruirne una di grado sufficientemente elevato da separare correttamente tutti i vettori (se vale l'ipotesi che non ne esistano due uguali con classificazioni diverse), come è anche possibile aumentare in qualche modo le dimensioni dello spazio degli attributi e trovare una superficie di decisione di grado minore che goda della stessa proprietà. In ogni caso si tratta di aumentare il numero dei parametri liberi per adattare la superficie ai dati.

Si è accennato in precedenza al rischio del sovradattamento, ed una superficie di decisione così costruita può molto facilmente incorrere in questo problema; non si può però rinunciare subito all'adattamento perfetto sull'insieme di apprendimento, decidendo così di perdere una parte dell'informazione disponibile, senza avere una precisa idea del problema e delle sue eventuali soluzioni.

Superfici di decisione

Cerchiamo quindi una superficie di decisione che operi una separazione perfetta dell’insieme di addestramento: in generale abbiamo una notevole libertà nella scelta dei valori dei parametri; useremo questa libertà per cercare una soluzione ottimale.

Per “ottimale” vogliamo intendere una soluzione che minimizzi l’errore di classificazione sui dati nuovi da classificare dopo l’addestramento, o, come si dice, che massimizzi la sua capacità di *generalizzazione* sui nuovi esempi: l’idea è quella di mappare i vettori in ingresso, attraverso una qualche funzione non lineare, in un nuovo spazio degli attributi con (molte) più dimensioni, e tale che la superficie di decisione sia in esso *lineare*, ossia un iperpiano (quando un iperpiano separa perfettamente un insieme di dati, tali dati vengono detti *linearmente separabili*). La linearità della superficie consentirà quindi di affrontare in maniera efficace il problema di ottimizzazione.

Per esempio si abbia uno spazio X costituito da vettori n -dimensionali di attributi, quindi gli esempi dell’insieme di addestramento saranno tutti del tipo (y, \mathbf{x}) , dove $\mathbf{x} = [x_1, \dots, x_n]$ è il vettore di attributi, mentre $y = \pm 1$ rappresenta la classificazione nota del vettore. Si cerchi in X una superficie di decisione polinomiale di grado 2: si crea allo scopo un nuovo spazio Z di attributi:

$$\begin{aligned} z_1 &= x_1, & z_2 &= x_2, & \dots, & z_n &= x_n, \\ z_{n+1} &= x_1^2, & z_{n+2} &= x_2^2, & \dots, & z_{2n} &= x_n^2, \\ z_{2n+1} &= x_1x_2, & z_{2n+2} &= x_1x_3, & \dots, & z_N &= x_{n-1}x_n, \end{aligned}$$

per un totale di $N = n(n+3)/2$ coordinate. Se in questo spazio i dati sono linearmente separabili, è qui che si cercherà l’iperpiano di decisione.

Se, per fissare le idee, consideriamo esempi caratterizzati da soli due attributi, quindi rappresentabili in un piano come punti bianchi o neri a seconda della loro classificazione, e supponiamo che tutti gli esempi positivi siano contenuti in una circonferenza e tutti quelli negativi siano esterni ad essa, otteniamo il caso di una “superficie” polinomiale di grado 2 (la circonferenza) che separa i dati.

È sufficiente ora sostituire al vettore originale $\mathbf{x} = [x_1, x_2]$ il nuovo vettore $\mathbf{z} = F(\mathbf{x}) = [z_1, z_2]$ di caratteristiche $z_1 = x_1^2$ e $z_2 = x_2^2$, per ottenere che nel nuovo piano $z_1 z_2$ i punti siano separati semplicemente da una retta.

In questo caso particolarmente semplice il nuovo spazio Z ha ancora due sole dimensioni, ma per superfici polinomiali di grado 2 possiamo arrivare ad avere spazi Z 5–dimensionali partendo da spazi X 2–dimensionali.

Il problema che ora si pone, nell’ipotesi di lineare separabilità, è quello di individuare un iperpiano che porti, se non alla migliore possibile, almeno ad una generalizzazione capace di prestazioni desiderabili.

Con l’aumentare degli attributi lo spazio Z può raggiungere dimensioni notevolissime, e definire un separatore lineare, che in uno spazio d dimensionale ha bisogno di d parametri, non è difficile quando il numero delle dimensioni è vicino al numero N dei dati a disposizione, ossia con $d \approx N$. Il sovradattamento in questo caso è un problema molto concreto, e si manifesta con l’incapacità di generalizzazione: non tutti gli iperpiani in grado di separare correttamente gli esempi di addestramento sono anche capaci di una buona classificazione con esempi nuovi.

Viene spontaneo il paragone col problema di trovare il polinomio interpolante un insieme di dati: oltre un certo grado la curva non descrive più il *fenomeno* da cui i dati sono generati, ma piuttosto i *dati* stessi. Molte le differenze, tuttavia il paragone è suggestivo.

Iperpiano ottimo

Si devono quindi affrontare in realtà due problemi: uno concettuale, ossia la ricerca dell’algoritmo per individuare l’iperpiano adatto; uno computazionale, cioè quello di come gestire questa ricerca in uno spazio di così elevate dimensioni.

Una soluzione del problema teorico è stata trovata nel caso del cosiddetto *iperpiano ottimo* [41].

L’iperpiano ottimo, per classi linearmente separabili, è definito come “la funzione di decisione lineare con i massimi margini (da entrambi i lati) tra i vettori delle due classi” [42].

In linea di principio: fissato un iperpiano separatore, si valutano le distanze tra esso ed i vettori, cercando quelli più vicini in ognuna delle due classi. Questi, per la ragione che in seguito vedremo, verranno chiamati *vettori di supporto*, mentre i due iperpiani, uno per ogni classe, passanti per i vettori di supporto e paralleli all'iperpiano separatore, saranno chiamati *margini*.

Al variare dei parametri che definiscono l'iperpiano, cambiano i vettori di supporto e conseguentemente anche gli iperpiani che li contengono: viene scelto un separatore che ne massimizza la distanza, fissando così i cosiddetti *margini ottimi* delle due classi, e si sceglie come iperpiano ottimo quello che opera la separazione con “massimo margine”, ossia quello che si situa in posizione intermedia tra i due margini ottimi.

La definizione di iperpiano ottimo è semplice, e intuitivamente espressiva: una superficie che sia il più possibile distante dalle classi definite dagli esempi di addestramento (anche se tale distanza non coinvolge direttamente tutti gli elementi della classe, ma solo quelli appartenenti ai margini) ci sembra sicuramente adatta a classificare un vettore ignoto; in più l'essere in uno spazio in cui questa risulta lineare, chiarisce e semplifica la procedura per definire e distanziare i margini.

Ma vi è un risultato molto più concreto che giustifica questo approccio: si può vedere, usando i dati dell'insieme di test con metodi propri della teoria dell'*apprendimento computazionale*, che quando i vettori di addestramento vengono separati *senza errori* da un iperpiano ottimo, allora il *valore di aspettazione della probabilità di commettere un errore* su un esempio di test risulta limitato dal rapporto tra il *valore di aspettazione del numero di vettori di supporto* ed il *numero dei vettori di addestramento*. In formula:

$$E[P(\text{errore})] \leq \frac{E[\text{numero dei vettori di supporto}]}{\text{numero dei vettori di addestramento}}$$

con significato chiaro.

Questo ci garantisce la qualità della generalizzazione, fornendocene anche una stima.

Vanno fatte a questo punto due importanti osservazioni. La prima è che, una volta definito, l'iperpiano ottimo viene descritto dai soli vettori di supporto (è proprio questo che giustifica il loro nome), comunque siano

disposti gli altri vettori di addestramento: questo può indicare una direzione da seguire per individuare tale superficie senza appesantire con troppi dati l'algoritmo. La seconda è che nella formula precedente *non* compare la dimensione dello spazio usato per la separazione; tutto ciò che conta è il numero atteso dei vettori di supporto rispetto a quelli di addestramento: questo vuol dire che per ottenere una grande capacità di generalizzazione non solo basta avere un grande insieme di addestramento a parità di vettori di supporto, come naturalmente ci aspetteremmo, ma è ugualmente importante costruire l'iperpiano col minimo numero possibile di supporti, a parità di numerosità in addestramento.

Resta da determinare come procedere praticamente per individuare l'iperpiano ottimo.

L'equazione dell'iperpiano ottimo si scrive nello spazio Z , in forma vettoriale:

$$\mathbf{w}_0 \cdot \mathbf{z} + b_0 = 0 ,$$

dove \mathbf{z} è il generico vettore di Z , mentre \mathbf{w}_0 rappresenta un vettore di pesi (i parametri dell'iperpiano) da determinare; lo scalare b è spesso chiamato *bias*.

Procedendo formalmente: dato un generico spazio Z di attributi, un insieme di esempi di addestramento

$$(y_1, \mathbf{z}_1), (y_2, \mathbf{z}_2), \dots, (y_N, \mathbf{z}_N)$$

con $y = \pm 1$, è detto *linearmente separabile* se esistono un vettore \mathbf{w} ed uno scalare b tali che valgano le disequazioni:

$$\begin{aligned} \mathbf{w} \cdot \mathbf{z}_i + b &\geq 1 && \text{per } y_i = 1 \\ \mathbf{w} \cdot \mathbf{z}_i + b &\leq -1 && \text{per } y_i = -1 \end{aligned}$$

per tutti i vettori dell'insieme. Queste disequazioni si possono scrivere in forma più compatta come:

$$y_i(\mathbf{w} \cdot \mathbf{z}_i + b) \geq 1, \quad \forall i$$

con l'uguaglianza che vale, per definizione, per i soli vettori di supporto.

Il vettore \mathbf{w} e lo scalare b definiscono un *iperpiano separatore* $\mathbf{w} \cdot \mathbf{z} + b = 0$, e il semispazio nel quale cade il generico vettore \mathbf{z} è indicato semplicemente dal segno di $\mathbf{w} \cdot \mathbf{z} + b$.

La distanza tra le due classi, che è funzione dell'iperpiano (\mathbf{w}, b) , si può scrivere sommando le distanze dall'iperpiano stesso dei due vettori più vicini, uno per classe. Allora:

$$d(\mathbf{w}, b) = \min_{\{\mathbf{z}: y=1\}} \frac{\mathbf{z} \cdot \mathbf{w}}{|\mathbf{w}|} - \max_{\{\mathbf{z}: y=-1\}} \frac{\mathbf{z} \cdot \mathbf{w}}{|\mathbf{w}|} = \frac{2}{|\mathbf{w}|}$$

poichè in corrispondenza dei vettori di supporto delle due classi i prodotti scalari valgono rispettivamente $1 - b$ e $-1 - b$.

L'iperpiano ottimo (\mathbf{w}_0, b_0) è quello che massimizza la distanza $d(\mathbf{w}, b)$, che risulta essere in questo caso:

$$d(\mathbf{w}_0, b_0) = \frac{2}{|\mathbf{w}_0|} = \frac{2}{\sqrt{\mathbf{w}_0 \cdot \mathbf{w}_0}},$$

con

$$\sqrt{\mathbf{w}_0 \cdot \mathbf{w}_0} = \min_{(\mathbf{w}, b)} \sqrt{\mathbf{w} \cdot \mathbf{w}}.$$

Esso è inoltre l'*unico* che minimizza $\mathbf{w} \cdot \mathbf{w}$ sotto la condizione prima definita di *lineare separabilità*.

Il problema dell'iperpiano ottimo si riduce quindi ad un problema di minimo su un prodotto scalare, e si affronta con tecniche di *programmazione quadratica*.

Si dimostra [42] che \mathbf{w}_0 si può scrivere nello spazio Z come combinazione lineare dei vettori di addestramento, cioè:

$$\mathbf{w}_0 = \sum_{i=1}^N y_i \alpha_i \mathbf{z}_i$$

dove, per di più, $\alpha_i = 0$ per tutti i vettori diversi da quelli di supporto, per i quali vale invece $\alpha_i > 0$. In definitiva \mathbf{w}_0 si scrive semplicemente:

$$\mathbf{w}_0 = \sum_{\text{vettori di supporto}} y_i \alpha_i \mathbf{z}_i$$

cioè come combinazione lineare dei soli vettori di supporto, che quindi rappresentano gli esempi essenziali, o critici, per la sua definizione.

Considerata ora l'equazione dell'iperpiano ottimo $\mathbf{w}_0 \cdot \mathbf{z} + b_0 = 0$, e ricordando quanto detto sui vettori che cadono da parti opposte di questo, abbiamo per la funzione di decisione:

$$h(\mathbf{z}) = \text{sgn}(\mathbf{w}_0 \cdot \mathbf{z} + b_0) = \text{sgn}\left(\sum_i y_i \alpha_i (\mathbf{z}_i \cdot \mathbf{z}) + b_0\right)$$

e infine

$$h(\mathbf{z}) = \operatorname{sgn}\left(\sum_{\text{vettori di supporto}} y_i \alpha_i (\mathbf{z}_i \cdot \mathbf{z}) + b_0\right).$$

Rimarchiamo che in questa formula sono coinvolti solo prodotti scalari con i vettori di supporto nello spazio Z , e questi ultimi sono normalmente in numero molto inferiore rispetto ai dati: la conseguenza è che il numero di parametri necessario a definire l’iperpiano ottimo, ed in definitiva l’intero classificatore addestrato, risulta molto minore del numero N di attributi, o dimensioni, in Z , con notevole diminuzione di costo computazionale. Questa caratteristica rende le macchine SVM molto meno soggette di altri metodi al sovradattamento dei dati [43], che come già detto si concretizza quando il numero di parametri che definiscono la superficie si avvicina alle dimensioni dello spazio.

Funzioni kernel

Per proseguire è ora necessario introdurre un risultato importante per lo sviluppo pratico dell’algoritmo: si dimostra [44] che la costruzione di una funzione di decisione può essere effettuata anche *scambiando l’ordine delle operazioni*. In altre parole, invece di effettuare una trasformazione non lineare dei vettori nello spazio X degli input seguita, come nella formula, da un prodotto scalare con i vettori di supporto nello spazio Z , si possono eseguire *prima*, ad esempio col prodotto scalare, le misure delle distanze ed i confronti relativi nello spazio X , e *dopo* applicare una trasformazione non lineare sul risultato.

Per chiarire questa procedura supponiamo di cercare un separatore lineare ottimo nello spazio X : usando i metodi di ottimizzazione della programmazione quadratica, il problema si riduce alla ricerca degli α_i che rendono massima [42] l’espressione:

$$\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

sotto le condizioni $\alpha_i \geq 0$ e $\sum_i \alpha_i y_i = 0$. Questa relazione ha l’importante proprietà di avere un singolo massimo globale che si può trovare con metodi di calcolo efficienti.

Ma ciò che qui più interessa è il fatto che in questa relazione i dati compaiono solo sotto forma di prodotti scalari di coppie di vettori, e questo vale anche per l’equazione finale di decisione, che in questo caso risulta

$$h(\mathbf{x}) = \operatorname{sgn}\left(\sum_i \alpha_i y_i (\mathbf{x} \cdot \mathbf{x}_i)\right) = \operatorname{sgn}\left(\sum_{\text{vettori di supporto}} \alpha_i y_i (\mathbf{x} \cdot \mathbf{x}_i)\right),$$

dove gli α_i sono quelli ottimi calcolati con l’espressione precedente, e valgono zero per tutti i vettori che non siano di supporto.

Questo nell’ipotesi di dati linearmente separabili in X , ma chiaramente non ci aspettiamo che in realtà essa si verifichi in tale spazio. In generale sappiamo che occorre rimappare l’ n -spazio X nel nuovo N -spazio Z , dove i vettori saranno del tipo $F(\mathbf{x})$, e la ricerca del separatore ottimo utilizzerà prodotti scalari del tipo $F(\mathbf{x}_i) \cdot F(\mathbf{x}_j)$. È però anche possibile calcolare *direttamente*:

$$F(\mathbf{x}_i) \cdot F(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j),$$

dove la funzione $K(\mathbf{x}_i, \mathbf{x}_j)$ è chiamata *funzione kernel*.

Grazie a questa proprietà possiamo cercare separatori lineari nello spazio Z calcolando solo i valori delle funzioni kernel, e non i valori di tutte le N caratteristiche di ogni vettore \mathbf{z} . L’espressione da usare nella ricerca degli α_i ottimi sarà infatti

$$\begin{aligned} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (\mathbf{z}_i \cdot \mathbf{z}_j) &= \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j F(\mathbf{x}_i) \cdot F(\mathbf{x}_j) \\ &= \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j). \end{aligned}$$

Così tutti i calcoli potranno essere eseguiti nello spazio di input X , e non nello spazio Z che è di dimensioni molto (ordini di grandezza) superiori: l’impatto sul tempo di calcolo è notevolissimo, considerando sia che ogni dimensione richiede una moltiplicazione ed una somma aggiuntive nel prodotto scalare, sia che non sono più necessarie le (costose) trasformazioni non lineari tra X e Y . Possiamo persino ignorare di quali trasformazioni realmente si tratti.

Il processo di ricerca di estremo nello spazio X usando le funzioni kernel è molto simile a quello necessario nello spazio Z con l’uso di prodotti scalari: l’unica sostanziale differenza è l’introduzione di un limite superiore C ai moltiplicatori α_i , e il cui valore va determinato sperimentalmente.

Un esempio semplice di funzione kernel si trova riprendendo il caso dei vettori del 2–spazio separati da una circonferenza: usando questa volta la trasformazione di caratteristiche $z_1 = x_1^2$, $z_2 = x_2^2$ e $z_3 = \sqrt{2}x_1x_2$, si ottiene facilmente $F(\mathbf{x}_i) \cdot F(\mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^2$ che è la funzione kernel $K(\mathbf{x}_i, \mathbf{x}_j)$ cercata.

Ad ogni kernel “ammissibile” (sono in realtà necessarie solo proprietà del tutto naturali come l’essere definito positivo) corrisponde una specifica trasformazione ed una conseguente superficie non lineare in X : la scelta della mappatura di X in Z equivale alla scelta del kernel più adatto ad ottenere la separazione lineare in Z dei vettori, ossia quello che porta alla migliore accuratezza sugli esempi di test.

Nella pratica la scelta si limita a poche classi dalle prestazioni note (altre classi sono attualmente oggetto di ricerca), le cui principali sono:

$$\begin{aligned} \text{kernel polinomiali di grado } h : K(\mathbf{x}_i, \mathbf{x}_j) &= (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^h \\ \text{kernel radiali di tipo gaussiano} : K(\mathbf{x}_i, \mathbf{x}_j) &= e^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2} \\ \text{kernel sigmoidali} : K(\mathbf{x}_i, \mathbf{x}_j) &= \text{tgh}(k\mathbf{x}_i \cdot \mathbf{x}_j - \delta) . \end{aligned}$$

Non esiste ancora un criterio per la scelta del kernel migliore, tuttavia le differenze di accuratezza tra quelli usati in pratica, sotto ipotesi sui dati molto generali, non sono grandi⁶.

Il metodo delle funzioni kernel, oltre che nella ricerca di separatori lineari ottimi, può essere utilizzato in ogni algoritmo che possa essere riformulato per operare solo su prodotti scalari di coppie di vettori.

Dati numerosi o non separabili

Un metodo computazionalmente molto efficiente, che si usa quando gli esempi sono numerosi, consiste nel suddividere i dati di addestramento in blocchi di numerosità ragionevole, e procedere come segue:

- si risolve il problema di programmazione quadratica sul primo blocco trovandone, se i dati sono linearmente separabili, l’iperpiano ottimo (se

⁶ Negli algoritmi esaminati in questo lavoro la scelta in genere è limitata a kernel lineari o gaussiani, e viene operata testandoli entrambi e scegliendo il più efficace; se non vi è differenza si preferisce quello lineare, computazionalmente più leggero.

ciò non fosse possibile allora l'intero insieme di addestramento sarebbe non separabile da un iperpiano);

- si confrontano i vettori del secondo blocco con la soluzione (\mathbf{w}_1, b) trovata per il primo, individuando quelli che non verificano la $y_i(\mathbf{w}_1 \cdot \mathbf{x}_i + b) \geq 1$;
- si crea un nuovo insieme di dati unendo questi ultimi vettori con i vettori di supporto, ed in questo insieme si cerca una nuova soluzione ottima (\mathbf{w}_2, b) ;
- si continua fino a coprire tutti i blocchi dell'insieme di addestramento.

Una volta completata, la procedura fornirà uno di due possibili risultati:

- (i) l'iperpiano finale è il separatore ottimo per tutto l'insieme di dati di addestramento;
- (ii) i dati non sono linearmente separabili senza errori.

In quest'ultimo caso si cerca un iperpiano che separi i dati di addestramento col minimo di errori. Per ottenere questo scopo si introduce un insieme di N variabili ξ_i non negative, e si considerano le disequazioni:

$$y_i(\mathbf{w} \cdot \mathbf{z}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

che vengono usate come vincoli nella minimizzazione del funzionale:

$$\Phi(\xi) = \sum_{i=1}^N \xi_i^\sigma$$

il quale esprime, per σ sufficientemente piccoli, il numero di errori commessi sull'insieme di addestramento.

Minimizzato $\Phi(\xi)$, si trova il sottoinsieme minimale di errori: questi ultimi vengono esclusi dall'insieme di addestramento, procedendo quindi alla ricerca del separatore coi vettori che restano; si ottiene così un iperpiano ottimo.

Questo procedimento ideale può essere sintetizzato nella ricerca formale del minimo del funzionale:

$$\frac{1}{2} \mathbf{w}^2 + C \cdot F \left(\sum_{i=1}^N \xi_i^\sigma \right)$$

con $F(\cdot)$ funzione monotona convessa e C costante, usando gli stessi vincoli di prima. Questo minimo, se C è abbastanza grande e σ abbastanza piccolo, determina direttamente l’iperpiano capace di *minimizzare* il numero di errori sull’insieme di addestramento, e contemporaneamente di *massimizzare* i margini di separazione sul resto dell’insieme.

Purtroppo, posto così, il problema è in generale di tipo NP-completo⁷. Per questa ragione si pone, in pratica, $\sigma = 1$, che rappresenta il più piccolo valore di σ per il quale si ha un’unica soluzione al problema computazionale; inoltre, diversamente dal caso $\sigma < 1$, per trovare tale soluzione si dispone anche di un metodo computazionalmente efficiente.

Per C abbastanza grande, l’iperpiano trovato con $\sigma = 1$ è ancora quello che massimizza i margini con i dati separati correttamente, mentre la minimizzazione non riguarda più il numero degli *errori* commessi, quanto piuttosto la somma degli *scarti* ξ .

3.4 Tecniche di diagnosi dell’AD con le immagini di risonanza magnetica

L’idea fondamentale su cui si sviluppa la ricerca sulla diagnosi di malattia di Alzheimer per mezzo di immagini MR, è la valutazione dell’atrofia cerebrale con essa correlata: questa si effettua sulla base di opportune caratteristiche presenti o estratte dalle immagini.

Dal punto di vista di tali caratteristiche, le strategie si possono raggruppare in tre grandi categorie:

- le caratteristiche sono definite al livello dei voxel, e comprendono le semplici intensità originali, più quelle ottenute dopo particolari manipolazioni e/o filtri, tra cui le probabilità di appartenenza ad uno dei tre tessuti fondamentali;
- le caratteristiche sono definite al livello del *vertice* (la parte più esterna) della corteccia cerebrale, di cui si cerca di valutare lo spessore;

⁷ Si intende con questo che appartiene ad una particolare classe di problemi di cui non si conoscono soluzioni deterministiche, ossia prive di passaggi ambigui, e inoltre non si sa se queste possano o meno esistere.

- lo studio si concentra sulla valutazione volumetrica e/o morfologica della struttura cerebrale detta *ippocampo*.

Tutte e tre le scelte hanno i loro presupposti teorici, e sono confortate da valutazioni di opportunità tecniche.

La prima è quella concettualmente più semplice, e offre il vantaggio di potersi estendere, a piacere, a tutto il cervello o a zone di interesse più specifiche: si affida soprattutto ad una precisa quantificazione, diretta o indiretta, delle probabilità di appartenenza ai tessuti principali, e può risultare sensibile alle deformazioni introdotte dalla tecnica di registrazione. Può disporre di software evoluto e ben collaudato, e dispone anche per questo del consenso di molti, ma è considerata da altri intrinsecamente limitata nella precisione, soprattutto quando comporta trasformazioni non lineari.

La seconda sfrutta le possibilità di una sofisticata registrazione elastica dell'immagine basata sul preciso riconoscimento dei principali *solchi* presenti sulla corteccia: sono strutture universalmente presenti e facilmente riconoscibili anche con procedure automatiche, consentendo un'ottima corrispondenza col riferimento; dalla deformazione risultante si può ricavare una stima precisa dello spessore del vertice corticale. È una tecnica complessa e computazionalmente pesante, ma offre il notevole vantaggio di usare come indice una conseguenza diretta dell'atrofia, appunto una misura di spessore invece di una semplice probabilità di appartenenza, ed è per questo ritenuta molto promettente; ha però il limite di essere applicabile solo su di un'area specifica, che potrebbe essere coinvolta in maniera non ancora evidente nella fase precoce della malattia.

La terza strategia si caratterizza non per i metodi, ma per la specifica area sotto osservazione: è dimostrato che l'*ippocampo*, unitamente alle zone limitrofe, è la struttura per prima e più profondamente interessata dai processi degenerativi causati dall'AD, ragion per la quale concentrare l'interesse su di essa rappresenta forse la più concreta speranza per una diagnosi precoce della malattia con metodi di neuroimaging. Per contro le ridotte dimensioni del volume coinvolto (per l'ippocampo basta una “scatola” di $30 \times 70 \times 30 \text{ mm}^3$ per contenerlo interamente) e i margini non precisamente identificabili, che ne rendono difficile la segmentazione, rendono impossibi-

le una misura morfometrica diretta, e costituiscono una sfida che impegna profondamente la ricerca.

3.4.1 Algoritmi basati sui voxel

Le strategie più usate e di più ampio respiro, e delle quali fondamentalmente si occupa questo studio, sono quelle basate sui voxel, e riguardano l'esame sia dell'intero cervello, sia del solo ippocampo con le zone limitrofe: esse condividono molte idee ed algoritmi, ma si differenziano a volte anche profondamente per il tipo di risposte che utilizzano per affrontare il problema generale.

La ricerca è molto viva in questo campo, e si arricchisce continuamente di idee e soluzioni originali, come è evidente dalla breve rassegna che segue.

VBM “diretta”

L'approccio più semplice in questa categoria di algoritmi è la classificazione diretta sulla base delle immagini (o *mappe*) di probabilità costruite nella prima fase dell'elaborazione.

Il metodo si affida alle potenzialità della Support Vector Machine, sfruttandone la capacità di lavorare su spazi ad altissimo numero di dimensioni.

Le MRI vengono elaborate per produrre, coi metodi già visti, una immagine di probabilità ben registrata nel volume di uno spazio comune, ogni voxel del quale costituisce una dimensione dello spazio matematico di lavoro della SVM, mentre l'intensità ne definisce, sempre per ogni immagine, la relativa coordinata.

Così facendo, nello “spazio-SVM” ogni immagine è ridotta ad un semplice punto, la cui posizione ne descrive le caratteristiche di probabilità, per ogni voxel, di corrispondere ad un elemento di materia grigia del cervello; una ulteriore variabile binaria ne contrassegna l'appartenenza ad uno dei due insiemi di classificazione.

La macchina a vettori di supporto procede quindi, con un kernel tipicamente lineare o radiale, alla ricerca dell'iperpiano ottimo di separazione (*optimal separating hyperplane*, OSH) tra le immagini contrassegnate come positive e quelle negative, col procedimento già esaminato nel paragrafo de-

dicato alla SVM; determinato l’OSH, sarà sufficiente descrivere ogni nuova MRI nello spazio-SVM e osservarne la collocazione rispetto all’iperpiano, per determinarne la classificazione.

Se il procedimento è concettualmente semplice, l’alto numero di dimensioni dello spazio-SVM lo rende computisticamente molto impegnativo, se non addirittura irrealistico: una parziale soluzione del problema risiede nell’accuratezza del processo di registrazione e normalizzazione.

La ragione di questo si comprende considerando, per ipotesi, due immagini di probabilità perfettamente corrispondenti geometricamente, e con livelli di intensità diversificati solo dalla diversa quantità di tessuto GM (quantificata dalla probabilità) causata dallo stato di salute o di AD del soggetto. In questo caso ideale solo i voxel del tessuto direttamente colpito da atrofia avranno, a parità di posizione nello spazio anatomico, intensità diverse nelle due immagini: la grande maggioranza dei voxel avrà invece uguale posizione anatomica e intensità, cioè eguale posizione nelle corrispondenti dimensioni nello spazio-SVM, con conseguente netta riduzione delle dimensioni effettive interessate dalla coppia di punti rappresentativi delle MRI. Lo spazio-SVM può quindi essere opportunamente ridimensionato prima di avviare la ricerca dell’OSH.

È quindi chiaro come solo con ottime registrazioni e normalizzazioni, che devono andare al di là del semplice caso lineare, si possa garantire una buona prestazione del processo: è in questo caso, infatti, che i punti rappresentativi delle MRI avranno una forte tendenza a raggrupparsi in un numero limitato di dimensioni.

In più, con questo procedimento, è possibile individuare i voxel più significativi per la classificazione, e assegnare loro un peso basato sul loro contributo nell’algoritmo; la distribuzione di questi nell’immagine del cervello offre un utilissimo ausilio nella localizzazione delle aree principalmente interessate dalla patologia, e soprattutto di quelle meglio utilizzabili come marcatori nelle MRI.

Questa metodologia la si trova proposta in [45], dove viene applicata sia all’intero cervello, sia in un volume di interesse localizzato nel lobo temporale mediale anteriore, comprendente anche l’ippocampo.

Da rilevare, in questo caso, l’assenza di procedure di smoothing.

VBM con aggregazione e selezione delle caratteristiche

Affidarsi alla sola registrazione per ridurre le dimensioni dello spazio-SVM può rivelarsi una strategia debole, per cui molti autori si dedicano, tra l'altro, alla ricerca di ulteriori modalità di selezione per le caratteristiche da usare nella macchina classificatrice.

Un approccio spontaneo consiste nell'eseguire una analisi preliminare delle mappe di probabilità tissutale (TPM), e stabilire delle semplici regole per eliminare subito quei voxel che non sembrano possano essere determinanti: dato che i processi di atrofia in esame interessano solo la materia grigia, è ragionevole ad esempio scartare i punti dove tale tessuto non è mai, o quasi, presente; inoltre si può sfruttare la capacità delle SVM di assegnare un peso ai voxel a seconda della loro importanza, per operare una selezione tra essi. In pratica si applica una prima SVM per assegnare tali pesi, individuare e quindi scartare i voxel meno importanti, e poi procedere con una seconda SVM che potrà operare in maniera più accurata sull'insieme di voxel così ridotto.

Questa idea richiede un'appropriata costruzione delle TPM, dalla quale dipende in definitiva l'efficacia delle fasi successive.

L'intera procedura, proposta in [46], può essere, in breve, così descritta:

- usando come modello iniziale le TPM fornite dal MNI, si costruiscono delle nuove mappe di probabilità basate su tutte e sole le MRI coinvolte nello studio;
- si normalizzano e segmentano quindi le immagini usando questi nuovi modelli TPM, ottenendo così le tre mappe tissutali di ogni immagine;
- si eliminano i voxel che presentano CSF in almeno metà delle immagini e comunque meno del 10% di tessuto, ove presente;
- si addestra con le restanti densità tissutali una SVM lineare per definirne i pesi;
- si opera, in base a questi, una ulteriore selezione dei voxel;
- si costruisce infine, con le probabilità tissutali dei voxel rimanenti, la SVM destinata alla classificazione definitiva.

Per rilevare possibili criticità del processo, sono utili alcune osservazioni:

- tutte le fasi di normalizzazione, correzione del bias e segmentazione, sono eseguite con moduli integrati nel pacchetto software “SPM5”, il che comporta una registrazione non lineare delle immagini, e quindi la necessità di una procedura di modulazione delle stesse per conservare le densità totali di tessuto in ogni soggetto;
- dopo entrambe le fasi di segmentazione, per la costruzione del nuovo template TPM e prima della selezione iniziale dei voxel, le mappe, con voxel di 1 mm^3 , vengono sottoposte a smoothing con un FWHM di 8 mm;
- la prima selezione viene in realtà operata su voxel ricampionati, con una semplice media, ad 8 mm^3 ;
- la prima SVM viene scelta lineare per assicurare ampi margini di separazione tra le classi;
- dopo la seconda selezione, per assicurare la coerenza spaziale, si includono tra gli attributi anche i voxel vicini a quelli prescelti.

VBM con selezione basata su atlante anatomico

L’idea base di questo tipo di tecnica, usata in [47] e [48], si fonda sulla suddivisione del cervello in regioni definite anatomicamente, ognuna delle quali viene considerata come una singola *feature* (o caratteristica): a ciascuna viene assegnato un valore, dopo una opportuna analisi, per poi procedere col classificatore SVM.

Quindi la riduzione delle caratteristiche si ottiene mediante il raggruppamento dei voxel su base anatomica.

Tale base anatomica è fornita da un atlante, sostanzialmente una immagine modello partizionata in regioni etichettate da esperti: la versione descritta in [49] comprende 116 parti, di cui 26 appartenenti al cervelletto, definite ed etichettate su una immagine di riferimento MNI, e quindi dispone di 90 caratteristiche utilizzabili per una valutazione di atrofia.

La procedura proposta in [47] prevede, per ogni MRI, una registrazione affine a 12 parametri sullo spazio MNI; quindi, coi parametri calcolati, si effettua la trasformazione inversa *del template* contrassegnato anatomicamente dallo spazio MNI a quello nativo dell’immagine, ottenendo così la corretta suddivisione anatomica in ROI della scansione in esame.

Il calcolo del valore da assegnare ai 90 attributi rappresentati dalle ROI parte dall’analisi dell’istogramma delle intensità di ognuna di esse: in molte regioni, e in tutti i soggetti, detto istogramma è chiaramente tri-modale, e può venire ben modellato come combinazione lineare di tre distribuzioni normali:

$$\alpha_1 N(\mu_1, \sigma_1^2) + \alpha_2 N(\mu_2, \sigma_2^2) + \alpha_3 N(\mu_3, \sigma_3^2) ,$$

con $\alpha_1 + \alpha_2 + \alpha_3 = 1$ e $\alpha_i \geq 0$, $i = 1, 2, 3$.

I coefficienti α_i rappresentano così, in ogni regione, le proporzioni relative di GM, WM e CSF, le cui distribuzioni, estratte dagli istogrammi con l’algoritmo EM [50], sono completamente descritte dalle terne $(\mu_i, \sigma_i, \alpha_i)$.

Considerato che i parametri gaussiani μ e σ sono condizionati dagli errori di acquisizione generati da macchine diverse più dei coefficienti α , e che l’atrofia coinvolge direttamente la materia grigia, si può scegliere come valore da assegnare ad ogni caratteristica proprio, e solo, α_1 , cioè il “peso” relativo di GM rispetto al totale con WM e CSF.

Per ogni soggetto, in definitiva, viene generato un vettore di 90 coefficienti α_1 , uno per regione; la successiva SVM, con kernel radiale $K(x, y) = e^{-\gamma|x-y|^2}$, potrà così operare su un numero di dimensioni facilmente gestibile.

Va aggiunto che la scelta del kernel radiale deriva dal confronto di prestazioni, effettuato dagli autori, con quello lineare; anche il potere discriminativo di α_1 è stato preventivamente accertato, usando la tecnica del *bootstrap* [51], con un test *t* confrontandone le distribuzioni tra malati di AD e controlli.

Altre strategie di estrazione delle caratteristiche

Sono stati proposti altri metodi per ridurre il numero di attributi, tramite l’individuazione di quelle inutili, ridondanti, o comunque meno efficaci; per esempio rendendoli adattativi alla patologia mediante l’analisi preventiva di MRI di soggetti malati [52] e la selezione automatica delle regioni del cervello interessate, oppure sperimentando nuovi sofisticati algoritmi di segmentazione [53][54] o di analisi statistica.

Le idee fondamentali si riconducono comunque:

– alla selezione delle feature più importanti, in base alla conoscenza a priori

e/o al punteggio generato da una classificazione preventiva;

– al loro raggruppamento, dopo un filtraggio tipicamente gaussiano che ne omogeneizza le distribuzioni spaziali, o registrandole su un riferimento già suddiviso in zone di interesse, oppure usando trasformazioni approssimate in serie di funzioni-base, che però perdono inevitabilmente informazioni morfologiche senza garantire alcun controllo su di esse.

Capitolo 4

Un nuovo algoritmo (pre-processing)

Quanto detto finora, se da un lato conferma le grandi aspettative riposte nelle tecniche di neuroimaging per la diagnosi dell'AD, dall'altro ben ne descrive le difficoltà intrinseche.

La VBM risulta una scelta pragmatica e relativamente economica, ma potrebbe essere sistematicamente viziata da alcuni problemi di principio: la normalizzazione iniziale introduce deformazioni che possono alterare significativamente i dati da classificare, inoltre quasi tutti gli approcci che derivano dalla VBM necessitano di pesanti filtraggi di smoothing, anche questi sospettati di accompagnare, agli effetti necessari per il calcolo successivo, una perdita di informazione forse decisiva per le procedure seguenti.

Infine, ma non meno importante, la catena di calcolo include molte diverse scelte (tipi di filtro, zone anatomiche su cui fissare l'attenzione, metodi di estrazione delle stesse dalle immagini, kernel per il classificatore, ecc.) operate dal ricercatore, selezionate e ottimizzate testandone direttamente i risultati, ma sempre necessariamente nel contesto dell'esperienza specifica sviluppata da ogni gruppo di ricerca, e interna ad esso: per quanto raffinato e migliorabile, tale contesto non è chiaramente replicabile. Una procedura il più possibile indipendente da fattori umani dovrebbe essere in grado di operare in modo autonomo alcune di queste scelte, almeno fin quando la ricerca non ne abbia accertato la validità.

Un'attenta analisi di questi, ed altri, problemi connessi, fornisce utili indicazioni per nuovi percorsi su cui indirizzare la ricerca.

È recentissima la pubblicazione di una nuova proposta [55] realizzata nell'ambito del progetto MAGIC-5 (Medical Application on a Grid Infrastructure Connection) dell'INFN, che comprende due ordini di novità rispetto alla letteratura precedente: da un lato sperimenta ulteriori miglioramenti nelle procedure "classiche", dall'altro introduce, nella catena di calcolo, nuovi recenti algoritmi che stanno dimostrando attitudini e potenzialità molto promettenti.

La procedura, completamente automatica, che costituisce la fase di pre-elaborazione delle immagini MR, si può schematizzare semplicemente in quattro sezioni:

- riduzione del rumore;
- registrazione affine;
- normalizzazione dei livelli di grigio;
- estrazione dei volumi di interesse (*Volumes Of Interest*, VOI) mediante registrazione rigida;

che saranno accuratamente descritte.

4.1 Riduzione del rumore

Un buon filtro antirumore (*de-noising*), come già detto in precedenza a proposito dello smoothing, è molto utile per uniformare le intensità dei tessuti e rendere più robusto qualunque successivo processo di registrazione e/o normalizzazione.

Il rumore in questione è presente soprattutto alle alte frequenze, poiché le distorsioni a bassa frequenza vengono già corrette, in funzione di ogni particolare scanner e modalità di acquisizione, già a cura degli esperti che curano il database ADNI (sui cui dati lavora l'algoritmo), inoltre la distorsione residua può essere efficacemente abbattuta in sede di normalizzazione: il de-noising viene quindi effettuato, in generale, con filtri del tipo passa-basso.

I classici filtri di sfumatura che spesso si trovano in letteratura, tuttavia, hanno anche il difetto di rendere meno riconoscibili i bordi e quindi le strutture anatomiche, perdendo così informazioni utili per segmentazione e classificazione: il filtro ideale deve quindi limitare il suo intervento al solo necessario, e per fare questo deve avere un'architettura calibrabile automaticamente.

4.1.1 Steerable pyramid filter

Il filtro piramidale “orientabile” [56] è un algoritmo multiscala invertibile di trasformazione delle immagini. Si basa su una struttura iterativa a banco di filtri che decompone l'immagine in bande, e le bande intermedie in componenti orientate, cambiando scala ad ogni iterazione.

La piramide decompone l'immagine in una serie di immagini componenti via via scalate, riducendole di un fattore due (lineare) ad ogni stadio dell'iterazione; dalla serie si può poi ricomporre l'immagine originale con un processo inverso.

Differentemente da altre trasformazioni discrete, ad esempio quelle modulate su una serie troncata di funzioni base, è però *invertibile*, cioè nessuna informazione viene persa durante il processo, e dopo la ricomposizione l'immagine torna identica all'originale: questo perché la scomposizione è tale da rendere essenzialmente priva di aliasing la componente che subisce il sottocampionamento. Grazie alla mancanza di aliasing, la trasformazione risulta anche invariante per traslazioni e rotazioni; la scomposizione è però anche sovracompleta per un fattore $4k/3$, con k uguale al numero di orientazioni, riducendone quindi l'efficienza al crescere delle orientazioni.

Utilizzando kernel di opportune proprietà (quelli utilizzati sono descritti nei dettagli in [57]), si costruiscono filtri H di tipo passa-alto, L di tipo passa-basso, e due o più filtri B passa-banda, uno per ogni orientazione¹; con questi filtri si costruisce un banco analogo a quello, con due orientazio-

¹ Questi filtri sono in pratica realizzati combinando passa-banda con derivate, col vincolo che il numero di orientazioni k deve superare di uno l'ordine di derivazione che si usa; nel nostro caso l'ordine è 1, quindi $k = 2$.

ni, illustrato in figura 4.1: estratta la componente $H_0^{(2,3)}$ (nodo 1), la parte residua selezionata dal filtro complementare L_0 viene ulteriormente suddivisa nelle due bande orientate B_1 e B_2 (nodi 2 e 3), il residuo L_1 viene quindi sottocampionato di un fattore 2 e inviato al nodo 4 dove potrà essere ulteriormente suddiviso in due sottobande orientate e un residuo, ulteriormente sottocampionato di un fattore 2, e così iterativamente fino a raggiungere la profondità voluta.

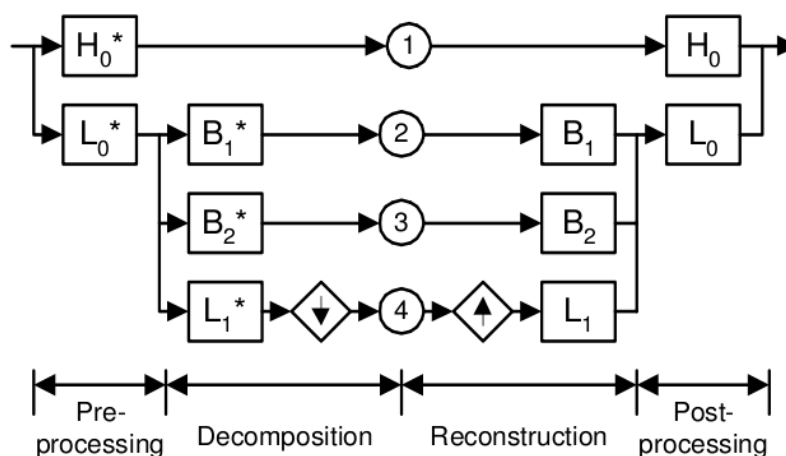


Figura 4.1: Diagramma a blocchi del filtro piramidale orientabile per $k=2$: l'intero banco di filtri è iterato in corrispondenza del nodo 4; le frecce verso l'alto e verso il basso indicano sotto- e sovra-campionamento.

Si dispone così di una decomposizione perfetta dell'immagine in sub-immagini base, con proprietà di invarianza per rotazioni, che può essere usata per numerose operazioni di filtraggio.

Per il de-noising si usano le peculiarità che distinguono il rumore dalle componenti effettive dell'immagine: il primo non ha direzioni preferenziali, le seconde presentano linee, contorni e tessiture anisotrope e ad alta energia locale. Un'analisi dei coefficienti della scomposizione consente di individuare quali parti della serie conviene attenuare, e questo con l'ausilio di una

² Indicherò con lo stesso simbolo sia il filtro che la banda da esso selezionata.

³ L'asterisco presente in figura nei filtri della sezione di ingresso indica solo che, nell'implementazione considerata, la funzione di trasferimento è ruotata di π radianti rispetto a quella senza asterisco.

funzione soglia, dipendente dall'energia del segnale, il cui valore può essere impostato dall'operatore per calibrare l'intervento del filtro [58].⁴

4.1.2 Structural Similarity Index

La calibrazione ideale di un filtro di de-noising è quella che unisce una decisa diminuzione del rumore con una perdita minima di segnale utile: una procedura automatica di filtraggio deve quindi poter disporre di una funzione in grado di quantificare il grado di intervento del filtro per raggiungere il risultato prefissato.

Un approccio classico si basa su misure di distanza, come l'errore quadratico medio (*Mean Square Error*, MSE), definito su una immagine di dimensioni $M \times N$ come:

$$\text{MSE} = \frac{1}{M \times N} \cdot \sum_{i=1}^M \sum_{j=1}^N (x_{ij} - y_{ij})^2,$$

dove l'immagine originale X e quella filtrata Y sono matrici di pixel di valori x_{ij} e y_{ij} rispettivamente.

Questa grandezza esprime, in questo caso, la distanza tra le due immagini intesa come funzione della differenza tra i toni di grigio di pixel corrispondenti, ma non è in grado di dare il peso corretto a tali differenze. Per esempio, un rumore casuale che provochi una semplice punteggiatura (analogo all'effetto neve della televisione analogica) senza pregiudicare il riconoscimento delle strutture presenti nell'immagine, può facilmente fornire lo stesso MSE di una forte sfocatura in grado di confondere in modo significativo i contorni degli oggetti: nel primo caso sarebbe opportuno aumentare il filtraggio, nel secondo la situazione potrebbe già risultare eccessiva.

Un approccio più moderno, invece, ricerca funzioni in grado di esprimere distanze che pesino in modo diverso i pixel in base alla loro rilevanza nel contesto in esame.

Un'idea innovativa, e che fornisce ottimi risultati, consiste nel prendere spunto dalla percezione umana: in questo caso si costruiranno funzioni di *somiglianza*, o *similarità*, nelle quali i pixel che descrivono strutture saranno più rilevanti di quelli "isolati", come sono quelli dovuti a rumore impulsivo.

⁴ Per ulteriori dettagli sulla progettazione di filtri orientabili si può vedere [59].

La difficoltà, ovviamente, consiste nel definire questo in termini matematici, soprattutto considerando che in questa fase di elaborazione non si dispone ancora, né si potrebbe, di alcuna descrizione strutturale esplicita.

L'indice di similarità strutturale SSIM (o anche SSI, *Structural Similarity Index*) [60] è un modo interessante di affrontare il problema: si individuano, in funzione delle due immagini da confrontare, tre grandezze legate a concetti che si rifanno alla percezione, ossia l per la luminosità, c per il contrasto e s per il riconoscimento di oggetti o strutture, e le si combina in un unico indice di somiglianza.

Si ha:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^\alpha \cdot [c(\mathbf{x}, \mathbf{y})]^\beta \cdot [s(\mathbf{x}, \mathbf{y})]^\gamma,$$

con $\alpha > 0$, $\beta > 0$, $\gamma > 0$, e con i vettori \mathbf{x} e \mathbf{y} che inglobano l'insieme di tutti i valori x_{ij} e y_{ij} dei pixel di X e di Y.

Le tre grandezze vengono poi definite basandosi su semplici indici statistici, scrivendo:

$$\begin{aligned} l(\mathbf{x}, \mathbf{y}) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \\ c(\mathbf{x}, \mathbf{y}) &= \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \\ s(\mathbf{x}, \mathbf{y}) &= \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \end{aligned}$$

dove le $\mu_x, \sigma_x^2, \mu_y, \sigma_y^2, \sigma_{xy}$ hanno l'usuale significato di media, varianza e covarianza di x e y , mentre le costanti C_1, C_2, C_3 hanno il solo scopo di stabilizzare la formula quando le grandezze a denominatore sono piccole.

In pratica la formula viene implementata con $\alpha = \beta = \gamma = 1$, per cui l'indice SSIM, sostituendo ed eseguendo i calcoli, in definitiva si può scrivere:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}.$$

Il valore comunemente assegnato alle costanti è $C_1 = (0.01 \cdot L)^2$, $C_2 = (0.03 \cdot L)^2$, avendo indicato con L la gamma dinamica dei pixel, tipicamente $(2^B - 1)$, con B che indica il numero di bit a disposizione per i valori dei pixel.

L'indice così definito varia tra 0 ed 1, e non necessita quindi di normalizzazione.

Osservando le definizioni, si può notare come esse siano in relazione con le caratteristiche percettive prima menzionate: $l(\mathbf{x}, \mathbf{y})$ è infatti funzione delle medie dei valori dei pixel, quindi delle luminosità; $c(\mathbf{x}, \mathbf{y})$ lo è delle varianze, quindi si può pensare legato al contrasto; infine $s(\mathbf{x}, \mathbf{y})$, trascurando la costante, corrisponde ad un coefficiente di correlazione, il che giustifica l'idea che descriva in qualche modo le altre componenti strutturali dell'immagine. Pur non essendo particolarmente immediate, tali relazioni si rivelano comunque efficaci per lo scopo che qui si vuole raggiungere.

Va aggiunto che in realtà si ottiene una migliore misura della similarità fornita dall'indice quando la formula viene applicata non direttamente sull'intera immagine, ma piuttosto localmente, calcolandola su piccole porzioni di immagine, e poi mediando i valori così ottenuti. In pratica si esamina il contenuto di una piccola finestra, tipicamente di 8×8 pixel, che si sposta di un punto alla volta spazzando l'intera immagine, e calcolandone prima l'SSIM, poi la media aritmetica di tutti i valori così raccolti: il risultato costituisce l'SSIM medio (*mean* SSIM, MSSIM), indicato nell'uso comune semplicemente SSIM, dato che questa è considerata oggi la procedura standard.

Per quanto un esame critico [61] dell'indice SSIM (basato comunque solo su grandezze statistiche semplici) suggerisca che le sue prestazioni non giustifichino l'entusiasmo espresso da alcuni autori [62][63], e per quanto la ricerca proponga sempre nuovi e più sofisticati indici di similarità di uso generale e fotografico, è indubbio che l'SSIM sia attualmente ritenuto ancora la scelta più affidabile tra quelle a disposizione, e di qualità più che adeguata al contesto in cui qui è usato.

4.1.3 Calibrazione automatica della soglia di rumore

Muniti dell'indice SSIM, lo si può utilizzare per il calcolo automatico del livello di soglia di rumore N_t che definisce il punto di intervento del filtro piramidale.

Per fare questo si opera con software sviluppato per immagini 2D, adattato all'uso in tre dimensioni come segue: procedendo indipendentemente sulle tre immagini bidimensionali giacenti sui piani mediali principali (assiale, coronale e sagittale) della MRI si filtra, per ogni direzione, l'immagine originale per diversi valori di N_t , e si calcola il relativo SSIM con l'immagine filtrata, fino ad ottenere un grafico di SSIM in funzione di N_t , sempre per ogni direzione [62].

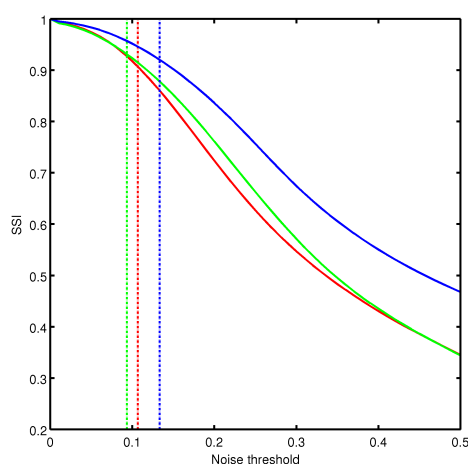


Figura 4.2: *Indice di similarità strutturale calcolato sulle tre sezioni mediali di una MRI (assiale in basso, coronale al centro, sagittale sopra); i valori selezionati come soglia di rumore sono rappresentati dalle linee tratteggiate (da sinistra: coronale, assiale, sagittale)*

La curva risulta monotona decrescente con andamento di tipo sigmoidale, come evidente in figura 4.2: inizialmente l'aumento di N_t , e quindi del potere filtrante, altera molto poco la similarità dell'immagine filtrata con l'originale, poi il peggioramento diviene più deciso; il valore scelto come ottimale di N_t corrisponde al punto in cui si osserva questo cambio di andamento, ossia sul ginocchio della curva.

In pratica, mediante il calcolo delle derivate prima e seconda della funzione SSIM, si stima come valore ottimale di N_t quello intermedio tra lo zero della derivata prima e il punto di flesso.

Ottenuti a questo punto i tre valori $N_{t,y}$, $N_{t,y}$ ed $N_{t,z}$, si filtra l'inte-

ra MRI separatamente in ogni direzione, ognuna col valore di soglia che le compete: le tre immagini filtrate vengono poi unite facendone la media, con ulteriore miglioramento del rapporto S/N.

La bontà del procedimento di calibrazione risulta evidente dall'esame della figura 4.3, che mostra una sezione dell'immagine MR di un individuo sano, prelevata dalla base di dati ADNI, confrontata con la sua versione filtrata mediante il processo descritto: la granulosità dovuta al rumore è pressoché sparita, mentre il livello di dettaglio visivo appare quasi inalterato.

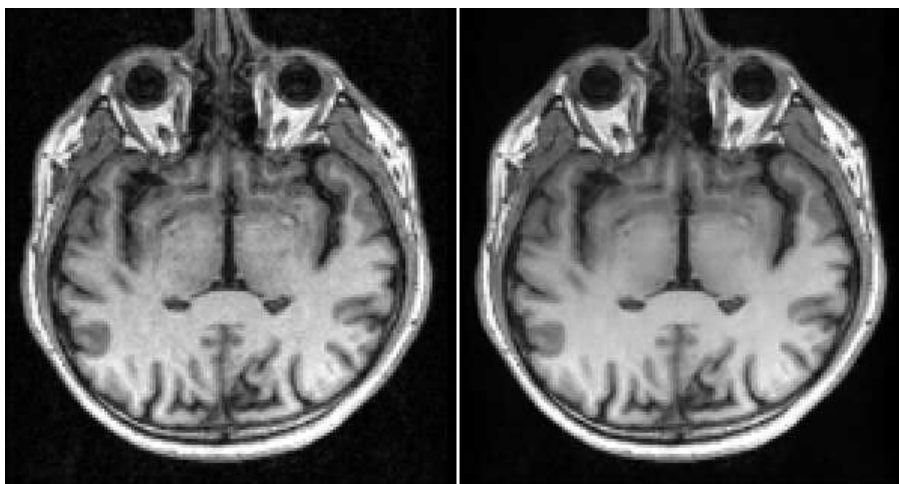


Figura 4.3: Immagini 2D su un piano assiale di un individuo sano: a sinistra l'originale prelevato dal database ADNI; a destra la stessa filtrata con filtro de-noise piramidale a calibrazione automatica.

4.2 Registrazione affine

Per quanto riguarda la fase di registrazione, si è già visto che gli algoritmi che si affidano a trasformazioni elastiche non lineari, che siano locali o di tipo diffeomorfo, comportano, oltre ad una certa “pesantezza” computazionale, il fondato rischio di alterare i dati grezzi al punto da perdere informazioni utili alla classificazione diagnostica finale: gli strumenti che tengono conto delle deformazioni introdotte (matrici jacobiane, modulazioni di intensità,

ecc.) sono sofisticati, e a volte essenziali proprio per la stessa analisi finale dei dati, come ad esempio negli algoritmi basati sulle misure di spessore corticale; tuttavia la loro efficacia, e anche solo la loro “innocuità” algoritmica, è tutt’altro che dimostrata per le procedure di morfometria basata sui voxel, le quali ultime cercano semplicemente di valutare, basandosi sulle intensità, il rapporto presente tra materia grigia ed altri tessuti. Inoltre, dati manipolati in maniera pesante potrebbero renderne più difficile anche l’interpretazione umana delle caratteristiche individuate come legate all’esordio della malattia.

Questo chiarisce l’esigenza di una registrazione che sia, per quanto possibile, semplice, robusta ed affidabile; e quindi la scelta di evitare algoritmi non lineari.

L’immagine filtrata viene perciò registrata utilizzando una trasformazione affine a 12-parametri: la registrazione rigida è insufficiente in questa fase iniziale, poichè le variabilità di forma e dimensioni dovute a pazienti e scanner diversi sono troppo grandi, e hanno una distribuzione troppo ampia; servirebbe invece almeno un certo grado di sovrapposizione delle regioni da esaminare per effettuare i passi successivi. Con dodici parametri è possibile, oltre che raggiungere un’ottima uniformità di forma e volume compensando le differenze globali, garantire una corrispondenza sufficiente, per le fasi successive, anche in regioni più limitate.

Una corretta registrazione ha bisogno di utilizzare una immagine di riferimento che rappresenti un “cervello medio” umano: una scelta affidabile e universalmente riconosciuta è rappresentata dal *template* fornito dal Montreal Neurological Institute [64], sul quale vengono quindi registrate tutte le immagini da analizzare utilizzate in questo lavoro.

Per rendere più accurato il procedimento, rispetto alla versione “classica”, la valutazione della corrispondenza tra MRI di riferimento ed MRI “bersaglio”, che guida l’algoritmo di trasformazione, è affidata a due indici diversi: la registrazione viene quindi effettuata per due volte, usando alternativamente e indipendentemente entrambe le misure, e ottenendo quindi due distinte registrazioni. I due risultati vengono poi confrontati col riferimento usando lo stesso indice, quindi il meglio corrispondente viene selezionato per proseguire nella catena di calcolo, mentre l’altro viene semplicemente

scartato.

Come misure di corrispondenza si sono scelti due indici entrambi ben collaudati in questo tipo di procedure: la *correlazione incrociata normalizzata*, e l'*informazione mutua normalizzata*.

La correlazione incrociata normalizzata viene definita, detti rispettivamente $f(x, y)$ e $t(x, y)$ i valori di intensità nell'immagine bersaglio ed in quella di riferimento, come:

$$\frac{1}{n-1} \sum_{x,y} \frac{(f(x, y) - \bar{f}) \cdot (t(x, y) - \bar{t})}{\sigma_f \sigma_t},$$

con n numero dei pixel in $f(x, y)$ e $t(x, y)$, \bar{f} e \bar{t} medie, σ_f e σ_t deviazioni standard.

L'informazione mutua normalizzata è invece definita da⁵:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p_1(x) p_2(y)} \right),$$

dove $p(x, y)$, $p_1(x)$, $p_2(y)$ sono la probabilità congiunta e le probabilità marginali di X e Y .

La scelta finale tra i due risultati migliori è effettuata misurando la somiglianza col riferimento mediante correlazione incrociata normalizzata.

Va notato che la presenza, nel seguito, di immagini registrate usando funzioni di misura diverse, non introduce errori sistematici in quanto la registrazione comporta solo la ricerca di parametri ottimali, mentre la trasformazione conseguente è sempre affine; inoltre la scelta dell'insieme di dodici parametri migliore è effettuata sempre tramite lo stesso indice, quindi con lo stesso criterio: la doppia registrazione ha quindi come unico effetto un irrobustimento del processo rispetto alla variabilità delle condizioni iniziali.

Ogni immagine viene infine effettivamente trasformata e ricampionata, coi metodi noti, su una griglia isotropa con celle di 1 mm^3 .

⁵ La definizione si può giustificare intuitivamente osservando che non vi è informazione "mutua" quando le probabilità sono disgiunte, nel qual caso $p(x, y) = p_1(x) \cdot p_2(y)$ rendendo l'indice nullo, mentre l'informazione mutua è massima quando $p(x, y) = p_1(x) = p_2(y)$, nel qual caso l'indice restituisce, a meno di una costante moltiplicativa, l'informazione totale secondo Shannon.

4.3 Normalizzazione dei livelli di grigio

Come in tutti gli algoritmi di morfometria basata sui voxel, anche in questo l'analisi finale si appoggia, in definitiva, sui livelli di grigio di ogni voxel, idealmente variabili in intervalli diversi e disgiunti per ognuno dei tre tipi fondamentali di tessuto cerebrale (materia grigia, materia bianca, liquido cerebrospinale).

Anche supponendo soddisfatti, in ogni MRI, tali requisiti ideali, resterebbe il problema della diversa posizione ed estensione dei tre principali intervalli di intensità al variare delle immagini da esaminare, con evidenti ambiguità ed effetti confondenti.

Qualcuno degli algoritmi esaminati nei capitoli precedenti è idealmente immune a questo tipo di problema, in quanto individua i livelli di riferimento all'interno stesso di ogni immagine, per esempio analizzando e modellandone l'istogramma delle intensità, e perché valuta prima il rapporto tra i diversi tessuti, e poi lo utilizza nel confronto con le MRI di altri soggetti. Al contrario, come si vedrà nel seguito, questo algoritmo evita ogni valutazione parziale fondata su singole immagini, riservando tutta la parte analitica e decisionale alle sezioni finali del processo.

Alla preelaborazione è quindi affidato il delicato compito di rendere tutti i dati uniformi tra loro, preservando l'informazione utile, con la massima accuratezza.

È fondamentale, perciò, che ai tre tessuti corrispondano circa gli stessi tipici valori di grigio, e con la massima aderenza possibile, per tutte le immagini. In buona sostanza, si tratta di individuare per ogni MRI i valori caratteristici di intensità di ogni tessuto, dopodiché renderli omogenei rispetto al passaggio tra una immagine e l'altra.

Come in precedenza per la registrazione affine, occorre per questo un'immagine modello che faccia da riferimento per i livelli di grigio, e naturalmente si utilizza lo stesso template che ha, in sostanza, definito la griglia spaziale di tutte le immagini. Tuttavia questa immagine, come ogni altra del resto, se presa per intero non può garantire una suddivisione nei tre tessuti fondamentali col livello di precisione desiderato, come già chiarito durante l'esame dei metodi VBM classici.

Anche per questa ragione si è preferito selezionare per la normalizzazione dei grigi solo una regione ristretta, relativamente facile da suddividere nei tre tessuti fondamentali grazie alle peculiarità anatomiche in essa incluse, e facilmente registrabile sulle immagini bersaglio: si tratta di una regione centrale orientata lungo l'asse maggiore del corpo calloso, la cui misura, nel riferimento MNI, è di $50 \times 120 \times 50 = 300'000$ voxel, ed evidenziata in figura 4.4 dai rettangoli scuri.

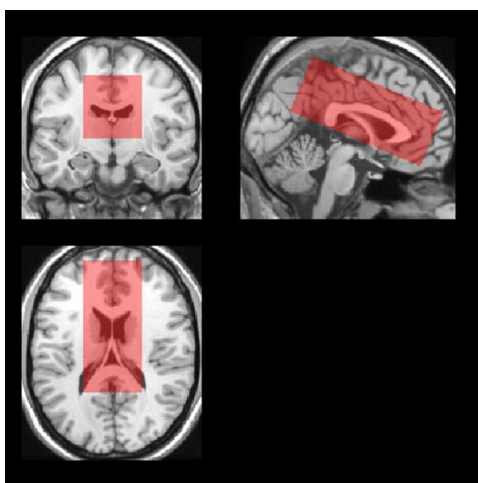


Figura 4.4: ROI per la normalizzazione dei toni di grigio, evidenziata dai rettangoli scuri, tracciata sull'immagine di riferimento del Montreal Neurological Institute.

La strategia utilizzata, una variazione di quella descritta in [40], è dunque la seguente: dopo aver selezionato manualmente, una volta per tutte, tale regione sulla MRI di riferimento, per mezzo di una registrazione rigida si mappa il riferimento su ognuna delle immagini bersaglio, individuando così su di esse la regione corrispondente a quella tracciata sul template. A questo punto si procede con la segmentazione, mediante l'algoritmo delle k-medie descritto più avanti, dell'MRI bersaglio, suddividendola in GM, WM e CSF; e così pure del riferimento (anche in questo caso una volta per tutte).

Infine, confrontando i relativi istogrammi, si scalano non linearmente, per ogni bersaglio, le intensità dei toni di grigio fino a raggiungere, per ogni tessuto, un valore medio coincidente a quello del riferimento. La mappatura

delle intensità viene completata estendendola agli altri valori di grigio con una curva polinomiale.

Si può notare la differenza di approccio rispetto ad altri metodi, i quali segmentano, tipicamente a mano, il template in GM, WM e CSF, trasferendo poi questa segmentazione sul bersaglio tramite registrazione: in questi casi è indispensabile ricorrere a pesanti deformazioni per ottenere buoni risultati.

Il metodo di classificazione che verrà qui usato ha invece bisogno di dati il meno possibile manipolati, ragione per la quale si è scelta una semplice registrazione rigida, il cui scopo è ora solo l'individuazione, con la dovuta precisione, della regione di interesse (ROI): la segmentazione deve essere perciò essere eseguita su ogni singola immagine in maniera automatica, ma solo sulla ROI appositamente scelta per questo scopo.

4.3.1 Algoritmo delle *k*-medie

L'algoritmo delle *k*-medie [65], è una tecnica di “*clusterizzazione*”, ossia di partizionamento degli elementi di un insieme in raggruppamenti detti *cluster*, semplice ma efficace, che utilizza una procedura iterativa basata anch'essa, come le procedure di registrazione, sulla minimizzazione di una funzione obiettivo.

Il funzionamento si può così sintetizzare: si fissa una funzione di misura della distanza tra gli elementi, si stabilisce il numero *k* di raggruppamenti che formeranno la partizione, dopodiché se ne determinano i centri, o *centroidi*, inizialmente scegliendo *k* dati per estrazione casuale, o in base a qualche informazione nota; a questo punto si esegue il ciclo:

- si assegnano i restanti elementi al cluster col centroide più vicino sfruttando la funzione distanza;
- per ogni cluster si calcola un nuovo centroide facendo la media degli elementi che vi appartengono;
- se i centroidi coincidono con i precedenti, il processo termina, altrimenti il ciclo riprende.

Più in generale, detti C_i i k centroidi dei cluster P_i , C il vettore che li contiene, e X_j gli N oggetti da raggruppare (rappresentabili come vettori di

attributi), si costruisce inizialmente e ad ogni iterazione una matrice logica \mathcal{A} di dimensione $k \times N$ detta *di appartenenza*, i cui elementi a_{ij} indicano lo stato di verità dell'appartenenza dell'oggetto j al cluster i ; fissata quindi una *funzione obiettivo* del tipo:

$$U(\mathcal{A}, C) = \sum_{i=1}^k \sum_{X_j \in P_i} \|X_j - C_i\|^2,$$

si calcola la \mathcal{A} che minimizza $U(\cdot, C)$, poi la C che minimizza $U(\mathcal{A}, \cdot)$; al raggiungimento di un criterio di convergenza, l'algoritmo si arresta.

I due tipici criteri di convergenza sono:

- matrice \mathcal{A} inalterata;
- differenza tra i valori della funzione obiettivo in due iterazioni successive al di sotto di un valore fissato.

Il pregio principale di questo algoritmo è la velocità, soprattutto in presenza di dati raggruppabili in cluster di numerosità simile, il che in pratica comporta spesso un numero di iterazioni addirittura inferiore al numero di punti da raggruppare.

I principali difetti riguardano il comportamento computazionale con insiemi di particolare difficoltà, nel qual caso il procedimento può rallentare anche di più ordini di grandezza [66][67]; la necessità di stabilire a priori il numero di cluster, il che richiede una qualche conoscenza iniziale sui dati; la dipendenza del risultato finale dalla condizione di partenza, che non fornisce quindi alcuna garanzia sulla qualità raggiunta.

In questo caso, però, il numero di cluster è noto, i valori di intensità sono ben distribuiti, inoltre è anche possibile fissare i centroidi iniziali avvalendosi di particolari localizzazioni anatomiche: si è quindi proprio nelle condizioni ideali per il buon funzionamento dell'algoritmo.

4.4 Estrazione dei volumi di interesse

Come visto nei capitoli precedenti, molti gruppi hanno scelto di focalizzare la loro ricerca su quelle particolari strutture cerebrali note per essere coinvolte per prime nella degenerazione causata dalla malattia di Alzheimer: se l'analisi allargata all'intero cervello fornisce molti più dati per la ricerca dell'atrofia, la quale coinvolge gradualmente tutta la materia grigia, una

scelta ristretta alle regioni più e prima colpite aumenta la frazione di dati significativi per la classificazione. Inoltre, quando l'obiettivo della ricerca è proprio la diagnosi *precoce* della malattia, l'analisi localizzata appare quasi una scelta obbligata.

Si è visto anche che la regione individuata dai medici come la più indicativa per l'esordio di AD, e coerente con la sintomatologia osservata, è la struttura chiamata *ippocampo*, con a seguire le regioni ad esso limitrofe: numerosi lavori concentrano perciò l'analisi di MRI proprio a queste.

La scelta degli autori si inserisce in questo filone di indagine, ma in una prospettiva più allargata: le regioni sotto esame sono ben dieci, cinque per emisfero, contenute in altrettanti parallelepipedi detti volumi di interesse (*Volumes of Interest*, VOI). Alcuni dettagli sui VOI prescelti sono presentati in tabella (uno di essi non è stato estratto a causa di un problema di calcolo).

VOI n°	Strutture anatomiche contenute	dimensioni in voxel
1	Ippocampo, corteccia entorinale dx	30×70×30
2	Ippocampo, corteccia entorinale sx	30×70×30
3	Amigdala dx	34×34×34
4	Amigdala sx	34×34×34
5	Giro temporale medio ed inferiore dx	30×50×30
6	Giro temporale medio ed inferiore sx	30×50×30
—	Insula, giro temporale superiore dx	—
7	Insula, giro temporale superiore sx	30×70×30
8	Corteccia rolandica dx	36×36×36
9	Corteccia rolandica sx	36×36×36

Questi VOI sono selezionati in modo da includere tutte le strutture del lobo temporale, tra cui l'ippocampo, che si sa essere coinvolte nelle fasi precoci di AD, con l'aggiunta di due regioni di controllo note invece per essere relativamente risparmiate dalla malattia. È così possibile approfondire la ricerca anche sull'efficacia di queste zone anatomiche come marcatori della

malattia (in figura 4.5 il VOI n° 1 e la principale struttura che contiene: l'ippocampo).

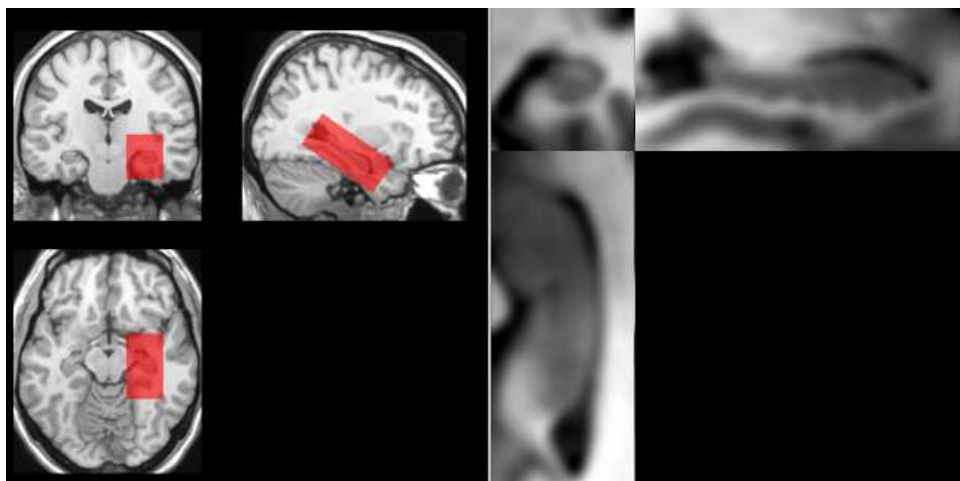


Figura 4.5: A sinistra: il VOI n° 1 evidenziato sul riferimento MNI. A destra: l'ippocampo, estratto dalla stessa immagine. Le sezioni sono, nell'ordine: coronale, sagittale, assiale.

Questa scelta incontra la necessità tecnica di disporre di volumi abbastanza piccoli da essere ben individuati sulle immagini MR con semplici registrazioni rigide, più affidabili per i motivi già esaminati, ma comporta un aumento di voxel, quindi di caratteristiche, da analizzare: il particolare processo di selezione delle caratteristiche adottato nelle fasi successive dell'elaborazione consente però di adottare questo tipo di soluzione senza particolari problemi.

La procedura di *estrazione* dei volumi di interesse, ossia la selezione in ogni immagine di tutti e soli i volumi anatomici da inviare alla elaborazione successiva, viene eseguita con un approccio del tipo basato su atlante.

L'atlante, in questo caso, è costituito da modelli (*template*) ricavati da un insieme misto costituito da pazienti di AD, soggetti in condizione di MCI, soggetti sani di controllo, dalle cui MRI sono state estratte le zone interessate in lavori precedenti [68]. Calcolando, da queste, delle immagini medie, viene costruito un atlante dei VOI dipendente dalla patologia, e quindi si hanno più template disponibili per ogni volume da estrarre.

Per estrarre il VOI da una immagine bersaglio, si considera ogni template relativo a quel VOI, e si procede con le seguenti operazioni:

- si registra rigidamente il template sul bersaglio, selezionando così un VOI-candidato;
- si calcola il coefficiente di correlazione tra template e VOI-candidato;
- si passa al template successivo e si riesegue il ciclo;

il VOI estratto sarà quel VOI-candidato che ha miglior correlazione.

La relativa complicazione della procedura di estrazione dei volumi di interesse, ha origine dalla scelta di evitare qualunque registrazione non lineare, anzi, in questo caso, di limitarsi ad una trasformazione rigida, con la conseguenza che un singolo template “medio” ora non riesce, in mancanza di opportune deformazioni, a coprire tutti i tipi di variabilità intersoggettiva. Il metodo proposto consente di usare per l'estrazione il template più idoneo al soggetto in esame, e lascia all'algoritmo la scelta di quale sia questo template.

Vale la pena di notare che la procedura complessiva di allineamento tra immagine di un soggetto e riferimento, comporta in definitiva due distinte registrazioni, una globale affine, e una locale rigida: questo processo in due fasi si dimostra capace di ridurre la variabilità posizionale del VOI [69], con vantaggi sulla qualità dei dati per le fasi successive.

Capitolo 5

Un nuovo algoritmo (calcolo e scelta delle caratteristiche, classificazione)

La classificazione, come si è già avuto modo di sottolineare nella descrizione della SVM, è in ultima analisi il processo di estrazione dai dati in *input* di una, o più, informazioni che supponiamo essi contengano in una qualche forma: la macchina classificatrice sfrutta un insieme di dati con *output* noto per costruire una relazione che la metta in grado di fornire previsioni anche su dati con output ignoto (*apprendimento supervisionato*).

Queste operazioni non comportano però, come invece accade con gli indici statistici, trasformazioni matematiche tali da creare grandezze più “espressive”, che meglio rappresentino l’informazione cercata, dai dati grezzi; piuttosto cercano in essi qualunque eventuale relazione, senza attribuirle o ricercarne alcun valore semantico, da utilizzare in base all’efficacia, e che contribuisca a costruire l’output desiderato.

Ne consegue che il compito, opportuno, di trattare i dati grezzi in modo da evidenziare eventuali caratteristiche utili che altrimenti potrebbero sfuggire ai meccanismi del classificatore, rappresenta un anello della catena di calcolo capace di fornire un contributo anche decisivo alle prestazioni globali.

5.1 Filtraggio

Questo trattamento consiste concretamente nel sottoporre le immagini provenienti dalle fasi precedenti ad una ulteriore azione di filtraggio, scelta e calibrata in base alle esperienze effettuate durante l'attività di ricerca.

Spesso questo filtraggio è svolto contestualmente ad altri passi del processo globale, tipicamente in fase di registrazione, e consiste in un qualche tipo di *smoothing* per fornire la necessaria uniformità richiesta dalle trasformazioni spaziali più complesse; altre volte, effettuata la registrazione, quella trasformazione di smoothing non viene poi effettivamente applicata, ma sostituita con un filtro diverso; altre ancora, infine, i filtri si sommano.

Raramente, comunque, si rinuncia all'introduzione di qualche filtro, la cui scelta è condizionata dalle scelte sul classificatore, e ne condiziona, a sua volta, la calibrazione dei parametri: nel complesso questo si risolve in un miglioramento globale delle prestazioni; non fornisce, però, informazioni definitive sui criteri generali da adottare. E non consente, soprattutto, di sperimentare vie diverse nello stesso esperimento, se non al costo di una moltiplicazione "esplosiva" delle caratteristiche, con costi computazionali che possono divenire inaccettabili.

D'altronde, un filtro capace anche di ottime prestazioni globali può facilmente mascherare caratteristiche utili evidenziabili da filtri diversi: l'opzione più opportuna, allora, sembra essere quella di non rinunciare all'ausilio dei filtri, ma piuttosto di munire il processo globale di un potente algoritmo di selezione delle caratteristiche utili, e utilizzare quindi (separatamente) i filtri che hanno dimostrato più efficacia.

È questa la scelta adoperata dagli autori. Si nota, per inciso, la coerenza di questo approccio con l'intervento pressoché *minimo* operato in fase di de-noising.

Benché si disponga oggi di un ventaglio pressoché infinito di possibilità, considerazioni computazionali e di efficacia hanno condotto alla selezione di sei tipi differenti di filtro, brevemente descritti qui di seguito (con la dicitura "nullo" si indica la mancanza di filtraggio), ognuno definito in intorni di diverse, e tipiche, estensioni.

Tipo di filtro	Intorni (voxel)				
	3×3×3	5×5×5	7×7×7	9×9×9	11×11×11
Nulla	—	—	—	—	—
Gaussiano	•	•	•	—	—
Media	•	•	•	—	—
Entropia	•	•	•	—	—
Intervallo	•	•	•	—	—
Deviazione std.	•	•	•	—	—
Sombrero	—	—	—	•	•

Filtro di media e filtro gaussiano

Sono i classici filtri di smoothing, e sono stati già descritti relativamente agli algoritmi generali di registrazione. Interessa qui notare, però, che nonostante la somiglianza “visiva” relativa al *tipo* di intervento, il loro comportamento è sostanzialmente diverso nel dominio delle frequenze, dove il filtro di media mostra sensibili oscillazioni nella sua risposta, assenti invece in quello gaussiano, la cui risposta in frequenza è ancora gaussiana, come evidente in figura 5.1, garantendone così un migliore effetto passa-basso. Questa ra-

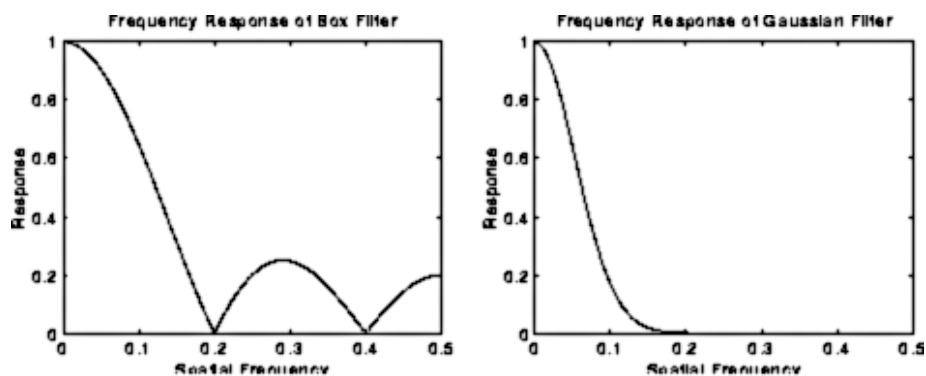


Figura 5.1: *Comportamento dei filtri di media, a sinistra, e gaussiano, a destra, nel dominio delle frequenze spaziali.*

gione è sufficiente per far preferire quest'ultimo nella maggior parte delle applicazioni pratiche.

Filtro ad “intervallo di variabilità”

Con questo filtro il valore di intensità di ogni voxel viene sostituito dalla differenza tra i valori massimo e minimo (intervallo di variabilità o *range*) riscontrati nel proprio intorno. È utile per trovare regioni di transizione come bordi e contorni.

Filtri a deviazione standard ed entropia

Il filtro a deviazione standard calcola questa grandezza sui voxel dell'intorno in esame, e assegna il valore calcolato al punto centrale dell'intorno; l'immagine risultante è in grado di evidenziare anch'essa alcuni tipi di transizione, ed è più robusta rispetto agli artefatti.

Il filtro ad entropia si calcola in modo analogo, ma valutando la quantità di informazione con la $H = -\sum_{k=0}^{L-1} p(k) \log_2 p(k)$, dove i livelli di intensità possibili sono L , le probabilità $p(k)$ sono stimate dalle frequenze relative.

Filtro a sombrero

Detto più correttamente filtro “laplaciano del gaussiano” (*Laplacian of Gaussian*, LoG), ma comunemente chiamato sombrero per la forma del suo grafico, si ottiene formalmente calcolando appunto il laplaciano di una gaussiana, e quindi ha un kernel del tipo $f(t) = K \cdot \left(1 - \frac{t^2}{\sigma^2}\right) e^{-\frac{t^2}{2\sigma^2}}$ (qui scritto per semplicità in una dimensione).

Le proprietà del laplaciano di individuare contorni, come le derivate prime, ma in maniera isotropica, diversamente da queste, vengono così combinate con quelle di un filtro gaussiano: mentre quest'ultimo riduce drasticamente l'intensità delle strutture a scale molto più piccole di σ , il primo evidenzia i contorni, ma senza più soffrire delle rapide transizioni causate dal rumore e alle quali è molto sensibile. Ai contorni è assegnato valore zero, e alle due regioni da essi separate valori negativi da un lato e positivi dall'altro: per mantenersi nell'intervallo positivo caratteristico delle immagini, i valori vengono spostati, rispettivamente, verso il grigio intermedio, il nero

e il bianco.

L'intervento dei filtri descritti sulle scansioni MR del cervello si possono osservare con chiarezza nella figura 5.2, dove sono applicati ad una sezione sagittale del volume contenente l'ippocampo, la formazione anatomica più significativa osservata nel lavoro.

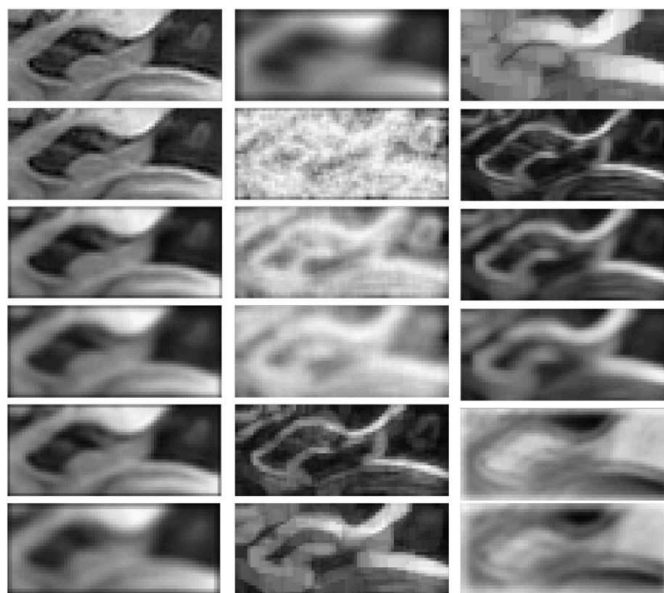


Figura 5.2: Applicazione dei filtri utilizzati nell'algoritmo ad una sezione sagittale dell'ippocampo. Nella colonna di sinistra, dall'alto: nullo (immagine originale), gaussiano $3 \times 3 \times 3$, $5 \times 5 \times 5$, $7 \times 7 \times 7$, media $3 \times 3 \times 3$ e $5 \times 5 \times 5$. In quella centrale: media $7 \times 7 \times 7$, entropia $3 \times 3 \times 3$, $5 \times 5 \times 5$, $7 \times 7 \times 7$, range $3 \times 3 \times 3$, $5 \times 5 \times 5$. A destra: range $7 \times 7 \times 7$, deviazione standard $3 \times 3 \times 3$, $5 \times 5 \times 5$, $7 \times 7 \times 7$, sombrero $9 \times 9 \times 9$, $11 \times 11 \times 11$.

Ognuno dei volumi di interesse, estratto per ogni soggetto dalla relativa MRI, viene quindi filtrato separatamente da 17 filtri diversi: i valori ottenuti per ogni voxel costituiscono 17 nuove caratteristiche che si aggiungono al valore di intensità originale, indicato come "filtro nullo", per un totale di 18 caratteristiche per voxel. Il numero totale di caratteristiche che si ottiene, detti V_z il numero di voxel del volume z e f il numero di caratteristiche per

ogni voxel, è:

$$F = f \times \sum_{z=1}^Z V_z ,$$

numero che raggiunge largamente l'ordine di 10^6 . E questo nonostante si sia scelto di sottoporre ad analisi, invece dell'intero cervello, solo cinque volumi di interesse per emisfero.

5.2 Scrematura dei dati

Idealmente il passo successivo da compiere, una volta *estratti* dai dati gli attributi, o caratteristiche, potenzialmente idonei a valutare la classe di appartenenza (sano o malato) di ogni immagine, consisterebbe nell'addestramento ed uso dell'algoritmo designato per la classificazione finale: la macchina a vettori di supporto, o SVM.

Il numero di caratteristiche estratte è però troppo elevato per garantire il miglior funzionamento della SVM, che finirebbe con l'avere così un addestramento molto più lento e difficile; inoltre si è visto [70] che l'eliminazione dall'input dei dati irrilevanti ne aumenta anche l'affidabilità delle prestazioni.

A differenza di altri lavori, si è evitato di operare ulteriori selezioni in base a conoscenza a priori o considerazioni euristiche, e soprattutto si sono evitate scelte rigide: piuttosto è sembrato opportuno ricorrere ad un nuovo strumento di *intelligenza artificiale* e *data mining*, capace di operare una selezione decisa ed affidabile, nonché completamente automatica, delle caratteristiche in input.

Si tratta di un classificatore che rappresenta una recente ed originale evoluzione dei cosiddetti "alberi di decisione".

5.2.1 Alberi di decisione

L'approccio più spontaneo al problema di decidere la classe di appartenenza di un vettore di caratteristiche, per mezzo delle quali si vuole pervenire alla risposta corretta, è esaminare ordinatamente il valore di ognuna di esse, alla luce della conoscenza acquisita, finché non si è in grado di assegnare la classe, o fino all'esaurimento del vettore.

Con *alberi di decisione* si intende la classe di algoritmi di decisione che procede eseguendo, sul vettore di caratteristiche (o attributi) in ingresso, una sequenza di test successivi: ognuno di essi, detto *nodo*, considera il valore di una delle proprietà disponibili, e determina in uscita una serie di risposte, una per ogni possibile risultato del test; ogni diramazione uscente dal nodo, detta *ramo*, può fornire già il risultato finale, quando cioè raggruppa vettori di una sola classe, e in tal caso il processo termina e si dice che il ramo conduce ad una *foglia*, oppure può portare ad un nuovo nodo, in corrispondenza del quale verrà testato un nuovo attributo, originando così nuovi rami.

Il processo termina quando il percorso seguito dal vettore raggiunge una foglia, ossia una *decisione*, oppure quando non ci sono più test disponibili (ogni attributo può essere usato una sola volta): in questo caso il risultato finale resta indeciso, e occorre qualche criterio aggiuntivo per definire la foglia.

Definita così la struttura generica dell'albero, si deve ora costruire la procedura di *addestramento*, ossia il passaggio da un insieme di vettori dalla classificazione nota, il cosiddetto *training set*, all'albero vero e proprio nella sua versione operativa, pronta a classificare vettori ignoti. Per semplicità si supponga che tutti gli attributi assumano, come è peraltro vero nelle immagini digitali, solo un numero finito di valori discreti, e che l'azione del test sia semplicemente quella di instradare in uno stesso raggruppamento tutti i vettori con identico valore della caratteristica sotto esame.

Una semplice soluzione al problema della costruzione di un albero di decisione, usando esclusivamente l'insieme di addestramento (training set), e cercando nel contempo una struttura il più possibile semplice, si può ottenere come segue [71] (la ricerca dell'albero "più piccolo" costituisce in realtà un problema in generale intrattabile, tuttavia un approccio di tipo euristico, come quello che descriviamo, porta a soluzioni ragionevolmente semplici):

- si effettua dapprima il test su un singolo attributo in tutto l'insieme di addestramento, e questo per ognuno degli attributi a disposizione: esaminando i raggruppamenti generati, quello che fornisce il miglior risultato, per esempio producendo subito da solo il maggior numero

di classificazioni corrette (raggruppamenti di vettori tutti della stessa classe), verrà scelto per costituire il primo nodo, o *radice* dell'albero;

- analoga procedura verrà eseguita sui restanti attributi per selezionare, per ogni ramo generato in precedenza, il miglior test di secondo livello;
- si prosegue così fino a completare l'intero albero.

Entrando più in dettaglio, e considerando da ora in poi solo alberi di decisione di tipo *booleano*, ossia con classificazione *finale* binaria (vero-falso, positivo-negativo, ecc.), si ha che ad ogni nodo possono verificarsi quattro eventualità:

- gli esempi da esaminare sono solo di un tipo (tutti positivi o tutti negativi), quindi si è raggiunta la classificazione e non resta che contrassegnare il nodo direttamente come foglia;
- gli esempi sono sia positivi che negativi, e sono disponibili attributi per un nuovo test: si sceglie l'attributo che meglio suddivide gli esempi, in base ad un preciso criterio prefissato, si esegue il test, e si procede con la costruzione dell'albero;
- non è presente alcun esempio, quindi la particolare combinazione di valori delle variabili lungo il percorso effettuato non è stata osservata nell'insieme di addestramento: viene assegnata una classificazione di default in base alle numerosità positivo-negativo del nodo genitore;
- gli esempi sono sia positivi che negativi, ma non sono disponibili altri attributi: abbiamo perciò esempi diversi che corrispondono ad identiche descrizioni in termini di attributi, e in tal caso la soluzione più semplice è considerare il risultato presente in maggioranza.

Il procedimento appena descritto cerca di limitare il numero dei test necessari per raggiungere la classificazione, pur mantenendo la massima efficacia: naturalmente nulla ci assicura che questo obiettivo venga raggiunto, né che la classificazione finale sia completa. Risposte parziali a queste domande possono essere fornite solo da verifiche sperimentali: è evidente a questo punto che l'importanza rivestita dalla fase di test supera la semplice misura delle prestazioni.

Problema del sovradattamento

Va ora considerato un problema tipico degli alberi di decisione, e che interessa in particolare i nostri scopi.

L'ultima eventualità prima menzionata riguarda la compresenza su un ramo terminale di istanze appartenenti a classi diverse, ma con identico vettore di caratteristiche: questo può accadere perché alcuni dati non sono corretti, e in tal caso si parla di *rumore* nei dati; oppure perché gli attributi non contengano tutta l'informazione necessaria a definire il problema, e quindi va migliorata o ampliata la loro definizione ed estrazione; infine il problema stesso può avere natura *non deterministica*.

Nel nostro caso, vettori identici ma di classe di appartenenza diversa indicano comunemente un difetto di informazione, al quale si può rispondere, in mancanza di nuovi attributi, assegnando alla foglia la classe più probabile, ossia la più numerosa ivi presente.

Quando però il numero di caratteristiche in input è elevato, può accadere che l'algoritmo riesca a costruire comunque un albero, ma che classifichi anche in base ad attributi del tutto estranei al problema: l'alta numerosità dei nodi può facilmente portare a test dal risultato corretto anche su caratteristiche indifferenti, per effetti casuali.

Siamo allora di fronte al cosiddetto problema, a cui si è già accennato, del *sovradattamento*, o *overfitting*, che colpisce in diversa misura ogni algoritmo ad apprendimento: la soluzione che cerchiamo ne deve essere il più possibile immune, essendo essa stessa deputata al compito di eliminare le caratteristiche non, o poco, significative.

Una tipica tecnica per affrontare la questione potrebbe essere la *potatura* dell'albero, ovvero la rimozione dei rami nati da attributi non o poco rilevanti, ma il problema della selezione dei rami da potare può essere delicato, mentre i fini medico-diagnostici che ci proponiamo di raggiungere richiedono piuttosto strumenti automatici, affidabili e robusti.

Guadagno di informazione

Un'altra questione tipica da affrontare nella costruzione di alberi di decisione è la scelta dell'attributo cosiddetto "migliore", ovvero di quell'attributo

che, in una determinata fase, possiede la miglior capacità classificatoria.

Sostanzialmente si tratta di provare tutti i test a disposizione, e vedere quale tra essi classifica meglio: a volte la distinzione è netta, e basta utilizzare quello che produce il maggior numero di risultati corretti; a volte però occorre stabilire un criterio più generale, in grado di “pesare” con precisione la bontà di ogni test.

Una risposta adeguata viene ancora una volta dalla *Teoria dell'Informazione* di Shannon [72], definendo l'informazione presente in un insieme in funzione delle probabilità q dei suoi singoli elementi come: $I = -\sum_i q_i \log_2 q_i$.

Se, nel semplice caso booleano che ci interessa, l'insieme da testare contiene p esempi positivi e n esempi negativi, l'informazione in esso contenuta si può scrivere, usando come stima di probabilità le frequenze relative:

$$I = -\frac{p}{(p+n)} \log_2 \frac{p}{(p+n)} - \frac{n}{(p+n)} \log_2 \frac{n}{(p+n)} .$$

Effettuando il test sull'attributo A avremo, supponendo che A possa assumere ν valori distinti, la suddivisione dell'insieme iniziale in ν insiemi disgiunti, uno per ogni esito del test: ognuno di essi conterrà p_i esempi positivi e n_i esempi negativi, quindi una informazione:

$$I_i = -\frac{p_i}{(p_i+n_i)} \log_2 \frac{p_i}{(p_i+n_i)} - \frac{n_i}{(p_i+n_i)} \log_2 \frac{n_i}{(p_i+n_i)} ,$$

che caratterizza quella diramazione. Sommando su tutte le diramazioni generate da A si otterrà:

$$I_A = \sum_{i=1}^{\nu} \frac{p_i+n_i}{p+n} \cdot I_i ,$$

dove $(p_i+n_i)/(p+n)$ rappresenta la “probabilità” di ogni ramo.

La quantità I_A è il *contenuto di informazione residuo* dopo l'azione del nodo A : minore è il suo valore, maggiore è l'informazione *estratta* dall'azione di A e migliore è il suo contributo alla classificazione.

Per vedere se il concetto di quantità di informazione corrisponde ai requisiti richiesti dall'intuizione per definire una misura della capacità di classificazione, ne esaminiamo il comportamento nei casi estremi.

Nel caso ideale in cui A completi già la classificazione, avremo che ognuna delle ν diramazioni conterrà solo esempi di un tipo, cioè $p_i+n_i = p_i$ oppure

$p_i + n_i = n_i$, o anche nessun esempio: in ogni caso i termini del generico I_i avranno probabilità nulla o, se non nulla, avranno nullo il logaritmo, e quindi sarà $I_i = 0 \forall i$, e in totale $I_A = 0$, confermando che tutta l'informazione disponibile nei dati è stata estratta.

Nel caso peggiore, invece, ogni raggruppamento conterrà le stesse proporzioni positivo-negativo dell'insieme di partenza, cioè sarà rispettivamente:

$$\frac{p_i}{(p_i + n_i)} = \frac{p}{(p + n)} \quad \text{e} \quad \frac{n_i}{(p_i + n_i)} = \frac{n}{(p + n)},$$

e l'informazione I_A sarà identica a quella I dell'insieme di partenza, indicando che A non è stato in grado di contribuire alla classificazione.

A questo punto è sufficiente definire il *guadagno di informazione* ottenuto dal test dell'attributo A con:

$$G(A) = I - I_A$$

per ottenere la valutazione quantitativa della capacità di classificazione cercata: si può così definire come “miglior” attributo quello col maggior guadagno tra quelli disponibili.

Punti di divisione

Un ultimo aspetto da considerare riguarda gli attributi a valori continui o interi.

Il caso che qui interessa riguarda sempre attributi con numero finito di valori discreti, potenzialmente in grado di generare ad ogni test un numero di rami diversi uguale all'intera gamma di valori disponibili: spesso questa elevata numerosità risulta di nessuna utilità, mentre il numero di ramificazioni che realmente occorre nel problema è molto minore, a volte ridotto solo a due, come è il caso del classificatore di cui si farà uso per la selezione dei dati.

Si tratta allora di trovare uno (o più) opportuno *punto di divisione*, o *split point*, che suddivida le uscite in due classi, e tale che il guadagno di informazione che ne risulta sia massimo.

Questa operazione, effettuata in fase di messa a punto dell'algoritmo, è concettualmente semplice, ma può diventare facilmente uno degli aspetti

computazionalmente più costosi del problema generale: per la ricerca di buoni punti di divisione si fa allora ricorso a metodi di programmazione dinamica, che si rivelano particolarmente efficienti nella ricerca di soluzioni adeguate.

5.2.2 Algoritmo Random Forest

Da quanto visto, si intuisce come gli alberi di decisione rappresentino una tecnica di classificazione estremamente rapida, una volta che l'albero è costruito, anche in presenza di un elevato numero di attributi: ogni esempio, o istanza, da classificare deve subire al massimo un numero di test, in genere molto semplici, pari alla dimensione del vettore di caratteristiche che lo rappresenta, per raggiungere la sua classificazione.

Altro discorso è invece la fase di costruzione-addestramento, che può essere molto lunga e complessa, e richiedere una buona dose di conoscenza a priori. Inoltre, quando le dimensioni dei vettori sono elevate, si concretizza il rischio rappresentato dal sovradattamento, la cui soluzione mediante tecniche di potatura si vuole assolutamente evitare.

Una evoluzione brillante verso la soluzione di questi problemi è rappresentata dall'algoritmo *Random Forest* (RF) [73].

L'idea fondamentale della Random Forest si basa sulla costruzione di numerosi alberi di decisione, che da ora in poi chiameremo *alberi di classificazione*, operata con la scelta di alcuni parametri su base *casuale*: costruita così una intera *foresta*, il generico vettore in ingresso viene inviato a tutti gli alberi che la compongono, e classificato da ciascuno di essi.

In generale, naturalmente, le singole classificazioni non coincidono: si dice allora che ogni albero *vota* per un particolare risultato, e la classificazione complessiva offerta dalla foresta è quella che ottiene più voti.

Ogni albero viene costruito come segue [74][75]. Sia dato un insieme di addestramento composto da N esempi diversi, ognuno caratterizzato da M variabili o attributi, allora:

- si estraggono a caso N esempi *con reimmissione*, che costituiranno un *nuovo* insieme di addestramento (questa tecnica per la costruzione di insiemi diversi sugli stessi dati viene chiamata *bootstrap*);

- ad ogni nodo si stabilisce un test a due uscite: vengono considerati solo m attributi estratti a caso tra gli M disponibili, con $m \ll M$, e si procede come al solito scegliendo per il test quello col miglior impatto sulla classificazione. Il valore di m è fissato una volta per tutte e vale per la costruzione dell'intera foresta; la selezione casuale delle m variabili si rinnova ad ogni nodo;
- l'albero viene fatto crescere fino alla massima estensione possibile, e non viene eseguita alcuna potatura.

Effettuando una nuova estrazione con reimmissione di N esempi tra gli N disponibili, si ottiene un altro diverso insieme con cui si addestra un nuovo albero, e si prosegue così creando una foresta di tanti alberi quanti si ritiene opportuno. Di fatto, la tecnica del bootstrapping genera versioni multiple del gruppo di addestramento, tutte contenute nello stesso insieme iniziale.

È evidente come questo approccio consenta di costruire alberi che coinvolgono un grande numero di attributi [73][76], essendo ogni nodo costruito solo su una piccola parte di essi.

L'utilizzo dell'informazione disponibile nei dati iniziali è, in ogni albero, limitata, ma si accresce aggiungendo nuovi alberi alla foresta grazie ai ripetuti bootstrap, mentre le estrazioni casuali consentono di evitare quelle decisioni sulla potatura che si renderebbe necessarie col proliferare di rami e nodi tipico degli algoritmi ad albero.

Anche la velocità della Random Forest risulta particolarmente elevata.

Si dimostra [73] che la quantità di errori commessi nella classificazione dipende essenzialmente da due fattori: il primo è ovviamente la capacità come classificatore, o *forza*, di ogni singolo albero; il secondo, più sottile, è la probabilità di errore legata alla *correlazione* tra coppie di alberi, che deve essere la minore possibile. Il parametro m , ossia la quantità di variabili da estrarre ad ogni nodo, determina in maniera diretta sia la correlazione che la forza: è l'unico parametro regolabile a cui l'algoritmo è sensibile, e il suo valore ottimale va cercato nel giusto equilibrio tra i due fattori di errore. In genere tale valore è alquanto elevato.

Important features map

Una caratteristica del classificatore Random Forest molto importante, e decisiva nello sviluppo della metodologia di classificazione diagnostica automatica mediante immagini, ha origine proprio dall'estrazione con reimmissione iniziale: essa infatti seleziona, per la costruzione di ogni albero, due terzi circa dei dati reali a disposizione, lasciandone fuori circa un terzo, disponibili quindi come dati di test.

Con questi dati, detti *out-of-bag* data o semplicemente *oob*, è possibile stimare, oltre l'errore commesso, anche la rilevanza di ogni attributo ai fini della classificazione [77], scopo principale di questa fase di elaborazione.

Fissata l'attenzione sull' i -esimo attributo, si procede in questo modo:

- si immettono in un albero classificatore i suoi esempi *oob* e si contano le classificazioni corrette;
- si permutano casualmente i valori della i -esima variabile tra tutti gli esempi *oob*, e si inseriscono nell'albero così modificati per classificarli nuovamente; si conta la nuova quantità di successi;
- si calcola la differenza tra i successi con i dati *oob* originali e quelli coi valori i -esimi permutati;
- si ripete per ogni albero della foresta;
- la media su tutti gli alberi delle differenze così calcolate fornisce una misura dell'importanza dell'attributo i -esimo.

L'informazione sull'importanza può essere quindi usata utilmente per selezionare gli attributi più indicati per la classificazione [78].

È comodo, a questo punto, costruire una mappa che riporta, per ogni attributo, la misura di importanza relativa fornita dall'algoritmo RF. Tale mappa, detta *important features map* (IFM), consente di individuare direttamente le caratteristiche più rilevanti stabilendo semplicemente un valore di soglia: gli attributi sotto tale valore nella IFM possono quindi essere scartati.

Si è verificato [55], costruendo un grafico delle prestazioni in funzione della soglia¹, che l'eliminazione degli attributi meno significativi può tranquillamente raggiungere la quota del 95% senza alcun deterioramento osservabile nella prestazione: per osservare un chiaro peggioramento bisogna arrivare a quote ben più alte, come ad esempio il 98%.

Va rilevato che questo parametro di soglia non risulta comunque critico.

Mantenendosi in un limite di tutta sicurezza, si può ottenere una riduzione delle caratteristiche da circa 10^6 a circa 10^4 , con un guadagno di ben due ordini di grandezza sulla dimensione dell'input per lo stadio seguente.

Un uso interessante della IFM si ha sovrapponendola direttamente, nella parte che riporta l'importanza relativa dei voxel, all'immagine MR che ha fatto da riferimento nella fase di registrazione (template MNI): vengono così rappresentate in modo diretto le regioni anatomiche effettivamente interessate dalla malattia, così come risultano all'algoritmo RF.

Si evidenzia, infine, che la Random Forest risulta particolarmente efficace in questo compito di *scrematura* dei dati proprio quando sono presenti numerose caratteristiche che risultano tra loro correlate [78], come appunto ci si aspetta dalle immagini tridimensionali rappresentanti strutture anatomiche.

5.3 Classificazione

La massiccia scrematura dei dati effettuata dalla Random Forest consente ora di sfruttare al meglio la potenza della Support Vector Machine, già descritta a proposito della Voxel Based Morphometry, a tutt'oggi il classificatore di elezione nella diagnosi di AD con neuroimmagini.

La combinazione di RF con bootstrap e SVM non è nuova in letteratura, per quanto molto recente (per esempio: [79][80][81]), ed è stata sperimentata con diverse strategie di selezione delle caratteristiche, sempre con un miglioramento delle prestazioni rispetto alla semplice SVM.

Addestramento e validazione² del classificatore SVM vengono eseguiti mediante una procedura di validazione incrociata su venti raggruppamenti

¹ I due gruppi presi in considerazione per il grafico sono costituiti, rispettivamente, da pazienti conclamati di AD e da controlli.

² Col termine "validazione" si intende qui la misura della prestazione mediante test.

(*20-fold cross-validation* [82]): si suddividono le N istanze già classificate di cui si dispone in venti raggruppamenti di $m \approx N/20$ elementi; si esclude il primo di questi gruppi, che farà da insieme di test, e si addestra il classificatore con gli $N - m$ esempi restanti; si testa con gli m elementi esclusi la macchina appena addestrata; si riesegue la procedura escludendo il secondo raggruppamento e addestrando con i rimanenti (compreso il primo); si procede allo stesso modo con tutti gli altri raggruppamenti.

Questo modo di operare, tipico dell'apprendimento supervisionato, si rende necessario nei casi dove l'insieme dei dati disponibili non è abbastanza ampio da consentire la costruzione di un insieme di test separato: tutta la conoscenza disponibile è nei dati, e rinunciare a parte di essi significherebbe perdere capacità di addestramento. La validazione incrociata affronta il problema suddividendo l'insieme in modo tale che ogni dato venga utilizzato una volta come test, e le altre volte nell'addestramento, mantenendo in quest'ultimo ruolo sempre la maggior parte degli elementi. Il caso estremo prevede un solo elemento di test e tutti gli altri di addestramento, e la procedura viene ripetuta tante volte quanti sono gli elementi: si parla in questo caso di procedura *leave-one-out*.

Anche con la SVM è necessario affrontare il problema del sovradattamento: il numero di caratteristiche originario, F , è stato ridotto con la scrematura ad \bar{F} , che è inferiore di circa due ordini di grandezza; tuttavia è ancora $\bar{F} \gg N$, ed effetti di sovradattamento, a cui questo classificatore non è immune, possono ancora verificarsi.

Questo problema viene risolto ricorrendo ancora ad una tecnica di bootstrapping: viene addestrato un numero n_{SVM} di macchine SVM, ognuna delle quali utilizza solo $\frac{N-m}{2}$ caratteristiche estratte a caso tra le \bar{F} selezionate dalla RF, mentre l'addestramento avviene con tutti gli $N-m$ elementi previsti dalla validazione incrociata, evitando così ogni possibile sovradattamento. Una media delle uscite (dicotomiche) fornite dalle n_{SVM} macchine su ognuno degli m elementi di test fornisce la sua classificazione finale, sotto forma di un *indice di classificazione* (CI).

Naturalmente il numero di attributi in ogni SVM è così troppo ridotto, esattamente di un fattore $\frac{\bar{F}}{(N-m)/2}$, per cui il numero n_{SVM} va scelto in modo da introdurre nel processo la maggior parte delle \bar{F} caratteristiche in gioco;

questo lo si ottiene scegliendo:

$$n_{\text{SVM}} = 1,2 \times \frac{2\bar{F}}{N - m}$$

In figura 5.3 è illustrato l'intero processo descritto in questo capitolo: si noti che le fasi di calcolo degli attributi mediante filtraggio, e di scrematura mediante RF, sono idealmente inserite, per chiarezza, all'interno dello schema di cross-validation.

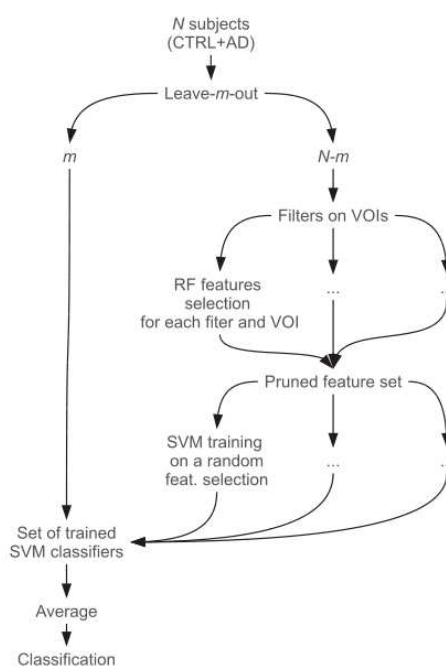


Figura 5.3: *Processo completo di preparazione delle caratteristiche e classificazione. Filtraggio e scrematura sono qui inserite nello schema di cross-validation.*

Va fatta in ultimo un'osservazione su un possibile errore sistematico presente nella procedura globale: i volumi di interesse selezionati durante il pre-processing sono stati delineati per ottenere la registrazione più accurata possibile in base alle particolarità anatomiche, e sono per lo più limitrofi: hanno quindi, inevitabilmente, delle zone di parziale sovrapposizione. Dopo la scrematura delle caratteristiche, l'esame della IFM mostra ancora una sovrapposizione di voxel di circa il 10%, che si concretizza in un piccolo grado

di ridondanza nei dati: così come è stato costruito l'algoritmo, però, questa ridondanza può solo aumentare lievemente la probabilità di alcuni attributi di essere estratti nella costruzione delle SVM, senza alcuna ripercussione pratica sul loro funzionamento.

Il processo di classificazione è così completo, e con esso l'intero algoritmo per la diagnosi.

Capitolo 6

Riepilogo e conclusioni

Fin dalla sua comparsa, la digitalizzazione delle immagini in campo medico ha annunciato nuove prospettive nel settore della diagnostica con metodi automatici, generando idee e introducendo paradigmi originali nei campi della matematica applicata e dell'informatica.

Il contemporaneo svilupparsi di nuove tecnologie per l'acquisizione di immagini per la diagnosi medica, unito all'imponente crescita dell'elettronica e della scienza dei calcolatori, ha portato quindi a rinnovate sinergie tra la Fisica e la Medicina che, nel campo delle neuroimmagini, hanno trovato uno spazio di ricerca ideale per le competenze del fisico.

Raffinando i processi di analisi delle scansioni di risonanza magnetica, si è potuta verificare la consistenza statistica dell'ipotesi di una possibile diagnosi per neuroimmagini, che ha trovato nella lotta alla malattia di Alzheimer un obiettivo di grandi rilevanza sociale ed interesse scientifico.

La comparsa degli studi sull'intelligenza artificiale ha poi aperto nuovi orizzonti, fornendo strumenti dedicati al problema della decisione automatica: la diagnosi per neuroimmagini si è così trasformata da semplice ipotesi a procedura di tipo operativo, stimolando e rinnovando l'interesse ed il lavoro dei ricercatori.

Le procedure oggi all'apice della ricerca si indirizzano su diversi filoni, tutti orientati alla valutazione dell'atrofia della materia grigia che si riscontra nei pazienti di AD.

La valutazione diretta dei volumi delle strutture anatomiche interessa-

te richiede procedure di allineamento e deformazione delle immagini molto elaborate, e riservate alla parte più esterna della corteccia cerebrale, che è l'unica che disponga di una morfologia idonea allo scopo: purtroppo essa è inadatta ad una diagnosi precoce, che invece è un interesse primario di questa ricerca.

Lo sviluppo e la sperimentazione si rivolgono perciò principalmente alle tecniche basate sull'intensità dei singoli punti di immagine, dalla quale dedurre la quantità di materia grigia da poi confrontare con un campione normale.

Questi approcci possono riguardare la globalità del cervello, oppure focalizzarsi su quelle sue strutture più e prima interessate dalla malattia: la principale di queste strutture è quella che i medici chiamano "ippocampo".

L'approccio globale dispone di molto tessuto cerebrale da analizzare, e consente anche un discreto riconoscimento di caratteristiche più locali, per cui con esso si può sottoporre l'immagine digitalizzata all'algoritmo classificatore, che è quello che emette l'ipotesi di diagnosi finale, in maniera anche quasi diretta.

L'approccio locale, invece, punta su una più precoce evidenza della malattia focalizzando l'attenzione sulle regioni più sensibili, ma i piccoli volumi anatomici interessati costringono ad elaborazioni molto più sofisticate per munire il classificatore di materiale sufficiente ad eseguire il suo compito.

Una idea promettente, nell'approccio locale, consiste nel deciso ampliamento delle caratteristiche da analizzare mediante un ampio ed accurato uso di filtri, aggiungendo inoltre un secondo diverso algoritmo di classificazione col compito di selezionare il materiale su cui agirà il classificatore finale.

Un confronto tra questi metodi, purtroppo, è problematico, non tanto per la diversità delle fonti dei dati, tutte comunque di qualità elevata, quanto per la diversità dei protocolli sperimentali seguiti, e il difficile paragone tra tutte quelle scelte che caratterizzano processi così complessi.

Un interessante lavoro in questo senso [83], comparso di recente in letteratura, tenta questo confronto replicando i procedimenti di altri gruppi ([45],[46],[47]) su materiale e protocollo comuni: qualche esempio dei risultati ottenuti è riportato in tabella, unitamente ai dati forniti nei riferimenti

originali; è riportato anche il dato originale relativo all'ultimo algoritmo ([55], di pubblicazione successiva) descritto in questa tesi.

	numerosità AD/CN	sensibilità %	specificità %
Klöppel (2008)	34 / 34	97,1	94,1
Cuignet (2011)	137 / 162	81	95
Vemuri (2008)	190 / 190	86	86
Cuignet (2011)	137 / 162	75	86
Magnin (2009)	16 / 22	91,5	96,6
Cuignet (2011)	137 / 162	75	93
Chincarini (2011)	144 / 189	89	94

In tabella: sensibilità (*malati ben classificati / totale malati*) e specificità (*controlli ben classificati / totale controlli*) dichiarati in lavori originali rappresentativi dei metodi visti, confrontati (prime tre coppie) coi risultati ottenuti da un gruppo indipendente che ne ha replicato gli esperimenti su dati comuni.

Le evidenti diversità riscontrabili confermano l'esigenza di stabilire procedure più omogenee e universalmente riconosciute, senza le quali i confronti numerici rischiano di non fornire ai ricercatori indicazioni sufficientemente utili.

Manca inoltre, nella maggior parte dei lavori, la prova più delicata: la classificazione e diagnosi tra i soggetti con la sindrome (detta MCI) che più denuncia una sospetta ma non conclamata malattia di Alzheimer. È chiaramente molto più difficoltosa della distinzione tra soggetti normali e malati conclamati, da tutti usata per testare le procedure, ma è proprio questa che rappresenta, dal punto di vista medico, l'obiettivo più importante e ambizioso per la ricerca.

In conclusione, allo stato attuale sono molti i risultati ottenuti: il processo di elaborazione dell'immagine in quanto segnale digitale, grazie alla quantità e qualità degli algoritmi dedicati al primo trattamento dei dati, fornisce già oggi risultati di qualità eccellente, ed in ulteriore continuo

perfezionamento.

La fase squisitamente diagnostica, o decisionale, ha mostrato prospettive di assoluto interesse, ma non ha invece ancora raggiunto le prestazioni desiderate, e richiede sicuramente ulteriori sforzi da parte della ricerca.

Molte sono però le idee in campo, ed il continuo prodursi di risultati incoraggianti garantisce, ai fisici impegnati in questa ricerca, nuovi ed importanti stimoli e obiettivi da raggiungere.

Bibliografia

- [1] Alzheimer's disease international, *World Alzheimer Report 2009*
- [2] Ferri, C.P. - Prince, M. - Brayne, C. e all. *Alzheimer's Disease International. Global prevalence of dementia: a Delphi consensus study*. Lancet 366, 2112–7, 2005
- [3] Ramani, A. - Jensen, J.H. - Helpren, J.A. *Quantitative MR imaging in Alzheimer disease*. Radiology 241, 2–44, 2006
- [4] Lambert, J.C. - Amouyel, P. *Genetic heterogeneity of Alzheimer's disease: Complexity and advances*. Psychoneuroendocrinology 32, 62–70, 2007
- [5] Kukull, W.A. - Ganguli M. *Epidemiology of dementia: concept and overview*. Neurol Clin Nov 2000, 18, 923–50, 2000
- [6] Alzheimer's Disease International, *World Alzheimer Report 2010. The Global Economic Impact of Dementia*. 2010
- [7] Irvine, G.B. - El-Agnaf, O.M. - Shankar, G.M. - Walsh, D.M. *Protein Aggregation in the Brain: The Molecular Basis for Alzheimer's and Parkinson's Diseases*. Mol. Med. 14(7–8), 451–464, 2008
- [8] Kidd, P.M. - *Alzheimer's Disease, Amnesic Mild Cognitive Impairment, and Age Associated Memory Impairment: Currentm Understanding and Progress Toward Integrative Prevention*. Altern. Med. Rev. 13(2), 85–115, 2005
- [9] Markesbery, W.R. - *Oxidative stress hypothesis in Alzheimer's disease*. Free Radical Biology & Medicine 23(1), 134–147, 1996
- [10] Castellani, R.J. - Lee, H. - Zhu, X. - Perry, G. - Smith, M.A. *Alzheimer's Disease pathology as a host response*. J. Neuropathol. Exp. Neurol. 67(6), 523–531, 2009

- [11] Mosconi, L. - Pupi, A. - De Leon, M.J. *Brain Glucose Hypermetabolism and Oxidative Stress in Preclinical Alzheimer's Disease*. Ann. N. Y. Acad. Sci. 1147, 180–195, 2008
- [12] Werner, P. - Korczyn, A.D. *Mild cognitive impairment: Conceptual, assessment, ethical, and social issues*. Clinical Interventions in Aging 3(3), 413–420, 2008
- [13] Chertkow, H. - Massoud, F. - Nasreddine, Z. *Diagnosis and treatment of dementia: 3.Mild cognitive impairment and cognitive impairment without dementia*. C.M.A.J. 178(10), 1273–1285, 2008
- [14] Satler Diniz, B. - Villela Nunes, P. - Yassuda, M.S. - Pereira, F.S. - Flaks, M.K. - Viola, L.F. - Radanovic, M. - Dutra de Abreu, I. - Borelli, D.T. - Gattaz, W.F. - Forlenza, O.V. *Mild cognitive impairment: cognitive screening or neuropsychological assessment?* Rev. Bras. Psiquiatr. 30(4), 316–21, 2008
- [15] Rosenberg, P.B. - Lyketsos, C.G. *Mild cognitive impairment: searching for the prodrome of Alzheimer's disease*. World Psychiatry 7, 72–78, 2008
- [16] Stephan, B.C.M. - Matthews, F.E. - Khaw, K.T. - Dufouil, C. - Brayne, C. *Beyond mild cognitive impairment: vascular cognitive impairment, no dementia*. V.C.I.N.D., 2009
- [17] Ries, M.L. - Carlsson, C. - Rowley, H. e all. *MRI characterization of brain structure and function in Mild Cognitive Impairment: a review*. J. Am. Geriatr. Soc. 56, 920–934, 2008
- [18] Tedeschi, G. - Cirillo, S. - Caltagirone, C. *Le neuroimmagini nelle demenze*. Critical Medicine Publishing Editore, 2005
- [19] Faulkner, W. *Basic Principles of MRI*. OutSource, Inc., 1996
- [20] Roth, C. *MR Safety*. OutSource, Inc., 1996
- [21] Garcia Sebastian, M.T. *Neuroimage Experimental Data Base Resources*. Grupo de Inteligencia Computacional, UPV/EHU GIC-RR-2009-03-03, 2009
- [22] Clifford, R.J.Jr. - Bernstein, M.A. - Fox, N.C. e all. *The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods*. J. Magn. Reson. Imaging. 27(4), 685–691, 2008
- [23] *Internet Brain Segmentation Repository (IBSR)*.
<http://www.cma.mgh.harvard.edu/ibsr/>

- [24] Marcus, D.S. - Wang, T.H. - Jamie Parker, J. - Csernansky, J.G. - Morris, J.C. - Buckner, R.L. *Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults*. J. Cogn. Neurosci. 19(9), 1498–1507, 2007
- [25] Buckner, R.L. - Head, D. - Parker, J. - Fotenos, A.F. - Marcus, D. - Morris, J.C. - Abraham Z Snyder, A.Z. *A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume*. Neuroimage 23(2), 724–738, 2004
- [26] Talairach, J. - Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain*. 1988
- [27] Zhang, Y. - Brady, M. - Smith, S. *Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm*. IEEE Trans. Med. Imaging 20(1), 45–57, 2001
- [28] *Biomedical informatics research network*.
<http://www.nbirn.net/>.
- [29] *Neuroscience database gateway (ndg)*.
<http://ndg.sfn.org/>.
- [30] *European Alzheimer's Disease Consortium*, 2010
<http://www.eadc.info/>
- [31] *Harmonization of protocols for the manual segmentation of the hippocampus: an EADC-ADNI joint effort*. 2011
<http://www.hippocampal-protocol.net>
- [32] Tzourio-Mazoyer, N. - Landeau, B. - Papathanassiou, D. - Crivello, F. - Etard, O. - Delcroix, N. - Mazoyer, B. - Joliot, M. *Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain*. NeuroImage 15, 273–289, 2002
- [33] *Laboratory of neuro imaging (loni), ucla*.
<http://www.loni.ucla.edu/>
- [34] *International Consortium for Brain Mapping (ICBM)*.
<http://www.loni.ucla.edu/ICBM/>

-
- [35] Mazziotta1, J. - Toga, A. - Alan Evans, A. e all. *A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM)*. Phil. Trans. R. Soc. Lond. 356, 1293–1322, 2001
- [36] Jain, A.K. *Fundamentals of digital image processing*. Prentice Hall, 1989
- [37] Eddy, W.F. - Fitzgerald, M. - Noll, D.C. *Improved image registration by using Fourier interpolation*. Magnetic resonance in medicine 36, 923–931, 1996
- [38] Gonzalez, R.C. - Woods, R.E. *Digital Image Processing*. Prentice Hall, 2008
- [39] Friston, K.J. - Ashburner, J. e all. *Spatial registration and normalization of images*. Human brain mapping 2, 165–189
- [40] Collins, D.L. - Neelin, P. - Peters, T.M. - Evans, A.C. *Automatic 3D Inter-subject Registration of MR Volumetric Data in Standardized Talairach Space*. Journal of Computer Assisted Tomography 18(2), 192–205, 1994
- [41] Vapnik, V.N. *Estimation of Dependences Based on Empirical Data*. Addendum 1, Springer-Verlag, New York, 1982
- [42] Cortes, C. - Vapnik, V. *Support-Vector Networks*. Machine Learning 20, 273–297, 1995
- [43] Han, J. - Kamber, M. *Data Mining*. Morgan Kaufmann Publishers, San Francisco, 2006
- [44] Boser, B.E. - Guyon, I. - Vapnik, V.N. *A training algorithm for optimal margin classifiers*. Proceedings of the Fifth Annual Workshop of Computational Learning Theory 5, 144–152, Pittsburgh, 1992
- [45] Klöppel, S. - Stonnington, C.M. - Chu, C. e all. *Automatic classification of MR scans in Alzheimer's disease*. Brain 131, 681–689, 2008
- [46] Vemuri, P. - Gunter, J.L. - Senjem, M.L. e all. *Alzheimer's Disease Diagnosis in Individual Subjects using Structural MR Images: Validation Studies*. Neuroimage 39(3), 1186–1197, 2008
- [47] Magnin, B. - Mesrob, L. - Kinkingnéhun, S. e all. *Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI*. Neuroradiology 51(2), 73–83, 2009
- [48] Lao, Z. - Shen, D. - Xue, Z. e all. *Morphological classification of brains via high-dimensional shape transformations and machine learning methods*. Neuroimage 21(1), 46–57, 2004

-
- [49] Tzourio-Mazoyer, N. - Landeau, B. - Papathanassiou, D. e all. *Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain*. Neuroimage 15, 273–289, 2002
- [50] Redner, R. - Walker, H. *Mixture densities, maximum likelihood and the EM algorithm*. SIAM Rev. 26, 195–239, 1984
- [51] Efron, B. - Tibshirani, R.J. *An introduction to the bootstrap*. Chapman and Hall, New York, 1993
- [52] Fan, Y. - Shen, D. - Gur, R.C. - Gur, R.E. - Davatzikos, C. *COMPARE: Classification of Morphological Patterns Using Adaptive Regional Elements*. IEEE Trans. on Medical Imaging 26(1), 93–105, 2007
- [53] Vincent, L. - Soille, P. *Watersheds in digital spaces: An efficient algorithm based on immersion simulations*. IEEE Trans. Pattern Anal. Mach. Intell. 13(6), 583–589, 1991
- [54] Grau, V. - Mewes, U.J. - Alcañiz, M. e all. *Improved watershed transform for medical image segmentation using prior information*. IEEE Trans. Med. Imag. 23(4), 447–458, 2004
- [55] Chincarini, A. - Bosco, P. - Calvini, P. e all. *Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease*. Neuroimage 58, 469–480, 2011
- [56] Simoncelli, E.P. - Freeman, W.T. *The Steerable Pyramid: A Flexible Architecture for Multi-Scale Derivative Computation*. 2nd Ann. IEEE Intl. Conf. on Image Processing, 1995
- [57] Castleman, K.R. - Schulze, M. - Wu, Q. *Simplified Design of Steerable Pyramid Filters*. Proceedings of the 1998 IEEE International Symposium 5, 329–332, 1998
- [58] Simoncelli, E.P. - Freeman, W.T. - Adelson, E.H. - Heeger, D.J. *Shiftable Multi-scale Transforms*. IEEE Trans. Information Theory 38(2), 587–607, 1992
- [59] Freeman, W.T. - Adelson, E.H. *The Design and Use of Steerable Filters*. IEEE Trans. Patt. Anal. and Machine Intell. 13(9), 891–906, 1991
- [60] Dosselmann, R. - Dong Yang, X. *A Formal Assessment of the Structural Similarity Index*. Technical Report TR-CS 2008-2, 2008

-
- [61] Piella, G. - Heijmans, H. *A new quality metric for image fusion*. 2003
- [62] Wang, Z. - Bovik, A.C. - Sheikh, H.R. - Simoncelli, E.P. *Image quality assessment: from error visibility to structural similarity*. IEEE Trans. Image Process. 13, 600–612, 2004
- [63] Wang, Z. - Bovik, A. *Mean squared error: love it or leave it? A new look at signal fidelity measures*. Signal Process. Mag. IEEE 26, 98–117, 2009
- [64] Mazziotta, L. *A probabilistic atlas of the human brain: theory and rationale for its development*. Neuroimage 2, 89–101, 1995
- [65] Lloyd, S.P. *Least Squares Quantization in PCM*. IEEE Transactions on Information Theory, 28, 2, 129–137, 1982
- [66] Arthur, D. - Vassilvitskij, S. *How Slow is the k-means Method?*. Proceedings of the 2006 Symposium on Computational Geometry, 2006
- [67] Vattani, A. *k-means requires exponentially many iterations even in the plane*. Proceedings of the 2009 Symposium on Computational Geometry, 2009
- [68] Calvini, P. - Chincarini, A. - Gemme, G. e all. *Automatic analysis of medial temporal lobe atrophy from structural MRIs for the early assessment of Alzheimer's disease*. Med. Phys. 36, 3737, 2009
- [69] Duchesne, S. *MICCAI brainstem segmentation protocol. From Statistical Atlases to Personalized Models (Workshop)*. MICCAI Society, Copenhagen, 2006
- [70] Guyon, I. - Vapnik, V. - Barnhill, J. - Stephen, W. *Gene selection for cancer classification using support vector machines*. Mach. Learn. 46, 389–422, 2002
- [71] Russel, S. - Norvig, P. *Artificial Intelligence. A Modern Approach*. Prentice Hall, 2003
- [72] Shannon, C.E. *Communication in the Presence of Noise*. Proceeding of the IRE 37, 10–21, 1949
- [73] Breiman, L. *Random Forest*. Machine Learning 45, 5–32, 2001
- [74] Breiman, L. *Random Forest – Random Features*. University of California, Berkeley, Technical Report 567, 1999
- [75] http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

- [76] Svetnik, V. - Liaw, A. - Tong, C. - Culberson, J.C. - Sheridan, R.P. - Feuston, B.P. *Random forest: a classification and regression tool for compound classification and QSAR modeling*. Journal of Chemical Information and Computer Sciences 43, 1947–1958, 2003
- [77] Bylander, T. *Estimating Generalization Error on Two-Class Datasets Using Out-of-Bag Estimates*. Machine Learning 48, 287, 2002
- [78] Archer, K.J. - Kimes, R.V. *Empirical characterization of random forest variable importance measures*. Computational Statistics & Data Analysis 52, 2249–2260, 2008
- [79] Ebina, T. - Toh, H. - Kuroda, Y. *DROP: an SVM domain linker predictor trained with optimal features selected by random forest*. Bioinformatics 27, 487–494, 2011
- [80] Larios, N. - Soran, B. - Shapiro, L.G. - Martinez-Munoz, G. - Lin, J. - Dietterich, T.G. *Haar Random Forest Features and SVM Spatial Matching Kernel for Stonefly Species Identification*. PR International Conference on Pattern Recognition. IEEE, 2624–2627, 2010
- [81] Waske, B. - Van Der Linden, S. *Classifying multilevel imagery from SAR and optical sensors by decision fusion*. IEEE Trans. Geosci. Remote Sens. 46, 1457–1466, 2008
- [82] Zhu, X. - Ambrose, C. - McLachlan, G.J. - *Selection bias in working with the top genes in supervised classification of tissue samples*. Stat. Method. 3, 29–41, 2006
- [83] Cuingnet, R. - Gerardin, E. - Tessieras, J. e all. *Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database*. NeuroImage 56, 766–781, 2011