

POLITECNICO DI TORINO

**Corso di Laurea Magistrale
in Ingegneria Biomedica**

Tesi di Laurea Magistrale

**Radiomica per Individuazione di Midollo
Osseo Emopoietico da Tomografia
Computerizzata**



Relatori

Prof. Gabriella Balestra
Prof. Samanta Rosati

Candidato
Armando De Chiara

Anno Accademico 2018/2019

Sommario

1. Introduzione	5
1.1. Stato dell'arte	6
1.1.1 MRI	6
1.1.2. PET	7
1.1.3. PET/CT e PET/MRI.....	9
1.2. Radiomica	11
1.3. Obiettivo del lavoro	11
2. Materiali e Metodi.....	12
2.1 Feature extraction	12
2.2. Costruzione del training set.....	13
2.3. Ottimizzazione dei classificatori – Algoritmo genetico	14
2.3.1. Decision Tree	15
2.3.2. K-Nearest neighbour	15
2.3.3. Neural network.....	16
2.3.4. Support vector machine	18
2.4. Funzione di fitness.....	19
2.5. Parametri dell'algoritmo genetico.....	19
2.5.1. Popolazione iniziale	19
2.5.2. Operatori genetici.....	20
2.5.3. Metodo di selezione	20
2.5.4. Condizioni di stop	20
2.5.5. Esecuzione dell'algoritmo genetico.....	20
2.6. Classificazione delle immagini	21
2.7. Post-processing.....	22
2.8. Valutazione dell'effetto della soglia SUV.....	23
2.9. Indici per la valutazione dei risultati.....	23
3. Risultati.....	26
3.1. Valutazione delle soluzioni dell'algoritmo genetico.....	26
3.3. Effetto del voting	29
3.2. Valutazione delle maschere	32
3.4. Postprocessing.....	41
3.5. Effetto della soglia SUV	48
3.6. Struttura LPBM e sue criticità.....	50
3.7. Valutazione della struttura LPBM con diversi training set	52

4. Conclusioni	57
Appendice.....	59
Bibliografia.....	64

1. Introduzione

Il midollo osseo è un tessuto molle disperso che occupa i canali delle ossa lunghe e la parte centrale delle ossa piatte, nelle cavità dell'osso spongioso. Si divide in due tipi di tessuto:

- Midollo osseo giallo; è costituito principalmente da tessuto adiposo, che ne determina il colore
- Midollo osseo rosso; detto anche ematopoietico o attivo, è un tessuto costituito da cellule ematopoietiche.

Il midollo osseo attivo è l'organo deputato alla produzione degli elementi del sangue, ovvero globuli bianchi, piastrine e soprattutto globuli rossi. La quantità di midollo osseo attivo presente è estremamente variabile da individuo a individuo e varia nel corso degli anni: alla nascita il midollo osseo è quasi interamente composto da midollo osseo rosso, ma col passare degli anni viene convertito in midollo osseo giallo, perdendo quindi la sua capacità di produzione di globuli rossi. Occasionalmente il midollo osseo giallo viene riconvertito in midollo osseo rosso, in particolari circostanze (gravi emorragie) in cui è necessario aumentare la produzione di elementi del sangue [3].

Sia la chemioterapia che la radioterapia contribuiscono a colpire e danneggiare sia le cellule ematopoietiche contenute nel midollo osseo attivo, dato che tali cellule sono caratterizzate da alta radiosensibilità[1]. Infatti, uno dei più tipici effetti collaterali legati alla tossicità di entrambe le terapie è l'anemia, in quanto la risultante minor quantità di midollo attivo diminuisce la capacità del corpo di produrre globuli rossi, assieme ad un maggior rischio di emorragie dovuto a piastrine basse e ad un generale indebolimento del sistema immunitario. Questa serie di effetti collaterali viene chiamata emato-tossicità. I danni causati dalla emato-tossicità sono particolarmente pericolosi e possono determinare l'interruzione delle terapie, con il rischio di rendere le terapie stesse meno efficaci e precludere la possibilità di guarigione. Inoltre tali danni possono rimanere cronicizzati, per cui in alcune zone del midollo osseo la produzione di globuli rossi e altre cellule rimane fortemente depressa anche una volta finita la terapia, riducendo la qualità di vita dei pazienti e rendendoli anemici e/o immunodepressi[2]. Essere in grado di ridurre l'impatto tossico di queste terapie sul midollo osseo è quindi importante sia per aumentare l'efficacia delle terapie stesse, sia per evitare danni permanenti per i pazienti.

Più di 50% del midollo osseo attivo in un adulto si trova tra pelvi e spina lombare; per questo il trattamento di tumori localizzati nei tessuti della zona bassa dell'addome (es. tumore anale, tumore cervicale, ecc.), causano spesso elevata emato-tossicità. I protocolli di trattamento utilizzati sono generalmente una combinazione di chemio e radioterapia, in grado di rilasciare elevate dosi di radiazione e farmaco nella regione pelvica, diminuendo la tossicità globale, ma colpendo il midollo osseo attivo localizzato nella medesima regione ed esponendo il paziente ai rischiosi effetti collaterali sopra citati[3].

Un metodo per ridurre gli effetti collaterali è di fare in modo di colpire con la radioterapia in maniera più precisa, evitando di irradiare con alte dosi di radiazione il midollo osseo attivo[4]. Questo permette da un lato di ridurre il danneggiamento dovuto ai protocolli di terapia radioterapica, riducendo sia l'emato-tossicità acuta sia eventuali danni cronici. La riduzione dei danni permanenti inoltre permette alle zone di

midollo osseo giallo di convertirsi in midollo attivo e contribuire a compensare la perdita di attività delle zone irradiate.

Attualmente, quando si pianifica radioterapia, si utilizzano immagini CT, le quali non sono in grado di rilevare (ad occhio nudo) la differenza tra midollo osseo attivo e midollo osseo giallo. Inoltre, pur essendo in grado di delineare il midollo osseo in maniera abbastanza precisa da poter essere usato nella pianificazione della radioterapia, la CT può causare la perdita di porzioni di tessuto che nella pianificazione vengono esclusi, rischiando di non irradiare in maniera precisa.

In questo lavoro abbiamo tentato di capire se sia possibile, sfruttando la radiomica, ovvero sfruttando le caratteristiche numeriche dei pixel delle immagini, di segmentare il midollo osseo attivo tramite CT.

1.1. Stato dell'arte

Attualmente esistono 3 metodi principali per l'individuazione del midollo osseo attivo: la biopsia, la MRI e la fPET.

Il metodo più preciso è ovviamente la biopsia: si estrae direttamente dal midollo osseo un campione e lo si analizza al microscopio per differenziare il tipo di midollo osseo. Si tratta tuttavia di un metodo impraticabile per la pianificazione di radioterapia, data la sua elevata invasività e l'impossibilità di avere informazioni spaziali (si ottengono informazioni solo riguardo alla regione di midollo in cui si è eseguita la biopsia)[5,11].

1.1.1 MRI

Un metodo utile all'individuazione di midollo osseo attivo è la risonanza magnetica (MRI). La principale differenza tra tessuto di midollo attivo e non attivo risiede nella quantità di grasso: il midollo osseo attivo contiene circa il 40% di tessuto adiposo, mentre nel midollo giallo questa percentuale è del 95%. Inoltre il midollo osseo attivo contiene un maggior numero di vasi sanguigni e capillari, che trasportano un abbondante quantità di sangue alle cellule ematopoietiche. Questo si traduce in una marcata differenza tra i due tessuti midollari in termini di quantità percentuale di grasso e di acqua.

La risonanza magnetica è una tecnica estremamente sensibile alle diverse proporzioni di acqua e grassi. Nei grassi i protoni sono infatti legati a gruppi CH₂, i quali permettono un tempo di rilassamento T₁ molto breve, risultante in un segnale intensità elevato nei tessuti adiposi. Viceversa, tali protoni hanno lunghi tempi di rilassamento T₂, che si traduce in un segnale molto debole di intensità, mentre i protoni legati a molecole d'acqua hanno elevata intensità di segnale. Per cui, nelle immagini T₁-weighted il tessuto adiposo appare molto brillante mentre le cavità più acquose appaiono più scure. Viceversa nelle immagini T₂-weighted appare più scuro il tessuto adiposo mentre appaiono più chiari i tessuti contenenti più acqua [6,11].

Grazie a questa sensibilità, diversi studi [6,7] hanno evidenziato la possibilità di segmentare semi-automaticamente il midollo osseo attivo.

Sfruttando le immagini T₁-weighted e T₂-weighted è anche possibile sviluppare un metodo di segmentazione del midollo osseo completamente automatico. Tale metodo, sviluppato da A.

Andreyvchenko et al., utilizza le sopracitate immagini T₁ e T₂ per calcolare il parametro di fat fraction (FF) [7]:

$$FF = \frac{fat\ image}{(fat\ image + water\ image)}$$

dove *fat image* e *water image* sono rispettivamente immagine T1 e immagine T2. Si tratta di un parametro che calcola la percentuale di grassi presente nel tessuto. Il contenuto in grassi è ciò che permette di distinguere i due tipi di midollo, per cui con una semplice doppia soglia del valore dei pixel è in grado di segmentare il midollo osseo attivo:

$$\text{if } S_{min} < P < S_{max}, \quad \text{then } P \in RM$$

Dove S_{min} e S_{max} sono le soglie, P è il valore del pixel e RM è il midollo attivo. Una volta segmentato il midollo osseo attivo, si è in grado di stimare anche la quantità percentuale.

Con questo metodo si è ottenuto una maschera di midollo osseo attivo che può essere utilizzata nella pianificazione di radioterapia, dopo un controllo di sicurezza da parte del radiologo, in grado di ridurre l'emotossicità della terapia stessa.

L'utilizzo della risonanza magnetica, oltre ad una elevata sensibilità e una elevata risoluzione spaziale, possiede anche il considerevole vantaggio di utilizzare radiazioni elettromagnetiche non ionizzanti, che non hanno alcun effetto dannoso sul paziente. Si tratta quindi del metodo di individuazione di midollo osseo attivo ideale.

Tuttavia l'utilizzo della risonanza magnetica, per le sue caratteristiche, da' informazioni di tipo strutturale, ovvero porta informazioni riguardo a come è strutturato il midollo osseo, ma non porta alcuna informazione riguardo al livello di attività del midollo osseo attivo. Questo si può tradurre in una certa misura di imprecisione, dato che possono essere presenti tessuti aventi all'incirca lo stesso valore di FF del midollo osseo attivo, ma che in realtà non possiedono attività ematopoietica e che non vengono scartati nemmeno dalle operazioni di post-processing dell'immagine. Certamente si tratta di un errore non molto importante ai fini della radioterapia, ma esiste e può avere entità non trascurabile, dato che limita la capacità di una precisa segmentazione delle zone tissutali effettivamente responsabili dell'ematopoiesi.

Inoltre, la MRI è un esame medico molto costoso e dispendioso in termini di tempo, e anche se è possibile usarlo nel workflow ospedaliero per la radioterapia, molto spesso sono macchine di difficile accesso.

1.1.2. PET

La tomografia ad emissione di positroni (PET) è una tecnica di imaging avente alta specificità e sensibilità, in grado di fornire informazioni quantitative riguardo ai processi biochimici fisiologici e non dei tessuti del corpo umano[8]. E' una tecnica inoltre in grado di produrre un segnale di imaging 3D direttamente legato alla quantità della fonte del segnale, e da questo deriva il suo grande valore clinico.

Esistono vari altri tipi di radiotraccianti utilizzabili per individuare le cellule del midollo osseo attivo, ognuno dei quali si basa su un diverso meccanismo di captazione [10]:

- Captazione da parte del reticolo endoteliale (RES) del midollo osseo attivo: un esempio è il SC (colloide zolfo-Tc-99m) utilizzato in scintigrafia. Tuttavia i radiofarmaci captati dal RES vengono assorbiti soprattutto a livello di reni e fegato, per cui solo una parte finisce effettivamente nel midollo osseo attivo.
- Captazione delle cellule eritropoietiche: un esempio è il citrato di Fe-52, il quale ha il vantaggio di non essere captato da reni o fegato, ma ha un costo proibitivo
- Captazione delle cellule granulopoietiche: si basano sul labeling di anticorpi o di cellule coinvolte nel sistema immunitario. Hanno lo svantaggio di essere complessi da produrre e possiedono alta captazione da parte di reni e fegato.

La tecnica più utilizzata per l'individuazione di midollo osseo attivo si basa sull'utilizzo del radioisotopo 2-fluoro-2-deossi-glucosio (FDG)[8,9], il quale viene captato dalle cellule emopoietiche in quantità maggiore rispetto al tessuto del midollo osseo giallo, in quanto analogo del glucosio, dato il maggior consumo di glucosio. L' FDG viene captato in quanto analogo del glucosio, ma non viene catabolizzato e rimane intrappolato all'interno della cellula nella forma FDG fino al decadimento del fosforo-18, che è il radioisotopo responsabile dell'emissione dei positroni. Siccome le cellule emopoietiche hanno un maggior consumo, ciò significa che captano maggiori quantità di FDG, per cui avranno un segnale di maggiore intensità. Una volta avvenuto il decadimento, la molecola si libera del fosforo in posizione 2, e la molecola di glucosio-6-fosfato viene metabolizzata normalmente, per cui non vi è alcun accumulo di biomolecole all'interno delle cellule.

L'utilizzo del FDG comporta una serie di vantaggi rispetto agli altri radiofarmaci: permette l'individuazione dei tessuti in base al grado di attività metabolica del midollo, e non in base alle caratteristiche strutturali, mentre altri farmaci mostrano un aumento di captazione nei tessuti del midollo funzionante e una minore captazione solo in presenza di lesioni [10].

Tramite questo radiotracciante è possibile quindi valutare la distribuzione di glucosio nel midollo osseo in maniera molto precisa e, data l'assenza di accumulo, in maniera molto pulita.

In letteratura vi sono vari studi che hanno dimostrato la possibilità di utilizzare le immagini di FDG-PET per la segmentazione automatica e semi-automatica del midollo osseo attivo. Queste tecniche si basano sul calcolo di un certo numero di parametri diversi [12].

Il più utilizzato di questi parametri è lo Standardized Uptake Value (SUV) , definito come la concentrazione di attività divisa per il peso del paziente e la quantità di attività iniettata. Viene utilizzato in praticamente tutte le implementazioni di metodi semi-automatici di visualizzazione e analisi d'immagine, data la facilità e la riproducibilità con cui si calcola. Ne esistono anche diverse varianti utilizzate:

- SUV_{mean} : pari alla media del valore di SUV in un volume di interesse.
- SUV_{max} : pari al valore massimo di SUV.
- SUV_{peak} : pari alla media del valore di SUV in un volume sferico di 1.2 cm di diametro. Rispetto a SUV_{mean} possiede una minore suscettibilità al metodo usato per la segmentazione del volume di interesse
- SUV_{LBM} : pari al valore di concentrazione di attività normalizzato con il valore di massa corporea magra del paziente. Rispetto alla SUV, non tiene conto della attività del tessuto adiposo, minimizzandola rispetto a quella di altri tessuti. E' una misura più stabile, meno variabile tra pazienti.

Altri parametri calcolabili da immagini FDG-PET sono la Total Lesion Glycolysis (TLG), il quale combina dati metabolici e dati volumetrici per ricavare una misura dell'attività cancerosa globale nel paziente e la Metabolic Tumor Volume (MTV), che combina valori di SUV_{mean} con valori volumetrici delle lesioni e li somma, in maniera da ricavare una misura dell'attività metabolica delle lesioni tumorali. Entrambi i parametri sono tuttavia non utilizzabili per l'individuazione di midollo osseo attivo [12].

Utilizzando la SUV, la segmentazione del midollo osseo attivo avviene tramite un semplice global thresholding, ovvero viene definito come midollo attivo il tessuto avente un valore SUV maggiore di un certo valore [4].

La FDG-PET è una tecnica che, rispetto alla MRI, genera informazioni legate all'attività metabolica del midollo osseo, e non legate alla struttura del tessuto stesso [12]. Questo è un indubbio vantaggio, dato che la segmentazione tramite immagine PET è più affidabile rispetto alla segmentazione tramite imaging legato a informazioni strutturali, oltre ad essere a queste complementari. Infatti la MRI fornisce informazioni solo sulla percentuale di tessuto adiposo nel midollo, ma non è in grado di individuare le zone di tessuto effettivamente attive.

Tuttavia la PET è una tecnica soggetta a molti problemi. Il primo problema riguarda le caratteristiche stesse dell'immagine. Solitamente la PET viene eseguita su tutto il corpo, per cui si ottengono immagini con informazioni sul consumo di glucosio di tutti i tessuti. Questo si traduce in un elevato rumore di fondo, che può rendere complicato un eventuale metodo automatico di segmentazione e può rendere difficile anche la corretta interpretazione dei risultati [12]. Inoltre alla PET mancano del tutto informazioni di tipo morfologico. Un altro problema è la bassa risoluzione spaziale, dovuta alla natura del segnale: il segnale fotonico viene infatti attenuato da effetti di assorbimento e di scattering. La bassa risoluzione spaziale, unito a algoritmi di ricostruzione di immagini non perfetti, si traducono in una quantificazione del radiofarmaco non ottimale, e alla presenza di vari effetti che pregiudicano la precisione delle immagini ottenute e al basso grado di standardizzazione dei risultati ottenuti con macchine e algoritmi diversi. Un esempio di effetto deleterio è il Partial Volume Effect (PVE) [12], per cui lesioni di ridotte dimensioni vengono tradotte in lesioni di dimensioni maggiori e intensità minore, oltre ad avere valori di contrasto minori. Di fatto, in sistemi PET si riescono ad avere risoluzioni spaziali di massimo 3-4 mm, e vengono correttamente visualizzate solo lesioni circa 2-3 volte maggiori della risoluzione massima. Il PVE può essere solo parzialmente corretto tramite l'impiego di alcuni algoritmi di ricostruzioni, ma rimane inevitabile.

Infine, la FDG-PET è un esame che necessita di radiofarmaci a bassa emivita, i quali possono essere prodotti solo tramite ciclotroni e, data la bassa emivita, una volta prodotti devono essere utilizzati immediatamente; macchinari in grado di eseguire PET devono essere necessariamente vicino a ciclotroni. Ciò contribuisce a ridurre grandemente la diffusione a livello territoriale, dato che solo pochi ospedali o università possono permettersi l'investimento necessario alla costruzione di un ciclotrone e i centri PET devono trovarsi nelle vicinanze, e ad aumentarne il costo; l'esame PET è un esame che in generale ha un costo molto maggiore rispetto alle altre tecniche di imaging.

1.1.3. PET/CT e PET/MRI

Per ovviare alla mancanza di informazioni morfologiche dell'imaging tramite medicina nucleare, si sono sviluppate delle tecniche che affiancano all'immagine PET, di tipo funzionale, una immagine delle caratteristiche strutturali del paziente. Questo tipo di tecnologia ibrida dovrebbe essere in grado di sopperire ai punti deboli delle singole tecnologie, fornendo allo stesso tempo delle informazioni più complete e dettagliate, facilitando lo sviluppo di metodologie e algoritmi automatici in grado di stimare la localizzazione e il volume del midollo osseo attivo. In letteratura si possono trovare diversi metodi sviluppati proprio sfruttando queste tecnologie ibride.

Un esempio è il metodo sviluppato da Perk et al. [13], chiamato QTMI. Questo algoritmo è in grado di sfruttare imaging da PET/CT in maniera tale da quantificare in maniera automatica l'attività totale del midollo osseo, sia globalmente (si prende in considerazione l'intero scheletro) che localmente (si prende in considerazione solo una regione dello scheletro). Si acquisiscono parallelamente una immagine da CT e un'altra da PET. L'immagine CT viene utilizzata per segmentare automaticamente le ossa dello scheletro, sfruttando un algoritmo di segmentazione "Atlas-based", ovvero un algoritmo che sovrappone all'immagine del paziente delle maschere di segmenti ossei pre-registrati che si deformano sino alla corretta

individuazione di tutte le regioni scheletriche. Sull'immagine PET vengono invece calcolati parametri di SUV, e le due immagini vengono moltiplicate, in maniera tale da evidenziare sull'immagine PET solo le regioni scheletriche. Infine viene applicato thresholding sul parametro SUV per distinguere il midollo osseo attivo. Una volta segmentato il midollo osseo attivo, è possibile calcolare tutti i parametri voluti per quantificare l'attività globale o regionale del midollo osseo.

Da letteratura si possono trovare anche altri esempi [9,14,15], ma sono tutti basati sul medesimo principio:

- 1. Acquisizione immagine CT e segmentazione delle ossa. Esistono numerosi metodi di segmentazione automatica, basati tecniche come level-set, procedure di soglia e altre.
- 2. Acquisizione immagine PET
- 3. Moltiplicazione delle immagini. Dalla moltiplicazione si ottiene una mappa dell'attività metabolica del midollo osseo
- 4. Individuazione del midollo osseo attivo e quantificazione

Nella tecnologia ibrida PET/CT, l'utilizzo della CT consente quindi di effettuare una precisa segmentazione delle ossa del corpo, sfruttando anche una maggiore risoluzione spaziale, che è una delle principali limitazioni della PET, oltre a poter essere utilizzata per attenuare la limitazione dovuto all'effetto di scattering fotonico. Dall'immagine CT, infatti, vengono estratti una serie di parametri correttivi, che consentono di correggere parzialmente gli effetti di scattering. Le correzioni di queste degradazioni sono generalmente già implementate nei software delle macchine: le immagini CT, infatti, forniscono una mappa dei coefficienti di attenuazione dei tessuti, direttamente legati al fenomeno di scattering.

Se da un lato si riesce ad avere una precisa segmentazione del midollo osseo, i macchinari PET/CT sono macchinari molto costosi, aventi tutti gli stessi problemi delle sole PET, e necessitano un importante investimento da parte degli ospedali, a cui si aggiunge un grande aumento delle dosi di radiazioni somministrate ai pazienti: sia CT che PET separatamente hanno un alto livello di radiazioni, e la tecnologia ibrida aumenta ulteriormente tale dose, diventando uno dei metodi d'indagine a più alto livello di radiazioni.

Accanto alla tecnologia PET/CT, si è sviluppata anche una tecnologia ibrida PET/MRI. Tale tecnologia unisce alla capacità tipiche della MRI di caratterizzare in maniera eccellente i tessuti molli con l'informazione funzionale dell'attività metabolica della PET [8]. Il principio di funzionamento nella PET/MRI è il medesimo visto per la PET/CT: si acquisisce una immagine tramite MRI, da questa immagine si segmenta il midollo osseo, e poi si moltiplica con l'immagine PET in maniera tale da ottenere una mappa dell'attività metabolica del midollo osseo. Rispetto alla PET/CT, l'utilizzo della MRI permette di ottenere subito una segmentazione del midollo osseo attivo, per cui alla moltiplicazione con l'immagine PET si ottiene subito una immediata mappa dell'attività metabolica del midollo osseo attivo. Dalla mappa metabolica è possibile ricavare parametri quantitativi sull'attività globale o regionale del midollo osseo attivo.

La tecnologia PET/MRI riduce notevolmente il livello di radiazioni a cui il paziente è sottoposto ed è in grado di aumentare di molto le informazioni che si possono ottenere dall'imaging, ma ha come svantaggio principale il costo estremamente elevato, soprattutto ai fini della sola individuazione del midollo osseo attivo per la pianificazione della radioterapia; per individuare il midollo osseo attivo rimane molto più conveniente utilizzare la sola MRI.

1.2. Radiomica

La radiomica sta rapidamente emergendo come tecnologia della medicina personalizzata ed è un campo di ricerca di sempre maggior interesse. Ogni immagine medica, che sia PET, MRI o ecografica, è una immagine che visualizza in modo non invasivo le caratteristiche dei tessuti del corpo e ne permette, da parte dei radiologi, di analizzare visivamente tali caratteristiche. Tuttavia queste immagini contengono anche un grandissimo numero di informazioni, nascoste nei dati numerici dei pixel che compongono il tessuto che stiamo osservando, e costituiscono un vero e proprio patrimonio di dati che la semplice analisi visiva da parte dei radiologi non è in grado di estrarre. La radiomica, quindi, nasce come strumento per lo sviluppo di supporto decisionale ed è basata su un'analisi quantitativa delle immagini mediche digitali, che vengono trattate come grandi matrici numeriche e da cui vengono estratti un elevato numero di parametri che riflettono le caratteristiche morfologiche e predittive mediante algoritmi di analisi più o meno automatici. Tale analisi permette quindi di ricavare delle informazioni quantitative che sono nascoste ad una analisi qualitativa dell'immagine. Tutti questi descrittori caratterizzano i pixel e le relazioni tra pixel e sono totalmente indipendenti dal tipo di immagine, cioè possono essere calcolati sempre, qualunque sia l'immagine. Questo tipo di analisi matematica viene anche detta texture analysis .

Questo elevato numero di dati può quindi essere sfruttato per l'integrazione in modelli predittivi, tramite anche l'uso di machine learning e intelligenza artificiale, per la gestione personalizzata del paziente.

La possibilità di estrarre informazioni nascoste da immagini mediche digitali è di particolare interesse anche in quanto può potenziare le capacità predittive degli algoritmi di segmentazione automatici esistenti: l'estrazione di nuove informazioni, precedentemente nascoste, può essere sfruttata per la segmentazione automatica delle immagini.

1.3. Obiettivo del lavoro

L'immagine CT sfrutta la diretta proporzionalità tra intensità e radiopacità dei tessuti per visualizzare internamente il corpo umano. Date le sue caratteristiche, il midollo osseo attivo è indistinguibile dal midollo osseo giallo, sia ad una analisi visiva, sia ad una prima analisi quantitativa dei pixel, e per questo non è un metodo utilizzato per la segmentazione del midollo osseo attivo. La radiomica, tuttavia, fornisce dei nuovi strumenti che possono essere in grado di distinguere in maniera automatica il midollo attivo da immagine CT.

Essere in grado di distinguere il midollo attivo dalla sola CT può portare numerosi vantaggi nella pianificazione della radioterapia: l'esame CT è un esame estremamente semplice da eseguire in ospedale e molto meno costoso rispetto sia MRI che PET. In effetti, tutti i pazienti che si sottopongono a cure in ambito oncologico, sono generalmente sottoposti a esame CT, dato che è uno esame standard nel workflow ospedaliero e nella pianificazione della terapia radioterapica. Essere in grado di utilizzare le stesse immagini CT, magari acquisite per altri motivi, è un notevole risparmio in termini di tempo e soldi, oltre ad essere una notevole semplificazione della procedura di pianificazione di radioterapia e ad essere molto più accessibile rispetto ai metodi utilizzati attualmente (PET e MRI). L'unico svantaggio che presenta è l'alta dose di radiazioni ionizzanti a cui si sottopone il paziente, ma questo livello è comunque più o meno equivalente alla dose della PET.

Nel lavoro svolto, applichiamo la radiomica all'individuazione di midollo osseo attivo da immagini CT: si prendono le immagini CT e, slice per slice, si applica la texture analysis e si estraggono tutti i parametri matematici calcolabili pixel per pixel. Tali parametri verranno utilizzati, con una applicazione di machine learning, per la classificazione pixel per pixel del midollo osseo, attraverso 4 tipi di classificatori diversi:

- Decision Tree (DT)
- K-Nearest Neighbors (KNN)
- Neural Network (NN)
- Support Vector Machine (SVM)

Ognuno dei classificatori viene ottimizzato sfruttando come algoritmo di ottimizzazione il Genetic Algorithm (GA), in maniera tale da ottenere i migliori classificatori possibili, e se ne valutano le prestazioni, per capire se sia possibile ottenere una precisa segmentazione del midollo osseo attivo con uno di questi classificatori.

L'obiettivo è di sviluppare un metodo di radiomica in grado di produrre segmentazioni di midollo osseo attivo per la pianificazione della radioterapia e di diminuire il grado di emato-tossicità sui pazienti.

2. Materiali e Metodi

Per lo studio si è utilizzata una popolazione di 25 pazienti, ognuno dei quali sottoposto ad un esame CT e un esame FDG-PET, in maniera tale da ottenere un dataset con 25 immagini CT a disposizione, con le 25 immagini PET prese come riferimento.

L'esame CT è stato eseguito utilizzando un Philips "BigBore" CT scanner, con un pixel spacing di 0,93 mm e 3 mm tra una slice e l'altra. L'esame è stato eseguito sui pazienti in posizione supina e con dei supporti per caviglie e ginocchia. La FDG-PET è stata eseguita con una Philips Gemini PET/CT tomography.

Le immagini PET usate come riferimento sono state segmentate in maniera semi-automatica, calcolando il valore SUV e, per ogni paziente, normalizzandolo con il valore SUV nel fegato e nella pelvi. Applicando poi come soglia SUV il valore SUV_{mean} del paziente, si sono classificati come midollo attivo tutti i pixel con valore SUV maggiore della media.

Da immagini CT si è eseguita segmentazione manuale delle ossa del bacino, a sua volta divisa in tre regioni separate: la cresta iliaca e la parte superiore del femore (Iliac Bone Marrow, IBM), la regione basso-pelvica (Lower Pelvis Bone Marrow, LPBM) e la regione lombo-sacrale (Lumbo Sacral Bone Marrow, LSBM). La classificazione dei pixel avviene separatamente per le tre regioni, per cui si eseguirà il learning dei 4 classificatori separatamente per le tre regioni.

L'immagine CT viene ulteriormente processata, in maniera tale da rimuovere anche i pixel appartenente all'osso compatto, il quale non contiene midollo osseo. Per ogni slice, si è eseguito un k-means clustering con $k=2$ in base all'intensità di ogni pixel, creando due cluster dei pixel del bacino. Siccome l'osso compatto è più denso dell'osso spongioso, appare più intenso nell'immagine, per cui si crea una maschera con i soli pixel del cluster a minore intensità, in maniera tale da mantenere solo i pixel dell'osso spongioso.

2.1 Feature extraction

Il primo passo è l'esecuzione della texture analysis su ogni slice delle immagini CT. Sulle immagini CT si divide l'immagine in tante regioni di 5x5 pixel, dette ROI, parzialmente sovrapposte tra loro, e si calcolano 36 parametri matematici su queste regioni. In questo modo si crea una grande matrice numerica dove ogni elemento corrisponde al pixel centrale della regione 5x5, e ogni elemento è estratto facendo spostare tale regione un pixel per volta, sino a coprire tutti i pixel dell'osso spongioso. I 36 descrittori si dividono in:

- 4 parametri statistici del primo ordine (media, deviazione standard, skewness e kurtosis dei pixel della regione)

- 32 parametri statistici del secondo ordine; di questi, 22 parametri vengono calcolati a partire dalla matrice di co-occorrenza (GLCM), 5 a partire dal Grey Level Difference Method (GLDM) e 5 dal Grey Level Run Length Method (GLRLM).

Quindi si ottiene una matrice di dimensioni $N \times 41$, dove N è il numero di pixel dell'osso spongioso della slice e dove le colonne contengono le coordinate dei pixel, le feature estratte e la classificazione del pixel (1-midollo attivo, 2-midollo non attivo). Si ottiene una tabella come segue (Tabella 1).

Coordinate pixel (x,y,z)			Feature			Label
406	-184	147	258	...	0,780	0
2230	-184	382	263	...	0,841	0
1918	-184	373	268	...	0,758	0
2524	-184	391	280	...	0,827	1
...

Tabella 1 – Estratto di una tabella d'esempio contenente, da sinistra a destra, le coordinate x,y,z di ogni pixel dell'osso spongioso, le feature e la classe d'appartenenza.

2.2. Costruzione del training set

I 4 classificatori eseguono il learning sullo stesso training set. Si esegue il learning sulle 3 strutture separatamente, per cui è necessario costruire 3 training set separati. Ognuno dei training set è stato costruito a partire dai primi 5 pazienti dei 25 a disposizione: dai primi 5 pazienti si sono estratte 5 slice appartenenti alla singola struttura, e da queste slice si sono estratti 1/5 dei pixel disponibili, in maniera casuale. L'estrazione dei pixel è avvenuta mantenendo una proporzione 1:1 tra pixel appartenenti al midollo osseo non attivo e pixel appartenenti al midollo osseo attivo. Tale proporzione è necessaria per evitare che i classificatori "over-fittino" verso una classe o l'altra. Utilizzando questo criterio si sono costruiti 3 training set, uno per ogni struttura. Dalla tabella 2 si possono osservare il numero di pixel contenuti in ogni training set e la percentuale di pixel estratti da quelli totali a disposizione.

	Numero di pixel della struttura	Numero di pixel del training set	Percentuale
IBM	194243	2400	0,0124
LPBM	279702	2200	0,0079
LSBM	181060	2500	0,0138

Tabella 2 – Tabella contenente il numero totale di pixel e il numero di pixel utilizzate per la costruzione dei training set delle strutture IBM, LPBM, LSBM.

2.3. Ottimizzazione dei classificatori – Algoritmo genetico

L'ottimizzazione dei classificatori avviene attraverso un algoritmo di ottimizzazione chiamato algoritmo genetico, e avviene attraverso apprendimento supervisionato. L'algoritmo genetico è strutturato in maniera tale da eseguire contemporaneamente feature selection per i classificatori che l'ottimizzazione di parametri dei classificatori. La feature selection è il processo che porta alla scelta di quali dei 36 parametri calcolati verranno effettivamente utilizzati dal singolo classificatore. Nel nostro caso, la feature selection avviene con metodo wrapper: iterativamente, viene scelto un subset di feature, si esegue il learning del classificatore con tale subset e si valutano le performance ottenute.

L'algoritmo genetico è un algoritmo di ottimizzazione ispirato alla selezione naturale e all'evoluzione biologica, che permette, in tempi ragionevoli, di risolvere problemi di ottimizzazione complessi e non risolvibili tramite normali algoritmi di ricerca lineari. Si basa sull'evoluzione di un insieme di soluzioni iniziali, che vengono ricombinate e ritrasformate casualmente, nel tentativo di trovare una soluzione "ottima". Data la natura casuale dell'algoritmo, il risultato ottenuto può variare pur partendo dalle stesse condizioni iniziali.

Per rendere possibile l'azione degli operatori che effettivamente fanno evolvere le soluzioni iniziali, è necessario che tali soluzioni siano codificate in vettori binari, attraverso l'uso di qualche tecnica.

I principali step di funzionamento dell'algoritmo genetico sono i seguenti

- 1. Generazione di una popolazione iniziale: attraverso una funzione, viene generata una popolazione di soluzioni iniziali in forma di vettori binari. L'algoritmo crea questa popolazione iniziale con soluzioni che sono tutte ammissibili.
- 2. Calcolo fitness della popolazioni iniziale: la fitness è un valore numerico che si assegna ad ogni soluzioni attraverso il calcolo di una funzione di fitness. Tale valore numerico valuta la "bontà" della soluzione al problema di ottimizzazione, e in base al tipo di funzione deve essere minimizzato o massimizzato.
- 3. Selezione dei genitori: dalla popolazione iniziale viene selezionato un subset di soluzioni, sfruttando un qualche metodo di selezione, che diverranno i "genitori" della prossima generazione di soluzioni.
- 4. Applicazione degli operatori genetici: gli operatori genetici sono responsabili della ricombinazione delle soluzioni, e sono quelli che fanno evolvere le varie generazioni. La mutazione produce un cambiamento su uno o più bit casuali della singola soluzione, complementandone il valore (da 0 a 1, o viceversa), in maniera da mantenere una certa diversità di soluzioni ed evitare che si uniformino gli individui della popolazione. Ogni soluzione possiede una certa probabilità di mutare detta pm (probabilità di mutazione). Il crossover invece avviene tra due soluzioni, che vengono tagliate in corrispondenza di uno o più punti di tagli, e le sotto-stringhe vengono scambiate tra loro. Ogni soluzione ha una certa probabilità di essere ricombinata tramite crossover, detta pc (probabilità di crossover).
- 5. Nuova generazione: ogni soluzione della nuova generazione viene valutata tramite il calcolo della funzione di fitness, e si ripete la selezione dei genitori della prossima generazione
- 6. Condizione di stop: l'algoritmo agisce iterativamente sino a che una determinata condizione di stop viene raggiunta.

L'algoritmo genetico utilizzato funziona cercando, per ognuno dei classificatori, il subset di feature ottimale e i parametri del classificatore stesso ottimali.

Ogni classificatore possiede parametri diversi su cui è possibile agire, per cui per ognuno dei classificatori è stata utilizzata una codifica in bit diversa. La codifica, in generale, è formata da una stringa di vettore binario divisa in due parti. La prima parte è un vettore di 36 bit che codifica il subset di feature scelto; ad ogni variabile corrisponde un bit del vettore, e al bit con valore 1 corrisponde la selezione del corrispondente parametro, mentre al bit 0 corrisponde la mancata scelta e l'esclusione dal subset. La seconda parte è invece costituito da un vettore binario di dimensioni variabili in base al tipo di classificatore, che codifica la scelta di parametri del classificatore e il valore di tali parametri.

2.3.1. Decision Tree

Il Decision Tree (DT) è un tipo di classificatore avente una particolare struttura ad albero, in cui ad ogni nodo corrisponde una determinata feature, e gli archi che collegano nodi successivi rappresentano determinati valori di tale feature. Partendo da un nodo iniziale, quindi, si percorrono i vari nodi in base ai valori delle variabili e in base alle regole di separazione di ogni nodo fino ad arrivare ad un nodo finale che corrisponde alla classe di destinazione. La struttura viene costruita iterativamente durante il training: ad ogni iterazione si ricerca la migliore splitting rule, ovvero la regola che permette di dividere il training set in partizioni che sono il più pure possibile, dove per puro si intende partizioni al cui interno sono presenti elementi appartenenti ad una sola classe. Ad ogni splitting rule migliore trovata corrisponde un nodo della struttura. La costruzione dell'albero si ferma quando si sono ottenute tutte partizioni pure e quando sono finite le variabili utilizzabili per ricavare splitting rule. Alla fine del training ai nodi finali vengono associate le classi di appartenenze, in base alla classe di appartenenza degli elementi della partizione; in particolare, la classe del nodo è la classe con maggiore rappresentanza all'interno della partizione.

Per il DT non vi sono parametri da ottimizzare per cui la codifica delle soluzioni da far evolvere con il GA è molto semplice: ogni soluzione è rappresentata da un vettore stringa binario, formato da 36 bit, uno per ogni parametro.

2.3.2. K-Nearest neighbour

Il KNN è un algoritmo di classificazione basato su una misura di distanza tra i vari elementi del training set e l'elemento da classificare. Tale misura di distanza può essere calcolata in vari modi, e il più usato è la distanza euclidea.

Quando si vuole classificare un nuovo elemento, si calcola la distanza tra l'elemento e il training set, e si ordinano i risultati in ordine crescente: si prendono i primi k elementi, e la classe di appartenenza sarà quella maggiormente rappresentata.

L'unico parametro da scegliere e ottimizzare è quindi il valore di K. La codifica delle soluzioni del GA è quindi formata da una prima parte di 36 bit per la selezione delle feature, e una seconda parte formata da un vettore di 6 bit, che rappresentano il valore in binario del parametro *bit*. In questo modo il GA esplora un totale di $2^6 = 36$ valori di K attorno ad un valore iniziale che è un valore pari a:

$$K_{in} = \sqrt{N}$$

Dove N è il numero di elementi del training set. Nel nostro caso, siccome ogni struttura ha un numero leggermente diverso di elementi nel training set, si è scelto come valore di N il numero di elementi del training set della struttura LPBM, per cui:

$$K_{in} = \sqrt{N} = 33 \quad \text{dove } N = 1100$$

Il parametro K sarà pari a:

$$K = K_{in} + bit - 36$$

Ogni soluzione è quindi rappresentata da un vettore binario di 42 elementi.

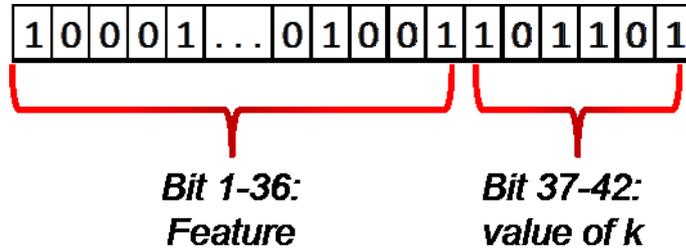


Figura 1 – Codifica binaria dei classificatori KNN

2.3.3. Neural network

Una rete neurale è un modello matematico ispirato alle reti neurali biologiche utilizzati per la risoluzione di problemi di vario tipo. E' un sistema adattativo, la cui struttura varia in base agli elementi passati durante la fase di apprendimento. Le rete neurale utilizzata nel nostro caso è la Feedforward Neural Network (FNN).

La FNN è formata da una serie di neuroni, cioè da una serie di funzioni matematiche, connesse tra loro da archi aventi diversi pesi. Ogni neurone è in grado quindi di processare informazioni in accordo al modello di McCulloch e Pitts .

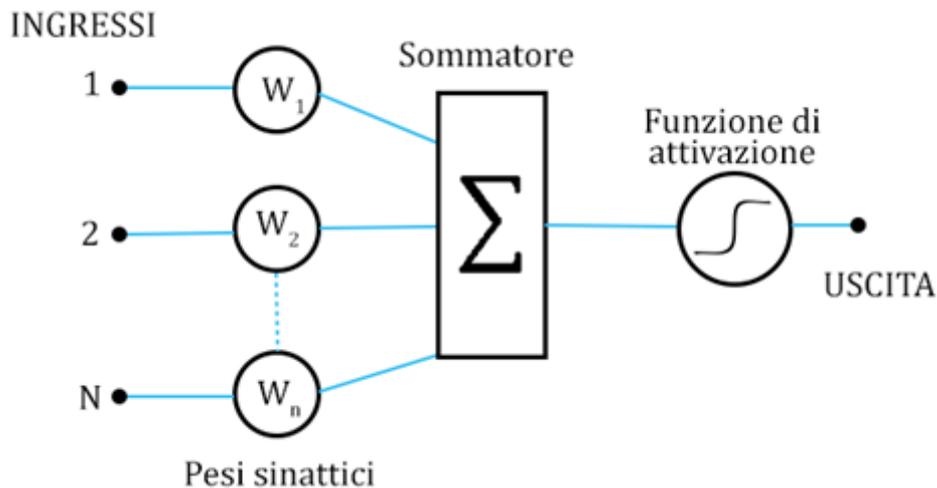


Figura 2 – Modello di neurone di layer nascosto. All'ingresso riceve input provenienti da altri neuroni opportunamente pesati, li somma e ottiene una uscita in base alla funzione di attivazione utilizzata.

I neuroni sono organizzati in layer, per cui i neuroni di un determinato layer sono connessi ai neuroni del layer precedente e successivo, ma non sono connessi ad altri neuroni dello stesso layer. La rete è organizzata con:

- un layer di input, contenente un neurone per ogni variabile di input .
- un layer di output, contenente un numero di neuroni variabile in base al numero di classi.
- un certo numero di layer nascosti, tra input e output, che processano gli elementi in entrata e il cui numero e il numero di neuroni che contengono variano in base al problema.

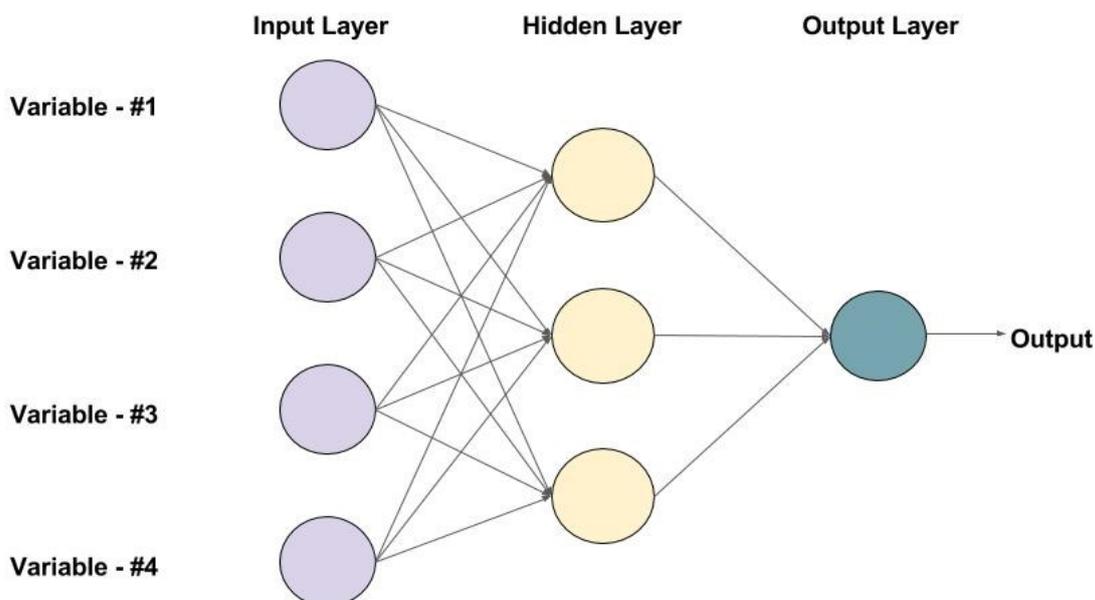


Figura 3 – Struttura di una rete neurale FNN

La rete neurale da noi utilizzata ha una struttura con:

- layer di input con 36 neuroni
- layer di output con 1 neurone, con una funzione di attivazione lineare. L'output della rete è lineare e ha un valore variabile, per cui, per ottenere come output 0 e 1 (corrispondenti a midollo attivo e non attivo) si applica la semplice condizione:

$$\text{if } Output < 0,5 \rightarrow Output = 0$$

$$\text{if } Output > 0,5 \rightarrow Output = 1$$

- un numero variabile di layer nascosti. Il numero di layer è il parametro da ottimizzare tramite GA, mentre il numero di neuroni di ogni layer è deciso in maniera tale da ottenere una struttura della rete piramidale. Il primo layer nascosto ha un numero di neuroni pari al layer di input, mentre i layer successivi hanno un numero di neuroni pari alla metà del layer precedente. Tutti i neuroni dei layer nascosti utilizzano come funzione di trasferimento la sigmoide.

Il GA ha soluzioni codificate sempre con un vettore formato da una prima parte di 36 bit per la selezione delle variabili, seguito da una seconda parte formata da 4 bit. Il valore decimale di questi bit codifica il numero di layer nascosti, per cui la GA esplora soluzioni con un numero di layer che va da 1 a 8.

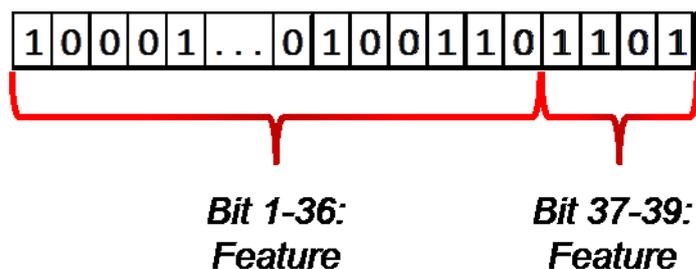


Figura 4 – Codifica binaria per i classificatori NN

2.3.4. Support vector machine

Una Support Vector Machine (SVM) è un classificatore binario, in grado di distinguere solo due classi. Dati gli elementi di un training set, ogni elemento può essere rappresentato come un punto in uno spazio multi-dimensionale. Una SVM crea un nuovo spazio multi-dimensionale, generalmente avente maggiore dimensionalità rispetto a quello originale, e crea un iperpiano in tale spazio che sia in grado di dividere gli elementi del training set in due classi distinte. L'iperpiano viene costruito in maniera tale da massimizzare la distanza tra il piano stesso i vettori supporto (detta margine), ovvero gli elementi delle due classi più vicini al piano di separazione. La creazione di un nuovo spazio permette di rendere qualsiasi problema di classificazione risolvibile tramite un classificatore lineare. La predizione a questo punto viene fatta semplicemente osservando in quale lato ricade ogni nuovo elemento.

Nella SVM, a differenza degli altri classificatori, vi sono due parametri su cui eseguire ottimizzazione:

- la funzione di kernel: è la funzione di trasferimento matematica che permette di mappare gli elementi nel nuovo spazio multi-dimensionale e quindi di permetterne la classificazione. Tale funzione può essere lineare o non lineare, a seconda del problema.
- il parametro C: regola il bilanciamento margine ampio e un errore di classificazione basso. Avere ampi margini, infatti, crea una distinzione molto precisa tra gli elementi del training set, ma aumenta la probabilità di errata classificazione di nuovi elementi. Per cui avere margini meno ampi può migliorare l'accuratezza complessiva del modello.

Data la presenza di due parametri, in questo caso la codifica di ogni soluzione è data da un vettore binario diviso in tre parti. La prima parte è un vettore di 36 bit per la selezione delle variabili. La seconda parte permette la scelta della funzione di kernel, ed è composta da un vettore di due bit, che permette di scegliere tra 4 funzioni di kernel diverse: kernel lineare, Gaussiano, polinomiale di ordine 2 e polinomiale di ordine 3.

BIT (valore decimale)	Funzione
0	Lineare
1	Gaussiana
2	Polinomiale di ordine 2
3	Polinomiale di ordine 3

Tabella 3 – Tabella contenente il valore dei bit codificanti la funzione di kernel e le funzioni codificate.

La terza parte, composta da un vettore di 4 bit, permette la scelta del valore del parametro C. Tale valore si ottiene calcolando il valore decimale degli ultimi 4 bit e tramite l'equazione:

$$C = \begin{cases} 0.5 & \text{if } C_{dec} = 0 \\ 1 & \text{if } C_{dec} = 1 \\ (C_{dec}-1) * 10 & \text{otherwise} \end{cases}$$

Dove C_{dec} è il valore decimale del vettore binario. Con questa equazione si esplorano valori tra 0,5 e 140.

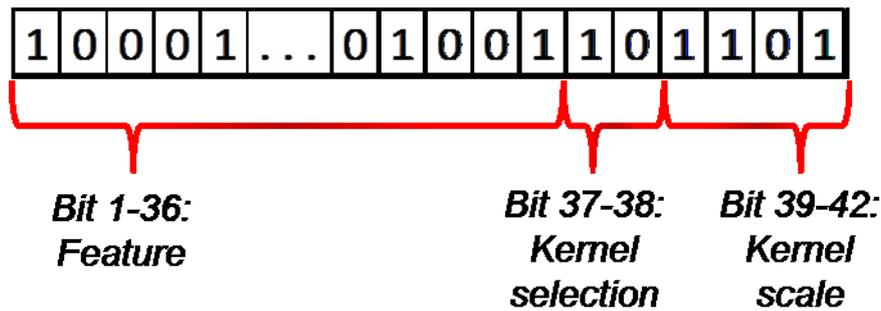


Figura 5 – Codifica binaria dei classificatori SVM

2.4. Funzione di fitness

Ognuna delle soluzioni della popolazione in evoluzione viene quindi valutata sfruttando una funzione di fitness. Tale funzione ha equazione:

$$f = 1 - accuracy + 0,3 \cdot |sensitivity - specificity|$$

Il valore di fitness è tanto migliore quanto è più basso. La fitness viene calcolata tramite i valori di accuratezza, sensibilità e specificità che si ottengono con il classificatore costruito con le feature selezionate e i parametri codificati dalla singola soluzione. Quindi per ogni soluzione si costruisce un classificatore e si esegue il training dello stesso utilizzando una partizione del training set estratto dai 5 pazienti, per cui metà del training set viene utilizzato per la fase di apprendimento. L'altra metà del training set viene invece utilizzato come test set per testare le performance del classificatore, per cui si tenta di predire la classe di appartenenza degli elementi del test set. Con i risultati delle predizioni si costruisce una confusion matrix, da cui si calcolano i valori di accuratezza, sensibilità e specificità, che a loro vengono utilizzate per il calcolo della fitness.

La funzione di fitness viene calcolata in maniera tale da valutare più positivamente non solo soluzioni aventi alta accuratezza, ma, in misura minore, favorisce quelle soluzioni aventi valori di specificità e sensibilità simili. Questa valutazione aggiuntiva garantisce che le soluzioni trovate abbiano valori bilanciati e siano in grado di classificare correttamente sia i veri negativi che i veri positivi. Le soluzioni aventi bassa specificità e alta sensibilità e viceversa, a parità di accuratezza, vengono penalizzate. Nella segmentazione del midollo osseo attivo, questo evita che i classificatori costruiti abbiano una generale tendenza a sotto-segmentare o sovra-segmentare.

2.5. Parametri dell'algoritmo genetico

2.5.1. Popolazione iniziale

Il GA agisce sempre a partire da una popolazione iniziale di soluzioni. Nel lavoro, è stata scelta una popolazione iniziale composta da 500 soluzioni create totalmente casualmente, con la sola condizione di avere, per i primi 36 bit, almeno due bit con valore 1, ovvero di avere soluzioni che selezionano almeno due variabili tra i descrittori possibili. E' stata creata una popolazione di 500 soluzioni per ogni tipo di classificatore, e varie ripetizioni dell'algoritmo genetico sono sempre partite dalla stessa popolazione iniziale.

2.5.2. Operatori genetici

Per gli operatori genetici si sono scelti i seguenti valori di probabilità di crossover (PC) e mutazione (PM):

- $PC = 1$
- $PM = 0,5$

Tali valori di probabilità di mutazione e crossover sono molto elevati. Questa scelta è dovuta all'altissimo numero di possibili soluzioni esistenti per ogni classificatore, da 2^{36} a 2^{42} possibilità che il GA deve poter esaminare. Dato l'enorme insieme da dover esplorare, alti valori di probabilità per gli operatori genetici consentono di avere una alta variabilità nelle soluzioni delle varie generazioni, consentendo di esplorare un numero abbastanza alto di soluzioni ed evitare che il GA converga troppo facilmente a soluzioni apparentemente ottimizzate. L'alta variabilità evita la convergenza, ma bisogna tenere conto che non consente che le soluzioni trovate siano molto ottimizzate, dato che eventuali buone soluzioni vengono sistematicamente cambiate. In sostanza, in questo modo il GA esplora un alto numero di soluzioni, ma non ottimizza molto le buone soluzioni incontrate.

2.5.3. Metodo di selezione

Ogni generazione è composta da 500 soluzioni. Di queste soluzioni vengono selezionate 350 che diverranno i genitori e su cui si applicheranno gli operatori genetici. Il metodo con cui vengono selezionati è il metodo la selezione a roulette, in cui ad ogni soluzione è assegnata una probabilità di essere scelta inversamente proporzionale al valore di fitness (minore il valore di fitness, maggiore la probabilità che venga scelta). Questo metodo di scelta consente anch'esso di esplorare maggiormente l'insieme delle possibili soluzioni, ma favorisce in minima misura al mantenimento durante le successive generazioni di soluzioni con minor fitness.

2.5.4. Condizioni di stop

Il GA viene eseguito per un numero di iterazioni pari a 100. Tuttavia viene aggiunta una condizione di stop aggiuntiva, per evitare che il GA esegua necessariamente 100 iterazioni ad ogni ripetizione, il che permette di risparmiare del tempo di calcolo. Si tenga conto, infatti, che il GA costruito in questo modo, dove ad ogni generazione si creano e valutano le prestazioni di 500 classificatori, il tempo computazionale è molto alto. Avere una condizione di stop aggiuntiva permette un notevole risparmio di tempo.

Come condizione di stop, si è deciso di interrompere la ripetizione se per 30 iterazioni consecutive non si sono osservati miglioramenti della soluzione con fitness migliore. Per cui ad ogni generazione si tiene conto della soluzione migliore e della sua fitness, e se non si incontra un individuo con fitness più bassa per 30 volte consecutive, allora si interrompe e si mantiene quella soluzione come migliore.

2.5.5. Esecuzione dell'algoritmo genetico

Come già detto in precedenza, il midollo osseo della regione pelvica è diviso in 3 strutture separate (IBM, LPBM, LSBM), ognuna delle quali aventi il proprio training set. Ogni struttura deve essere segmentata da 4 classificatori diversi (DT, KNN, NN, SVM), ognuno dei quali deve essere costruito e ottimizzato tramite esecuzione del GA. Data la natura stocastica dell'algoritmo di ottimizzazione, pur partendo da una medesima popolazione iniziale, la soluzione finale non è garantita che sia la stessa in esecuzioni diverse. Perciò per ogni struttura e ogni classificatore il GA viene eseguito per 5 ripetizioni, in maniera tale da ottenere 5 soluzioni separate.

		STRUTTURE		
		IBM	LPBM	LSBM
CLASSIFICATORI	DT	sol 1	sol 1	sol 1
		sol 2	sol 2	sol 2
		sol 3	sol 3	sol 3
		sol 4	sol 4	sol 4
		sol 5	sol 5	sol 5
	KNN	sol 1	sol 1	sol 1
		sol 2	sol 2	sol 2
		sol 3	sol 3	sol 3
		sol 4	sol 4	sol 4
		sol 5	sol 5	sol 5
	NN	sol 1	sol 1	sol 1
		sol 2	sol 2	sol 2
		sol 3	sol 3	sol 3
		sol 4	sol 4	sol 4
		sol 5	sol 5	sol 5
	SVM	sol 1	sol 1	sol 1
		sol 2	sol 2	sol 2
		sol 3	sol 3	sol 3
		sol 4	sol 4	sol 4
		sol 5	sol 5	sol 5

Tabella 4 – Risultati dell’algoritmo GA. Per ogni combinazione di struttura e classificatore si ottengono 5 soluzioni diverse.

Di ogni soluzione vengono sempre calcolati i valori di specificità, sensibilità e accuratezza, oltre che del valore di fitness loro derivato. Di questi, i parametri utilizzati per la valutazione dei singoli classificatori per ogni struttura sono sensibilità, specificità e accuratezza, calcolati a partire dal test set. Con questi parametri siamo in grado di osservare le prime indicazioni utili sulla capacità di predizione dei classificatori, ma non sono sufficienti per valutare effettivamente le performance della segmentazione. Per fare ciò si rendono necessari altri passi di processing.

2.6. Classificazione delle immagini

Dopo aver ricavato le soluzioni del GA, bisogna passare alla segmentazione vera e propria del midollo osseo. La segmentazione avviene classificando slice per slice tutti i pixel appartenenti all’osso spongioso delle immagini relative a tutti e 25 i pazienti.

Prima di applicare i singoli classificatori per la segmentazione, si rende necessaria una operazione aggiuntiva che ne migliori le capacità predittive. Ad ogni soluzione corrisponde un diverso classificatore, ottimizzato per avere un diverso subset di feature utilizzate e diverso valore dei parametri, il quale classificherà i pixel diversamente dalle altre soluzioni per lo stesso classificatore e struttura, e in particolare sbaglierà la classificazione di pixel diversi e in misura diversa. Siccome ogni classificatore sbaglia diversamente, e siccome l’algoritmo genetico ha ottimizzato le soluzioni in maniera che siano in qualche modo favorite le segmentazioni bilanciando sensibilità e specificità, se si combinano i risultati ottenuti è possibile migliorare la precisione complessiva escludendo eventuali pixel che sono stati classificati in un certo modo da un solo classificatore, i quali è più probabile che siano errori di sovra-segmentazione o sotto-

segmentazione, mentre pixel classificati allo stesso modo da più classificatori hanno una probabilità più alta di essere correttamente classificati.

La combinazione dei risultati si chiama voting: per ogni pixel si controlla il valore ottenuto dalla classificazione nei vari classificatori, e si attribuisce il valore avente maggiore rappresentanza.

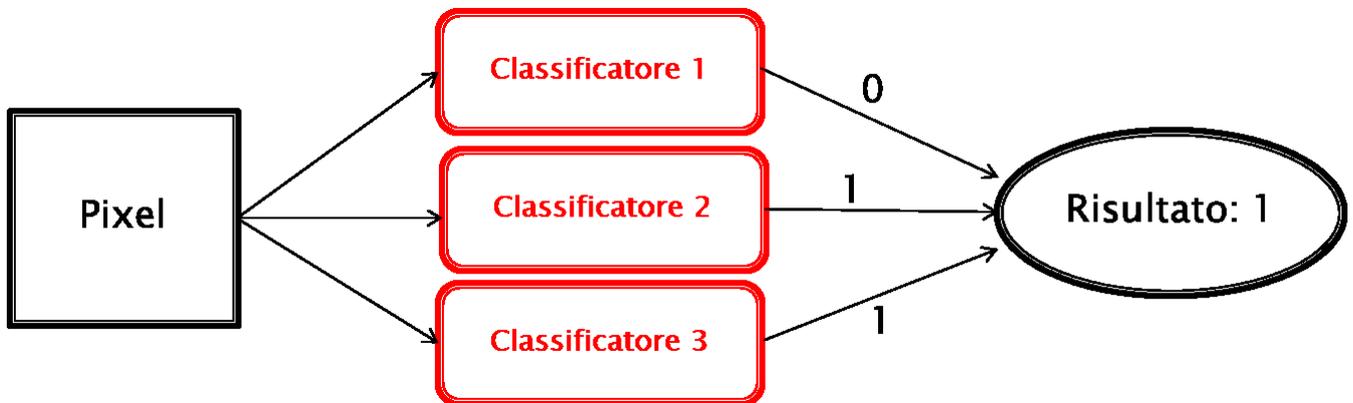


Figura 6 – Schema di voting. Il pixel si riferisce alla ROI. La ROI viene classificata dai tre classificatori, e si ottengono due risultati positivi e uno negativo. Il risultato finale è positivo.

Nel lavoro si è scelto di utilizzare un sistema di voting con tre classificatori:

- Classificatore con valore di fitness più bassa
- Classificatore con valore di sensibilità più alta
- Classificatore con valore di specificità più alta

Per ogni singola slice delle immagini dei pazienti si eseguono i seguenti passaggi:

- Si costruiscono 3 classificatori a partire dalle 5 soluzioni trovate dal GA
- I pixel della slice appartenenti al midollo osseo vengono classificati, in maniera tale da estrarre 3 maschere diverse di midollo osseo attivo
- Si esegue il voting con le 3 maschere e si ottiene la maschera finale

Queste operazioni si ripetono per ogni slice fino alla segmentazione di tutte le slice dell'immagine, per ogni struttura e per ogni classificatore. Per ogni slice si applica una certa soglia minima di pixel positivi necessari a mantenere la stessa slice; se il numero di pixel è minore di 64, allora tali pixel vengono rigettati e si tratta tale slice come non contenente midollo osseo attivo. Tale soglia è estratta dalla letteratura, ed è conseguenza della natura dell'immagine PET. E' legata in particolare alla differenza di risoluzione spaziale tra PET e CT, in cui un singolo pixel equivale a 64 pixel di CT. Dato che è impossibile che il riferimento PET possa distinguere zone di midollo osseo attivo aventi meno di 64 pixel di CT, a causa della bassa risoluzione, si fissa una soglia minima al di sotto del quale non si considerano tali slice. Questo evita che eventuali zone di midollo più piccole vengano tenute in conto nel momento di valutare le performance della segmentazione e ne diminuiscano il valore.

2.7. Post-processing

La maschera ottenuta in questo modo ha le caratteristiche di essere ottenuta tramite la classificazione di ogni singolo pixel, per cui è una segmentazione ottenuta solo tramite informazioni locali dell'immagini. Perciò la maschera che si ottiene ha sempre dei bordi molto frastagliati e sono presenti spesso piccoli

buchi. Per migliorare l'aspetto e eliminare questi buchi, che sono spesso dovuti ad una cattiva classificazione di pochi pixel, viene eseguita sulla maschera di midollo attivo una operazione morfologica di chiusura. Tramite chiusura, quindi, si uniforma la maschera e i bordi della segmentazione.

La maschera così ottenuta è la segmentazione automatica delle immagini 3D da CT.

2.8. Valutazione dell'effetto della soglia SUV

Le performance dei classificatori dipendono non solo dall'ottimizzazione operata dal GA, ma dipendono ovviamente anche dalle maschere di riferimento. Le maschere di riferimento sono anch'esse state ricavate tramite un algoritmo semi-automatico, con un meccanismo basata sul valore soglia di SUV. Tale valore di soglia è stato preso come valore medio di SUV del paziente. Tuttavia tale valore di soglia può essere modificato, e ciò può di conseguenza modificare i risultati della segmentazione, pur non variando di fatto di molto la capacità dell'algoritmo semi-automatico di segmentare correttamente il midollo attivo. In tal modo si vuole controllare quanto il metodo usato per la creazione di maschere di riferimento da PET sia robusto nella sua capacità di segmentare le 3 strutture di midollo osseo attivo.

Per capire quale sia l'effetto del SUV, si è ripetuta l'operazione di segmentazione 3 volte, con i seguenti valori di soglia di SUV per il riferimento:

- SUV_{mean}
- $SUV_{mean} + 10\%$
- $SUV_{mean} - 10\%$

Se il metodo di segmentazione dei riferimenti è robusto, le performance dei classificatori variano poco. Al contrario, tale performance variano molto.

2.9. Indici per la valutazione dei risultati

Le maschere di segmentazione ottenute, per ogni struttura e ogni classificatore, vengono valutate tramite il calcolo di una serie di parametri descrittivi della concordanza tra le maschere di riferimento e le maschere ottenute. Tali parametri vengono calcolati a partire dalle maschere da valutare e dalle maschere di riferimento, e sono:

- Dice coefficient
- Recall
- Precision
- Mean Distance to Conformity (MDC)
- Over-Contouring (OC)
- Under-Contouring (UC)

Il Dice coefficient è un parametro che calcola il grado di sovrapposizione tra due maschere. E' un parametro facile da calcolare, e da una misura della sovrapposizione tra maschere di riferimento e maschere trovate da algoritmo. Si calcola come il rapporto tra il doppio del numero di pixel appartenenti ad entrambe le maschere e la somma del numero di pixel totale di entrambe le maschere:

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

Dove X è la maschera di riferimento e Y è la maschera ottenuta. Ad un valore pari a 1 corrisponde la perfetta sovrapposizione delle maschere. Il Dice è un parametro molto utilizzato nella valutazione del grado

di sovrapposizione tra due maschere binarie data la facilità di calcolo e la facilità d'interpretazione dei valori ottenuti, ma presenta alcuni problemi. In particolare, il parametro non presenta alcuna informazione riguardo alla morfologia delle maschere confrontate, cioè dal parametro non si ottiene alcuna informazione riguardo alla forma delle maschere. Il Dice viene calcolato tenendo conto solo del numero di pixel presenti nelle maschere, ma la struttura stessa degli insiemi non viene tenuta in conto, e tale mancanza può portare ad una interpretazione dei risultati non esatta. Inoltre il parametro di Dice è molto sensibile alle dimensioni delle maschere confrontate, essendo calcolato tramite il numero di pixel. Quando le maschere confrontate diminuiscono di dimensioni, diminuisce anche il numero complessivo di pixel di cui si tiene conto, e a questa diminuzione consegue che eventuali pixel non appartenenti alla sovrapposizione hanno un peso maggiore nella determinazione del valore stesso del Dice, che diminuisce più facilmente. Per cui, alla diminuzione delle dimensioni delle maschere, i valori di Dice tendono a diminuire di valore.

Il Recall è un parametro di valutazione di classificazione binaria, definito come il rapporto tra il numero di veri positivi e la somma complessiva dei positivi. E' una misura della percentuale di pixel positivi correttamente classificati come tali. In imaging, il Recall si traduce nel rapporto tra la sovrapposizione delle due maschere e la maschera di riferimento:

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

Il Recall è una misura legata all'errore di sotto-segmentazione della maschera, in quanto un valore basso significa che una percentuale dei pixel appartenenti al midollo attivo non è stato individuato dalla maschera.

La Precision è definita come il rapporto di tra il numero di veri positivi e il numero totale di pixel classificati come positivi, ed è una misura della probabilità che un pixel classificato come positivo sia effettivamente un positivo. In imaging, la Precision è calcolata come il rapporto tra la sovrapposizione delle maschere e la maschera segmentata:

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

E' una misura legata all'errore di sovra-segmentazione della maschera, in quanto un valore basso corrisponde ad un numero alto di pixel individuati che non sono effettivamente appartenenti al midollo attivo.

Oltre al valore di Dice, anche Precision e Recall sono soggetti agli stessi problemi di dipendenza dal numero di pixel totale, per cui ne diminuisce il valore al diminuire dei pixel totali considerati, e non tengono conto in alcun modo delle caratteristiche morfologiche della maschera.

In alternativa al parametro Dice, è stato usato, per la valutazione, anche un nuovo metodo di misura del grado di sovrapposizione di due maschere, in grado anche di fornire informazioni riguardo alla morfologia delle maschere confrontate e con minore dipendenza dal numero di pixel. Tale nuova misura è chiamata Mean Distance to Conformity (MDC), sviluppata da Jena et al [16].

La MDC è definita come la distanza media che i pixel delle due maschere non coincidenti, ovvero i pixel appartenenti ad una sola delle maschere, devono percorrere per rientrare nella maschera di riferimento. La MDC è un parametro composto dalla somma di due componenti:

- La componente di Over-Contouring (OC), che è la misura dell'errore dovuto alla sovra-segmentazione della maschera segmentata.
- La componente di Under-Contouring (UC), che è la misura dell'errore dovuto alla sotto-segmentazione della maschera segmentata.

E' possibile calcolare separatamente UC e OC, per cui la MDC presenta anche il vantaggio di fornire informazioni riguardo a che tipo di errore viene commesso dalla maschera segmentata, e in quale proporzione.

Rispetto al parametro Dice, facile e immediato da calcolare, la MDC è computata attraverso un algoritmo più complicato. Ottenuta la maschera 3D segmentata, viene creata una maschera vuota, e a seconda dei valori dei corrispondenti voxel del volume di riferimento e il volume da controllare, a ogni elemento viene assegnato un valore di stato diverso:

- 1. Stato 0: il voxel non appartiene a nessuno dei volumi da confrontare
- 2. Stato 1: il voxel appartiene al solo volume di riferimento, non al volume di da segmentazione. Questi voxel causano quindi l'errore di sotto-segmentazione
- 3. Stato 2: il voxel appartiene al solo volume da valutare, non al riferimento. Sono i pixel che causano l'errore di sovra-segmentazione
- 4. Stato 3: il voxel appartiene ad entrambi i volumi

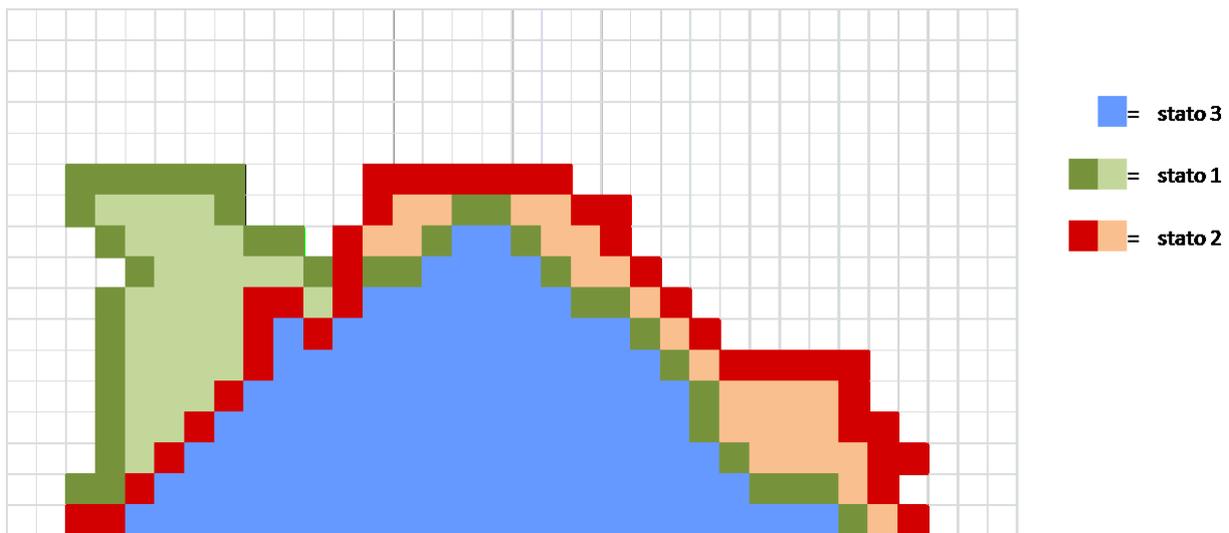


Figura 7 – A sinistra, la sovrapposizione di una maschera di riferimento e una maschera di cui bisogna valutare il grado di sovrapposizione. A destra, la legenda contenente il codice colore. Ad ogni colore corrisponde uno degli stati. I pixel bianchi corrispondono allo stato 0.

Il calcolo della MDC avviene considerando i pixel appartenenti agli stati 1, 2 e 3. Una volta creato la griglia 3D degli stati dei voxel, si procede al calcolo delle componenti OC e UC nel seguente modo:

- Per ogni elemento di stato 1, si calcola la distanza tra l'elemento e gli elementi del volume da valutare e si memorizza la distanza minima in un vettore contenente tutte le distanze minime

$$OC = \frac{\sum_{i=1}^N d(p_i^2, A)}{N}$$

- Per ogni elemento di stato 2, si calcola la distanza tra l'elemento e gli elementi del volume di riferimento e si tiene da parte la distanza minima un altro vettore di distanze minime

$$UC = \frac{\sum_{i=1}^N d(p_i^1, B)}{N}$$

Ottenuti i due vettori, si calcolano le componenti OC e UC come le medie delle distanze dei rispettivi vettori:

- Dal vettore degli elementi di stato 1 si ottiene la componente UC, corrispondente all'errore di sotto-segmentazione.
- Dal vettore degli elementi di stato 2 si ottiene la componente OC, corrispondente all'errore di sopra-segmentazione.

La MDC sarà pari a:

$$MDC = OC + UC$$

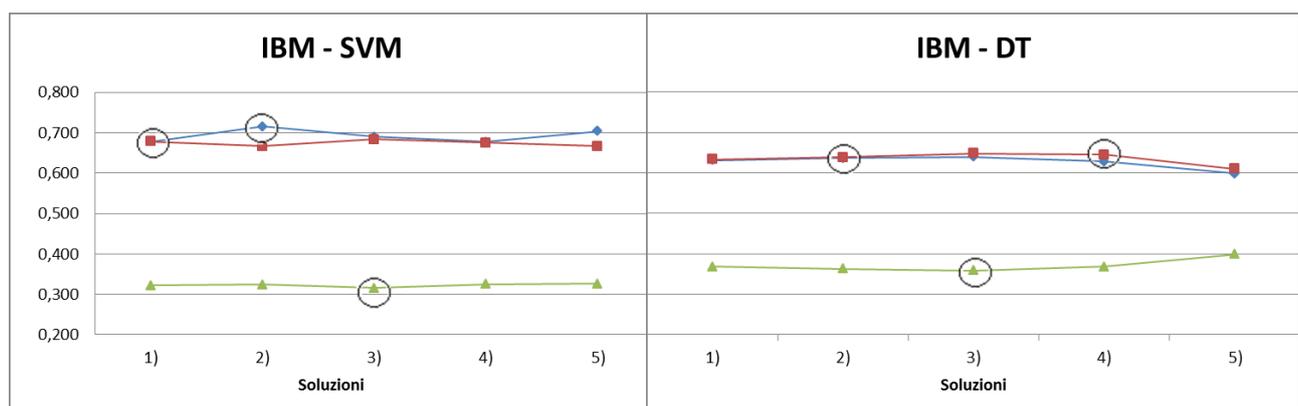
e si misura in mm.

La MDC è quindi una misura che viene calcolata attraverso l'informazione di distanza tra vari pixel, e questo aggiunge informazioni riguardo alla struttura dei volumi confrontati. Inoltre, rispetto alla misura di Dice, è un indicatore molto più sensibile, per cui a piccole variazioni tra volumi possono esservi anche grandi differenze in valore. Per esempio, nel caso di volumi segmentati lontani dal riferimento: pochi pixel sovra-segmentati in una zona più lontana del riferimento ha poca differenza di valore di Dice, ma il calcolo delle distanze fa sì che il valore MDC vari di molto. Una alta variazione di questo tipo è indicativa quindi della presenza di pixel ad elevata distanza dal riferimento. Infine, attraverso la valutazione di OC e UC, è facilmente valutabile quanto il valore è dovuto a sovra o sotto-segmentazione.

3. Risultati

3.1. Valutazione delle soluzioni dell'algorithm genetico

Dall'esecuzione del GA, per la selezione delle feature e la costruzione dei classificatori, si sono ottenute 5 soluzioni per ogni classificatore e per ogni struttura, che sono state valutate tramite il valore di fitness, sensibilità e specificità. Tali risultati sono mostrati nelle seguenti immagini.



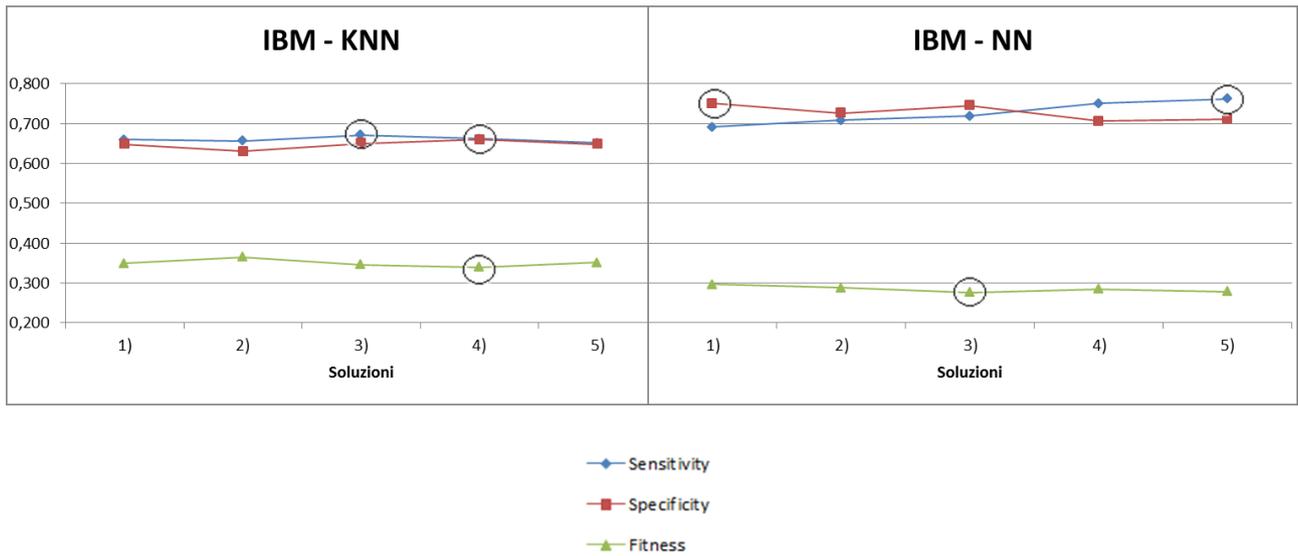
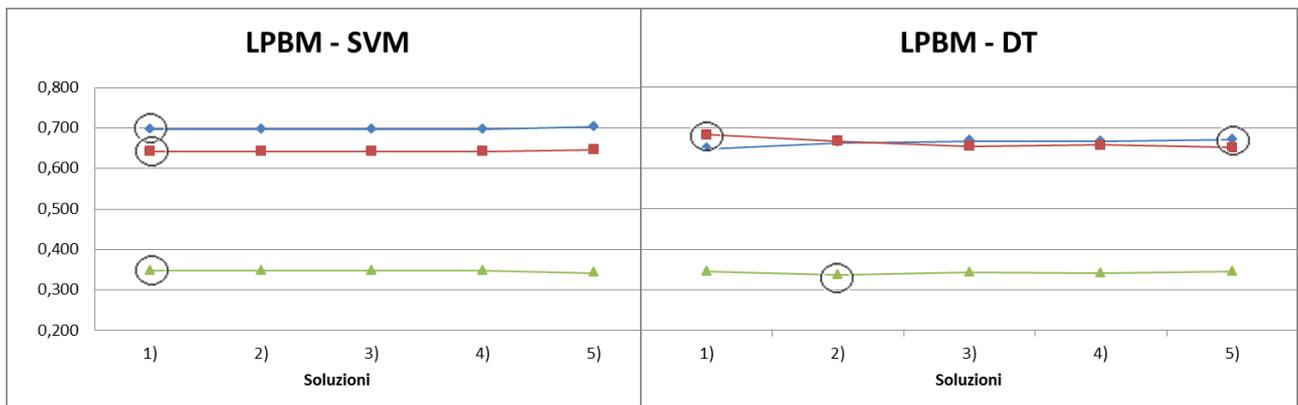


Figura 8 – Risultati del GA per la struttura IBM. L’ascissa contiene il numero della soluzione, l’ordinata è il valore percentuale di sensibilità, specificità e accuratezza. In basso, la legenda: in blu la sensibilità, in rosso la specificità, in verde la fitness. Con i cerchietti neri sono segnalate le 3 soluzioni prese per eseguire il successivo voting.

Dall’osservazione dei risultati da struttura IBM, si può osservare come i risultati non abbiano valori molto elevati per nessuno dei 4 classificatori. I migliori risultati si sono ottenuti attraverso il classificatori di reti neurali, dove l’accuratezza ha raggiunto poco più del 70%. Questo è un risultato prevedibile, data la capacità delle reti neurali di essere in grado di processare informazioni anche molto complesse. Il DT invece, possiede le performance peggiori, con una accuratezza che non supera il 65%. Questo riflette la semplicità dell’algoritmo con cui tali classificatori funzionano, che quindi non sono in grado di prevedere con precisione la classe di appartenenza. Per ogni classificatore si può infine notare che le 5 soluzioni hanno valori di specificità e sensibilità relativamente variabili, mentre l’accuratezza rimane sostanzialmente costante. Questa può essere una indicazione che ognuna delle 5 soluzioni costruisce classificatori diversi, informazione cruciale per la successiva adozione del voting.

Nelle immagini sono segnalate con cerchi le soluzioni che sono state prese per eseguire la successiva classificazione dei pixel e la segmentazione tramite voting. Anche avendo valori di fitness molto simili, ogni soluzione possiede caratteristiche diverse, e creeranno quindi maschere diverse. Si noti che in alcuni casi le soluzioni selezionate sono prese più di una volta. In questi casi particolari il voting non ha alcun effetto, dato che prendere la medesima soluzione genera maschere uguali, le quali impediscono il funzionamento del meccanismo di voto.



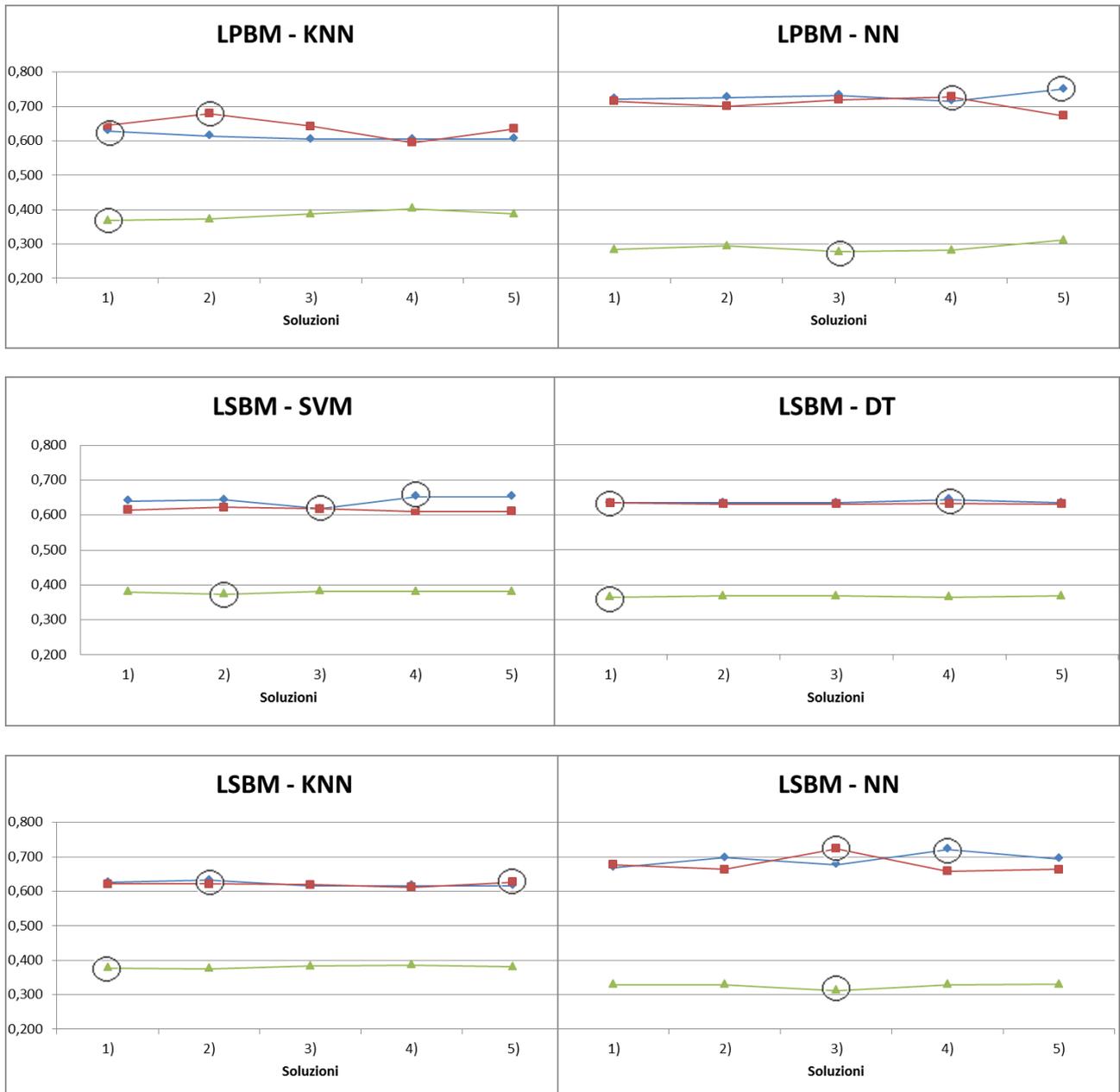


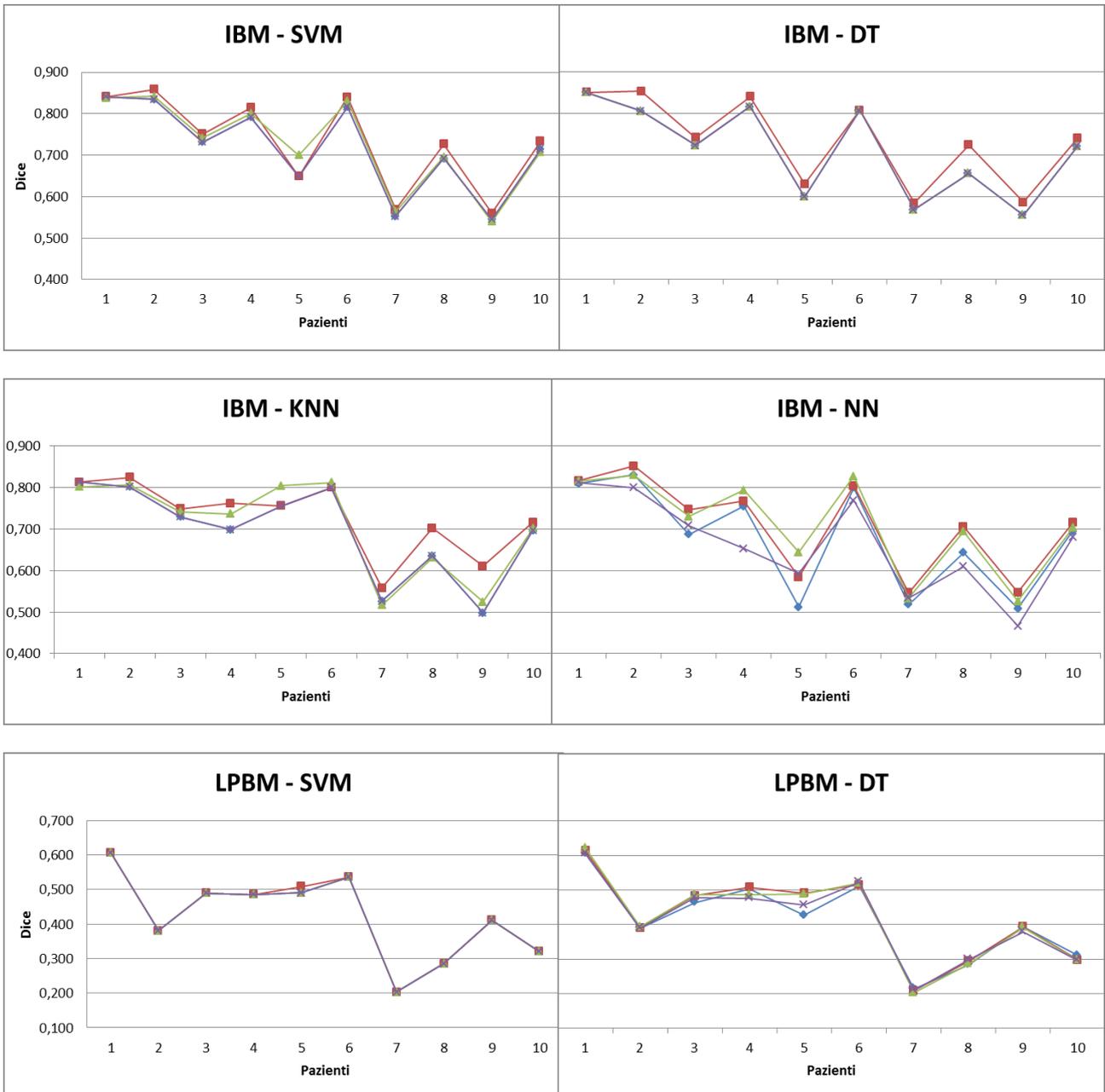
Figura 9 – Risultati dell’algoritmo GA per le strutture LPBM e LSBM. In ascissa vi sono le soluzioni, in ordinata la percentuale. In blu, rosso e verde si osservano rispettivamente i valori di sensitività, specificità e fitness, come nella legenda in figura 8. Con i cerchietti neri sono segnalate le 3 soluzioni prese per eseguire il successivo voting.

Le stesse osservazioni fatte per la struttura IBM valgono anche nelle strutture LPBM e LSBM: i classificatori presentano soluzioni con performance non molto alte, che al più superano di poco la soglia del 70%. In ogni struttura sono sempre le reti neurali a risultare i migliori classificatori ottenuti attraverso il GA. Le soluzioni dei classificatori, anche al variare della struttura, hanno valori molto simili tra loro. Per cui dall’algoritmo GA si ottengono soluzioni diverse, ma tutte aventi prestazioni simili.

Tutte le soluzioni hanno valori di sensitività e specificità molto bilanciate, testimone dell’azione della funzione di fitness, che ha favorito soluzioni con valori bilanciati di sensitività e specificità.

3.3. Effetto del voting

L'effetto del voting si può osservare nel dettaglio nei seguenti grafici, che mettono a confronto le segmentazioni ottenute tramite i classificatori singoli con la segmentazione ottenuta con voting. Per semplicità, si è scelto di osservare l'effetto del voting solo per i primi 10 pazienti, per avere una idea di quale sia l'effetto mantenendo la leggibilità dei grafici, e di esprimere i risultati in Dice.



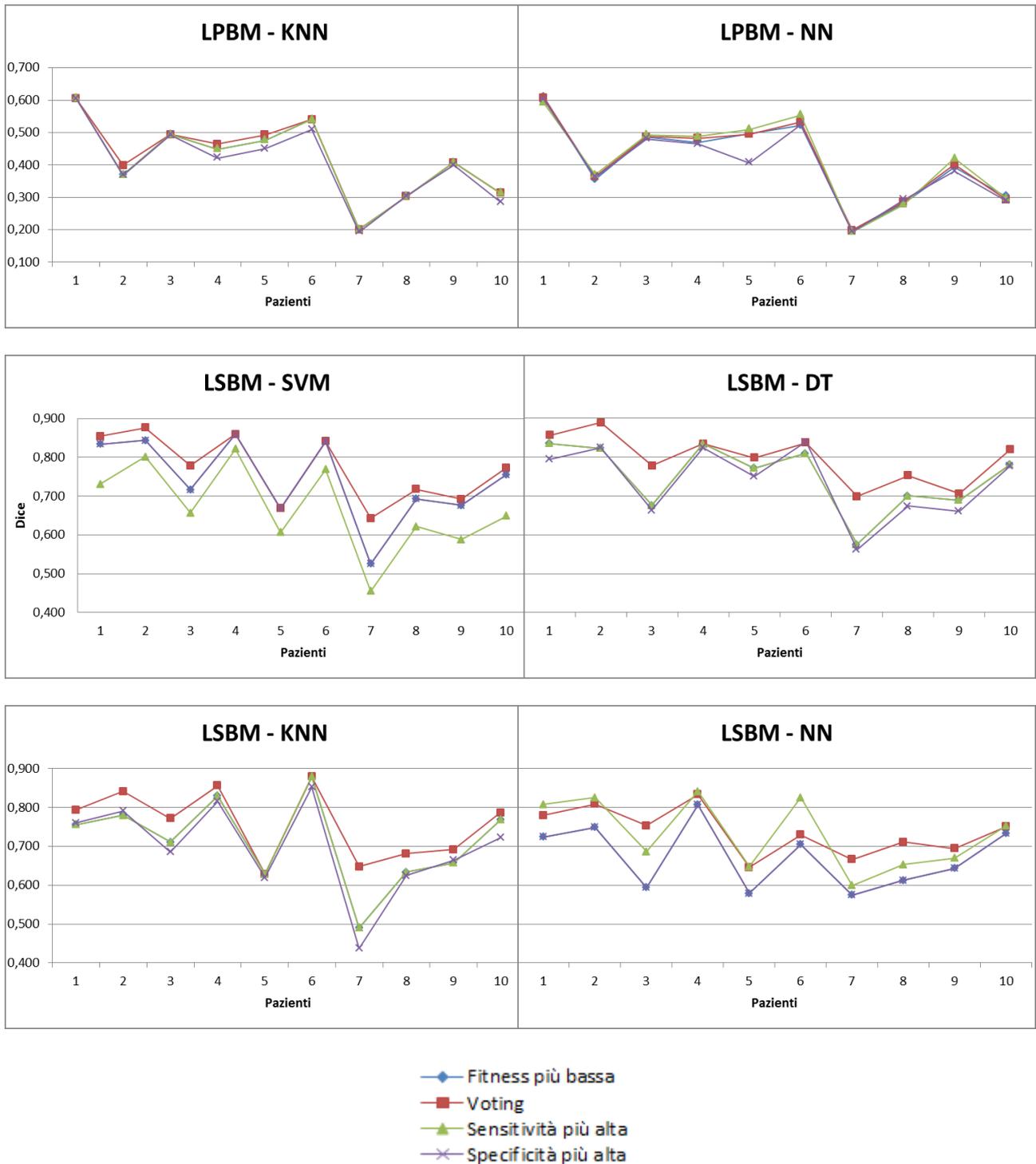
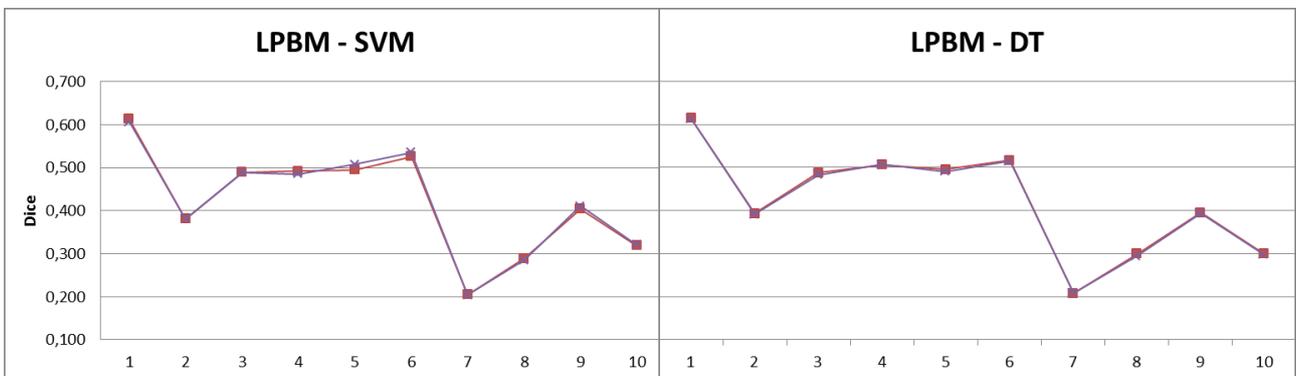
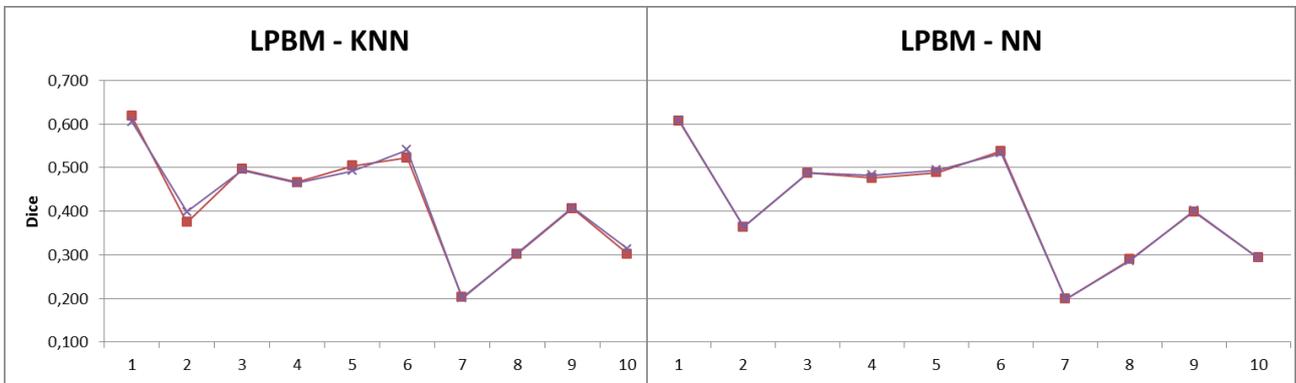
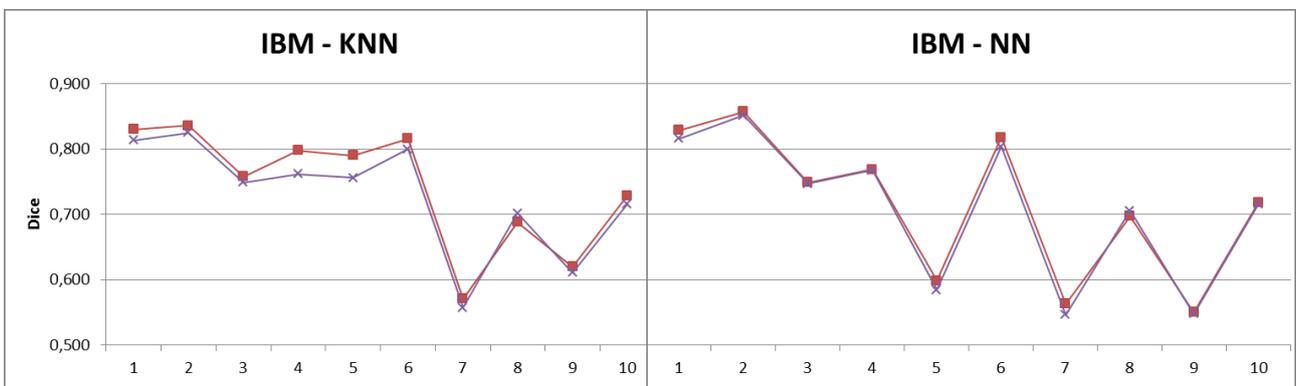
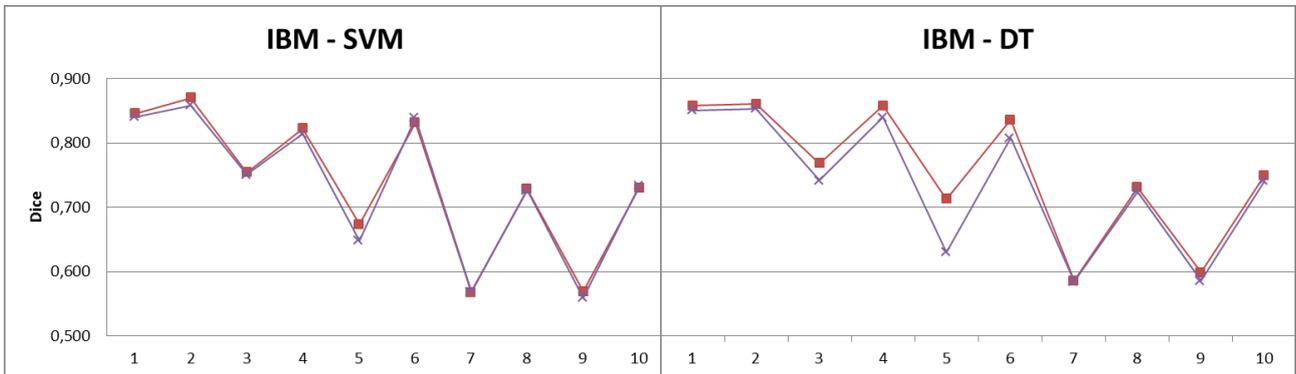


Figura 10 – Grafici che mostrano le performance espresse in Dice dei singoli classificatori e del voting. In ascissa vi sono i pazienti, in ordinata il valore Dice espresso in percentuale.

Si può osservare come l'azione del voting porti ad un generale miglioramento delle prestazioni della segmentazione. Tale miglioramento è piccolo nella struttura IBM, quasi nullo nella struttura LPBM, mentre per la struttura LSBM assume maggiore importanza: in alcuni casi rende segmentazioni buone ottenute con i classificatori singoli di pessime prestazioni. È importante far notare tuttavia che non sempre il voting porta miglioramenti di prestazione; in alcuni casi porta leggeri peggioramenti delle prestazioni, ma di entità molto piccola e poco frequentemente. I risultati confermano quindi che l'utilizzo di voting può migliorare la precisione dei classificatori singoli.

Dato che da GA si ottengono 5 soluzioni diverse, si è anche pensato di eseguire il voting non solo con tre soluzioni, ma con tutte le soluzioni disponibili. I risultati si possono osservare di seguito, sempre per i primi 10 pazienti.



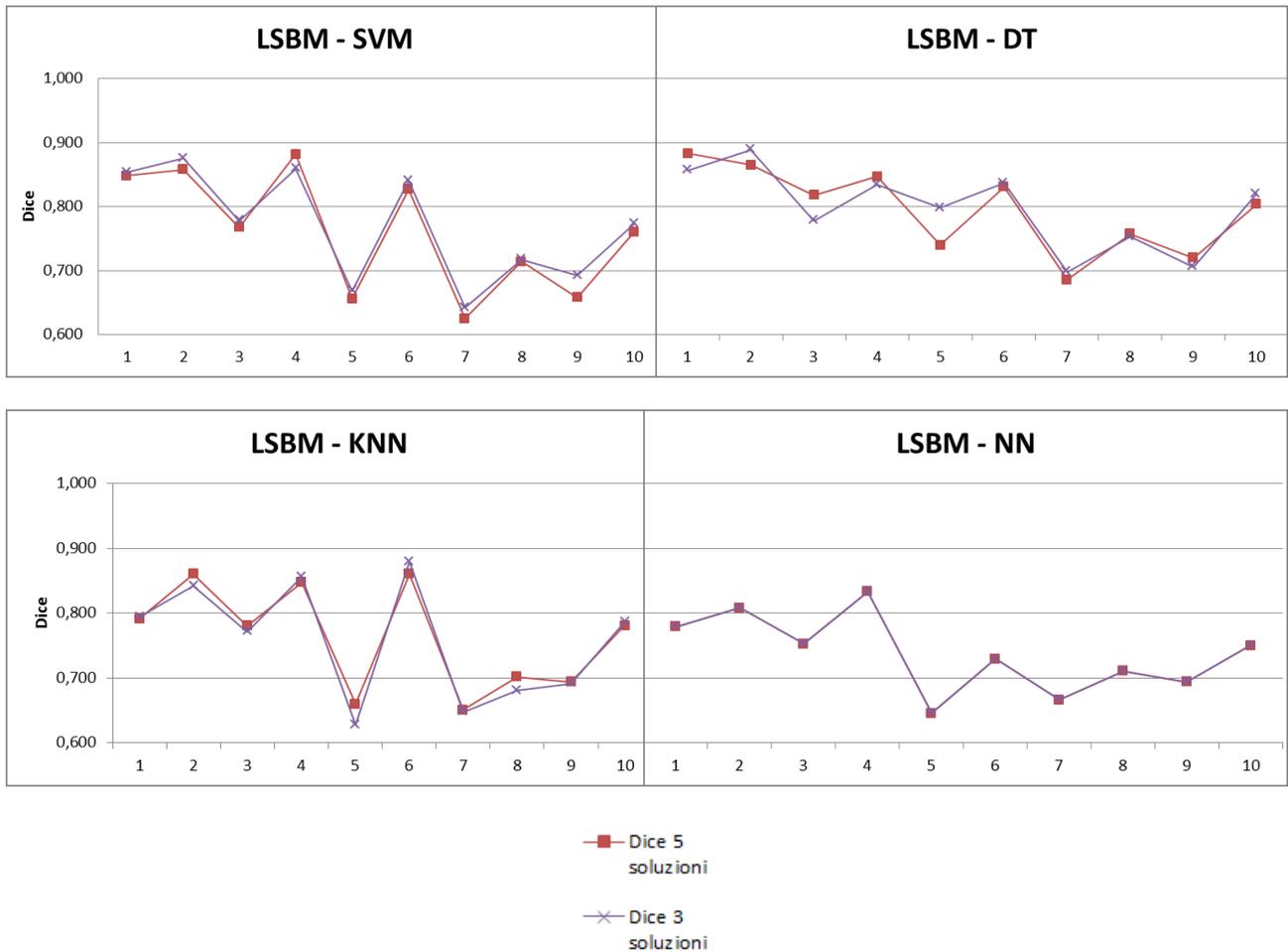


Figura 11 – Grafici che mostrano i risultati del voting con 3 soluzioni (viola) e 5 soluzioni (rosso). In ordinata vi sono i valori Dice dei pazienti mostrati in ascissa.

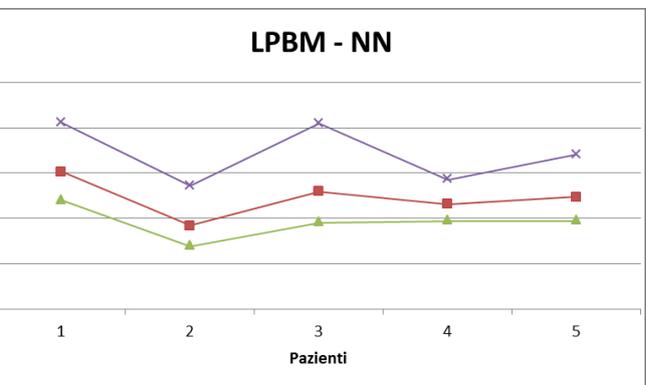
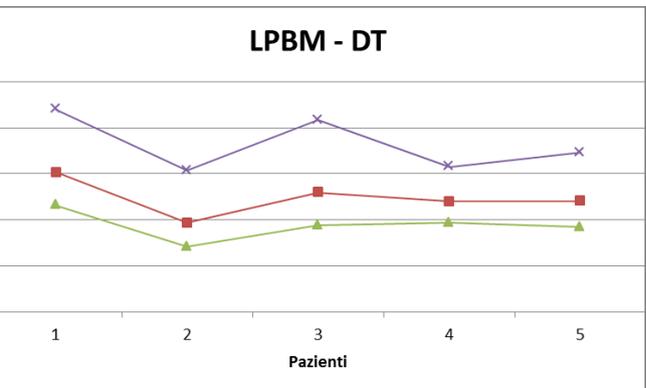
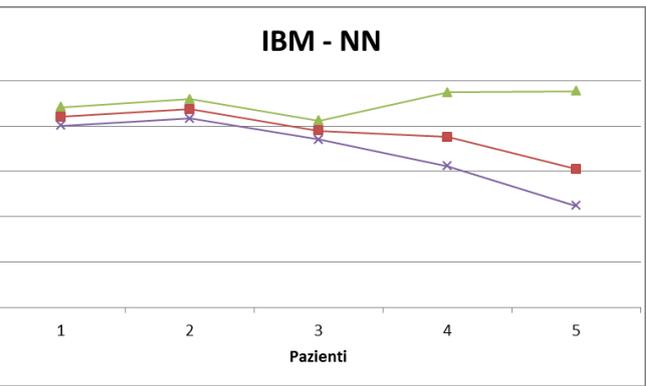
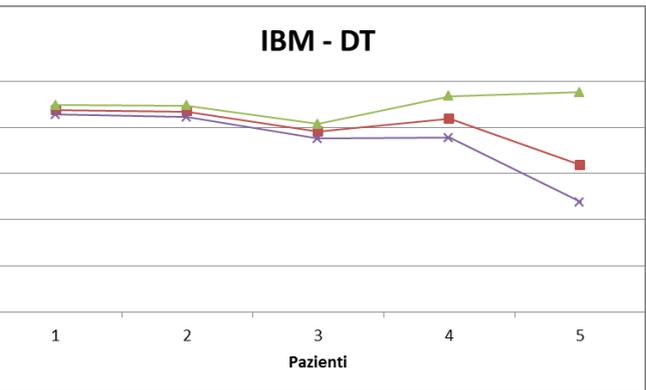
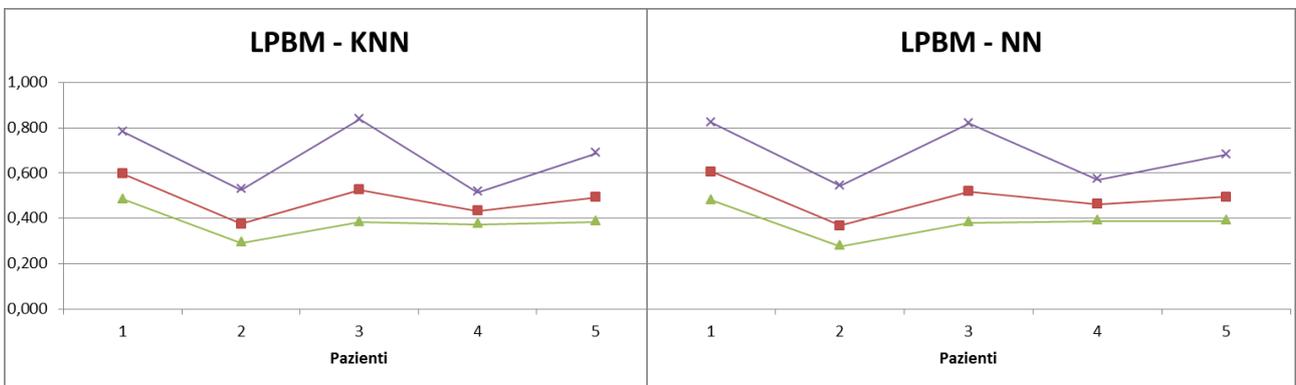
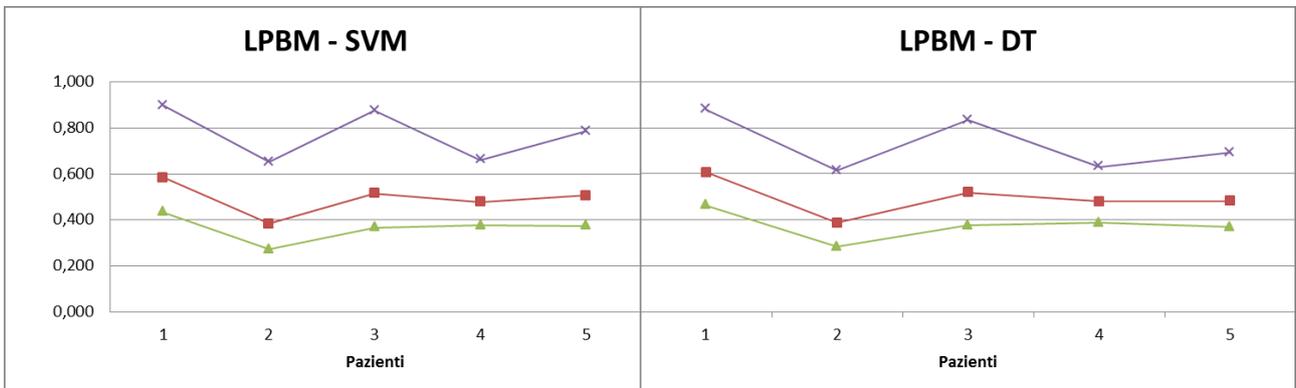
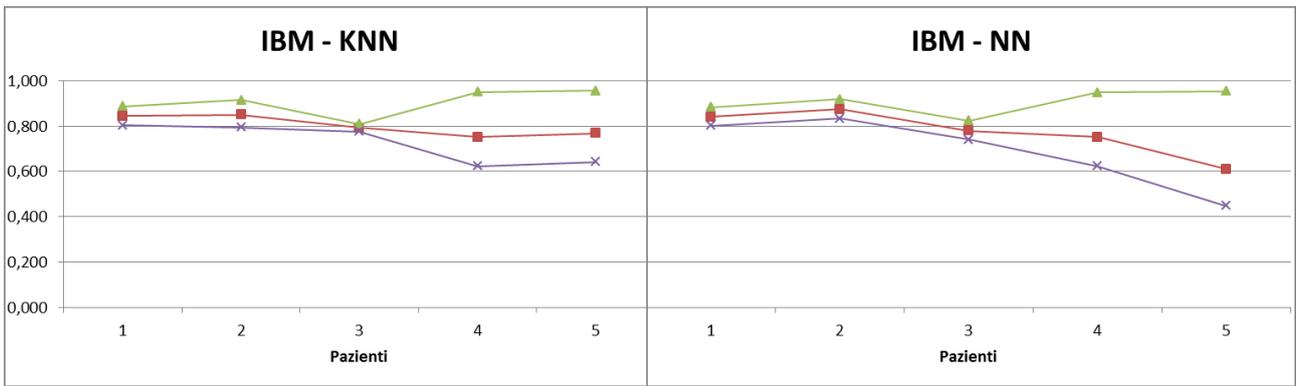
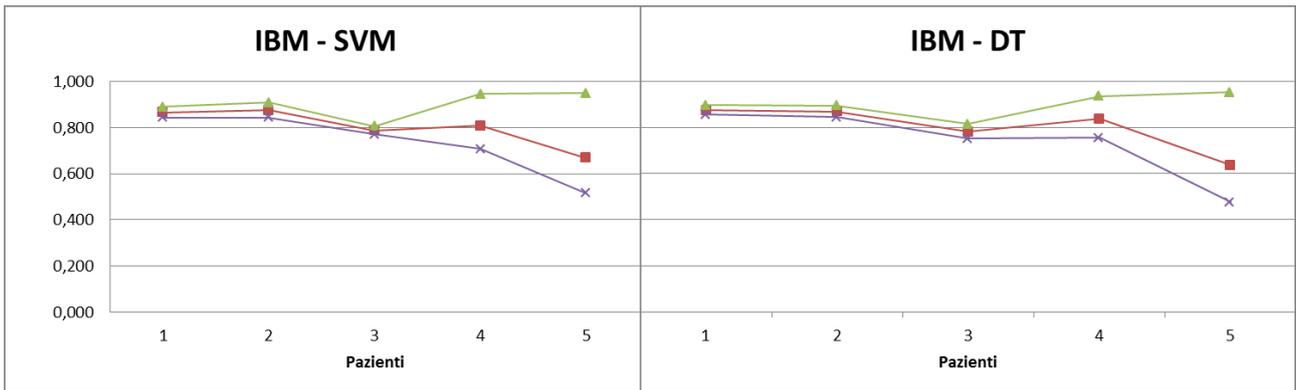
Da questi risultati è immediatamente osservabile che, in termini di Dice, l'utilizzo di tutte le 5 soluzioni da GA non porta ad alcun vantaggio rispetto al metodo adottato per la segmentazione, perciò non si rende necessario l'utilizzo di più soluzioni, che rendono più pesante il carico computazionale.

3.2. Valutazione delle maschere

Una volta scelte le soluzioni per il voting, per ogni struttura e tipologia di classificatore si è proceduto alla segmentazione del midollo osseo attivo con il metodo esposto. I seguenti risultati mostrano le performance ottenute con l'algoritmo, dopo che è stato eseguito il post-processing.

I risultati sono mostrati separatamente per i primi 5 pazienti, da cui si è ottenuto il training set sfruttato dal GA, e per i successivi 20 pazienti.

PAZIENTI 1-5



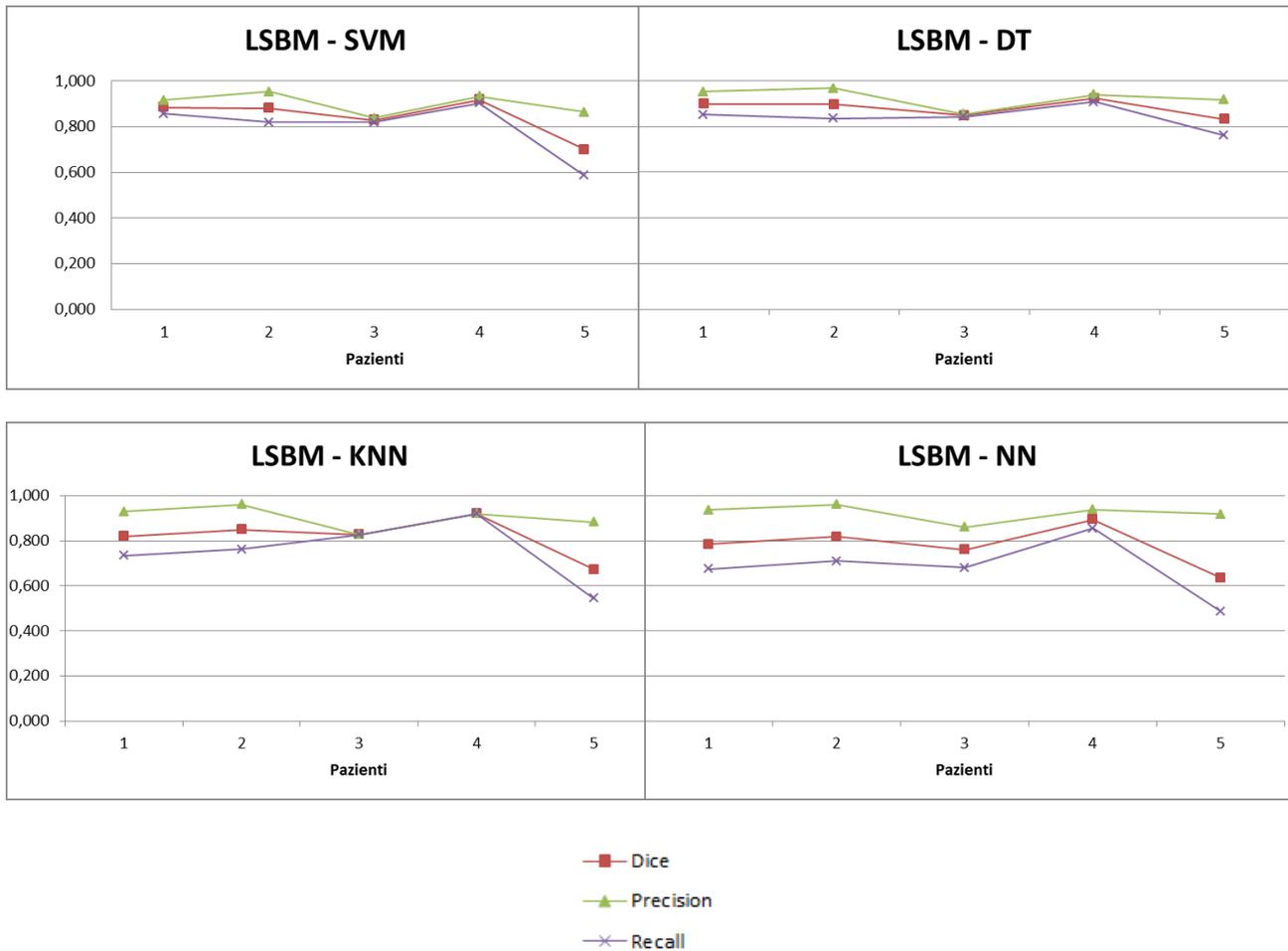


Figura 12 – Risultati della classificazione con voting per i primi 5 pazienti. In ascissa sono presenti i pazienti, in ordinata i valori di Dice, Precision e Recall, rispettivamente in rosso, verde e viola.

I grafici in figura 10 mostrano le performance ottenute tramite il calcolo dei parametri di Dice, Recall e Precision. Per le strutture di IBM e LSBM si può osservare un valore di Dice nei pazienti intorno all'80%, che nel caso di LSBM arriva anche al 90%, con l'unica eccezione del paziente 5, in cui tali performance sono sensibilmente peggiori. In queste strutture si può anche osservare un valore molto alto di Precision più basso di Recall. Questa è una indicazione di sotto-segmentazione del midollo attivo, per cui i risultati ottenuti, per ogni classificatore, sono in grado di escludere con precisione le zone di midollo giallo, al costo di non essere in grado di individuare tutti i pixel appartenenti al midollo attivo.

Per la struttura LPBM, invece, si ottengono performance nettamente peggiori, con valori di Dice che oscillano tra il 60% e il 40%. Inoltre, contrariamente alle altre due strutture, in questo caso si ottengono buoni risultati di Recall, ma pessimi di Precision. La struttura viene quindi enormemente sovra-segmentata, qualunque siano i classificatori.

Si può osservare inoltre un altro interessante risultato: a differenza dei risultati ottenuti dopo il GA, il classificatore con cui si ottengono le migliori performance non sono più le NN, ma sono SVM, KNN e anche DT. La situazione, rispetto ai risultati precedenti, si è ribaltata. Tale cambiamento è dovuto all'azione del voting. Attraverso il voting, se le 3 maschere hanno pixel diversi classificati come midollo, tendono ad aumentare la superficie di midollo attivo correttamente segmentato. Ogni maschera commette errori in maniera diversa, e tali errori vengono corretti. Se invece le maschere dei 3 classificatori sono molto simili, gli eventuali errori non vengono corretti, e le maschere semplicemente si sovrappongono tra loro. Nel caso

specifico, i classificatori SVM, KNN e DT tendono ad avere soluzioni piuttosto diverse, e attraverso il voting migliora la precisione di classificazione. Nel NN, invece, le soluzioni ottenute sono molto simili, e il voting influisce di meno.

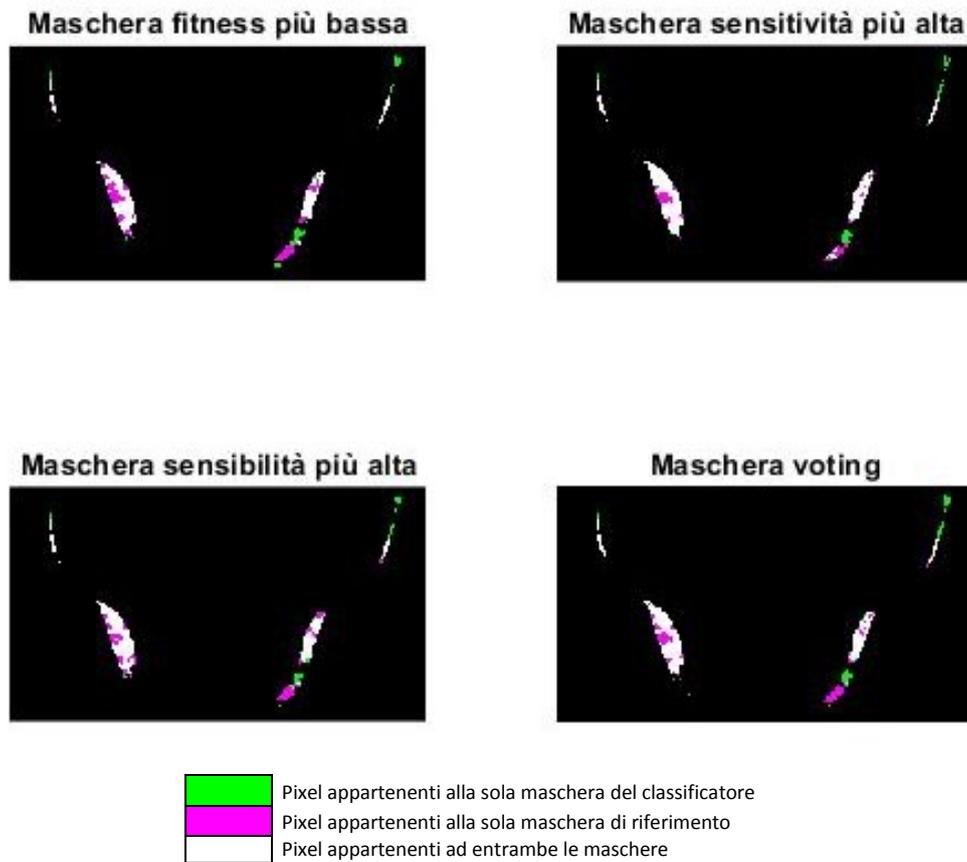
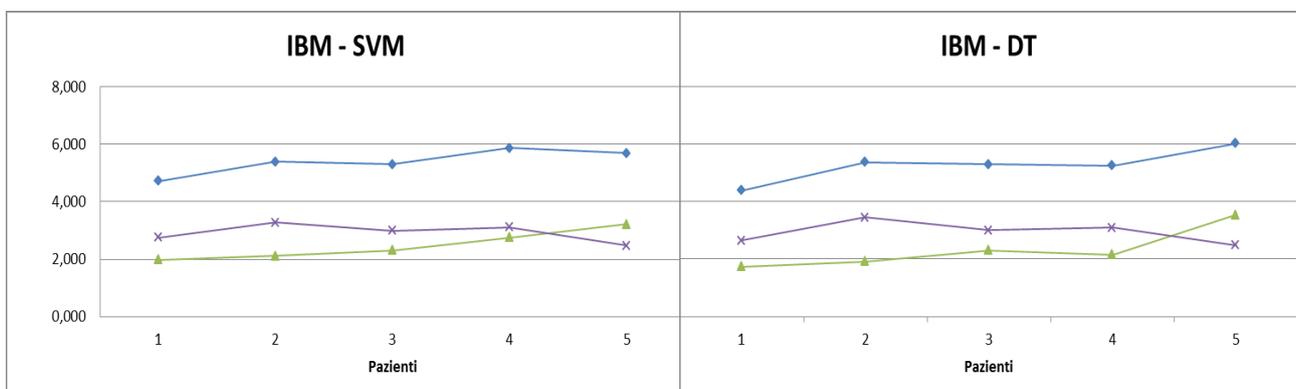
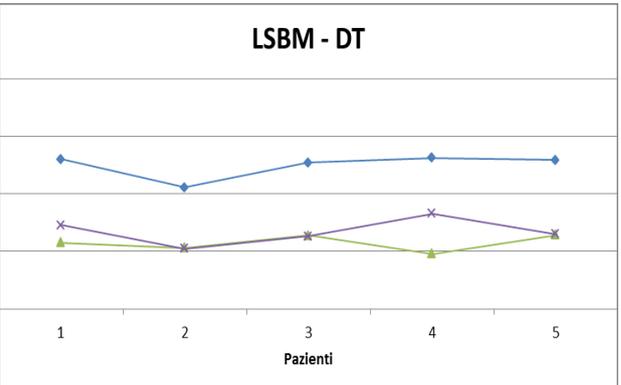
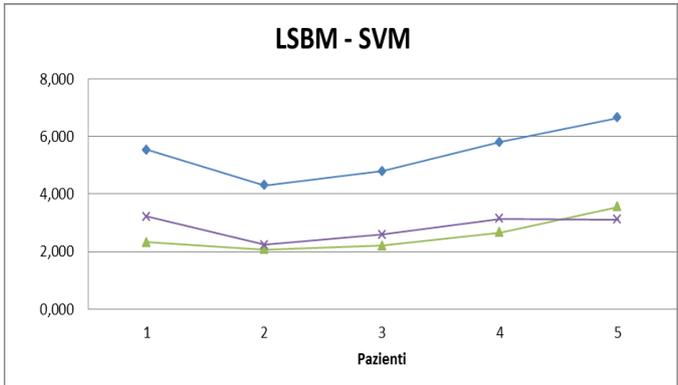
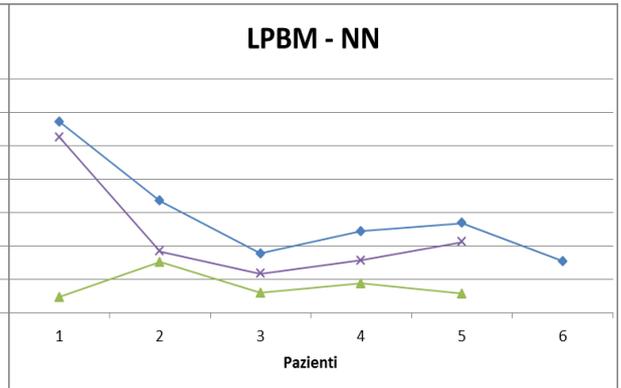
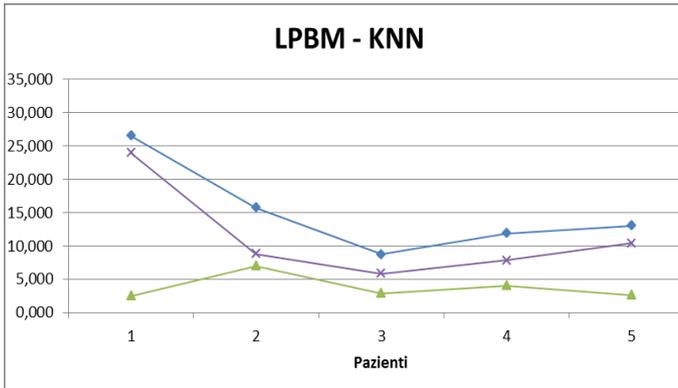
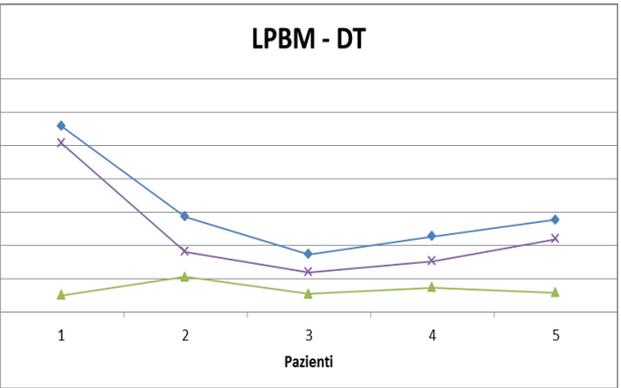
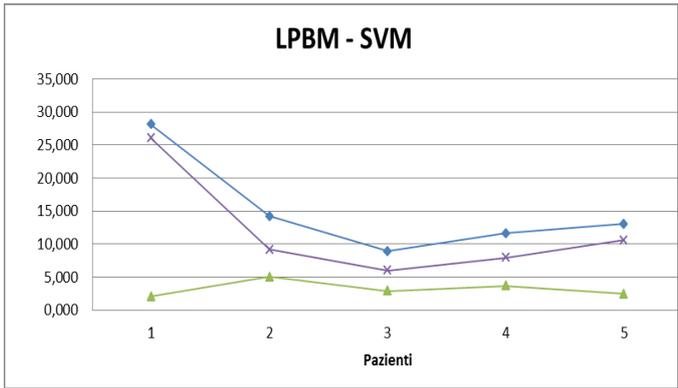
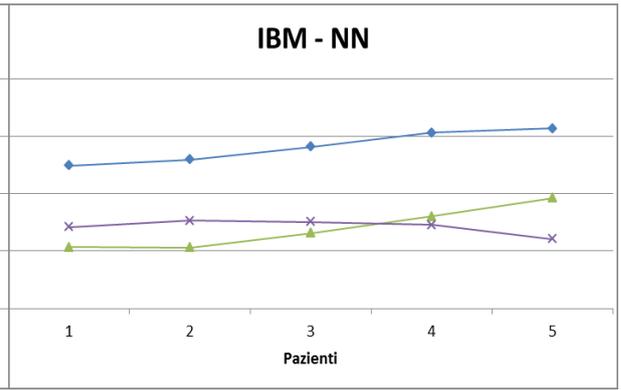
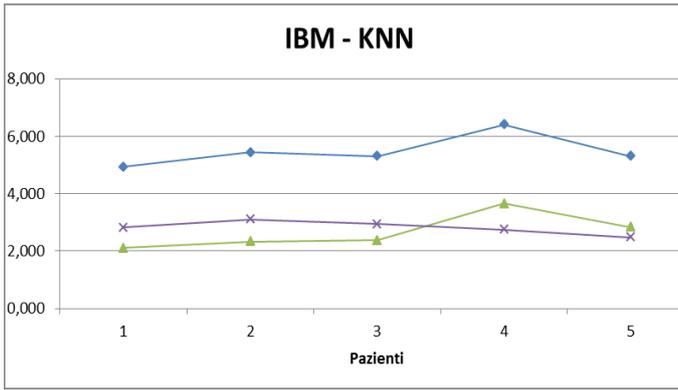


Figura 13 – Esempio dell’effetto del voting. In basso, la legenda. I pixel in verde sono i pixel appartenenti alla sola maschera del classificatore, i pixel viola alla sola maschera di riferimento, i pixel bianchi appartengono a entrambe. Le figure mostrano le 3 maschere prese dai singoli classificatori e la maschera del voting; si può osservare come il grado di sovrapposizione migliora nello specifico esempio, anche se di poco.

Di seguito vi sono gli stessi grafici visti in precedenza, dove si osservano invece i risultati espressi in MDC, OC e UC.





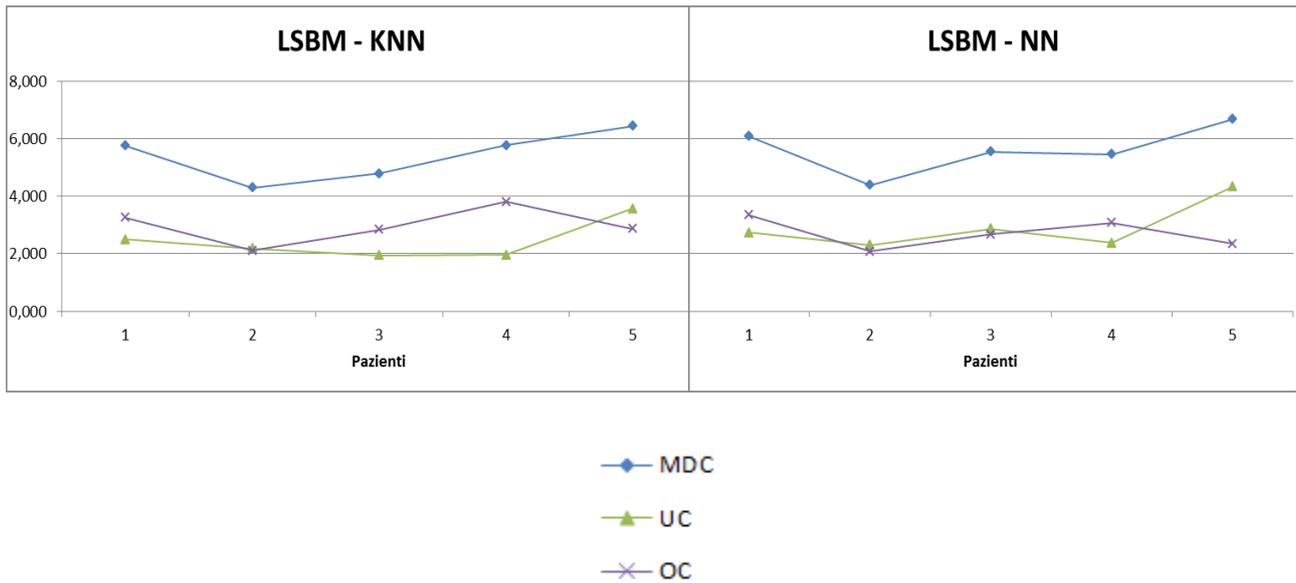
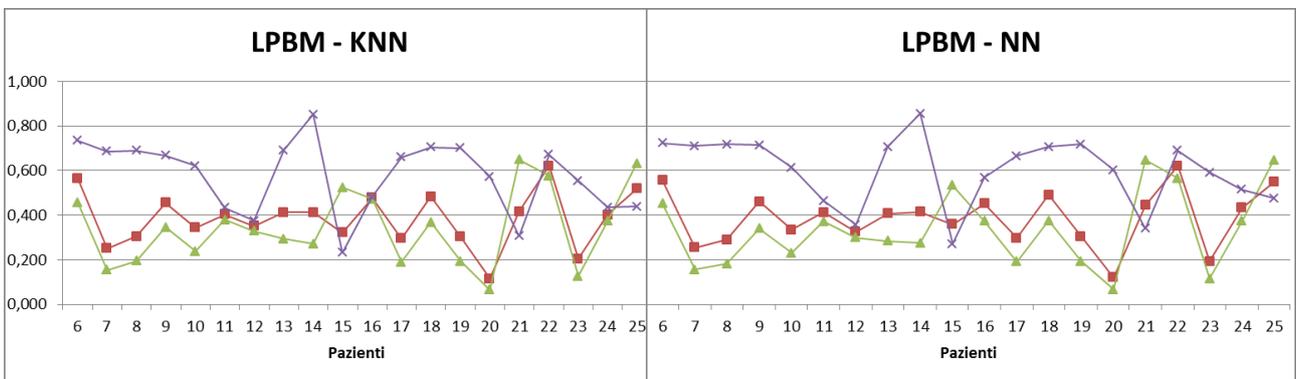
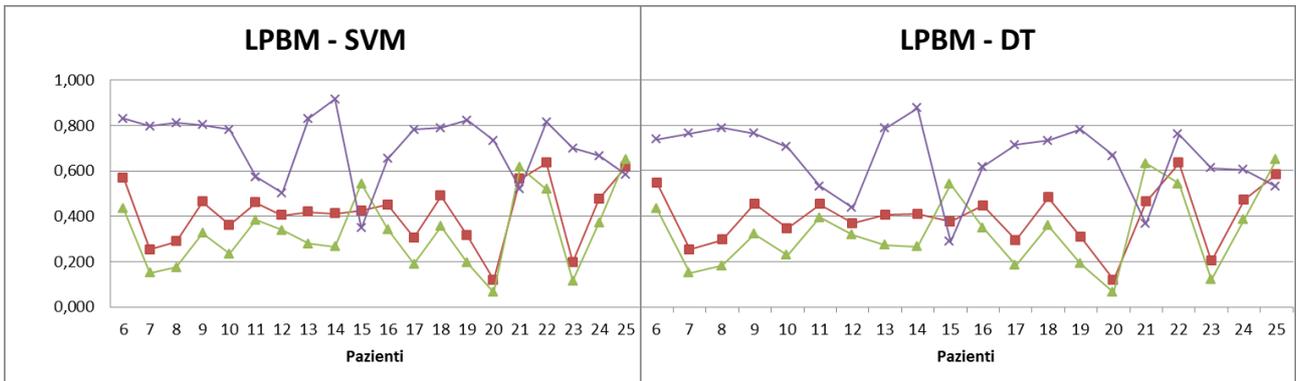
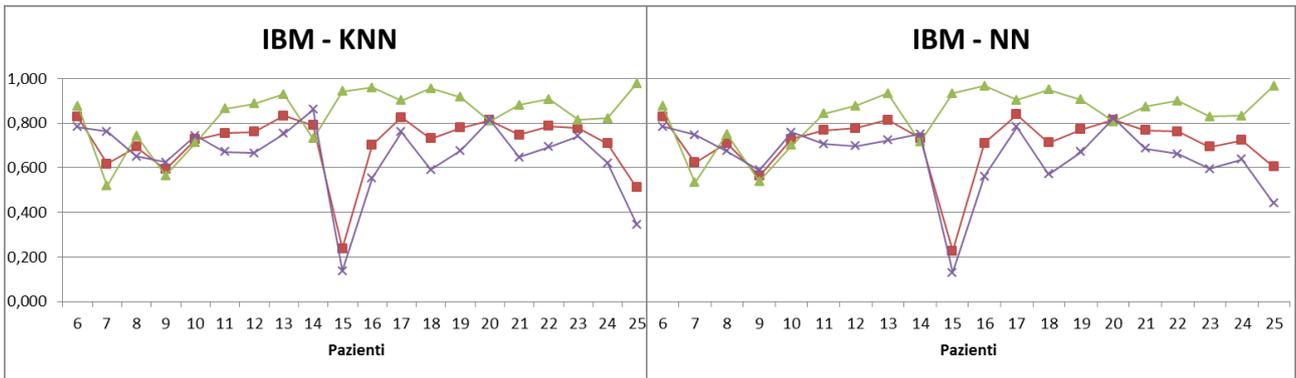
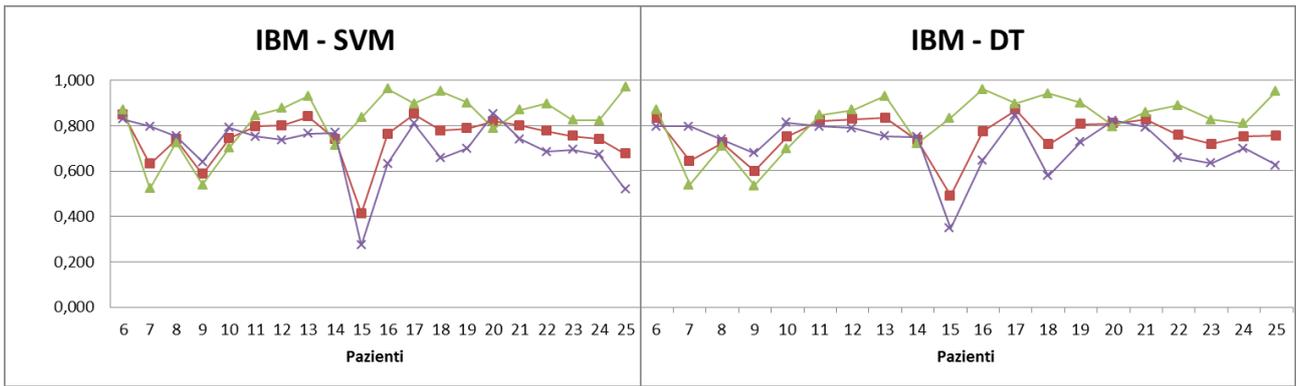
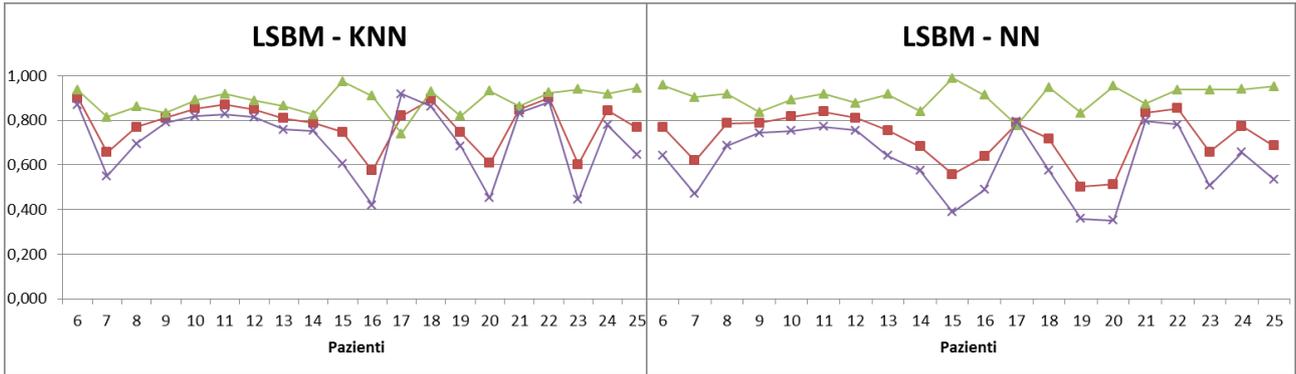
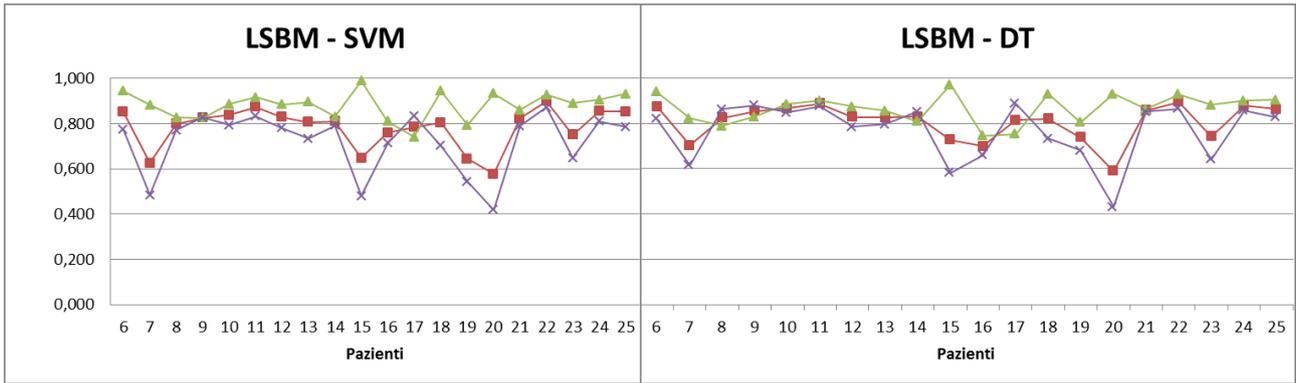


Figura 14 – Risultati della classificazione delle immagini per i primi 5 pazienti. I grafici hanno i pazienti in ascissa e i valori di MDC (blu), UC (verde) e OC (viola) espressi in mm.

I medesimi risultati osservati attraverso i valori di Dice, Recall e Precision si osservano anche per MDC, UC e OC. Le strutture IBM e LSBM hanno valori piuttosto buoni, segno di buone performance della segmentazione, mentre la struttura LPBM ha valori decisamente peggiori. Se per IBM e LSBM la MDC ha valori intorno ai 5 mm, nella LPBM si arriva fino ai 15 mm. Per il paziente 1 in LPBM si osserva una alta impennata, con MDC di più di 30 mm. Se per i valori di MDC la differenza in termini di valutazione risultati non varia molto rispetto al Dice, si osserva che i valori di UC e OC danno una interpretazione diversa. Si può osservare come nelle strutture IBM e LSBM i classificatori hanno valori generalmente maggiori di OC rispetto a UC, che suggerisce una tendenza a sovra-segmentare che contraddice i risultati visto con Precision e Recall. Questo è dovuta alla natura morfologica di OC e UC. Questo significa che in termini di numero di pixel classificati, vi sono più pixel sotto-segmentati (falsi negativi) che sovra-segmentati (falsi positivi), ma che la forma della maschera è tale che i pixel sovra-segmentatisi trovano a distanze maggiori rispetto ai pixel sotto-segmentati. A distanze maggiori corrisponde un valore maggiore di OC.

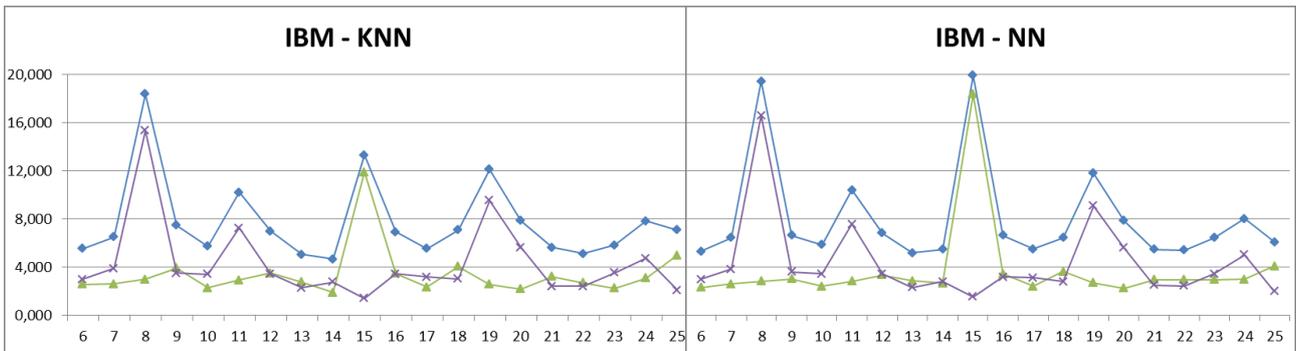
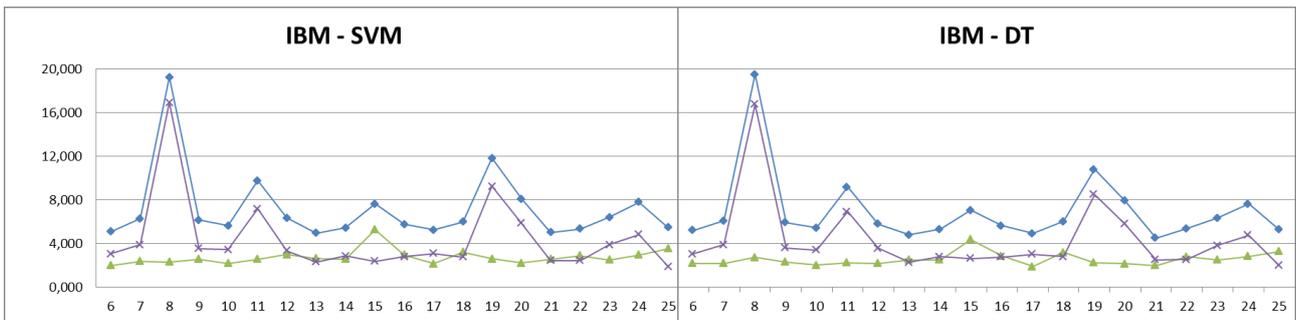
PAZIENTI 6-25





- Dice
- ▲ Precision
- × Recall

Figura 15 – Grafici che mostrano i risultati di Dice, Precision e Recall per i pazienti 6-25.



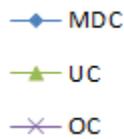
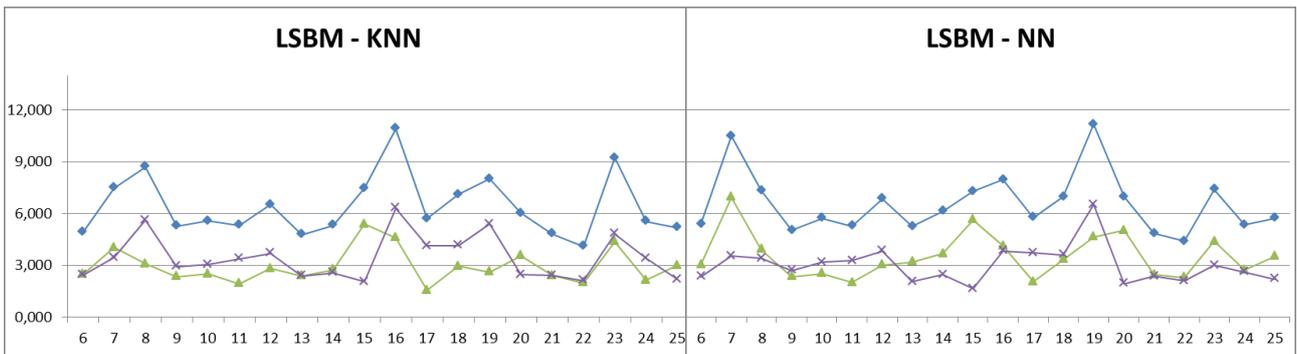
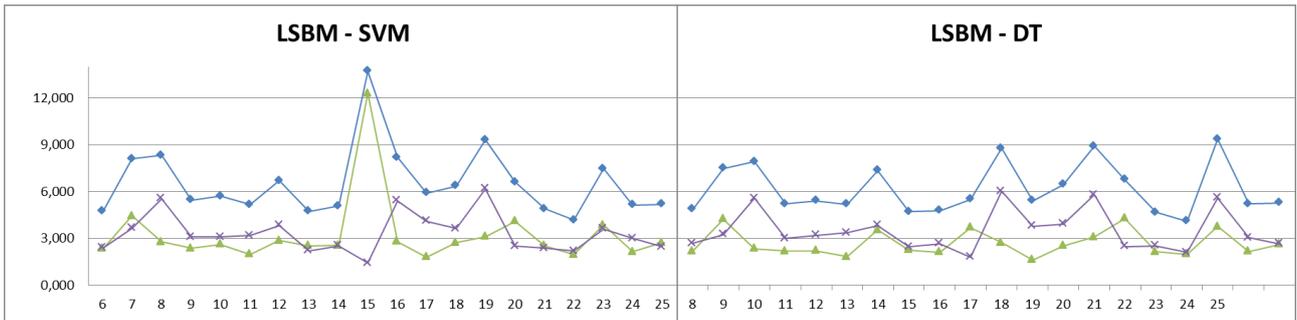
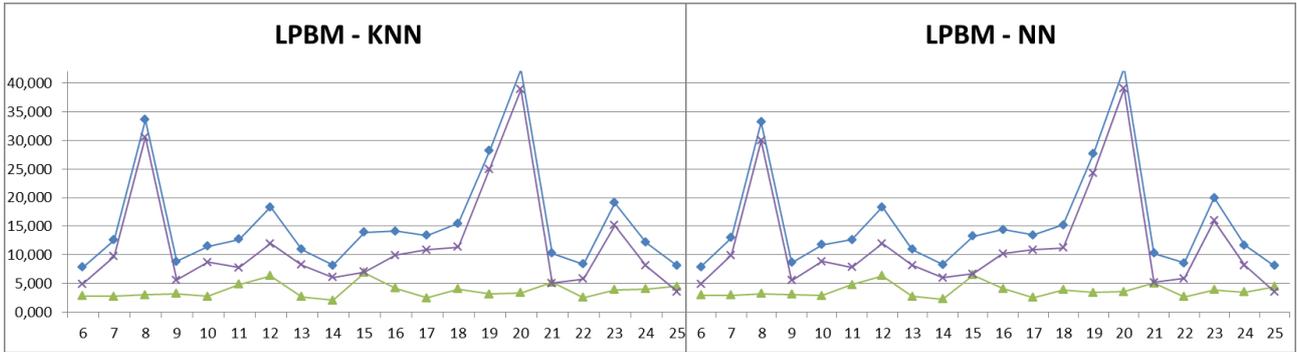
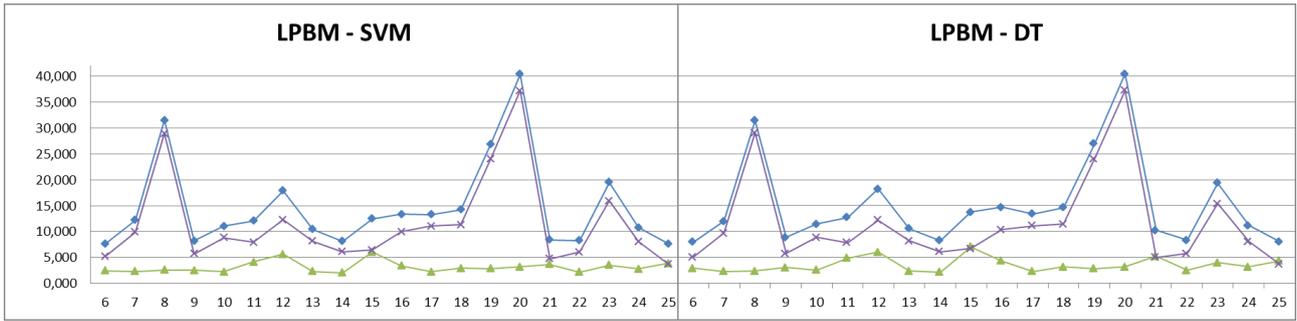


Figura 16 – Grafici che mostrano i risultati espressi in MDC, OC, UC dei pazienti 6-25.

Nei grafici si osservano i risultati ottenuti con i pazienti da 6 a 25, che sono i pazienti non utilizzati per la creazione del training set. I risultati rimangono sostanzialmente in linea con i pazienti 1-5, ma alcuni pazienti possiedono risultati con performance nettamente peggiori. Si osserva che la struttura LPBM viene enormemente sovra-segmentata da ogni classificatore, mentre IBM e LSBM hanno performance migliori.

In generale, i pazienti 6-25 hanno risultati di segmentazione tendenzialmente meno precisi. Questo è un risultato comprensibile, dato che il training set sfrutta solo i primi 5 pazienti, e altri pazienti possono avere caratteristiche diverse.

Alcuni pazienti, come visto prima, possiedono un valore di MDC estremamente più alto. Un esempio è il paziente 8 nella struttura IBM.

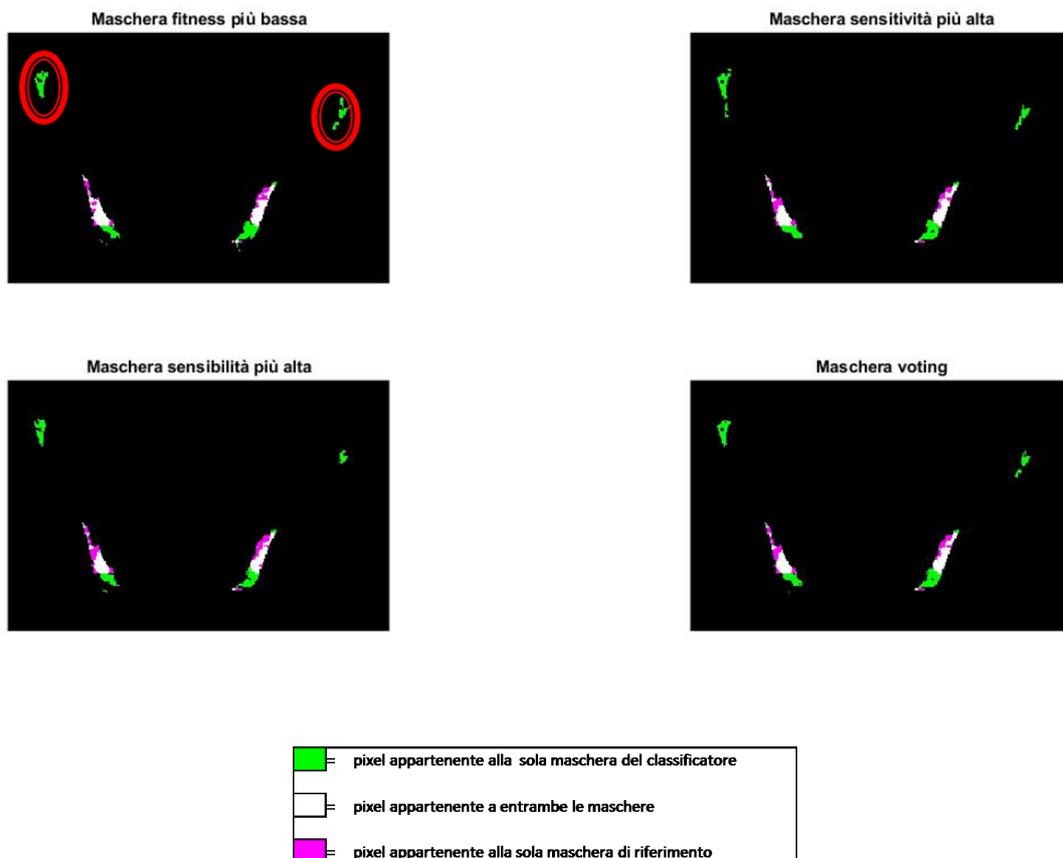


Figura 17 – Maschera ottenuta per struttura IBM del paziente 8. I pixel in verde sono i pixel appartenenti alla sola maschera del classificatore, i pixel viola alla sola maschera di riferimento, i pixel bianchi appartengono a entrambi.

Si può osservare che il classificatore individua delle piccole zone di pixel al di fuori del riferimento di midollo attivo. Tali zone non influiscono molto sul valore Dice, ma la maggiore distanza influisce sulla MDC in maniera molto più marcata, aumentandone di molto il valore. L'esempio è una prova del maggior numero di informazioni che consente di fornire il parametro MDC, il quale possiede valori che variano anche al variare della forma assunta dalla segmentazione. Inoltre, dato che ad essere calcolata è una misura di distanza, anche pochi pixel possono influenzare molto il valore di MDC, rendendo il parametro più sensibile rispetto al Dice.

3.4. Post-processing

Di seguito un esempio dell'effetto delle operazioni per il post-processing. Si può osservare dall'immagine come l'aspetto della segmentazione migliora, diventando meno frastagliata e richiudendo parte dei buchi.

Oltre a migliorare l'aspetto, in generale migliorano anche i valori di Dice e MDC, grazie al riempimento dei buchi.

DT

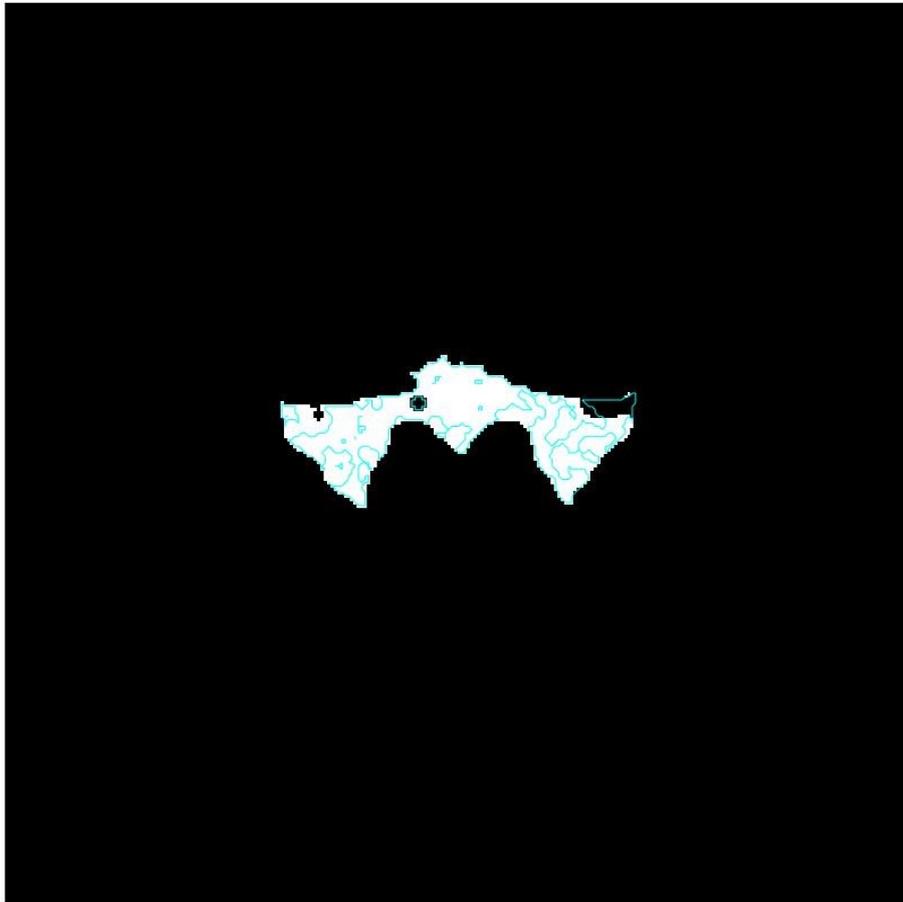


Figura 18 – Segmentazione di una slice della struttura LSBM del paziente 14, senza applicazione di chiusura della maschera.

DT

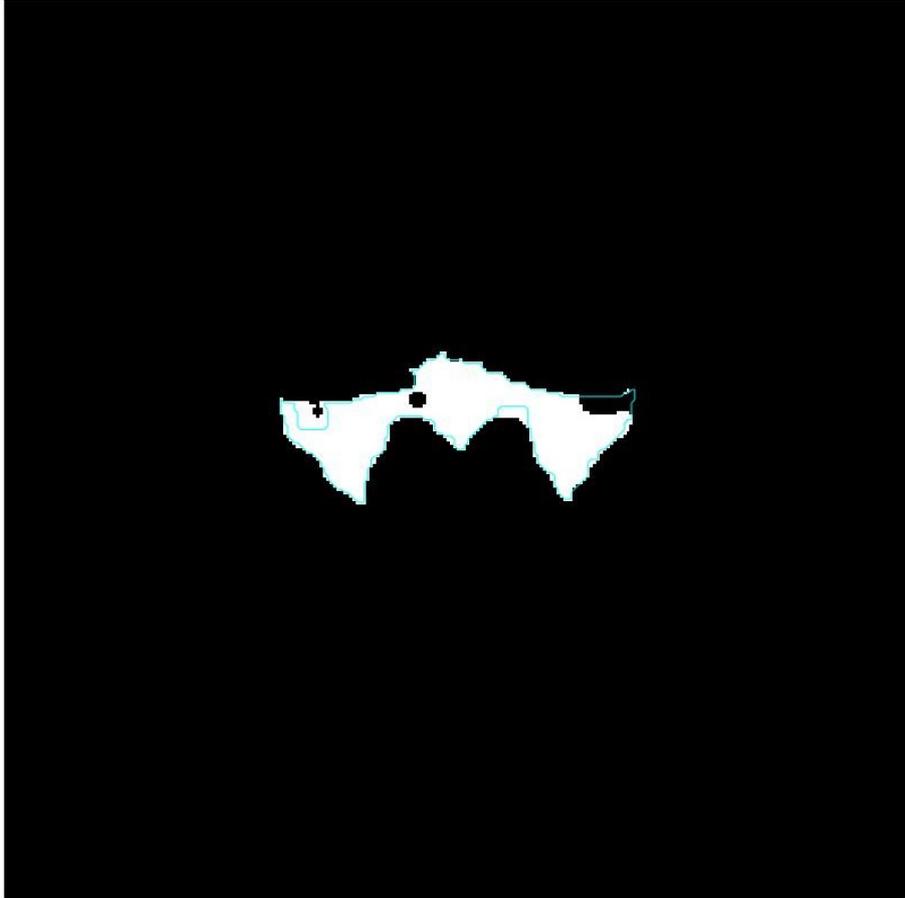


Figura 19 - Segmentazione di una slice della struttura LSBM del paziente 14, con applicazione di chiusura della maschera

Di seguito sono mostrate delle tabelle contenenti i valori numerici medi, la deviazione standard e i valori massimi e minimi dei parametri calcolati, separatamente per ogni classificatore. Tali valori sono stati calcolati alla fine del post-processing

DT

Pazienti 1:5		DT					
		Dice	Precision	Recall	MDC	UC	OC
IBM	Media	0,799	0,899	0,737	5,478	2,437	3,041
	Std	0,099	0,053	0,153	0,643	0,767	0,400
	Max	0,875	0,953	0,855	6,292	3,728	3,544
	Min	0,635	0,815	0,477	4,500	1,765	2,564
LPBM	Media	0,494	0,375	0,730	15,680	3,618	12,061
	Std	0,079	0,064	0,120	7,390	1,232	7,809
	Max	0,606	0,462	0,880	28,229	5,529	25,614
	Min	0,387	0,282	0,614	8,864	2,615	6,098
LSBM	Media	0,880	0,926	0,839	5,166	2,390	2,776
	Std	0,038	0,044	0,052	0,482	0,292	0,525
	Max	0,923	0,967	0,907	5,557	2,697	3,544
	Min	0,832	0,854	0,761	4,329	2,013	2,124

Pazienti 6:25		DT					
		Dice	Precision	Recall	MDC	UC	OC
IBM	Media	0,752	0,819	0,714	6,941	2,560	4,381
	Std	0,091	0,122	0,115	3,345	0,583	3,347
	Max	0,869	0,960	0,844	19,503	4,400	16,769
	Min	0,490	0,536	0,347	4,507	1,903	1,996
LPBM	Media	0,395	0,330	0,653	15,129	3,552	11,577
	Std	0,127	0,167	0,155	8,584	1,372	8,738
	Max	0,633	0,650	0,875	40,407	7,054	37,198
	Min	0,119	0,065	0,287	7,988	2,177	3,701
LSBM	Media	0,806	0,866	0,767	6,170	2,680	3,490
	Std	0,080	0,064	0,126	1,600	0,805	1,287
	Max	0,895	0,971	0,886	9,372	4,281	6,034
	Min	0,589	0,746	0,431	4,104	1,632	1,799

Tabella 5 - Tabelle che mostrano i valori numerici dei risultati per i classificatori DT con voting, per i pazienti 1-5 (a sinistra) e i pazienti 6-25 (a destra). Le tabelle mostrano la media, la deviazione standard, il massimo e il minimo dei risultati.

KNN

Pazienti 1:5		KNN					
		Dice	Precision	Recall	MDC	UC	OC
IBM	Media	0,801	0,904	0,727	5,735	2,810	2,925
	Std	0,044	0,060	0,088	0,752	0,780	0,238
	Max	0,850	0,956	0,804	7,035	4,107	3,215
	Min	0,752	0,809	0,623	5,085	2,162	2,556
LPBM	Media	0,485	0,384	0,669	15,664	4,040	11,624
	Std	0,085	0,068	0,146	6,766	1,999	7,246
	Max	0,598	0,484	0,836	26,837	7,330	24,228
	Min	0,375	0,292	0,515	8,976	2,610	6,008
LSBM	Media	0,819	0,905	0,758	5,647	2,529	3,118
	Std	0,090	0,051	0,139	0,969	0,717	0,757
	Max	0,920	0,962	0,920	6,707	3,740	4,211
	Min	0,674	0,828	0,544	4,388	1,988	2,133

Pazienti 6:25		KNN					
		Dice	Precision	Recall	MDC	UC	OC
IBM	Media	0,711	0,836	0,655	7,749	3,423	4,326
	Std	0,139	0,127	0,165	3,384	2,130	3,202
	Max	0,833	0,979	0,861	18,354	11,902	15,348
	Min	0,238	0,518	0,136	4,683	1,894	1,407
LPBM	Media	0,383	0,342	0,576	15,502	3,770	11,732
	Std	0,122	0,169	0,162	9,133	1,289	9,220
	Max	0,620	0,649	0,852	42,281	6,882	38,869
	Min	0,116	0,064	0,232	7,840	2,137	3,606
LSBM	Media	0,783	0,887	0,720	6,403	2,950	3,453
	Std	0,101	0,058	0,154	1,759	0,984	1,263
	Max	0,902	0,974	0,919	10,925	5,393	6,318
	Min	0,575	0,739	0,421	4,103	1,566	2,065

Tabella 6 - Tabelle che mostrano i valori numerici dei risultati per i classificatori KNN con voting, per i pazienti 1-5 (a sinistra) e i pazienti 6-25 (a destra). Le tabelle mostrano la media, la deviazione standard, il massimo e il minimo dei risultati.

NN

Pazienti 1:5		NN					
		Dice	Precision	Recall	MDC	UC	OC
<i>IBM</i>	Media	0,771	0,906	0,689	5,904	2,948	2,956
	Std	0,102	0,054	0,157	0,707	0,839	0,271
	Max	0,874	0,955	0,833	6,747	4,011	3,161
	Min	0,610	0,824	0,448	5,139	2,172	2,496
<i>LPBM</i>	Media	0,490	0,383	0,687	16,440	4,214	12,226
	Std	0,087	0,072	0,132	7,599	2,287	8,237
	Max	0,606	0,479	0,823	28,968	7,978	26,602
	Min	0,367	0,277	0,544	9,026	2,366	5,984
<i>LSBM</i>	Media	0,779	0,924	0,682	5,863	3,058	2,805
	Std	0,094	0,038	0,131	0,895	0,858	0,563
	Max	0,895	0,962	0,855	6,938	4,530	3,435
	Min	0,636	0,862	0,487	4,501	2,373	2,128

Pazienti 6:25		NN					
		Dice	Precision	Recall	MDC	UC	OC
<i>IBM</i>	Media	0,709	0,832	0,650	8,065	3,691	4,374
	Std	0,135	0,127	0,154	4,315	3,488	3,419
	Max	0,839	0,967	0,822	19,924	18,384	16,595
	Min	0,227	0,533	0,129	5,190	2,262	1,540
<i>LPBM</i>	Media	0,386	0,333	0,601	15,508	3,780	11,729
	Std	0,125	0,169	0,153	9,123	1,172	9,146
	Max	0,621	0,647	0,856	42,623	6,514	39,015
	Min	0,120	0,066	0,271	7,914	2,341	3,618
<i>LSBM</i>	Media	0,719	0,906	0,614	6,572	3,550	3,022
	Std	0,108	0,052	0,150	1,764	1,292	1,074
	Max	0,853	0,990	0,798	11,173	6,956	6,531
	Min	0,501	0,777	0,351	4,411	2,020	1,646

Tabella 7 - Tabelle che mostrano i valori numerici dei risultati per i classificatori NN con voting, per i pazienti 1-5 (a sinistra) e i pazienti 6-25 (a destra). Le tabelle mostrano la media, la deviazione standard, il massimo e il minimo dei risultati.

SVM

Pazienti 1:5		SVM					
		Dice	Precision	Recall	MDC	UC	OC
IBM	Media	0,801	0,900	0,736	5,625	2,590	3,034
	Std	0,083	0,058	0,136	0,567	0,580	0,348
	Max	0,875	0,949	0,844	6,387	3,361	3,366
	Min	0,669	0,806	0,516	4,864	2,025	2,549
LPBM	Media	0,494	0,364	0,774	15,627	3,389	12,239
	Std	0,073	0,058	0,115	7,440	1,261	8,041
	Max	0,585	0,434	0,897	28,428	5,242	26,290
	Min	0,384	0,272	0,651	9,150	2,138	6,182
LSBM	Media	0,842	0,900	0,797	5,670	2,690	2,979
	Std	0,086	0,048	0,122	1,034	0,655	0,492
	Max	0,917	0,954	0,903	6,945	3,718	3,402
	Min	0,699	0,837	0,587	4,393	2,117	2,276

Pazienti 6:25		SVM					
		Dice	Precision	Recall	MDC	UC	OC
IBM	Media	0,744	0,822	0,704	7,189	2,767	4,422
	Std	0,103	0,126	0,128	3,333	0,705	3,453
	Max	0,851	0,970	0,853	19,257	5,278	16,923
	Min	0,412	0,523	0,274	4,975	2,015	1,919
LPBM	Media	0,411	0,327	0,713	14,707	3,143	11,564
	Std	0,136	0,164	0,143	8,745	1,099	8,736
	Max	0,634	0,650	0,916	40,334	6,001	37,162
	Min	0,119	0,065	0,347	7,642	2,030	3,741
LSBM	Media	0,782	0,881	0,719	6,556	3,232	3,324
	Std	0,090	0,061	0,133	2,218	2,246	1,234
	Max	0,898	0,989	0,871	13,727	12,312	6,197
	Min	0,578	0,740	0,418	4,174	1,813	1,415

Tabella 8 – Tabelle che mostrano i valori numerici dei risultati per i classificatori SVM con voting, per i pazienti 1-5 (a sinistra) e i pazienti 6-25 (a destra). Le tabelle mostrano la media, la deviazione standard, il massimo e il minimo dei risultati.

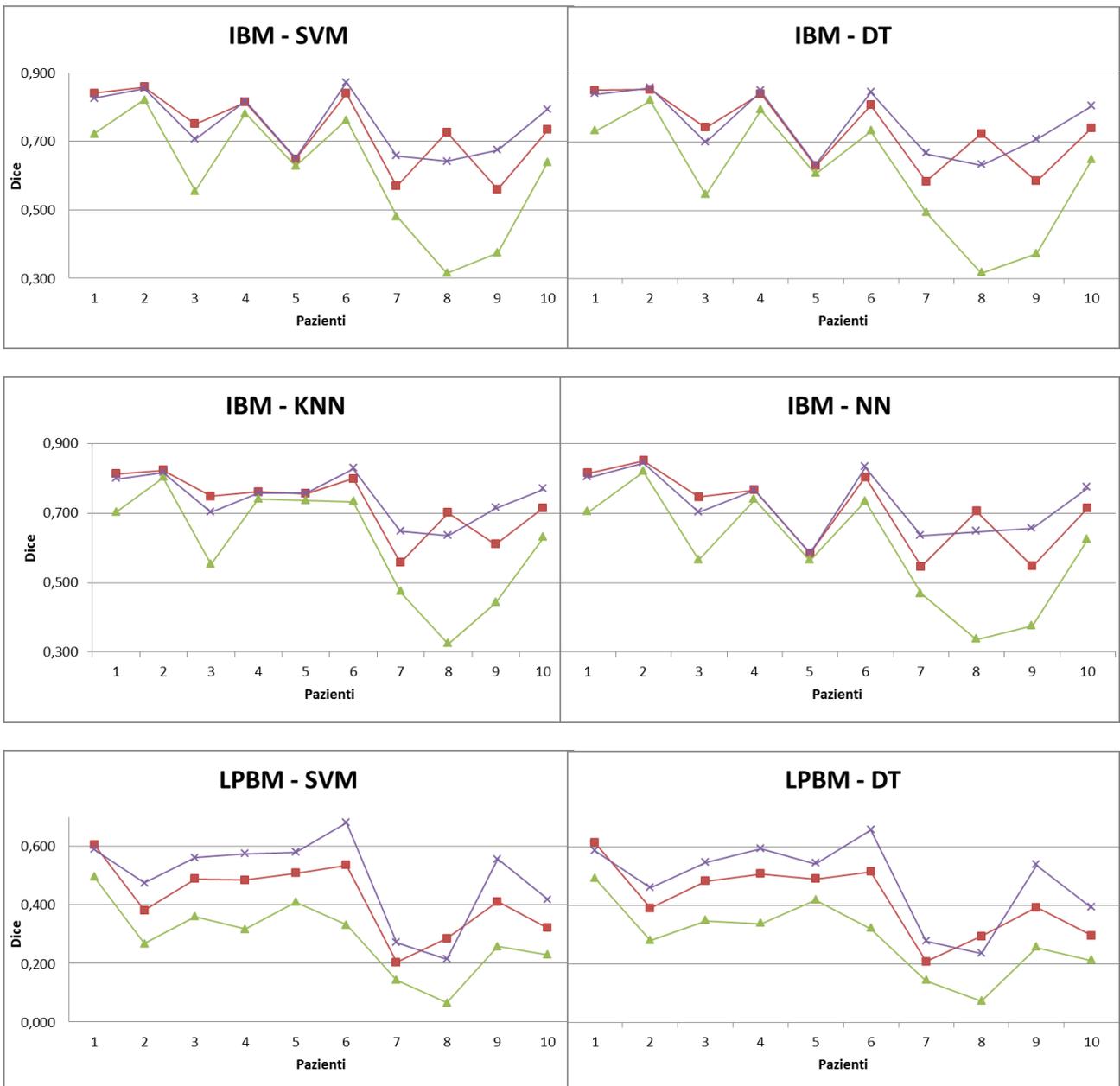
Osservando i valori numerici si possono confermare le osservazioni fatte attraverso i grafici. I valori di Dice tendono ad essere leggermente maggiori nei primi 5 pazienti, i cui pixel sono stati usati per il training set. Inoltre presentano anche deviazione standard mediamente più bassa, per cui i primi 5 pazienti hanno anche risultati che sembrano essere più stabili. Tali differenze tuttavia sono di bassissima entità, di fatto trascurabili, e sono probabilmente dovute a singoli pazienti aventi valori di Dice bassissimi, che diminuiscono la media e aumentano la deviazione standard. Per avere una idea più precisa, si è ragionato sui valori massimi e minimi ottenuti. Osservando i valori minimi, si può vedere come nei pazienti non sfruttati per il training set vi sono valori estremamente bassi di Dice. Questi influiscono sul valore di media, ma data la bassa differenza con i valori medi dei primi 5 pazienti, significa che le prestazioni ottenute nei due sotto insiemi di pazienti sono confrontabili, a meno di poche eccezioni. Osservando i massimi, invece, non si notano differenze significative, anche se nei pazienti non utilizzati per il training tali valori di massimo sono sempre leggermente minori. La tendenza dei pazienti 1-5 di avere risultati migliori è prevedibile, dato che sono stati utilizzati per il training. Gli altri pazienti tuttavia non evidenziano grosse differenze, a meno di alcuni casi, e questa è una indicazione della buona capacità predittiva dell'algoritmo al di fuori del training set.

Lo stesso si può osservare per i valori di Precision e Recall: i valori di media e deviazione standard sono rispettivamente più bassi e più alti, ma la differenza è trascurabile ed è dovuto a poche eccezioni tra i pazienti 6:25, dove si hanno valori minimi molto bassi, mentre i valori massimi rimangono solo leggermente più bassi. Queste osservazioni valgono per tutte e 3 le strutture.

Il parametro MDC è in grado di fornire qualche informazione aggiuntiva. I valori medi hanno lo stesso comportamento, differenza tra i due gruppi molto bassa (non più di 2 mm), ma si può osservare molto meglio la presenza di alcuni pazienti con prestazioni molto più basse nel gruppo 6:25. Questo è dovuto alla maggiore sensibilità del parametro MDC, per cui pazienti anche con poche differenze in termini di Dice posso avere differenze in termini di MDC molto più marcate. I valori di deviazione standard sono quindi più alti di qualche mm nel gruppo 6:25.

3.5. Effetto della soglia SUV

I risultati che si ottengono dalla segmentazioni, oltre che dalle prestazioni dei classificatori e dalle strutture esaminate, dipendono anche dalle caratteristiche del riferimento. Di seguito si può osservare come i risultati variano al variare della soglia SUV utilizzata, per i primi 10 pazienti.



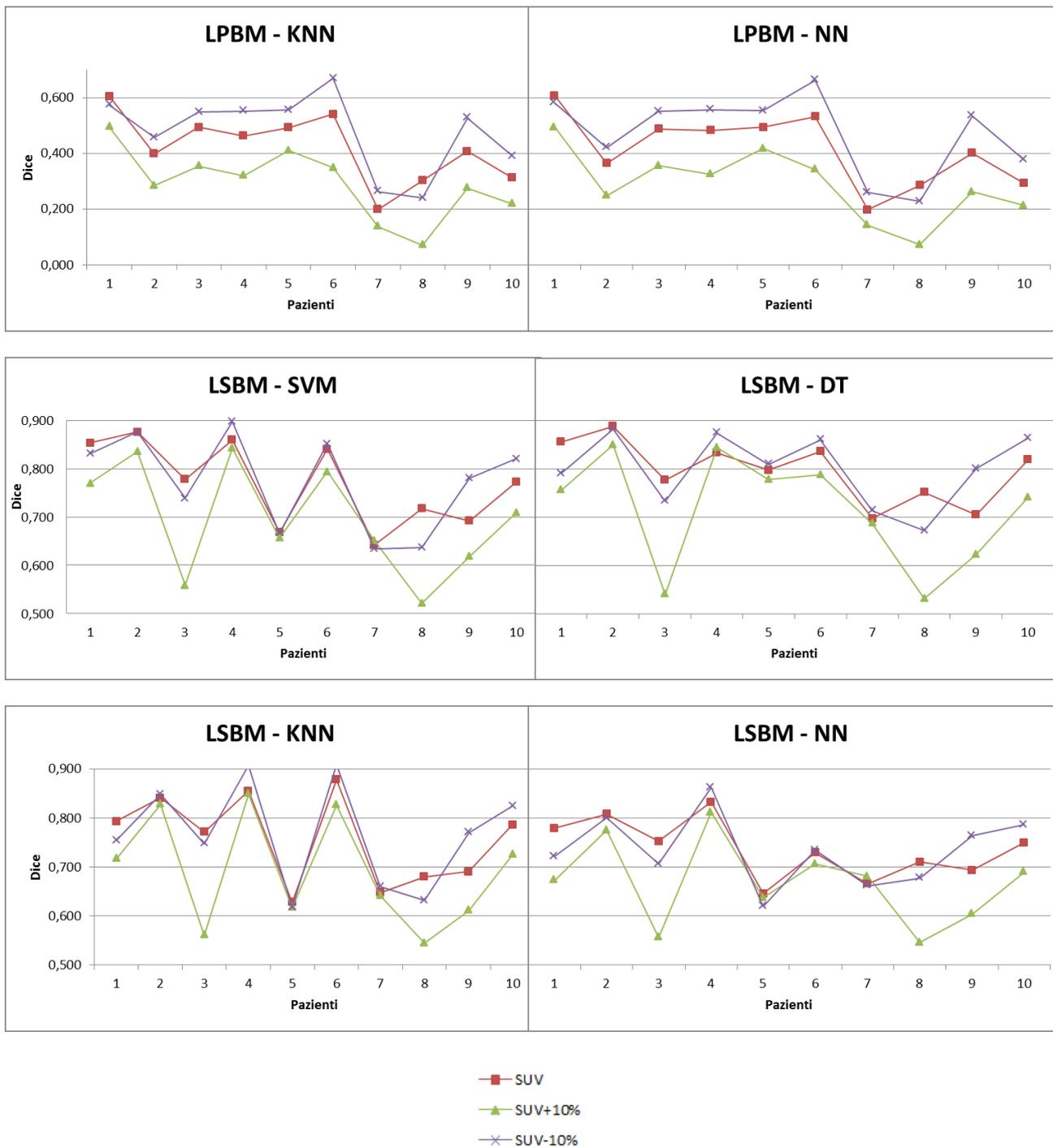


Figura 20 – Grafici che mostrano i risultati dei primi 10 pazienti, espressi in Dice, al variare della soglia SUV da cui si è estratta la maschera di riferimento. In ascissa vi sono i pazienti e in ordinata i valori Dice.

Le strutture IBM e LSBM hanno valori di Dice che non variano tra la soglia SUV normale e la soglia SUV-10%, mentre con la soglia SUV+10% diminuiscono leggermente. Per queste due strutture il metodo usato per la segmentazione del riferimento è quindi piuttosto robusto, e le maschere ottenute hanno aspetto piuttosto simile tra loro. Diminuendo la soglia si ottengono maschere piuttosto sovrapponibili a quelle ottenute con soglia normale, mentre all'aumentare tali maschere sono più piccole, per cui tendono ad esservi più errori di sovra-segmentazione.

Tali considerazioni non valgono per la struttura LPBM, dove viene meno la robustezza dell'algoritmo e si osservano prestazioni migliori al diminuire della soglia. In questo caso la diminuzione del valore di soglia

aumenta sensibilmente le dimensioni della maschera. Dato che dai risultati precedenti si è osservata una grande sovra-segmentazione delle maschere da valutare per la struttura LPBM, avendo maschere di riferimento di dimensioni maggiori diminuisce l'errore e aumenta il valore di Dice.

3.6. Struttura LPBM e sue criticità

Tutti i risultati ottenuti durante i vari passi dell'algoritmo di segmentazione mostrano sempre che le prestazioni dei classificatori sono pessime e sono dovute ad un enorme sovra-segmentazione della struttura stessa. Questi risultati non sono in linea con i risultati che si sono ottenuti nell'ottimizzazione del GA: in tal caso i classificatori ottenuti possiedono le stesse prestazioni delle strutture IBM e LSBM in termini di accuratezza. Questo significa che la capacità dei classificatori di classificare correttamente i pixel del training set non si traduce in una buona capacità di classificare correttamente i pixel della struttura stessa.

Per capire il perché di questo fenomeno, si sono osservate le dimensioni delle tre strutture di midollo osseo attraverso il calcolo dei pixel a loro appartenenti.

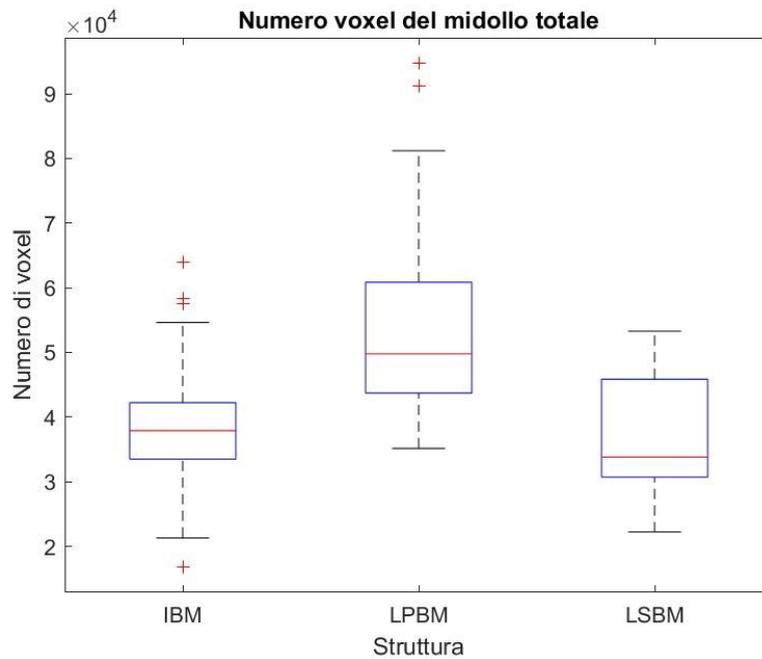


Figura 21 – Boxplot che mostra i valori del numero di voxel totali presenti nel midollo osseo per ogni struttura.

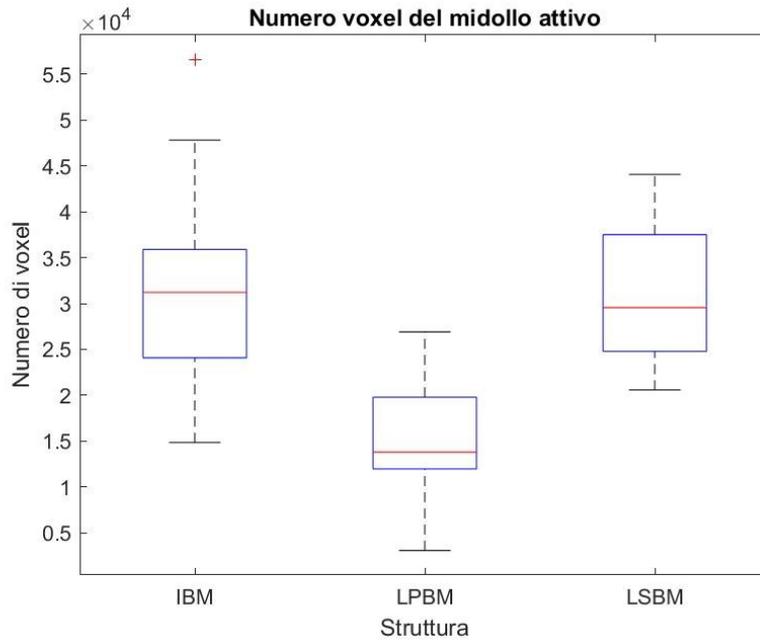


Figura 22 - Boxplot che mostra i valori del numero di voxel totali presenti nel midollo osseo attivo per ogni struttura.

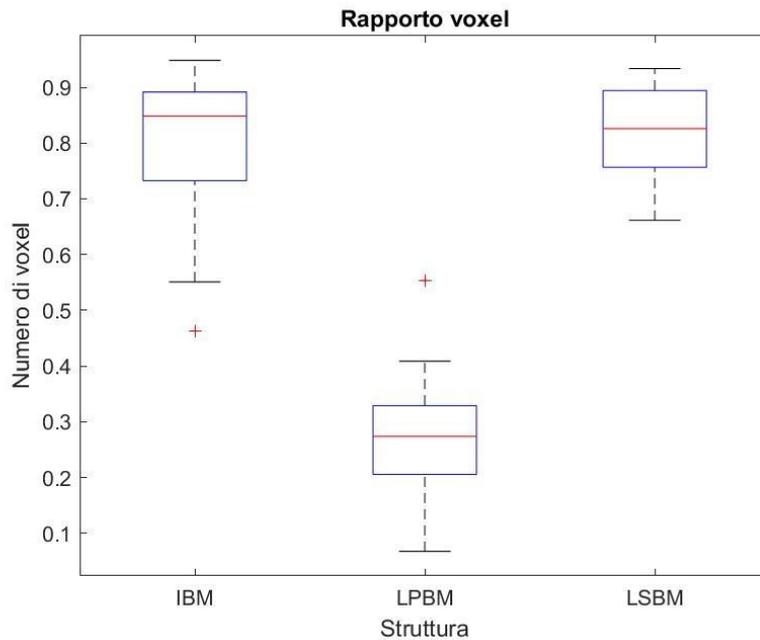


Figura 23 - Boxplot che mostra i valori del rapporto tra numero di voxel totali presenti nel midollo osseo e numero di voxel di midollo osseo attivo per ogni struttura.

I boxplot mostrano le differenze in termini di numero di pixel delle strutture. Si può osservare come le caratteristiche delle strutture sono diverse. Nelle strutture IBM e LSBM, il numero di pixel del midollo in totale sono minori rispetto alla struttura LPBM. Quindi la struttura LPBM è più grande delle altre due. Tuttavia, il numero di pixel della struttura LPBM effettivamente appartenenti al midollo osseo attivo è molto minore rispetto alle strutture IBM e LSBM. Quindi, anche se la struttura totale è più grande, le

dimensioni del midollo attivo sono molto minori. Si calcola il rapporto tra il numero di pixel totali e il numero di pixel di midollo attivo come segue:

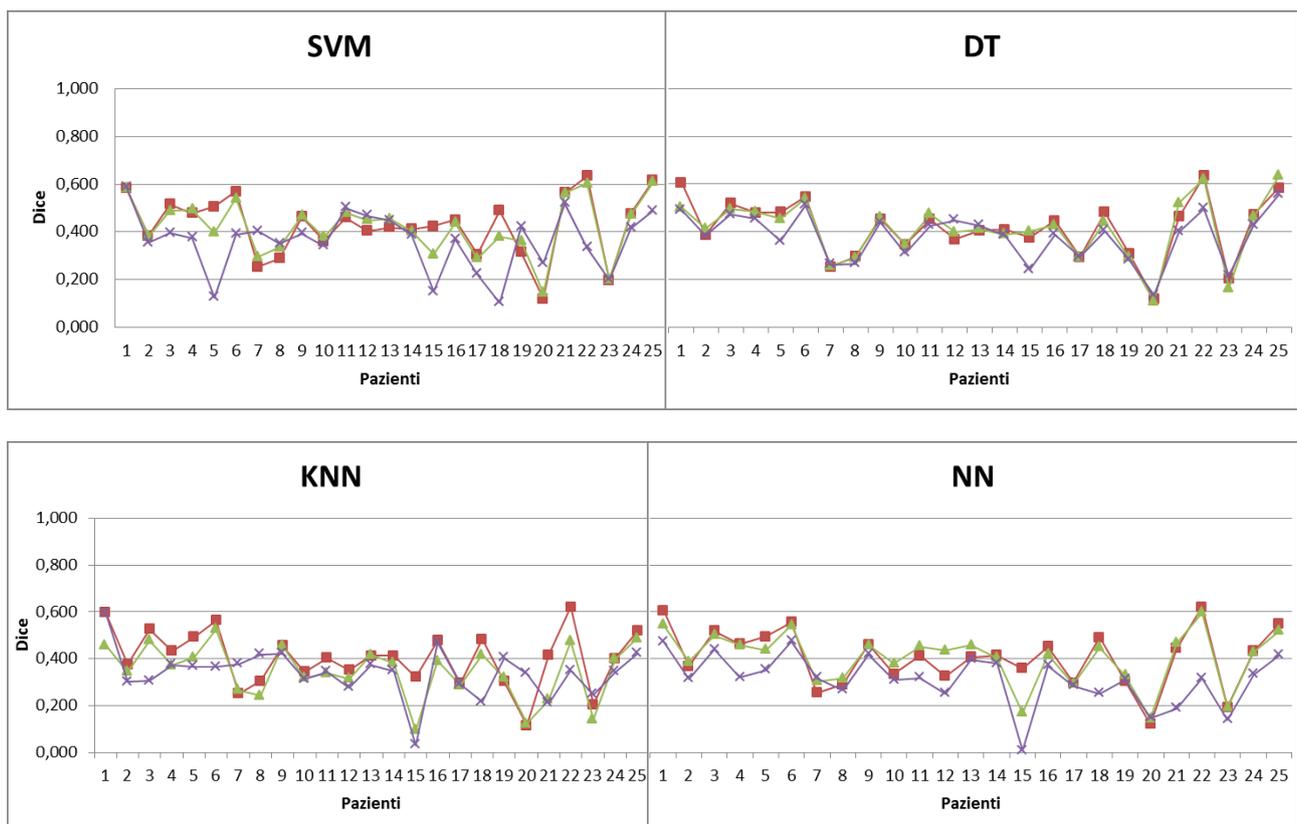
$$R = \frac{N_{rm}}{N_{bm}} \quad \text{dove}$$

N_{rm} = numero di pixel di midollo attivo, N_{bm} = numero di pixel di midollo totale

Si nota subito che il numero di pixel della struttura LPBM appartenenti al midollo attivo è molto più basso del numero di pixel totale, mentre per IBM e LSBM la maggior parte appartengono al midollo attivo. Questo suggerisce un legame tra il numero basso di pixel di midollo attivo e le difficoltà, da parte dei classificatori, di individuare correttamente il midollo osseo attivo nella struttura LPBM. Queste difficoltà peggiorano quindi le performance di segmentazione.

3.7. Valutazione della struttura LPBM con diversi training set

Per la struttura LPBM, data la sua criticità e l'incapacità dei classificatori di segmentare correttamente la struttura, si è deciso di eseguire nuovamente l'algoritmo genetico e gli algoritmi di segmentazione, per verificare il comportamento al variare del training set. Si sono costruiti quindi altri 2 training set, con le stesse dimensioni del training set già utilizzato, e con le stesse proporzioni tra pixel positivi e negativi. I risultati ottenuti con i due nuovi training set sono stati confrontati con il training set originale nei seguenti grafici, per tutti i 25 pazienti.



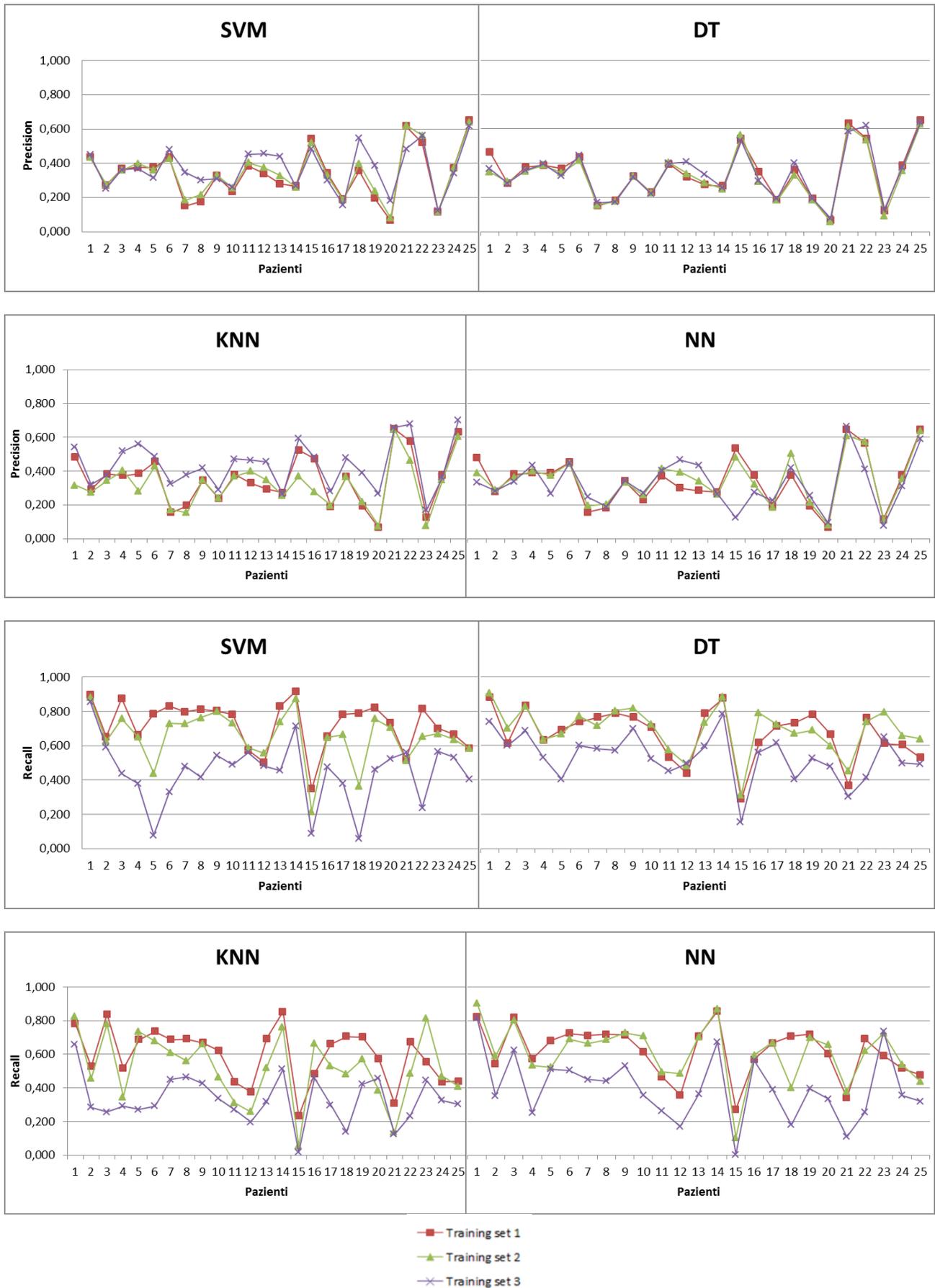
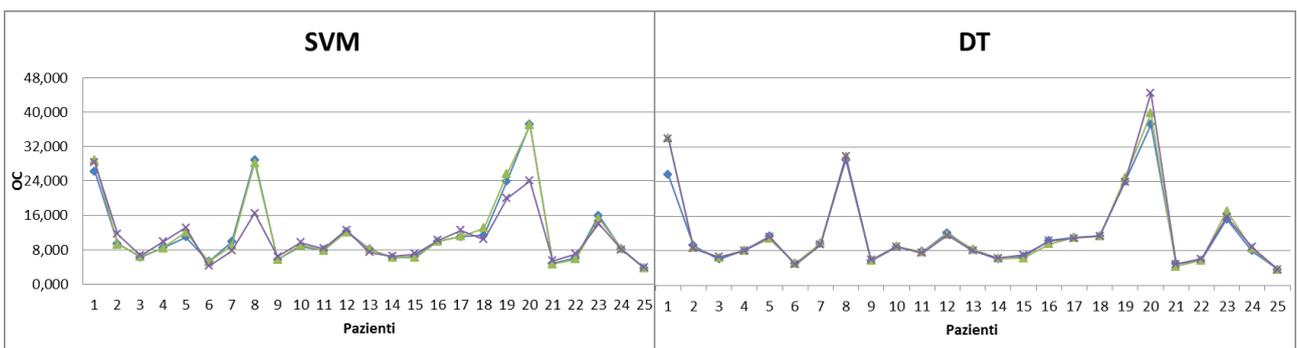
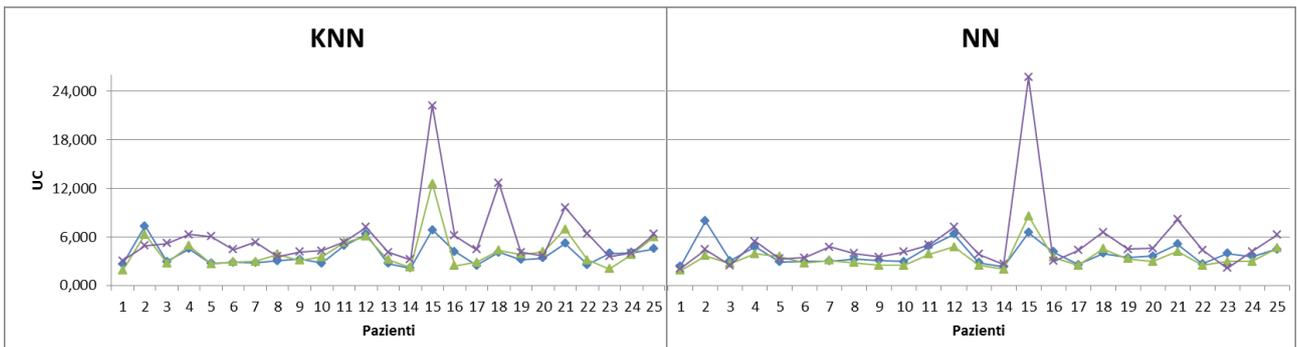
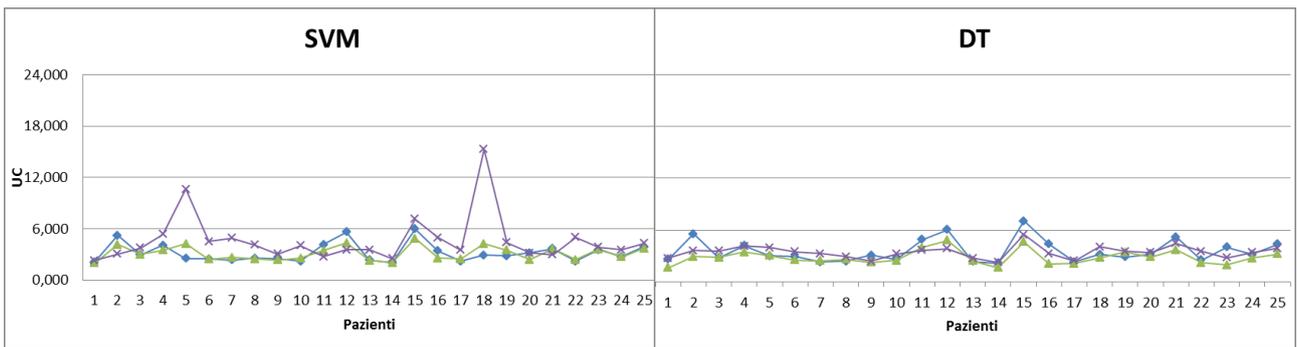
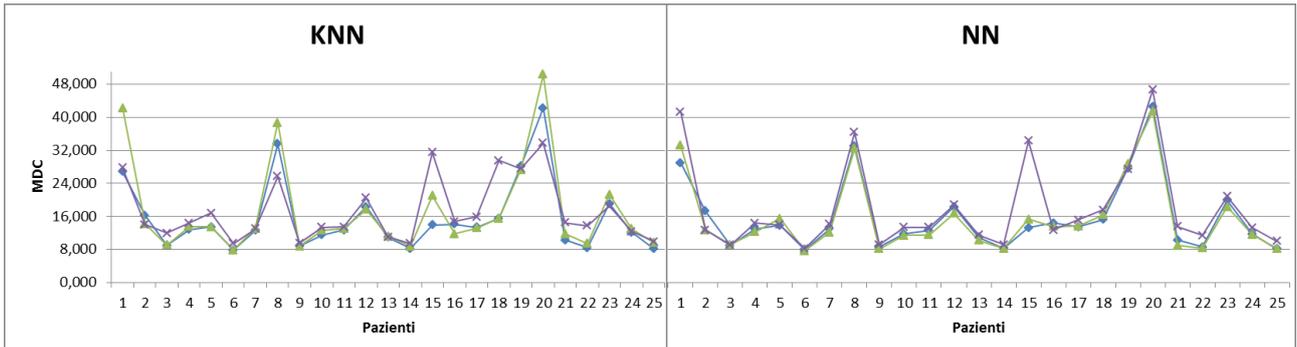
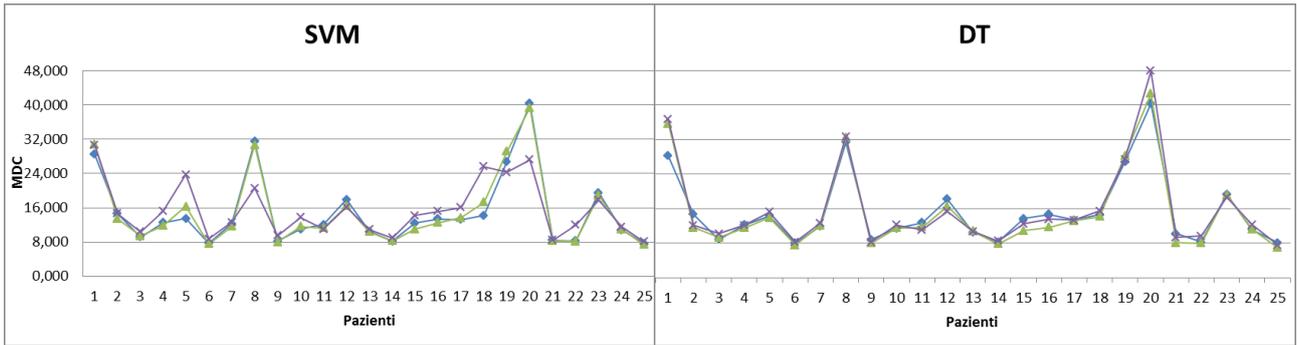


Figura 24 – Grafici che mostrano i risultati di Dice, Precision e Recall al variare del training set utilizzato nel GA. In basso, la legenda dei training set: il training set 1 è il training set utilizzato nei precedenti risultati.



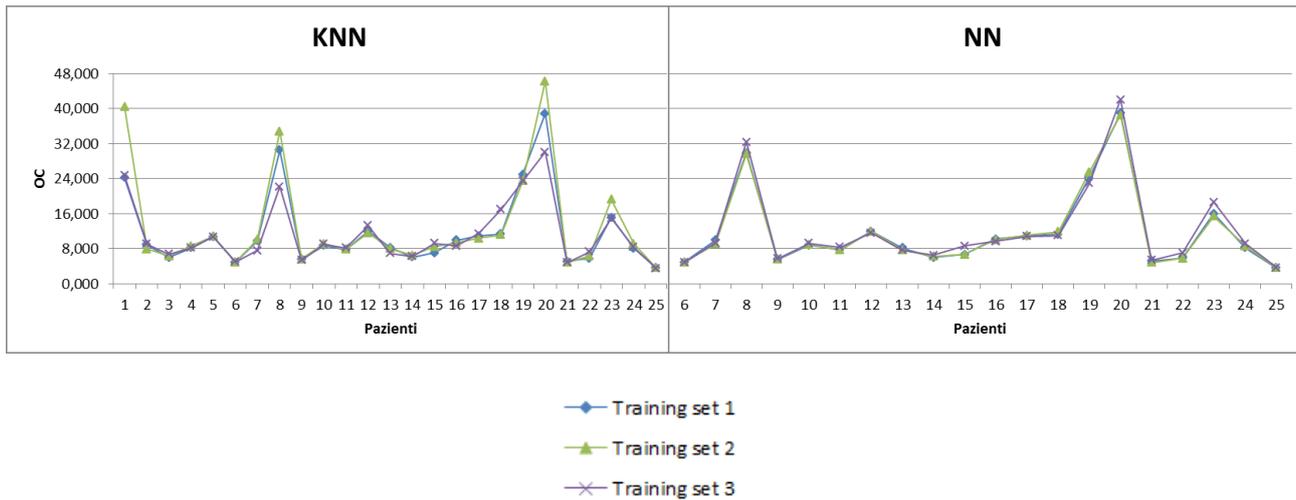


Figura 25 - Grafici che mostrano i risultati di MDC, UC e OC al variare del training set utilizzato nel GA. In basso, la legenda dei training set.

Si può osservare da questi risultati che variare il training set non porta alcun beneficio alle performance dei 4 classificatori. Questo significa che i cattivi risultati della struttura LPBM non sono dovuti ad un training set particolarmente sfortunato, e che questi siano dovuti alla natura stessa della struttura, che, come abbiamo già visto, ha un numero di pixel appartenenti al midollo osseo attivo basso, molto inferiori alle altre due strutture considerate. Di seguito sono riportati in tabella i risultati numerici dei parametri calcolati, in termini di media, deviazione standard, valore massimo ottenuto e valore minimo ottenuto.

Pazienti 1:25		DT					
		Dice	Precision	Recall	MDC	UC	OC
Training set 1	Media	0,415	0,339	0,668	15,239	3,566	11,674
	Std	0,125	0,152	0,150	8,215	1,321	8,405
	Max	0,633	0,650	0,880	40,407	7,054	37,198
	Min	0,119	0,065	0,287	7,988	2,177	3,701
Training set 2	Media	0,411	0,326	0,693	15,010	2,822	12,189
	Std	0,126	0,149	0,135	9,504	0,818	9,608
	Max	0,634	0,627	0,907	42,712	4,777	39,859
	Min	0,107	0,059	0,312	6,927	1,627	3,733
Training set 3	Media	0,379	0,340	0,535	15,773	3,430	12,343
	Std	0,104	0,149	0,137	9,938	0,707	10,065
	Max	0,555	0,637	0,785	47,906	5,437	44,506
	Min	0,134	0,078	0,156	7,564	2,240	3,691

Tabella 9 - Tabelle che mostrano i valori numerici dei risultati ottenuti tramite classificatori DT per la struttura LPBM e al variare del training set utilizzato, per i pazienti 1-25. Le tabelle mostrano la media, la deviazione standard, il massimo e il minimo dei risultati ottenuti

Pazienti 1:25		KNN					
		Dice	Precision	Recall	MDC	UC	OC
Training set 1	Media	0,403	0,350	0,594	15,535	3,824	11,710
	Std	0,121	0,154	0,161	8,583	1,412	8,721
	Max	0,620	0,649	0,852	42,281	7,330	38,869
	Min	0,116	0,064	0,232	7,840	2,137	3,606
Training set 2	Media	0,347	0,319	0,519	17,157	4,155	13,002
	Std	0,116	0,138	0,202	11,105	2,251	11,312
	Max	0,526	0,646	0,824	50,378	12,592	46,239
	Min	0,094	0,073	0,054	7,854	1,870	3,644
Training set 3	Media	0,340	0,437	0,330	17,308	6,012	11,296
	Std	0,104	0,139	0,139	7,501	3,985	6,985
	Max	0,595	0,701	0,658	33,791	22,229	30,132
	Min	0,031	0,170	0,016	9,459	3,076	3,525

Tabella 10 - Tabelle che mostrano i valori numerici dei risultati ottenuti tramite classificatori KNN per la struttura LPBM e al variare del training set utilizzato, per i pazienti 1-25. Le tabelle mostrano la media, la deviazione standard, il massimo e il minimo dei risultati ottenuti

Pazienti 1:25		NN					
		Dice	Precision	Recall	MDC	UC	OC
Training set 1	Media	0,407	0,343	0,618	15,695	3,867	11,828
	Std	0,124	0,154	0,151	8,698	1,411	8,807
	Max	0,621	0,647	0,856	42,623	7,978	39,015
	Min	0,120	0,066	0,271	7,914	2,341	3,618
Training set 2	Media	0,403	0,350	0,608	15,436	3,402	12,035
	Std	0,116	0,145	0,170	8,965	1,343	9,166
	Max	0,597	0,639	0,902	41,455	8,585	38,502
	Min	0,144	0,081	0,103	7,689	1,860	3,613
Training set 3	Media	0,311	0,327	0,398	17,953	5,202	12,752
	Std	0,109	0,141	0,194	10,691	4,521	10,411
	Max	0,475	0,664	0,812	46,682	25,668	42,084
	Min	0,007	0,077	0,004	8,266	2,081	3,744

Tabella 11 - Tabelle che mostrano i valori numerici dei risultati ottenuti tramite classificatori NN per la struttura LPBM e al variare del training set utilizzato, per i pazienti 1-25. Le tabelle mostrano la media, la deviazione standard, il massimo e il minimo dei risultati ottenuti

Pazienti 1:25		SVM					
		Dice	Precision	Recall	MDC	UC	OC
Training set 1	Media	0,428	0,335	0,725	14,891	3,192	11,699
	Std	0,129	0,148	0,138	8,361	1,110	8,442
	Max	0,634	0,650	0,916	40,334	6,001	37,162
	Min	0,119	0,065	0,347	7,642	2,030	3,741
Training set 2	Media	0,420	0,346	0,651	14,964	3,110	11,854
	Std	0,119	0,143	0,151	8,560	0,828	8,681
	Max	0,609	0,638	0,883	39,359	4,893	36,968
	Min	0,143	0,080	0,216	7,393	2,006	3,683
Training set 3	Media	0,360	0,369	0,444	15,481	4,647	10,834
	Std	0,124	0,127	0,184	6,416	2,790	6,006
	Max	0,589	0,616	0,857	30,596	15,281	28,286
	Min	0,102	0,121	0,057	8,089	2,310	3,811

Tabella 12 - Tabelle che mostrano i valori numerici dei risultati ottenuti tramite classificatori SVM per la struttura LPBM e al variare del training set utilizzato, per i pazienti 1-25. Le tabelle mostrano la media, la deviazione standard, il massimo e il minimo dei risultati ottenuti

4. Conclusioni

Il metodo descritto è un primo tentativo di rendere la CT in grado di fornire informazioni riguardo alle caratteristiche del midollo osseo attraverso la radiomica, informazioni molto utili nella pianificazione della radioterapia e non di facile acquisizione tramite i sistemi d'imaging utilizzati negli ospedali. I risultati ottenuti sono molto promettenti per le strutture IBM e LSBM, strutture per le quali è probabilmente possibile, attraverso una ottimizzazione più attenta, ottenere performance di segmentazione che potrebbero renderle utilizzabili in pianificazione di radioterapia.

Le criticità osservate per la struttura LPBM, invece, necessitano di uno studio più approfondito, sia sulle cause che portano al fenomeno di sovra-segmentazione, sia sui metodi utilizzabili per rendere la classificazione più corretta (calcolo di nuove feature, utilizzo di altri classificatori per l'esecuzione del voting).

Appendice

La GLCM (Grey-Level Co-occurrence Matrix) è una matrice contenente le frequenze P_{ij} , ovvero il numero di volte che due pixel ad una certa distanza L assumono valori di grigio rispettivamente pari a i e j [17].

Questa matrice può essere calcolata lungo 4 direzioni: 0° , 45° , 90° , 135° . Dalla GLCM si sono calcolate le seguenti feature: autocorrelation[18], correlation1[19], correlation2[18], contrast[18], cluster prominence[18], cluster shade[18], dissimilarity[18], energy[18], entropy[18], homogeneity1[19], homogeneity2[18], maximum probability[18], sum of squares[17], variance[17], sum average[17], sum variance[17], sum entropy[17], difference variance[17], difference entropy[17], information measure of correlation1[17], information measure of correlation2[17], inverse difference normalized[19], inverse difference moment normalized[19]. Siccome si hanno 4 matrici, i valori delle feature sono dati dalla media dei valori nelle 4 direzioni.

Il GLRLM (Grey-Level Run Length Method) è un metodo di calcolo basato sui grey-level run. Un grey-level run si calcola come il numero di occorrenze nell'immagine della presenza di set di pixel adiacenti di lunghezza L aventi lo stesso valore j [20]. Scegliendo una certa lunghezza L , si ottengono delle matrici nelle 4 direzioni (0° , 45° , 90° , 135°). Dalle matrici si estraggono 5 feature: short run emphasis, long run emphasis, grey level distribution, run length distribution, run percentages. I valori vengono mediati nelle 4 direzioni

Il GLDM (Grey-Level Difference Method) è un metodo basato sulla differenza tra i valori di due pixel. Data una distanza L e una certa direzione, il metodo calcola le differenze tra i valori dei pixel posti ad una distanza L [20]. La matrice che si ottiene contiene la frequenza con cui si ottengono valori pari a $i-j$. Dalle matrici nelle 4 direzioni si ottengono: contrast, entropy, mean, angular second moment, inverse difference moment.

In GLCM, GLRLM e GLDM si è sempre utilizzata una distanza di un pixel ($L=1$). Di seguito vi sono le formule utilizzate per l'estrazione dei parametri.

Notazione

P = matrice GLCM

R = matrice GLRLM

D = matrice GLDM

$p(i, j)$ = elemento della matrice P normalizzata ($P(i, j)/R$)

$r(i, j)$ = elemento della matrice R normalizzata ($P(i, j)/R$)

$d(i)$ = elemento della matrice D

$p_x(i)$ = elemento della matrice di probabilità marginale ottenuta sommando le righe di P

N_g = numero di livelli di grigio dell'immagine

PARAMETRI DA GLCM

Autocorrelation

$$f_1 = \sum_i \sum_j (ij)p(i, j)$$

Correlation2

$$f_2 = \frac{\sum_i \sum_j (ij) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

Contrast

$$f_3 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \mid |i - j| = n \right\}$$

Cluster prominence

$$f_4 = \sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i, j)$$

Cluster shade

$$f_5 = \sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i, j)$$

Dissimilarity

$$f_6 = \sum_i \sum_j |i - j| \cdot p(i, j)$$

Energy

$$f_7 = \sum_i \sum_j p(i, j)^2$$

Entropy

$$f_8 = - \sum_i \sum_j p(i, j) \log(p(i, j))$$

Homogeneity2

$$f_9 = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i, j)$$

Maximum probability

$$f_{10} = \text{MAX}_{i, j} p(i, j)$$

Correlation1

$$f_{11} = \sum \frac{(i - \mu_x)(j - \mu_y) p(i, j)}{\sigma_x \sigma_y}$$

Homogeneity1

$$f_{12} = \sum p(i,j)^2$$

Inverse difference normalized

$$f_{13} = \sum \frac{p(i,j)}{1 + |i - j|}$$

Inverse difference moment normalized

$$f_{14} = \sum \frac{p(i,j)}{1 + (i - j)^2}$$

Sum of squares

$$f_{15} = \sum (i - \mu)^2 p(i,j)$$

Sum average

$$f_{16} = \sum_{i=2}^{2N_g} i p_{x+y}(i)$$

Sum entropy

$$f_{17} = - \sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\}$$

Sum variance

$$f_{18} = \sum_{i=2}^{2N_g} (i - f_{17})^2 p_{x+y}(i)$$

Difference entropy

$$f_{19} = - \sum_{i=0}^{N_g-1} p_{x-y}(i) \log\{p_{x-y}(i)\}$$

Difference variance

$$f_{20} = \text{variance of } p_{x-y}$$

Information measure of correlation1

$$f_{21} = \frac{f_8 - HXY1}{\max\{HX, HY\}}$$

$$HXY1 = - \sum_i \sum_j p(i,j) \log(p_x(i)p_y(j))$$

$HX = \text{entropy of } p_x$

$HY = \text{entropy of } p_y$

Information measure of correlation2

$$f_{22} = (1 - \exp[-2.0(HXY2 - f_8)])^{1/2}$$

$$HXY2 = - \sum_i \sum_j p_x(i)p_y(j) \log(p_x(i)p_y(j))$$

PARAMETRI DA GLRLM

Short run emphasis

$$f_{23} = \frac{1}{T_R} \sum_{i=0}^{N_g-1} \sum_{j=1}^{N_R} \frac{r(i,j)}{j^2}$$

$$T_R = \sum_{i=0}^{N_g-1} \sum_{j=1}^{N_R} r(i,j)$$

$N_R = \text{numero di run della matrice}$

Long run emphasis

$$f_{24} = \frac{1}{T_R} \sum_{i=0}^{N_g-1} \sum_{j=1}^{N_R} j^2 r(i,j)$$

Grey level distribution

$$f_{25} = \frac{1}{T_R} \sum_{i=0}^{N_g-1} \left[\sum_{j=1}^{N_R} r(i,j) \right]$$

Run length distribution

$$f_{26} = \frac{1}{T_R} \sum_{j=1}^{N_R} \left[\sum_{i=0}^{N_g-1} r(i,j) \right]$$

Run percentages

$$f_{27} = \frac{1}{T_P} \sum_{i=0}^{N_g-1} \sum_{j=1}^{N_R} r(i,j)$$

$T_P = \text{numero di punti dell'immagine}$

PARAMETRI DA GLDM

Contrast

$$f_{28} = \sum_{i=0}^{N_g-1} i^2 d(i)$$

Angular second moment

$$f_{29} = \sum_{i=0}^{N_g-1} [d(i)]^2$$

Entropy

$$f_{30} = \sum_{i=0}^{N_g-1} d(i) \log(d(i))$$

Mean

$$f_{31} = \sum_{i=0}^{N_g-1} id(i)$$

Inverse difference moment

$$f_{32} = \sum_{i=0}^{N_g-1} \frac{d(i)}{i^2 + 1}$$

Bibliografia

- [1] – P. Franco et al. Dose to specific subregions of pelvic bone marrow defined with FDG-PET as a predictor of hematologic nadirs during concomitant chemoradiation in anal cancer patients. *Medical Oncology*, 2016, Vol.33(7), pp.1-10
- [2] - P. Franco et al. Dosimetric predictors of acute hematologic toxicity during concurrent intensity-modulated radiotherapy and chemotherapy for anal cancer. *Clinical and Translational Oncology*, 2017, Vol.19(1), pp.67-75
- [3] – P. Franco et al. Hematologic toxicity in anal cancer patients during combined chemo-radiation: a radiation oncologist perspective. *Expert Review of Anticancer Therapy*, 2017
- [4] – P. Franco et al. Incorporating 18-FDG-PET-defined pelvic active bone marrow in the automatic treatment planning process of anal cancer patients undergoing chemo-radiation. *BMC Cancer*, 01 November 2017, Vol.17(1), pp.1-11
- [5] – Carr R, Barrington S F, Madan B, O’doherly M J, Saunders C A, van der Walt J and Timothy A R 1998 Detection of lymphoma in bone marrow by whole-body positron emission tomography. *Blood*913340–6
- [6] – Judy S. Blebeal et al. Structural and Functional Imaging of Normal Bone Marrow and Evaluation of Its Age-Related Changes. *Seminars in Nuclear Medicine*, 2007, Vol.37(3), pp.185-194
- [7] – A. Andreyvchenko et al . The feasibility of semi-automatically generated red bone marrow segmentations based on MR-only for patients with gynecologic cancer. *Radiotherapy and Oncology*, April 2017, Vol.123(1), pp.164-168
- [8] – Kogan et al. Applications of PET-MRI in Musculoskeletal Disease. *Journal of Magnetic Resonance Imaging*, July 2018, Vol.48(1), pp.27-47
- [9] – G. Sambuceti. Estimating the whole bone-marrow asset in humans by a computational approach to integrated PET/CT imaging. *European Journal of Nuclear Medicine and Molecular Imaging*, 2012, Vol.39(8), pp.1326-1338
- [10] – Desai AG, Thakur ML. Radiopharmaceuticals for spleen and bone marrow studies. *Semin Nucl Med*. 1985;15:229–38
- [11] – Kugel H, Jung C, Schulte O, et al: Age- and sex-specific differences in the 1 H-spectrum of vertebral bone marrow. *J Magn Reson Imaging* 13:263-268, 2000
- [12] – H. Zaidi, A. Abass, et al. Novel Quantitative PET Techniques for Clinical Decision Support in Oncology. *Seminars in Nuclear Medicine*, November 2018, Vol.48(6), pp.548-564
- [13] – T. Perk, et al. A method for quantitative total marrow imaging (QTMI) with PET/CT. *Biomedical Physics & Engineering Express*, 2016, Vol.2(5), p.055006 (13pp)
- [14] – N.E. Makris et al. An automatic delineation method for bone marrow absorbed dose estimation in 89-Zr PET/CT studies. *Huisman, M.C. EJNMMI Physics*, 2016, Vol.3(1), pp.1-9
- [15] – P. Leydon et al. Automatic bone marrow segmentation for PETCT imaging in multiple myeloma. *Physica Medica*, 09/2016, Vol.32, S3, p.242

[16] – R. Jena, et al. A novel algorithm for the morphometric assessment of radiotherapy treatment planning volumes. The British journal of radiology, January 2010, Vol.83(985), pp.44-51

[17] - R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," IEEE Trans. Syst. Man. Cybern., vol. 3, no. 6, pp. 610–621, Nov. 1973

[18] - L.-K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," IEEE Trans. Geosci. Remote Sens., vol. 37, no. 2, pp. 780–795, Mar. 1999.

[19] - D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," Can. J. Remote Sens., vol. 28, no. 1, pp. 45–62, Jun. 2014.

[20] - R. W. Conners and C. a Harlow, "A theoretical comparison of texture algorithms.," IEEE Trans. Pattern Anal. Mach. Intell., vol. 2, no. 3, pp. 204–222, 1980.