

POLITECNICO DI TORINO

Corso di laurea magistrale
in Ingegneria Biomedica

Tesi di Laurea Magistrale

Rilevazione automatica di tumori
neuroendocrini in immagini 68Ga-DOTATOC
PET/CT full-body



Relatore: Filippo Molinari

Correlatore: Bruno De Santi

Candidato: Simone Scanu

Anno accademico 2020/2021

SOMMARIO

Abstract	5
Introduzione	5
Tumori neuroendocrini.....	5
Imaging	6
Rilevazione del tumore.....	7
Stato dell'arte	8
Obiettivo (thesis outline).....	9
Materiali e Metodi.....	10
Materiali	10
Pazienti	10
Imaging	10
Metodi	10
Caricamento del volume.....	11
Supervoxel	11
Registrazione volume CT	13
Analisi di tessitura multimodale	14
Analisi delle componenti principali	16
Classificatori.....	17
Manova.....	17
Support Vector Machine	18
Reti neurali	19
Addestramento e validazione.....	20
Training.....	21
Test	22
Metriche	22
Risultati.....	24
Supervoxel	24
Analisi delle componenti principali	25
Classificazione.....	26
Discussione	30
Conclusione	36
Riferimenti.....	38

ABSTRACT

I tumori neuroendocrini sono una famiglia di neoplasie a bassa incidenza ma elevata prevalenza che interessano le cellule del sistema neuroendocrino, presenti in tutto il corpo. L'imaging multimodale PET/CT con l'utilizzo di radiofarmaci a base di ^{68}Ga -DOTA-peptidi è impiegato in clinica per la diagnosi e la caratterizzazione dei tumori neuroendocrini. Tuttavia, la diagnosi è resa complessa dall'elevata variabilità delle caratteristiche dei tumori neuroendocrini, inoltre, la valutazione visiva può richiedere molto tempo ed è soggetta a variabilità inter ed intra operatore. In questo studio viene esplorato un metodo automatico per la rilevazione di tumori neuroendocrini all'interno di volumi multimodali ^{68}Ga -DOTATOC PET/CT. Il metodo proposto fa uso di un algoritmo supervoxel per la segmentazione dei volumi in zone con caratteristiche omogenee, di analisi di tessitura per la caratterizzazione di tali zone e di un classificatore per identificare le zone contenenti una lesione. La classificazione finale è eseguita tramite un classificatore basato sull'analisi della varianza multivariata (MANOVA) e fornisce un grado di confidenza, in una scala da 0 a 4, con cui ogni zona può essere considerata di natura neoplastica. Il grado indica la certezza con cui il classificatore ritiene che le caratteristiche della porzione di tessuto descrivano una lesione neoplastica neuroendocrina. Il metodo viene addestrato su 15 pazienti con 16 lesioni neuroendocrine e validato su 5 pazienti con 5 lesioni neuroendocrine tramite due metriche: il coefficiente di Sørensen-Dice per confrontare le segmentazioni ottenute con le segmentazioni manuali di operatori esperti e un confronto tra i gradi medi, assegnati alle zone per le due classi di interesse, per valutare la qualità della predizione del modello. Si è registrato un valore medio DSC = 81.48 ± 9.92 % per la segmentazione e un grado medio maggiore per le lesioni (grado medio = 4.00 ± 0) rispetto alle zone corrispondenti a tessuti sani (grado medio = 1.13 ± 0.40). In futuro tale metodologia potrebbe fornire al medico uno strumento accurato e robusto nella valutazione di pazienti con neoplasie neuroendocrine, riducendo le tempistiche della diagnosi.

INTRODUZIONE

TUMORI NEUROENDOCRINI

I tumori neuroendocrini sono una famiglia di neoplasie che interessa tessuti neuroendocrini, complessa da caratterizzare per via della loro elevata eterogeneità. Ad oggi non esistono standard univocamente accettati per la descrizione e gradazione di tutte le tipologie di tumori che fanno parte dei NET, e molte sono le caratteristiche di questi tumori che ancora sono oggetto di discussione, a partire dalla nomenclatura utilizzata. Il termine *neuroendocrino* trova origine nell'ipotesi che le cellule interessate da questa tipologia di tumore derivassero dalla cresta neurale embrionale. Nonostante questa ipotesi sia stata confutata, e quindi la terminologia *endocrino* sia più corretta, la nomenclatura neuroendocrino è ormai comunemente utilizzata e pertanto riconosciuta[1]. Sempre riguardo alla nomenclatura, un'altra questione fondamentale è l'utilizzo complementare fatto dei termini *neoplasia*, *tumore* e *carcinoma*. Il termine *neoplasia* è infatti più adatto a descrivere le caratteristiche di tale famiglia, data la natura più o meno aggressiva che può caratterizzarle, ma poco utilizzato rispetto a *tumore*, che sarebbe più adatto a descrivere un sottoinsieme di queste neoplasie, storicamente usato per indicare la totalità di queste neoplasie. Le linee guida dell'Agenzia Internazionale per la Ricerca sul Cancro (IARC) dell'Organizzazione Mondiale della Sanità (OMS), pubblicate nel 2018 hanno recentemente fornito una direzione precisa sotto questo aspetto. Esse si riferiscono a questa famiglia di neoplasie come *neoplasie neuroendocrine* (NEN), a loro volta suddivise in *carcinomi neuroendocrini* (NEC) e *tumori neuroendocrini* (NET).

L'elevata eterogeneità che contraddistingue le NEN, e che rende complesso caratterizzarle, è principalmente dovuta alla varietà di tessuti e organi in cui hanno origine. Le NEN interessano cellule neuroendocrine, che presentano caratteristiche simili sia alle cellule endocrine che a quelle nervose, presenti in tutti i tessuti del

corpo e nei vari organi, in cui svolgono compiti differenti. La neoplasia possiede quindi delle caratteristiche peculiari a seconda dell'organo/tessuto di origine, pertanto è comune suddividere questi tumori in base al sito anatomico in cui si sviluppano. Secondo la Fondazione AIRC per la ricerca sul cancro l'incidenza in Italia è di 4-5 casi ogni 100.000 abitanti, per un totale di 27.000 casi annui, di cui "I più frequenti sono i tumori che riguardano il tratto gastro-entero-pancreatico (60-70%); seguono quelli che colpiscono polmoni e apparato respiratorio (20-30%) o altre regioni del corpo (10%) come cute, tiroide, paratiroide e surreni.". Simili percentuali d'incidenza sono riportate in letteratura[2].

Oltre al tipo di tessuto in cui hanno origine, esistono altre due caratteristiche generalmente utilizzate per categorizzare queste neoplasie: il grado di differenziazione ed il grado di proliferazione o aggressività. Questi due fattori sono connessi ma non sono coincidenti e forniscono informazioni diverse sul tumore. Il grado di differenziazione fa riferimento alla somiglianza che intercorre tra le cellule neoplastiche e la loro controparte sana nel tessuto interessato. Le NEN possono essere suddivise in *ben differenziate* e *poco differenziate*. Lesioni ben differenziate presentano una struttura caratteristica delle cellule che va a costituire un organoide, con cellule uniformi e un'elevata produzione di granuli neurosecretori. Lesioni poco differenziate al contrario presentano meno somiglianze con le cellule sane del tessuto, hanno una struttura più schiacciata e distribuita, con nuclei irregolari e una minore risposta a marcatori immunologici. Il grado di proliferazione è invece indicatore dell'aggressività della lesione. Generalmente si riconoscono tre gradi: i tumori con basso grado sono quelli più indolenti mentre quelli ad alto grado sono più attivi ed aggressivi, i tumori con grado medio presentano caratteristiche intermedie tra i due estremi. I tumori ben differenziati presentano solitamente un grado di proliferazione basso o intermedio, mentre quelli scarsamente differenziati sono di norma considerati ad alto grado di proliferazione[1]. Il grado di proliferazione viene misurato tramite la conta di mitosi per area della lesione o come percentuale di risposta al marcatore immunologico Ki-67.

Un esempio di classificazione in base a questi parametri, prendendo in considerazione le neoplasie neuroendocrine con maggiore occorrenza, cioè le pancreatiche, è quello ottenuto seguendo le indicazioni delle linee guida dell'OMS del 2018 (NEN-WHO-2018). La prima grossa distinzione è legata alla morfologia ed è tra lesioni ben differenziate, chiamate *tumori neuroendocrini pancreatici* (Pan-NET), e lesioni scarsamente differenziate, indicate come *carcinomi neuroendocrini pancreatici* (Pan-NEC). I tumori neuroendocrini pancreatici vengono a loro volta suddivisi in tre gruppi osservando il grado di proliferazione, indicati come G1, G2 e G3, seguendo i valori della conta mitotica e dell'indice Ki-67. I carcinomi neuroendocrini del pancreas, così come il resto dei NEC, vengono sempre considerati ad elevata aggressività e quindi non vengono ulteriormente suddivisi in categorie a seconda del grado di proliferazione e rientrano tutti in G3.

IMAGING

L'imaging delle neoplasie neuroendocrine è basato su una combinazione di tecniche morfologiche, generalmente tomografia computerizzata (CT) o risonanza magnetica (MRI), e funzionali, come la tomografia ad emissione di singolo fotone (SPECT) o la tomografia ad emissione di positroni (PET). La tomografia computerizzata e la risonanza magnetica sono le tecniche di riferimento per la rilevazione di lesioni e la valutazione del volume del tumore[3][4], per via dell'ampia disponibilità e della elevata risoluzione spaziale. La CT fornisce informazioni anatomiche importanti per delineare l'area di interesse per la terapia ma il basso contrasto può compromettere la qualità della definizione dei margini delle lesioni[4]. L'impiego della risonanza magnetica può pertanto essere preferito, dato il migliore contrasto e l'utilizzo di energie non ionizzanti. D'altro canto l'MRI è soggetto ad artefatti come distorsioni spaziali e disomogeneità delle intensità, oltre a non fornire informazioni sulla densità elettronica dei tessuti come la CT, utile per la programmazione delle dosi per la radioterapia[4]. Sebbene CT e MRI siano storicamente alla base dell'imaging per la rilevazione di neoplasie neuroendocrine[3][4][5] ad oggi esse sono generalmente affiancate a tecniche di medicina nucleare, quali SPECT e PET.

Lo studio della fisiologia delle neoplasie neuroendocrine e la comprensione dei loro processi metabolici, in particolare la sovraespressione dei recettori per la somatostatina (SSR), ha promosso l'utilizzo di tecniche di imaging funzionale nella caratterizzazione di questi tumori. Queste tecniche vengono utilizzate sotto forma di imaging ibrido, in cui procedure come la PET e la SPECT sono eseguite in contemporanea con la TC, attraverso scanner PET/CT e SPECT/CT rispettivamente, e in alcuni casi con la risonanza magnetica PET/MRI[5][6]. L'utilizzo di tali approcci ibridi consente di ottenere migliori risultati e ha un effetto positivo sulla confidenza con cui l'operatore legge le immagini[6]. Inoltre, l'utilizzo dei recettori per la somatostatina come target consente di rilevare una quantità più elevata di metastasi rispetto a quelle individuate tramite tecniche CT o MRI, e permette di valutare l'eleggibilità del paziente per radioterapie tramite analoghi della somatostatina, *peptide receptor radiotherapy* (PRRT)[6]. Ad oggi la tecnica di riferimento per l'imaging di neoplasie neuroendocrine è la PET/CT con radiofarmaci marcati al ^{68}Ga -DOTA, che ha sostituito la scintigrafia tramite SPECT e traccianti ^{111}In -pentetretotide (Octreoscan), per via della maggiore risoluzione spaziale garantita dalla PET rispetto alla SPECT, del miglior contrasto tra lesione e tessuto sano e della minore dose di radiazioni somministrate al paziente[6].

In alternativa ai traccianti con ^{68}Ga -DOTA, può essere preferibile l'impiego di traccianti che sfruttano il metabolismo del glucosio, come il ^{18}F FDG. L'espressione dei recettori per la somatostatina è infatti legata al grado di proliferazione tumorale, con un andamento che decresce fino ad annullarsi con l'aumentare del grado di proliferazione, risultando quindi poco efficace per la rilevazione di neoplasie con scarso grado di differenziazione[6][7]. Al contrario l'affinità del ^{18}F FDG cresce all'aumentare del grado di proliferazione, rendendo l'imaging ^{18}F FDG-PET/CT più adatto allo studio dei carcinomi neuroendocrini[7][8]. Una soluzione ottimale consisterebbe nell'eseguire l'indagine utilizzando entrambi i tipi di radiofarmaci, così da sfruttare la complementarità per migliorare la sensibilità dell'imaging[7][8]. Non sempre entrambi i traccianti sono disponibili all'interno delle strutture, che spesso scelgono di utilizzare uno solo dei due[6]. In tale situazione è importante valutare quale dei due traccianti utilizzare, in base al tipo di lesione che ci si aspetta.

RILEVAZIONE DEL TUMORE

L'utilizzo di tecniche di imaging ibride ed in particolare della PET/CT porta miglioramenti nella rilevazione delle neoplasie neuroendocrine[5][6] ma introduce nuove complicazioni. Uno dei problemi principali è costituito dalla complessità delle tecniche che è necessario applicare per delineare i contorni delle zone di interesse all'interno di immagini rumorose e con bassa risoluzione spaziale come quelle ottenute dall'imaging funzionale[4]. La segmentazione è quindi un aspetto fondamentale nella rilevazione dei tumori neuroendocrini. La tecnica maggiormente utilizzata a questo scopo è il thresholding, in cui viene individuato un valore soglia con cui i voxel dell'immagine vengono assegnati allo sfondo o alla lesione a seconda del loro valore di intensità. L'unico approccio utilizzato con pari frequenza del thresholding è la segmentazione manuale operata da medici e oncologi con esperienza. Questo metodo possiede però numerosi svantaggi: per sfogliare i volumi una fetta alla volta e delineare i margini delle lesioni richiede molto tempo; il risultato è suscettibile al metodo con cui vengono visualizzati i volumi e alle caratteristiche degli schermi; la variabilità inter-operatore e intra-operatore sono elevate[4][9]. Il thresholding viene pertanto adoperato perché, essendo un metodo semi-automatico e di semplice implementazione, permette di superare queste problematiche. Le modalità per la determinazione della soglia ottimale sono varie e vengono generalmente definite sui valori SUV dell'immagine. Lo Standardized Uptake Value (SUV), è comunemente utilizzato in medicina nucleare per fornire un'indicazione della quantità relativa di radiofarmaco assorbita da una porzione di tessuto. Viene calcolato come:

$$SUV = \frac{c_{im}}{(c'_{inj}/w)}$$

Dove c_{im} è la concentrazione di radioattività misurata in una porzione di volume, c'_{inj} è la concentrazione di radiofarmaco iniettato, corretta per tenere conto del decadimento dei radionuclidi, e w è il peso del paziente. Soglie tipicamente utilizzate per il thresholding sono calcolate come percentuale (40-50%) del valore SUV massimo[4][10], anche se alcuni studi hanno mostrato che la percentuale ideale varia in relazione al volume delle lesioni e del rapporto tra intensità dell'immagine e dello sfondo [11][12][13]. La disomogeneità delle neoplasie neuroendocrine rende in generale difficile individuare una soglia ottimale e pertanto sono stati proposti algoritmi thresholding con soglia automatica calcolata in base al rapporto tra oggetti e sfondo[13][10]. Queste tecniche richiedono però la calibrazione dei parametri dell'algoritmo, solitamente svolta attraverso dei fantocci che simulano la distribuzione del radiofarmaco specifico all'interno del corpo utilizzando delle sfere, di diametri differenti a seconda della dimensione delle lesioni che si sta considerando, con una distribuzione di radioattività modellata su quella fisiologica del radiofarmaco. Questo presuppone che si conoscano le caratteristiche delle lesioni che si desidera rilevare.

Esistono poi un insieme di tecniche che sfruttano le variazioni delle intensità ed i gradienti delle immagini per ottenere la segmentazione. Alcune di queste tecniche si basano sui *deformable active contours* [10][14] in cui curve o superfici vengono definite nello spazio delle immagini e vengono deformate da forze, correlate alle caratteristiche del modello e dell'immagine, in un bilancio energetico che ha come obiettivo quello di trasportare i punti del modello al di sopra delle zone dell'immagine con variazioni decise di intensità, andando idealmente ad individuare i margini delle lesioni. Queste tecniche necessitano però l'inizializzazione dei parametri del modello e hanno difficoltà a seguire le variazioni topologiche degli oggetti, come lo sdoppiamento o la creazione di fori sulla superficie. Un altro gruppo di tecniche è quello che sfrutta i *level set* [14], in cui il contorno delle lesioni è definito dall'intersezione di una funzione implicita ϕ , definita sui gradienti dell'immagine, e il piano dell'immagine. Sebbene questa famiglia di tecniche superi la necessità di inizializzazione e le limitazioni topologiche, risultano in egual modo sensibili al rumore [4].

Anche i metodi di learning, generalmente adoperati in compiti di classificazione, possono essere utilizzati allo scopo di segmentare le lesioni dal resto dell'immagine. La distinzione tra voxel con elevato assorbimento di radiofarmaco ed il resto dell'immagine viene ottenuto valutando un insieme di caratteristiche estratte dall'immagine. Metodi come *k-nearest neighbor* (KNN), *artificial neural network* (ANN) e *support vector machine* (SVM), comunemente utilizzati in compiti di classificazione su immagini MRI o CT, risultano poco efficaci nel caso di immagini PET, data l'elevata eterogeneità delle lesioni che complica la ricerca di pattern con un numero limitato di esempi su cui basarsi[15]. Più frequentemente utilizzati sono gli algoritmi basati su *k-means*[4][15], in cui a partire da un numero di punti iniziali che costituiranno i centri dei cluster, scelti casualmente o manualmente sull'immagine, vengono individuati dei cluster di voxel tramite delle metriche di distanza. Iterativamente i centri vengono aggiornati e così i voxel assegnati ad ogni cluster. Queste tecniche risultano sensibili al metodo con cui vengono inizializzati i cluster, al rumore e alla disomogeneità spaziali delle immagini PET.

STATO DELL'ARTE

Numerose tecniche sono state adottate per la delineazione e la caratterizzazione delle neoplasie neuroendocrine all'interno di volumi PET/CT. Nonostante la letteratura riguardante la PET con fluoro-deossiglucosio sia più vasta, dal momento che si tratta del radiotracciante più utilizzato[6], l'impiego di peptidi ^{68}Ga -DOTA per l'imaging funzionale ha mostrato maggiore sensibilità nell'individuazione di lesioni primarie e metastatiche con alto grado di differenziazione[6][8], rendendo necessario lo sviluppo di metodi che siano efficaci su tale tipologia di immagini. La tecnica più utilizzata ad oggi resta ancora la segmentazione manuale effettuata da un esperto, sebbene questo sia un procedimento laborioso e temporalmente dispendioso; oltre ad essere soggetto ad un'elevata variabilità intra e inter operatore[9]. Le tecniche di thresholding sono

considerate come una valida alternativa per via della facile implementazione e della possibilità di automatizzazione; per contro la decisione di una soglia da applicare all'imaging PET è complessa e può necessitare di opportune calibrazioni dei parametri degli scanner e dell'algoritmo. Metodi con maggiore complessità, come reti neurali o support vector machine, sono stati applicati per sfruttare la capacità degli algoritmi di learning, supervisionati e no, di individuare pattern nella delineazione delle lesioni ma l'elevata complessità computazionale introdotta da questi metodi e la necessità di training, valutate in rapporto alla validità dei metodi ottenuti, non è sempre soddisfacente.

OBIETTIVO (THESIS OUTLINE)

In questo lavoro è stato sviluppato un modello per la detezione di tumori neuroendocrini completamente automatico, con lo scopo di offrire un approccio innovativo che fornisca supporto al personale oncologico e radiologico nello studio di volumi ^{68}Ga -DOTATOC PET/CT con sospetta presenza di neoplasie neuroendocrine, per i quali poche metodiche sono state ad oggi riportate in letteratura. L'obiettivo è delineare in maniera precisa le lesioni dal tessuto sano circostante con la stessa qualità ottenuta dalle segmentazioni manuali ma superando l'elevata variabilità inter e intra operatore che caratterizza l'attuale gold standard. Il metodo, inoltre, si propone di essere di semplice implementazione, con pochi parametri scelti dall'operatore e bassa complessità computazionale. Le componenti essenziali del metodo proposto sono: un algoritmo supervoxel per la delineazione dei margini delle lesioni; un'analisi di tessitura multimodale per la caratterizzazione dei tessuti; un classificatore basato sull'analisi della varianza multivariata (MANOVA) per la determinazione della natura dei tessuti di interesse individuati all'interno del volume PET/CT. L'applicazione di tale metodo fornisce all'operatore una mappa delle porzioni di volume che presentano caratteristiche riferibili a tessuti neoplastici neuroendocrini individuati dai loro margini e a cui è assegnato un grado di confidenza con cui si può considerare tale tessuto una lesione.

MATERIALI E METODI

MATERIALI

PAZIENTI

All'interno di questo lavoro sono stati utilizzati i volumi PET/CT di 20 pazienti che rispondessero a tali criteri: presenza di NET a seguito di esame istologico; essersi sottoposti a imaging 68Ga-DOTA-TOC PET/CT prima del trattamento per la determinazione dello stadio della neoplasia; aver firmato il modulo per l'espressione del consenso informato. Ognuno dei 20 pazienti presenta una lesione, queste sono sia primarie che metastatiche. Le lesioni primarie sono di origine pancreatica (40%), gastroenterica (35%), polmonare (25%). Per ognuno dei pazienti sono disponibili i volumi 68Ga-DOTATOC PET e CT e i volumi di interesse (VOI) ottenuti tramite segmentazione manuale di uno specialista delle lesioni all'interno del relativo volume.

IMAGING

Lo scanner PET/CT utilizzato è uno scanner analogico 3D PET (Philips Gemini Dual-slice EXP PET Allegro system con Brilliance CT—Philips Medical Systems, Cleveland, OH). Riferendosi alle linee guida indicate dalle procedure per l'imaging tramite 68Ga-DOTA-TOC PET/CT [16], l'attività del radiofarmaco iniettato era 145.1 ± 25.3 MBq. Dopo 60 minuti dall'assorbimento si è proceduto all'acquisizione delle immagini. L'algoritmo *Ordered Subset Expectation Maximization* (OSEM) è stato utilizzato per la ricostruzione dell'immagine PET con i seguenti parametri: 4 iterazioni, 8 subset e FOV di 576 mm. La risoluzione delle immagini PET è di $4 \times 4 \times 4$ mm³, mentre quella delle immagini CT è $1.1719 \times 1.1719 \times 6.5$ mm³.

METODI

Il modello sviluppato esegue una segmentazione dei volumi PET tramite algoritmo supervoxel per definire le zone di interesse, all'interno delle quali si opera l'analisi di tessitura. Questa viene eseguita sia sul volume PET che sul volume CT, in seguito alla loro registrazione, e pertanto si parla di analisi di tessitura multimodale. Per evitare che eventuali collinearità tra le feature estratte condizionino i risultati della classificazione, viene eseguita un'analisi delle componenti principali. Le nuove variabili individuate dalla PCA sono fornite al classificatore che assegnerà ad ogni porzione di volume di interesse una classe e un grado di confidenza di appartenenza a tale classe. I passi del modello sono schematizzati nella pipeline in figura.

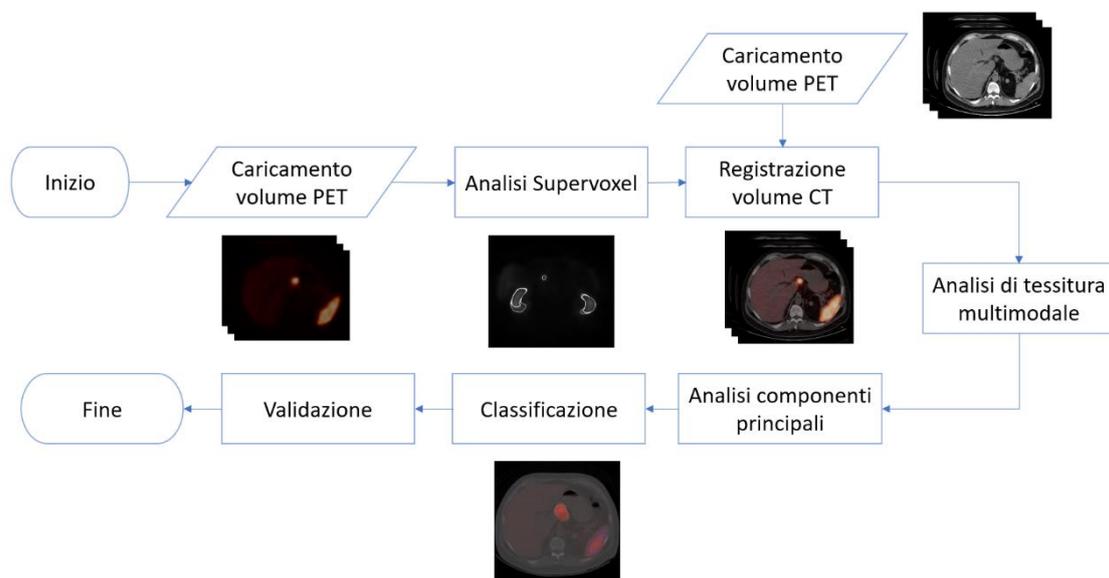


FIGURA 1 PIPELINE DEL METODO PROPOSTO

CARICAMENTO DEL VOLUME

Il modello sviluppato prevede che l'analisi supervoxel, che ha lo scopo di individuare e segmentare le regioni di interesse all'interno del volume, non lavori sul volume PET grezzo, ma su una sua conversione in valori SUV.

L'imaging PET misura la concentrazione di radioattività nel paziente, che sarà correlata, oltre che alle caratteristiche dei tessuti osservati, a fattori di esecuzione dell'esame, come la quantità di radiofarmaco iniettato, e ad altre caratteristiche del paziente, in particolare il peso[17]. Questi parametri introducono variabilità all'interno delle immagini ottenute che idealmente devono essere rimosse, in quanto impediscono il confronto tra pazienti con caratteristiche fisiche diverse o che sono stati esaminati con parametri diversi. Per compensare tale variabilità viene adottata la conversione in valori SUV delle immagini. In questo lavoro la conversione è stata ottenuta utilizzando i parametri contenuti nel file DICOM relativo ai volumi PET (SUV scale factor).

SUPERVOXEL

Il passo successivo dell'algoritmo proposto prevede l'individuazione di zone omogenee del volume che possano essere considerate facenti parte della stessa porzione di organo o tessuto, andando a clusterizzare i voxel e in sostanza segmentare tutti gli oggetti di interesse che si trovano all'interno del volume, come lesioni, organi o loro parti. Per ottenere questo risultato è stato utilizzato un algoritmo di tipo Supervoxel.

Per supervoxel si intende l'agglomerato di un certo numero di voxel che presentano caratteristiche di omogeneità valutate tramite un'opportuna metrica sulle proprietà dei voxel o del loro vicinato, variabile a seconda dell'algoritmo utilizzato. La suddivisione in supervoxel porta quindi alla riorganizzazione dell'immagine in regioni uniformi percettivamente più significative[18].

Gli algoritmi supervoxel vengono generalmente classificati in due categorie: "graph-based" e "gradient-ascent based". Gli algoritmi "graph-based" partono considerando ogni voxel come nodo di un grafo in cui i pesi tra due nodi vengono assegnati proporzionalmente alla somiglianza tra voxel confinanti. I supervoxel vengono quindi ottenuti ottimizzando una funzione di costo associata al grafo. Gli algoritmi "gradient ascent" partono invece da una prima suddivisione approssimativa dei voxel in cluster, questi vengono poi perfezionati iterativamente fino a quando non è raggiunto un criterio di convergenza.

Lavorare su supervoxel, superpixel nel caso di immagini 2D, piuttosto che su voxel presenta significativi vantaggi. Il primo vantaggio è una forte riduzione della complessità computazionale[2][20][19][21]. Le immagini mediche 3D, in particolare quelle volumetriche come le full-body CT/PET, contengono dalle centinaia di migliaia alle centinaia di milioni di voxel. La possibilità di contenere un'elevata quantità di informazioni si ottiene quindi al costo di dover gestire un'altrettanto elevata complessità computazionale in fase di elaborazione dei volumi, di visualizzazione ed analisi degli stessi. L'elaborazione in supervoxel assorbe la ridondanza delle immagini riducendo significativamente il costo computazionale e in termini di memoria di diversi ordini di grandezza. Un secondo vantaggio è rappresentato dal maggior contenuto informativo dei supervoxel rispetto ai voxel originari[19].

Per rappresentare un vantaggio sia come strumento di riduzione della complessità computazionale in fase di pre-processing che come strumento di segmentazione, un algoritmo supervoxel dovrebbe possedere alcune caratteristiche base:

- Dovrebbe occupare poca memoria, essere veloce nel calcolo dei supervoxel e facile da usare
- Dovrebbe ottenere segmentazioni con elevati risultati in tempi rapidi, con supervoxel che seguano bene i contorni degli oggetti

Tenendo a mente queste caratteristiche ottimali, si è quindi scelto di utilizzare l'algoritmo *simple linear iterative clustering* (SLIC) che consente di ottenere sia superpixel che supervoxel. L'algoritmo SLIC presenta tutte le caratteristiche di velocità di esecuzione e qualità di segmentazione elencate in precedenza, con elevate capacità di aderenza ai contorni degli oggetti. Inoltre consente di controllare semplicemente il numero di supervoxel generati e la loro compattezza[2][4].

```

Algorithm 1. SLIC superpixel segmentation
  /* Initialization */
  Initialize cluster centers  $C_k = [l_k, a_k, b_k, x_k, y_k]^T$  by sampling
  pixels at regular grid steps  $S$ .
  Move cluster centers to the lowest gradient position in a  $3 \times 3$ 
  neighborhood.
  Set label  $l(i) = -1$  for each pixel  $i$ .
  Set distance  $d(i) = \infty$  for each pixel  $i$ .

  repeat
    /* Assignment */
    for each cluster center  $C_k$  do
      for each pixel  $i$  in a  $2S \times 2S$  region around  $C_k$  do
        Compute the distance  $D$  between  $C_k$  and  $i$ .
        if  $D < d(i)$  then
          set  $d(i) = D$ 
          set  $l(i) = k$ 
        end if
      end for
    end for
    /* Update */
    Compute new cluster centers.
    Compute residual error  $E$ .
  until  $E \leq \text{threshold}$ 

```

FIGURA 2 PSEUDOCODICE ALGORITMO SLIC PER LA CREAZIONE DI SUPERPIXEL [2]

SLIC si basa su un metodo di clustering k-means, con alcune distinzioni importanti riguardo al numero di distanze calcolate, che viene limitato ad una porzione di spazio intorno al supervoxel, e alla metrica utilizzata per calcolare le distanze, che tiene conto sia della distanza spaziale tra i voxel che della loro intensità[18]. L'unico parametro richiesto in ingresso dall'algoritmo è il numero di supervoxel che si vuole generare, anche se è possibile fornire altri valori come la compattezza dei supervoxel finali.

Una volta definito il numero di supervoxel che si vuole ottenere vengono individuati i centri dei supervoxel, distanziati in maniera regolare. I centri vengono poi spostati sul più vicino minimo locale del gradiente. I voxel vengono quindi assegnati al cluster con il centro più vicino. I centri dei cluster vengono ricalcolati e i voxel riassegnati iterativamente. In post-processing si verifica che tutti i supervoxel creati siano connessi e i voxel orfani vengono assegnati al supervoxel vicino più simile.

A partire dai volumi PET/CT si è quindi applicato l'algoritmo SLIC per effettuare la sovra-segmentazione. Per ogni paziente il primo passo è il calcolo di una soglia sul valore SUV del volume PET, questo poiché si vuole selezionare solo quelle porzioni del volume in cui si ha una presenza rilevante di segnale e che quindi sono significative nella ricerca di eventuali lesioni. Le porzioni di volume al di sopra della soglia vengono memorizzate in una maschera binaria. Si applica quindi l'algoritmo SLIC e si esegue il prodotto tra il volume sovra-segmentato e la maschera appena calcolata. Si ottiene quindi un volume PET/CT in cui sono presenti solo porzioni con indice SUV elevato suddivise in regioni omogenee. I parametri forniti in ingresso sono il numero di supervoxel desiderato N , scelto pari a 2000, e la compattezza ideale dei supervoxel che controlla la regolarità della forma dei supervoxel ottenuti, scelta pari a 0.1, valore consigliato per ottenere una forma

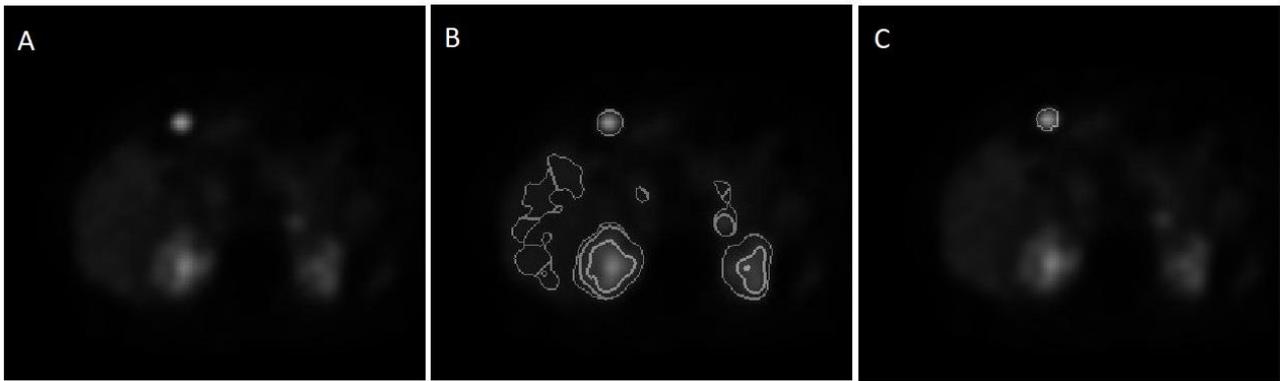


FIGURA 3 ESEMPIO DI CLUSTERIZZAZIONE OTTENUTA TRAMITE ALGORITMO SLIC: A) FETTA DI VOLUME PET CONVERTITO IN VALORI SUV B) FETTA DI VOLUME PET CON CONTORNI DEI SUPERVOXEL INDIVIDUATI C) FETTA DI VOLUME PET CON CONTORNI DELLA SEGMENTAZIONE MANUALE

regolare ma che segua abbastanza flessibilmente i gradienti dell'immagine. Viene applicato un operatore di apertura 3D per ripulire il volume ottenuto da eventuali porzioni di volume che vengono selezionate in quanto presentano valori di SUV sopra soglia, ma le cui dimensioni sono sicuramente inferiori a quelle di una lesione. Su ognuna di queste regioni saranno quindi calcolate le feature/texture di nostro interesse.

Viene inoltre eseguito un controllo sulla corrispondenza spaziale tra la porzione di volume interessata da ogni supervoxel e la maschera binaria che individua la lesione ottenuta tramite segmentazione manuale. Se si individua una sovrapposizione, allora le feature calcolate in quel supervoxel verranno salvate come feature relative ad una lesione. Questo in vista della classificazione che verrà eseguita successivamente.

REGISTRAZIONE VOLUME CT

Una delle componenti fondamentali del modello proposto in questo lavoro è l'analisi di tessitura multimodale, cioè eseguita sulle caratteristiche di più modalità di imaging (PET e CT). Per poter essere certi che le zone del volume su cui vengono calcolati i parametri di tessitura corrispondano alla stessa porzione di tessuti del paziente in entrambe le modalità di imaging è necessario registrare i volumi delle due modalità. La registrazione di immagini è un insieme di tecniche utilizzata nell'elaborazione delle immagini mediche allo scopo di allineare immagini ottenute con diverse modalità di imaging, o diverse condizioni di acquisizione, al fine di poter confrontare le informazioni contenute.

In questo lavoro, la registrazione dei volumi PET e CT è stata ottenuta tramite una trasformazione affine, in cui quindi si conservano le rette ed il parallelismo degli elementi delle immagini, che si basa sul passaggio da un sistema di coordinate basate sull'immagine (ICS) ad un sistema di coordinate basate sul paziente (PCS) per entrambi il volume PET ed il volume CT. Il sistema di coordinate basato sull'immagine è quello comunemente utilizzato per indicizzare i voxel di un volume: i tre assi i , j e k , hanno origine nel primo voxel acquisito, che generalmente si trova nell'angolo in alto a sinistra dell'immagine, e sono diretti verso destra l'asse i , verso il basso l'asse j , e all'indietro l'asse k . Il sistema di coordinate basato sul paziente definisce invece le direzioni degli assi utilizzando le direzioni anatomiche: l'asse x incrementa verso la mano sinistra del paziente, l'asse y verso il piano posteriore, l'asse z verso i piedi. Gli indici PCS sono comuni ad entrambe le modalità di imaging. I parametri per poter effettuare questa trasformazione sono contenuti nei file DICOM dei volumi. A partire da questi si può definire una matrice di trasformazione affine 3D $M_{ICS \rightarrow PCS}$.

La trasformazione è quindi ottenuta come:

$$\begin{bmatrix} P_x^{PET} \\ P_y^{PET} \\ P_z^{PET} \\ 1 \end{bmatrix} = M_{ICS \rightarrow PCS}^{PET} \begin{bmatrix} r^{PET} \\ c^{PET} \\ s^{PET} \\ 1 \end{bmatrix}$$

e

$$\begin{bmatrix} P_x^{CT} \\ P_y^{CT} \\ P_z^{CT} \\ 1 \end{bmatrix} = M_{ICS \rightarrow PCS}^{CT} \begin{bmatrix} r^{CT} \\ c^{CT} \\ s^{CT} \\ 1 \end{bmatrix}$$

Ora che i voxel delle immagini esistono all'interno di un sistema di riferimento comune è possibile proseguire con l'analisi della tessitura su entrambi i volumi PET e CT.

ANALISI DI TESSITURA MULTIMODALE

Il passo successivo dell'approccio adottato in questo lavoro prevede l'estrazione di feature/texture da ogni porzione del volume PET/CT individuata dai supervoxel. L'obiettivo è quello di caratterizzare i tessuti tramite le loro caratteristiche di tessitura e feature globali per individuare una relazione tra tali caratteristiche e la natura del tessuto, andando potenzialmente a discriminare tra tessuti di origine tumorale e porzioni di tessuto sano.

L'analisi di tessitura è l'insieme dei metodi utilizzati per estrarre le caratteristiche di tessitura di un'immagine, o di una sua porzione, dove con caratteristica di tessitura si intende, in senso statistico, la misura quantitativa o semi-quantitativa della distribuzione dei valori di intensità di tale porzione dell'immagine. Complessivamente vengono calcolate 94 feature/texture per ogni supervoxel sia sul volume PET che sul volume CT. Non tutte le caratteristiche calcolate sono di tessitura, in particolare alcune riguardano la forma globale della porzione di tessuto individuata dal supervoxel (solidità e sfericità), mentre altre riguardano la posizione all'interno del volume (coordinate del centroide). Il resto delle feature estratte, per un totale di 89, sono caratteristiche di tessitura scelta all'interno di 5 famiglie: texture globali, GLCM, GRLRM, GLSZM, NGTDM. Per una breve descrizione di ogni texture estratta fare riferimento alla Tabella 1.

Per l'analisi vengono forniti in ingresso il volume PET e CT separatamente, entrambi in seguito ad un pre-processing che ha lo scopo di ricampionarli così da ottenere una risoluzione isotropica dei voxel. Questo per via della natura stessa delle texture misurate che prevedono l'utilizzo di istogrammi 3D, nel caso delle feature globali, o il calcolo di distanze 3D nelle texture matriciali. Tutte le caratteristiche vengono inoltre calcolate in seguito ad una quantizzazione dei volumi su un totale di 128 livelli di grigio.

Tabella 1

GLOBALI Statistiche del primo ordine
<p>Si considera l'istogramma dei livelli di grigio del volume P. Con $P(i)$ si indica quindi il numero di voxel con intensità pari all'i-esimo livello di grigio, mentre con N_g il numero di livelli di grigio. L'istogramma normalizzato è calcolato tramite:</p> $p(i) = \frac{P(i)}{\sum_{i=1}^{N_g} P(i)}$ <p>Le caratteristiche calcolate all'interno di questa classe sono: Varianza, Skewness, Kurtosis.</p>
GLCM Gray-Level Co-occurrence Matrix
<p>Si considera P la GLCM del volume. Con $P(i, j)$ si indica quindi il numero di volte che voxel con intensità pari all'i-esimo livello di grigio si trovano nel vicinato di voxel con il j-esimo livello di grigio, mentre con N_g il numero di livelli di grigio.</p> $p(i, j) = \frac{P(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j)}$ <p>Le caratteristiche calcolate all'interno di questa classe sono: Energia, Contrasto, Correlazione, Omogeneità, Varianza, Somma della media, Entropia, Dissimilarità, Autocorrelazione.</p>
GRLRM Gray-Level Run-Length Matrix
<p>Si considera P la GRLRM del volume. Con $P(i, j)$ si indica quindi il numero di file di voxel con intensità pari all'i-esimo livello di grigio di lunghezza j, con N_g il numero di livelli di grigio e con L_r la lunghezza della fila più lunga.</p> $p(i, j) = \frac{P(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{L_r} P(i, j)}$ <p>Le caratteristiche calcolate all'interno di questa classe sono: Short Run Emphasis (SRE), Long Run Emphasis (LRE), Grey-Level Nonuniformity (GLN), Run-Length Nonuniformity (RLN), Run Percentage (RP), High Gray-Level Run Emphasis (HGRE), Short Run Low Gray-Level Emphasis (SRLGE), Short Run High Gray-Level Emphasis (SHRGE), Long Run Gray-Level Emphasis (LRLGE), Long Run Gray-Level Emphasis (LRHGE), Gray-Level Variance (GLV), Run-Length Variance (RLV).</p>
GLSZM Gray-Level Size Zone Matrix
<p>Si considera P la GLSZM del volume. Con $P(i, j)$ si indica quindi il numero di zone 3D con intensità pari all'i-esimo livello di grigio di dimensione j, con N_g il numero di livelli di grigio e con L_z la dimensione della zona più grande.</p> $p(i, j) = \frac{P(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{L_z} P(i, j)}$ <p>Le caratteristiche calcolate all'interno di questa classe sono: Small Zone Emphasis (SZE), Large Zone Emphasis (LZE), Gray-Level Nonuniformity (GLN), Zone-Size Nonuniformity (ZSN), Zone Percentage (ZP), Low Gray-Level Zone Emphasis (LGZE), High Gray-Level Zone Emphasis (HGZE), Small Zone Low Gray-Level Emphasis (SZLGE), Small Zone High Gray-Level Emphasis (SZHGE), Large Zone Low Gray-Level Emphasis (LZLGE), Large Zone High Gray-Level Emphasis (LZHGE), Gray-Level Variance (GLV), Zone-Size Variance (ZSV).</p>

NGTDM
Neighbourhood Gray-Tone Difference Matrix

Si considera P la NGTDM del volume. Con $P(i)$ si indica quindi la somma delle differenze di grigio tra tutti i voxel con intensità pari all' i -esimo livello di grigio e il valore medio di intensità del corrispettivo insieme 26 connesso 3D, con N_g il numero di livelli di grigio predefinito e con $(N_g)_{eff}$ il numero effettivo di livelli di grigio. L' i -esimo valore è calcolato tramite:

$$P(i) = \begin{cases} \sum_{\text{all voxels} \in \{N_i\}} |i - \bar{A}_i| & \text{if } N_i > 0, \\ 0 & \text{if } N_i = 0. \end{cases}$$

Si definisce inoltre:

$$n_i = \frac{N_i}{N}$$

Le caratteristiche calcolate all'interno di questa classe sono: Coarseness, Contrasto, Busyness, Complessità, Forza.

Le definizioni sono riportate da [22].

ANALISI DELLE COMPONENTI PRINCIPALI

In seguito all'analisi di tessitura il nostro dataset sarà caratterizzato da un'elevata dimensionalità, con possibili collinearità tra le feature che lo costituiscono. Questi sono due fattori che influenzano la qualità della classificazione, poiché due gruppi di feature possono portare informazioni ridondanti o in contrasto tra loro, e necessitano quindi di essere presi in considerazione[23][24]. A questo scopo viene eseguita l'analisi delle componenti principali (PCA).

L'analisi delle componenti principali è una tecnica che permette di diminuire la dimensionalità di un dataset, favorendo la conservazione delle componenti significative e minimizzando la perdita di informazioni[9][10]. Questo viene ottenuto andando a creare delle nuove variabili, scorrelate tra loro, come combinazione lineare delle variabili del dataset, con l'obiettivo di conservare la massima quota di variabilità possibile. Le nuove variabili sono chiamate componenti principali (PC).

Normalmente i dataset su cui viene applicata la PCA sono delle matrici $n \times p$ dove per ogni osservazione/individuo n sono presenti p variabili numeriche. Individuare le componenti principali equivale a calcolare gli autovalori/autovettori della matrice di covarianza o di correlazione della matrice del dataset. Data la natura progressiva della ricerca delle componenti principali, che prevede che per ogni nuova componente principale si massimizzi la variabilità del dataset presa in considerazione, ogni nuova componente principale spiegherà una percentuale di quest'ultima inferiore a quella della precedente. La variabilità spiegata da ogni componente principale viene generalmente espressa sotto forma di percentuale della variabilità totale del dataset originale. Questo parametro diventa quindi fondamentale per scegliere il numero di componenti principali che si utilizzerà.

Nonostante la PCA non abbia richieste particolari in termini delle caratteristiche dei dati forniti in ingresso, normalmente viene eseguita una normalizzazione degli stessi:

$$z_{ij} = \frac{x_{ij} - X_j}{\sigma_j}$$

L'osservazione normalizzata z_{ij} viene quindi ottenuta a partire dall'osservazione originale x_{ij} sottraendo la media di tutte le osservazioni per la variabile considerata X_j e dividendo per la deviazione standard della stessa σ_j . Nel nostro dataset sono presenti 94 variabili, cioè le feature estratte dai supervoxel, mentre le osservazioni hanno un numero variabile corrispondente al numero di supervoxel individuati all'interno del volume del paziente. Ogni colonna (variabile) viene inoltre normalizzata rispetto al valore massimo assunto dalla variabile stessa.

Il numero di componenti principali prese in considerazione al termine del processo è quello necessario a spiegare almeno il 90% delle variabilità originale.

CLASSIFICATORI

Il dataset, ora costituito dalla proiezione delle osservazioni sul nuovo spazio costituito dalle componenti principali, è pronto per essere classificato. A questo scopo sono stati valutati tre diversi modelli di classificazione, il primo è un classificatore basato su un'analisi della varianza multivariata (MANOVA), il secondo su una support vector machine (SVM), mentre il terzo modello è una rete neurale.

MANOVA

A partire dai dati in ingresso, questo modello di classificatore effettua un'analisi MANOVA così da individuare un nuovo insieme di variabili, dette variabili canoniche, che permette di massimizzare la separazione tra i dati delle due classi. A partire dalle distribuzioni delle proiezioni dei dati sul nuovo sistema di riferimento per le due classi, apprese in fase di training, il classificatore è in grado di assegnare una classe alle nuove osservazioni in base alla loro proiezione nello spazio delle variabili canoniche.

L'analisi della varianza multivariata è una procedura statistica che permette di verificare delle ipotesi riguardo alle medie di due o più popolazioni. Si tratta di una forma di analisi della varianza (ANOVA) utilizzata nel caso siano presenti più variabili dipendenti, nel nostro caso le feature estratte dai supervoxel, e una o più variabili indipendenti o gruppi, nel nostro caso la distinzione è tra lesioni e non lesioni. I dati che vengono forniti in ingresso a MANOVA sono costituiti dagli score della PCA, cioè le proiezioni delle osservazioni sulle componenti principali.

La MANOVA è un'estensione dell'analisi della varianza (ANOVA) ad una situazione con più variabili dipendenti. Per verificare la relazione tra le medie di due gruppi, l'analisi della varianza ripartisce la varianza dei dati in componenti a seconda dell'origine di tale varianza, che può essere inter o intra gruppo. Definisce poi un test statistico, il rapporto F (F-ratio), come rapporto tra la varianza tra i due gruppi e la varianza all'interno dei gruppi.

$$F = \frac{\text{varianza tra gruppi}}{\text{varianza nei gruppi}}$$

Un valore elevato del rapporto indica che le differenze tra i due gruppi sono maggiori del rumore o variabilità all'interno dei singoli gruppi. Tale statistica viene utilizzata per valutare l'ipotesi nulla. L'estensione di questo test al caso multivariato non è immediata ma sono presenti altri test statistici, come la lambda di Wilks, la Traccia di Pillai, la traccia di Hotelling-Lawl e l'autovalore massimo di Roy, che possono essere utilizzati, anche se solitamente, dato che le distribuzioni di tali statistiche sono complesse da calcolare, queste vengono trasformate così da riportarsi ad una distribuzione di tipo F.

L'analisi MANOVA è interessante perché, come anticipato, fornisce in uscita delle nuove variabili, le variabili canoniche, che sono combinazione lineare delle variabili fornite in ingresso, che vanno a massimizzare la separazione tra le due classi. Questo rende più semplice, osservando la proiezione del dataset sul nuovo spazio, effettuare una separazione lineare delle osservazioni nelle due classi. In questo lavoro si è considerata solamente la prima variabile canoniche, poiché si è osservato che essa risulta sufficiente per separare le due classi mentre variabili canoniche successive non aggiungono capacità discriminativa al modello.

SUPPORT VECTOR MACHINE

L'SVM è una tecnica di apprendimento supervisionato utilizzato nella classificazione e analisi di dati. Uno dei pregi di questa tecnica è la possibilità di ottenere classificazioni non lineari.

Considerando i dati iniziali come variabili p -dimensionali, quindi punti in uno spazio di dimensioni finite p , lo scopo di un algoritmo di classificazione è generalmente quello di individuare un piano di dimensioni $p-1$ in grado di separare le due classi di dati. Questo tipo di separazione lineare non è sempre facile da individuare e può essere necessario ricorrere a separazioni non-lineari.

Gli algoritmi SVM ri-mappano lo spazio in cui si trovano i dati in un nuovo spazio di dimensioni superiori o infinite, facilitando la ricerca del piano, o gruppo di piani, in grado di ottenere la separazione tra le due classi di dati. Inoltre, poiché i vettori che definiscono il nuovo spazio sono calcolati in termini di funzioni kernel, che possono essere non lineari, la separazione che si ottiene può essere a sua volta non lineare. La separazione in due classi può però essere ottenuta attraverso più possibili piani, l'obiettivo è quindi quello di individuare il piano di che possiede la massima distanza possibile dai punti di entrambe le classi. Questo permette di migliorare la qualità della classificazione.

Come nel caso del classificatore MANOVA, anche l'SVM viene applicato sugli score della PCA, considerando un numero di componenti principali capace di spiegare almeno il 90% della variabilità totale. L'uscita del classificatore SVM fornisce oltre ad una classe per ogni nuova osservazione fornita in ingresso al modello, un grado di appartenenza a tale classe sotto forma di percentuale.

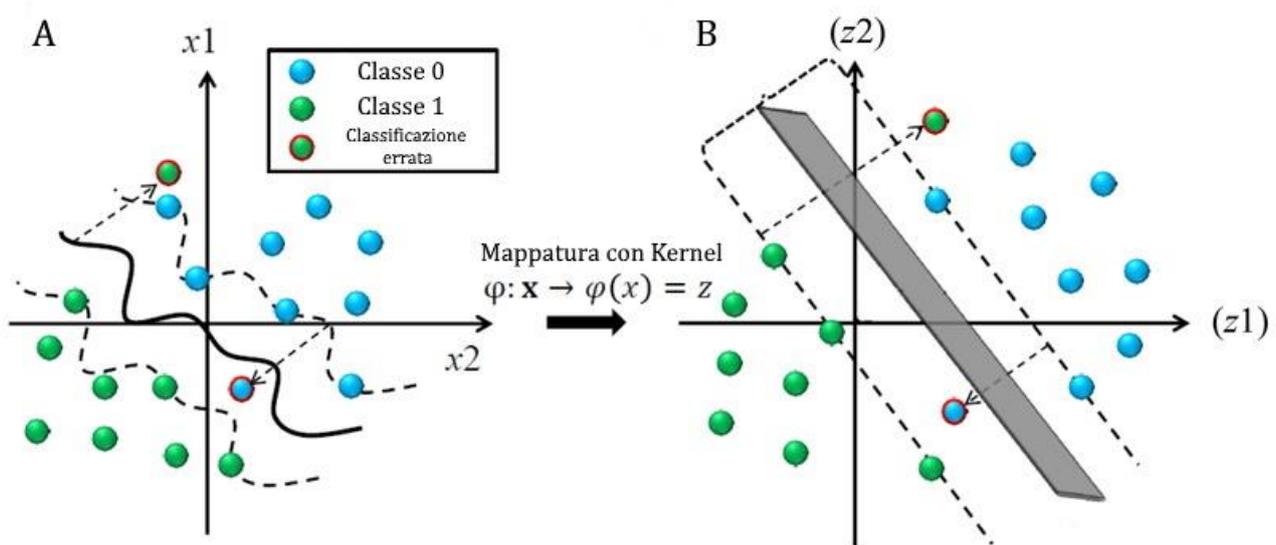


FIGURA 4 TRASFORMAZIONE DEL DATASET ESEGUITA DA UN ALGORITMO SVM TRAMITE MAPPATURA CON KERNEL DALLO SPAZIO ORIGINALE (A) AL NUOVO SPAZIO (B)

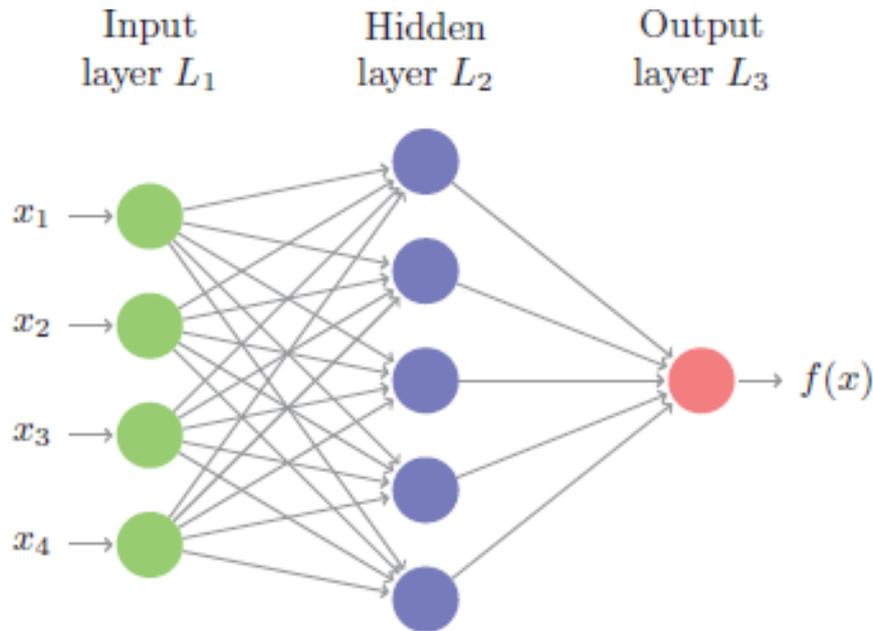


FIGURA 5 STRUTTURA TIPICA DI UNA RETE NEURALE FEEDFORWARD

RETI NEURALI

Le reti neurali sono una famiglia di algoritmi di machine learning, ispirati dal funzionamento delle connessioni neurali, ampiamente utilizzati in problemi di classificazione[27]. Esistono diverse tipologie di reti neurali che si distinguono tra loro per l'architettura della rete, per le caratteristiche delle unità base di tale rete, i nodi o neuroni, e per ulteriori fattori come la tipologia di apprendimento su cui si basano o il metodo con cui le informazioni vengono trasmesse all'interno della rete.

In questo lavoro si è sviluppato una rete neurale feedforward multistrato. Alla base della rete neurale ci sono i neuroni, o nodi, organizzati in strati e legati tra loro attraverso connessioni pesate il cui valore viene determinato in fase di addestramento della rete. Ogni neurone possiede più ingressi, alimentato da dati esterni o da neuroni di strati precedenti, ed una sola uscita, che può essere fornita ad uno o più neuroni dello strato successivo. Un neurone è rappresentato da una funzione, lineare o meno, che determina, in base ai valori ottenuti in ingresso, quale sarà l'uscita del neurone stesso. Il risultato complessivo della rete è costituito dall'uscita dei neuroni dell'ultimo strato, chiamato Output Layer, che determina la risposta finale del sistema. La struttura tipica, come riportata in figura, prevede un primo strato, chiamato Input Layer, che costituisce l'ingresso della rete, uno o più strati nascosti (Hidden layer) ed uno strato finale, l'Output layer, che costituisce l'uscita della rete. Le informazioni si spostano all'interno della rete in una sola direzione (feedforward), attraversando i neuroni dei diversi strati in ordine sequenziale a partire dallo strato di input. La correttezza dell'uscita calcolata dalla rete viene poi verificata confrontando il risultato prodotto con il valore atteso, utilizzando una tecnica di apprendimento supervisionato chiamata backpropagation, in cui a partire dall'errore misurato grazie a tale confronto, i pesi delle connessioni tra i neuroni dei diversi strati vengono aggiornati procedendo a ritroso lungo la struttura della rete, in modo tale da ridurre l'errore.

Il numero di neuroni in ogni strato ed il numero di strati utilizzati non è predefinito ma varia a seconda delle necessità. In questo lavoro è stata utilizzata una rete con tre hidden layer, rispettivamente contenenti 90, 60 e 30 neuroni ciascuno. L'Input layer è costituito da un numero di neuroni pari al numero di componenti principali scelte, questo perché alla rete verranno forniti come ingressi le proiezioni delle variabili di ogni osservazione sullo spazio delle componenti principali. L'uscita è invece costituita da un solo neurone, che assegnerà una classe alle nuove osservazioni, oltre ad una probabilità di appartenenza a tale classe.

ADDESTRAMENTO E VALIDAZIONE

Al fine di addestrare e validare il modello proposto in questo lavoro, il dataset iniziale, composto da 20 pazienti con lesione primaria, è stato separato in un set di training di 15 pazienti ed uno di test, contenente i restanti 5 pazienti. Per entrambi i subset l'elaborazione si è svolta seguendo la stessa procedura dalla fase di caricamento dei volumi PET fino all'analisi delle componenti principali. L'approccio differisce per quanto riguarda lo step di classificazione.

Dato lo sbilanciamento del dataset, formato da molti più oggetti di origine non tumorale, ai classificatori viene fornito in ingresso un dataset ribilanciato. A partire dal dataset originale si identificano tutti gli oggetti che sono indicati come lesioni nelle segmentazioni manuali eseguite dai medici, si sceglie quindi un numero di oggetti non lesioni tale che il rapporto lesioni/non lesioni sia di circa 3:1 per ridimensionare lo

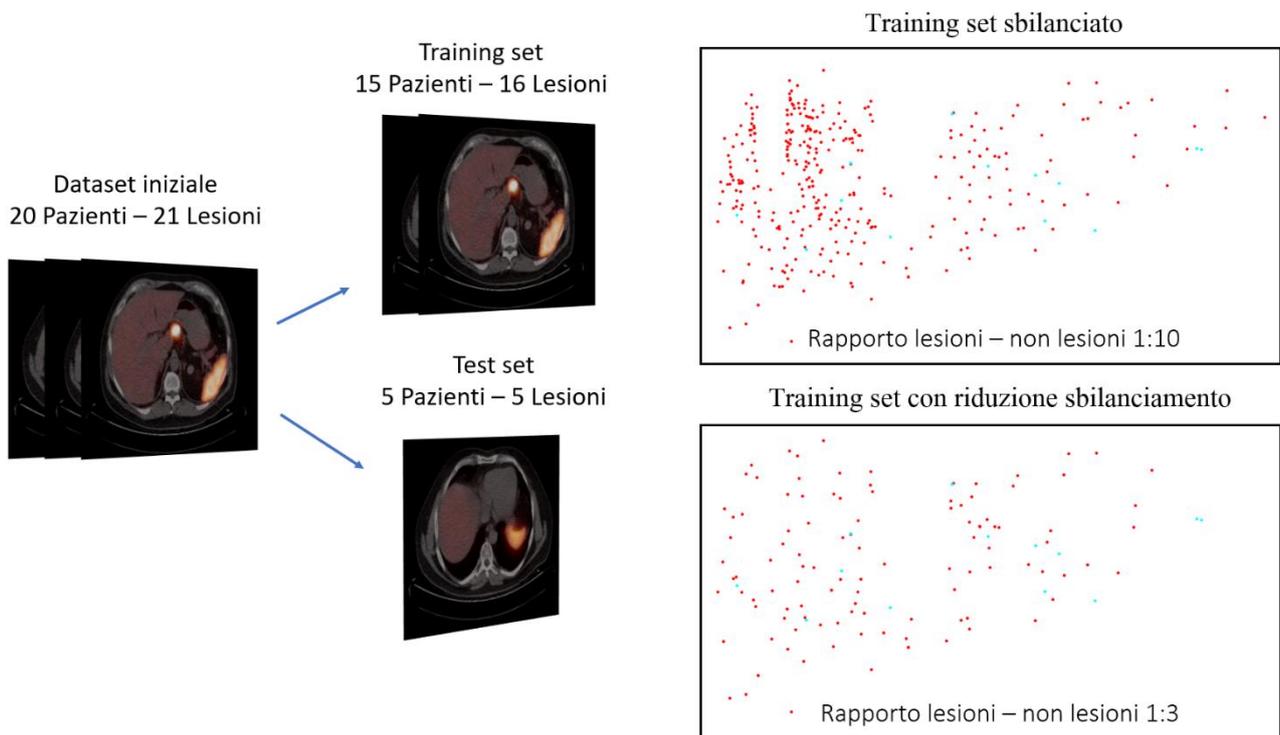


FIGURA 6 IL DATASET INIZIALE VIENE SCOMPOSTO IN DUE SUBSET AI FINI DELLA VALIDAZIONE. IL TRAINING SET È COSTITUITO DA QUINDICI PAZIENTI, IL TEST SET DAI RESTANTI CINQUE. IL TRAINING SET VIENE INOLTRE RIDIMENSIONATO IN MODO DA RIDURRE LO SBILANCIAMENTO TRA IL NUMERO DELLE ZONE NON LESIONALI E QUELLE LESIONALI UTILIZZATE PER L'ADDESTRAMENTO DEI CLASSIFICATORI.

sbilanciamento ma non annullarlo completamente. Tale numero di oggetti viene quindi scelto casualmente all'interno del dataset originale. Il dataset così generato viene utilizzato per la fase di addestramento dei classificatori. Poiché la scelta è quindi casuale si pone il problema di come questa possa influire sulla classificazione. Gli oggetti non lesionali sono infatti altamente eterogenei, provenendo da organi e tessuti diversi, e alcuni sottoinsiemi potrebbero presentare delle caratteristiche altamente favorevoli, o sfavorevoli, alla classificazione. Per ridurre tale possibilità, si è scelto di valutare ogni paziente del test set cinque volte, così da poter valutare anche la consistenza della classificazione al variare del sottoinsieme del training set utilizzato.

TRAINING

Durante la fase di addestramento l'obiettivo è quello di addestrare il classificatore, attraverso degli esempi, a distinguere le lesioni dagli altri oggetti segmentati all'interno del volume nel modo più accurato possibile. Tutti i metodi di classificazione adottati sono tecniche di apprendimento supervisionato, cioè necessitano di conoscere la classe di appartenenza degli oggetti le cui caratteristiche vengono fornite in ingresso.

Per quanto riguarda il classificatore basato sulla MANOVA in fase di training si considerano le distribuzioni delle proiezioni degli score della PCA sulle variabili canoniche delle due classi separatamente, si calcolano le distribuzioni gaussiane che meglio le approssimano e si misura la media e la deviazione standard delle due gaussiane ottenute. Si estraggono così quattro valori di riferimento, media e deviazione standard della distribuzione riferita alle non lesioni e media e deviazione standard della distribuzione riferita alle lesioni. Questi quattro valori vengono quindi memorizzati per essere utilizzati in fase di test. Oltre ai valori peculiari delle distribuzioni viene salvato l'insieme di matrici che definiscono le trasformazioni effettuate durante la PCA e la MANOVA sul training set.

Il training del classificatore SVM coincide con la creazione del modello SVM che verrà utilizzato in fase di test. In particolare, il modello viene addestrato in modo tale che fornisca in uscita, oltre alla classe di appartenenza della nuova osservazione, una probabilità di appartenenza a tale classe. Sia la creazione del modello SVM che il suo addestramento sono stati eseguiti tramite funzioni Matlab.

Per quanto riguarda la rete neurale, in fase di training viene creata la struttura della rete, specificando il numero e la dimensione di ogni layer nascosto ed essa viene poi allenata fornendo una alla volta tutte le osservazioni che compongono il training set. Il numero di neuroni del layer di ingresso è legato al numero di componenti principali scelto precedentemente. I pesi e i parametri della rete vengono quindi salvati per la fase di test.

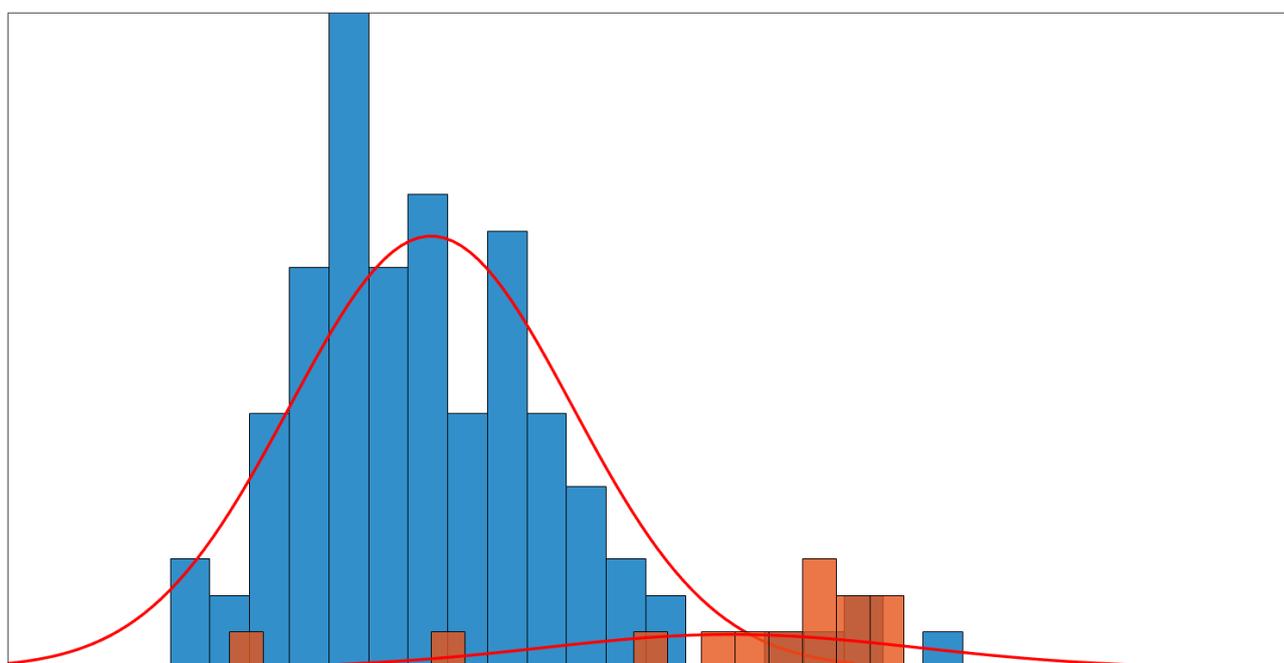


FIGURA 7 ESEMPIO DI DISTRIBUZIONI DELLE PROIEZIONI DEGLI SCORE DELLA PCA SULLE VARIABILI CANONICHE PER DUE CLASSI DI DATI. IN ROSSO L'ANDAMENTO DELLE DISTRIBUZIONE GAUSSIANA CHE MEGLIO APPROSSIMA LE DUE DISTRIBUZIONI

TEST

Durante la fase di test è invece il classificatore a determinare, seguendo quanto appreso in fase di training, quale sarà la classe di appartenenza della nuova osservazione.

In fase di test del classificatore MANOVA, a partire dalle matrici salvate, vengono proiettate le nuove osservazioni, i.e. l'insieme delle feature delle nuove porzioni di volume, sulle componenti principali e, in seguito, sulle variabili canoniche. Il valore della prima variabile canonica di ogni osservazione viene quindi confrontato con i parametri delle distribuzioni calcolate in fase di training, per classificare il nuovo dato. A seconda della distanza dalla media delle due classi, viene assegnata una classe alla nuova osservazione. Inoltre, se la distanza dalla media di una delle due classi si trova all'interno di un range definito dalla deviazione standard di tale classe, l'assegnazione di classe viene effettuata con un maggiore grado di certezza. Questo viene indicato all'interno del classificatore tramite un valore numerico intero che identifica per l'appunto il grado di certezza con cui una porzione di tessuto può essere considerata lesione.

La fase di test per i restanti due classificatori, SVM e Reti Neurali, risulta più semplice, in quanto è sufficiente fornire in ingresso al modello o alla rete già addestrati la nuova osservazione per ottenere una classe e una percentuale di confidenza della classificazione.

METRICHE

All'interno di questo lavoro sono state utilizzate due metriche per valutare i risultati ottenuti: il coefficiente Sørensen-Dice per valutare la qualità della segmentazione ottenuta dall'algorithm supervoxel, e la media del grado di classificazione delle due classi per valutare la qualità delle classificazioni ottenute.

Il coefficiente Dice è una metrica comunemente utilizzata per valutare la somiglianza tra due segmentazioni e viene calcolata come il rapporto tra due volte l'area di sovrapposizione di due immagini, o maschere binarie nel caso di questo lavoro, e la somma dei pixel totali che compongono le due immagini. I valori della metrica variano tra 0 ed 1.

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

Considerando singolarmente ogni oggetto individuato dall'algorithm supervoxel, il coefficiente Dice è stato utilizzato per valutare la sovrapposizione tra i contorni ottenuti dal supervoxel e quelli individuati manualmente dal medico, che costituiscono la nostra ground truth. Nel caso l'oggetto valutato non fosse una lesione, o una sua parte, il valore Dice ottenuto sarà molto basso, tendenzialmente nullo. Tali casi non saranno considerati nell'analisi dei risultati e si osserveranno solamente i valori ottenuti per le segmentazioni delle lesioni.

Ogni classificatore oltre ad assegnare una classe ad ogni nuova osservazione, valuta anche il grado di confidenza con cui tale classe viene assegnata. Come conseguenza i risultati di classificazione non si limitano a definire una porzione di tessuto come lesione o no, ma forniscono un'indicazione della sicurezza con cui il modello afferma che un'oggetto è di origine tumorale. In particolare, il modello usa una scala su cinque gradi di pericolosità, partendo da "sicuramente non lesione" fino a "sicuramente lesione". Il grado intermedio corrisponde ad un'incertezza di classificazione, per cui la porzione di tessuto valutata presenta caratteristiche

inconsistenti con l'una o l'altra classe e pertanto non gli può essere assegnata una classe in maniera univoca. Per costruire la metrica si separano gli oggetti in base alla classe reale di appartenenza, che viene determinata basandosi sulle segmentazioni manuali del volume, e si mediano i gradi assegnati a tutti gli oggetti di ciascuna classe. Dal momento che ogni paziente viene valutato cinque volte in fase di test, si otterranno cinque valori medi per entrambe le classi. Ai fini della valutazione del risultato questi cinque valori vengono a loro volta media, così da poter valutare quanto consistente è il classificatore e ripulire i risultati da eventuali fluttuazioni casuali dei risultati che potrebbero verificarsi.

RISULTATI

SUPERVOXEL

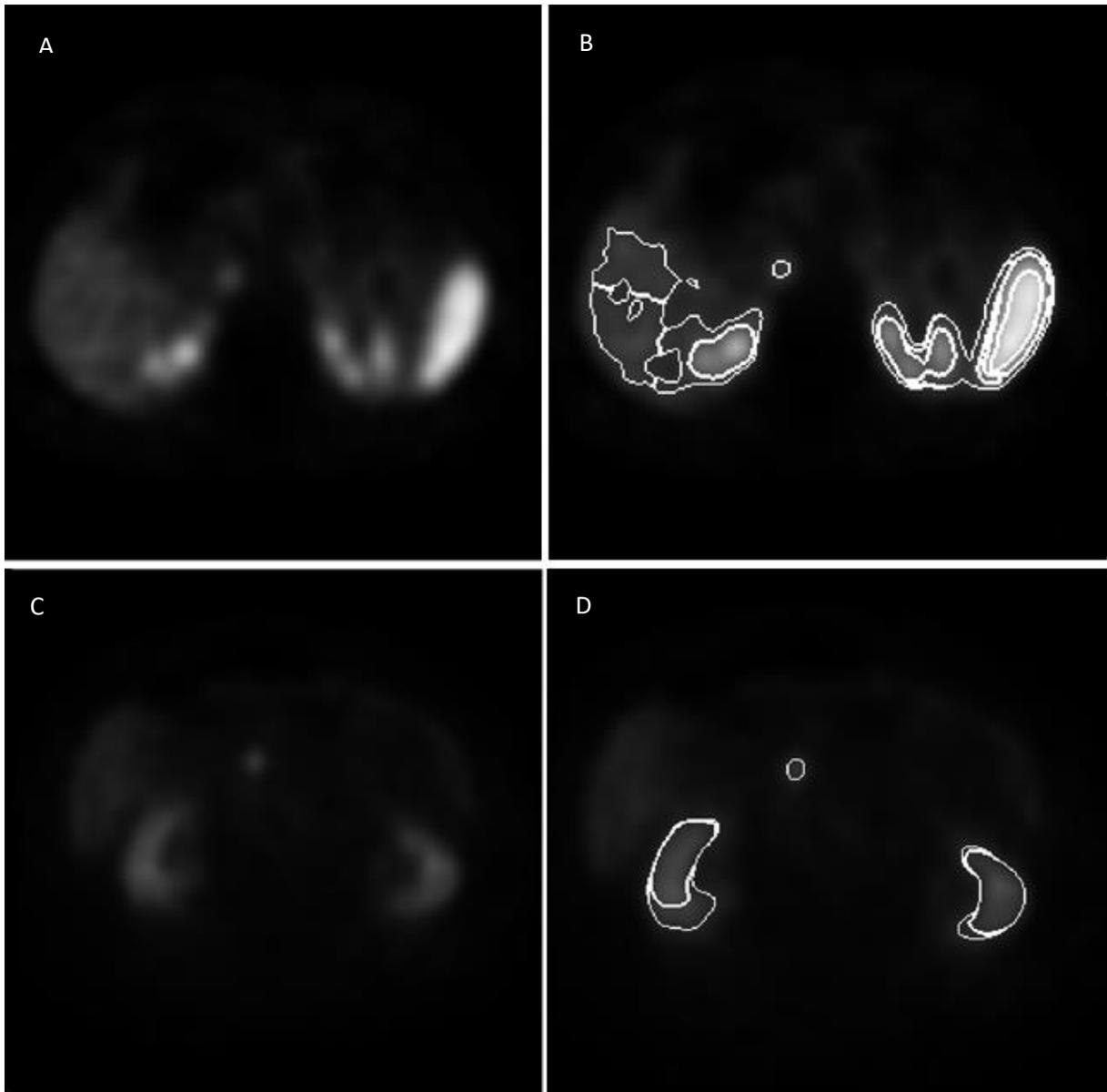


FIGURA 8 RISULTATI OTTENUTI CON ALGORITMO SUPERVOXEL: A & C) VOLUMI PET DEI PAZIENTI CONVERTITI IN VALORI SUV. B & C) CONTORNI DEI SUPERVOXEL OTTENUTI PER LE PORZIONI DI VOLUME CON VALORE SUV SOPRASOGLIA

Per definire le zone di interesse, e quindi individuare e segmentare le porzioni rilevanti del volume PET, è stato utilizzato un algoritmo supervoxel. A partire da una conversione in valori SUV del volume paziente, elaborato tramite thresholding con una soglia definita sulla distribuzione di tali valori SUV per escludere dalla ricerca le zone con indice di assorbimento del radiofarmaco non significativo, l'algoritmo supervoxel è stato in grado di clusterizzare i voxel che compongono il volume così da individuare e segmentare le porzioni di organi, tessuti e lesioni all'interno dell'immagine PET.

Come si può osservare in figura, i contorni dei supervoxel si adattano efficacemente ai contorni degli oggetti presenti nell'immagine, guidati dai gradienti dei tessuti, ma sono sufficientemente compatti da non essere sensibili al rumore presente sull'immagine così da seguire le caratteristiche rilevanti della forma degli oggetti.

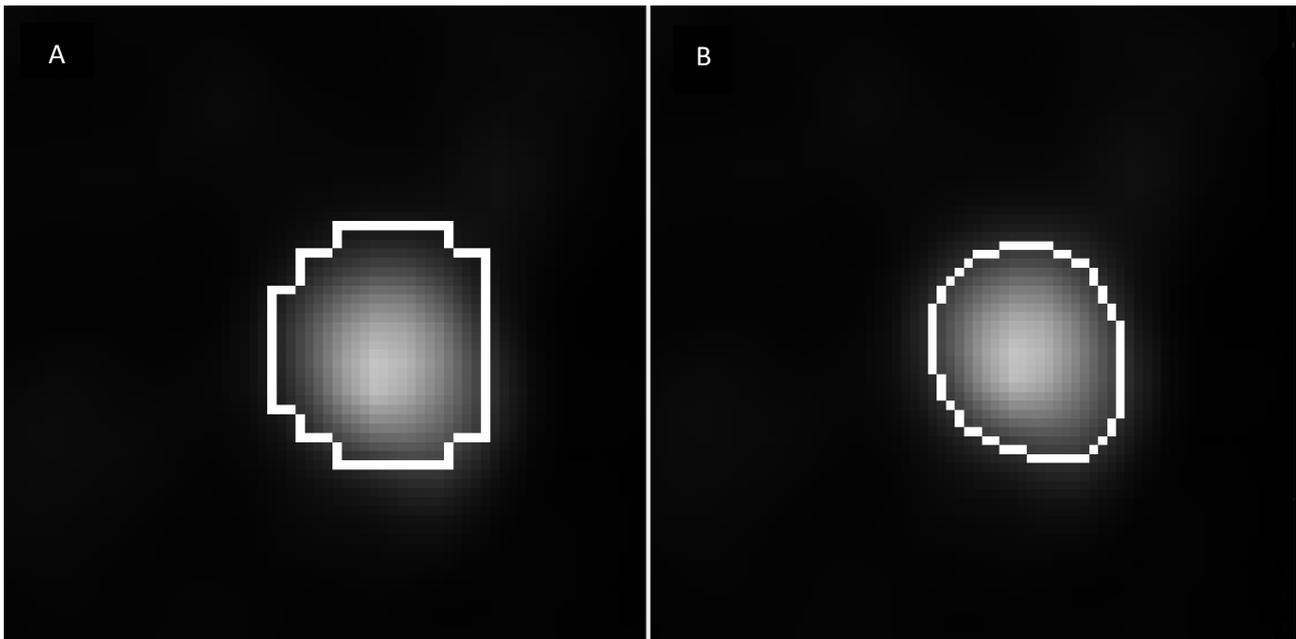


FIGURA 9 CONTORNI DELLA LESIONE DI UNO DEI PAZIENTI: A) SEGMENTAZIONE MANUALE - GROUND TRUTH B) SEGMENTAZIONE AUTOMATICA OTTENUTA TRAMITE ALGORITMO SUPERVOXEL

La qualità della segmentazione ottenuta è stata valutata calcolando il coefficiente Dice, al fine di misurare la similitudine tra la segmentazione delle lesioni ottenuta tramite metodo automatico e quella manuale utilizzata come ground truth. Il valore medio ottenuto su tutte le lesioni del test set è $DSC = 81.48 \pm 9.92 \%$

Si riportano i valori Dice calcolati per le lesioni dei cinque pazienti valutati nel test set.

Tabella 2. Valori percentuali del coefficiente Dice per i pazienti del test set

	Paziente 1	Paziente 2	Paziente 3	Paziente 4	Paziente 5
DSC (%)	79.94 %	77.54 %	95.73 %	68.89 %	85.33 %

ANALISI DELLE COMPONENTI PRINCIPALI

Data l'elevata dimensionalità del dataset generato dell'analisi di tessitura e la possibilità che le feature calcolate possano essere tra loro correlate è stata eseguita un'analisi delle componenti principali al fine di creare delle nuove variabili, scorrelate tra loro, come combinazione lineare delle variabili del dataset, con l'obiettivo di conservare la massima quota di variabilità possibile. Le nuove variabili sono chiamate componenti principali (PC). Il numero delle componenti principali generate dall'analisi è pari al numero di variabili presenti nel dataset originario e ognuna di esse è responsabile di una quota della variabilità totale. Questo è dovuto alla natura progressiva della ricerca delle componenti principali, che prevede che per ogni nuova componente principale si massimizzi la variabilità del dataset presa in considerazione, ogni nuova componente principale spiegherà quindi una percentuale di quest'ultima inferiore a quella della precedente. Il numero di componenti principali prese in considerazione in questo lavoro è quello necessario a spiegare almeno il 90% delle variabilità originale, che si è osservato essere generalmente contenuto tra le 6 e le 9 componenti principali.

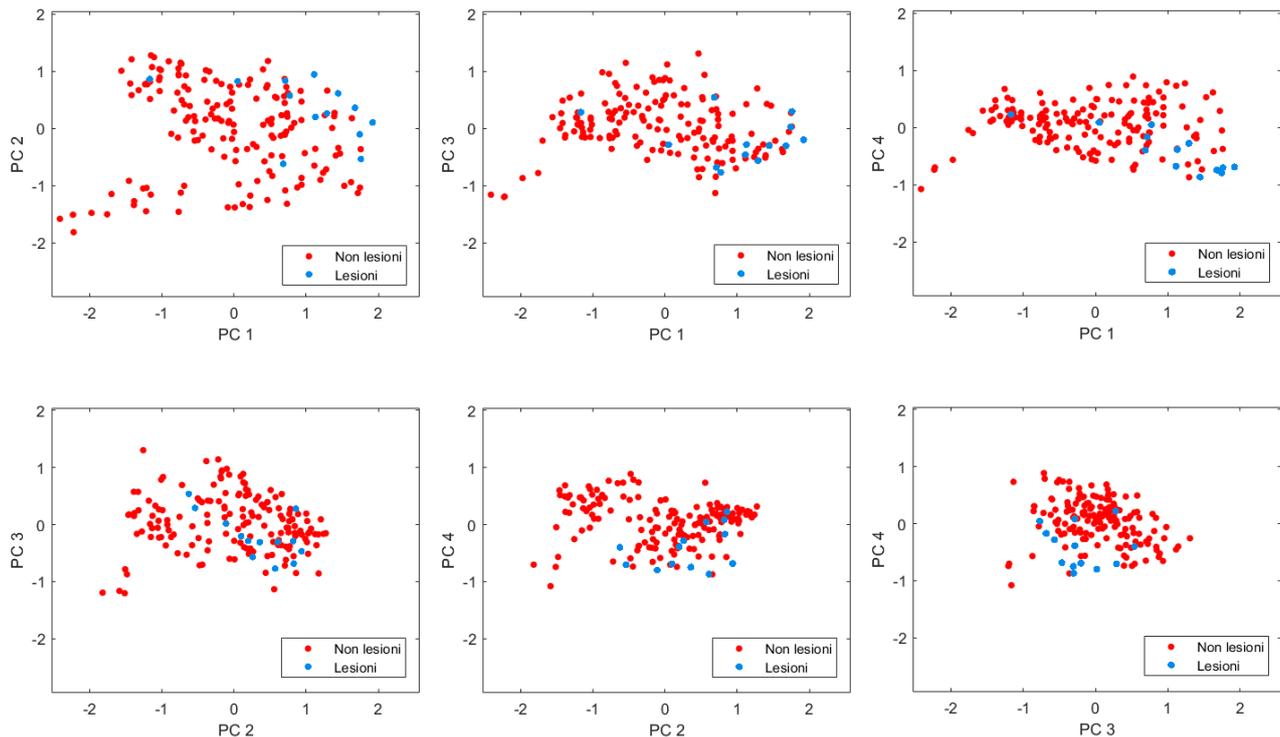


FIGURA 10 CONFRONTO DEGLI SCORE DELLA PCA DI UNA PORZIONE DEL DATASET PER LE PRIME QUATTRO COMPONENTI PRINCIPALI

CLASSIFICAZIONE

Tre tipologie di classificatori sono state impiegate all'interno di questo lavoro: un classificatore lineare basato sull'analisi della varianza multivariata (MANOVA); un classificatore non lineare basato su SVM; una rete neurale feedforward. Tutti e tre i classificatori lavorano sugli score delle variabili dell'analisi di tessitura multimodale ottenute in seguito all'analisi delle componenti principali e tutti e tre forniscono in uscita una classe per ogni porzione di volume di interesse ottenute tramite algoritmo supervoxel ed un grado che esprime la confidenza con cui tale classe è stata assegnata. Alcuni esempi dei risultati ottenuti con i tre metodi sono riportati in figura.

Ad ogni oggetto individuato viene assegnata una classe ed un grado, indicato come un valore intero, che esprime il grado di confidenza con cui tale classe è stata assegnata. Il grado varia su una scala da 0 a 4, dove 0 indica una assegnazione alla classe delle non lesioni netta e 4 un'assegnazione alla classe delle lesioni con netta. Un grado pari a 2 indica invece un'incertezza da parte del classificatore nel decidere la classe dell'oggetto. Il grado degli oggetti è rappresentato graficamente tramite una scala colore applicata sui margini della segmentazione di tale oggetto ottenuta tramite supervoxel. Nelle immagini di esempio in figura 9, sono riportate due fette di uno dei pazienti con la classificazione ottenuta tramite i tre metodi. Le sezioni riportate raffigurano una la lesione ed uno dei reni del paziente, l'altra la vescica. Tali porzioni sono state scelte poiché sono zone con elevato indice di assorbimento relativo (SUV) e che quindi sono sempre presenti e ben evidenziate all'interno dei volumi. Queste zone sono quindi sempre considerate di interesse durante l'applicazione del metodo proposto ed è dunque rilevante che siano sempre classificate in maniera adeguata.

Per quantificare la qualità della classificazione ottenuta con i tre metodi si è deciso di osservare il grado medio assegnato da ogni classificatore agli oggetti delle due classi. Separando le porzioni di volume nelle due classi utilizzando come riferimento le segmentazioni manuali, si è calcolata la media dei gradi assegnati agli oggetti

che ne fanno parte. Dal momento che la classificazione di ogni paziente viene ripetuta cinque volte in fase di validazione, il valore medio viene ottenuto considerando l'insieme dei gradi assegnati agli oggetti di ognuna delle due classi (lesioni e non lesioni) per tutte le ripetizioni. Lo stesso insieme di valori viene utilizzato per valutare la deviazione standard, che quindi tiene conto non solo della variabilità dei gradi all'interno della classe ma anche all'interno di classificazioni ripetute.

Tabella 3. Gradi medi assegnati dai classificatori ai pazienti del test set per le due classi di interesse.

	Manova		SVM		Rete neurale	
	Lesioni	Non lesioni	Lesioni	Non lesioni	Lesioni	Non lesioni
Paziente 1	4.00 ± 0	1.50 ± 0.56	3.00 ± 0	0.40 ± 0.30	3.00 ± 0	0.50 ± 0.40
Paziente 2	4.00 ± 0	0.25 ± 0.25	4.00 ± 0	0 ± 0	2.00 ± 0	0 ± 0
Paziente 3	4.00 ± 0	1.00 ± 0.54	2.00 ± 0	0.80 ± 0.80	0 ± 0	1.60 ± 0.97
Paziente 4	4.00 ± 0	2.11 ± 0.48	3.00 ± 0	0.44 ± 0.44	2.00 ± 0	1.33 ± 0.66
Paziente 5	4.00 ± 0	0.83 ± 0.16	4.00 ± 0	0 ± 0	4.00 ± 0	0 ± 0
Media	4.00 ± 0	1.13 ± 0.40	3.20 ± 0	0.33 ± 0.31	2.20 ± 0	0.68 ± 0.40

In tabella 3 sono riportati i valori ottenuti per tale metrica per i cinque pazienti che costituiscono il set di validazione. I valori sono riportati separatamente per le due classi e per i metodi di classificazione utilizzati. Un valore medio dei risultati ottenuti per l'insieme dei cinque pazienti è riportato alla fine, come indicazione dei risultati ottenuti dallo specifico classificatore per le due classi di interesse.

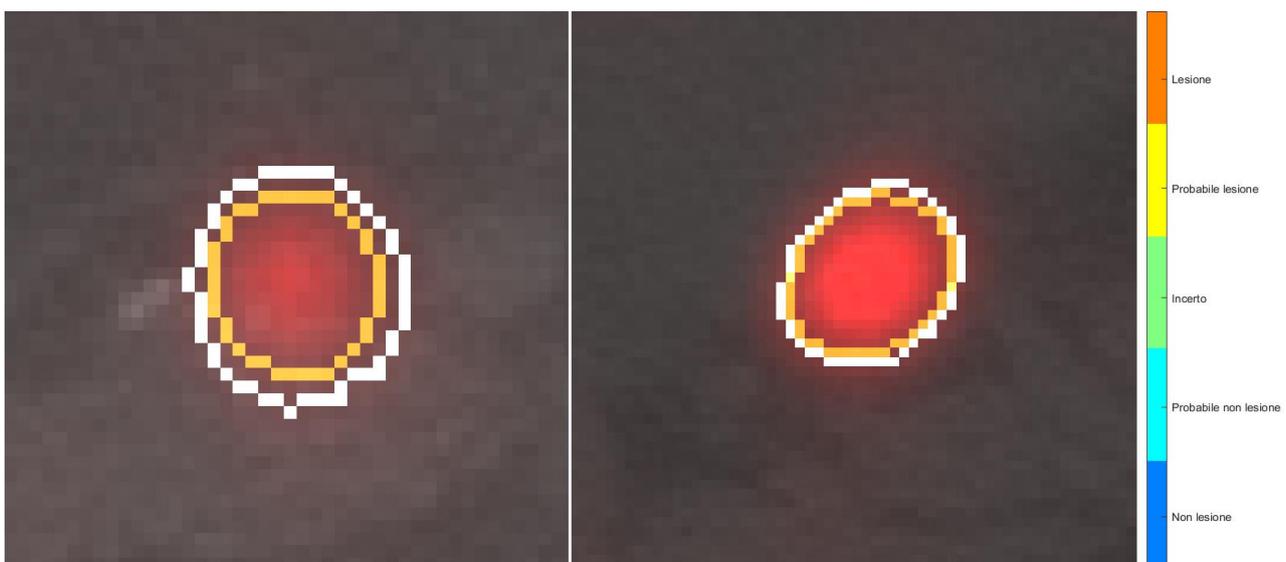


FIGURA 11 ESEMPI DI CLASSIFICAZIONE OTTENUTE TRAMITE CLASSIFICATORE MANOVA SU DUE LESIONI DEL TEST SET. IN BIANCO I CONTORNI DELLA SEGMENTAZIONE MANUALE, IN SCALA COLORE I MARGINI DELLA SEGMENTAZIONE AUTOMATICA CON IL GRADO DI CLASSIFICAZIONE INDICATO DAL COLORE

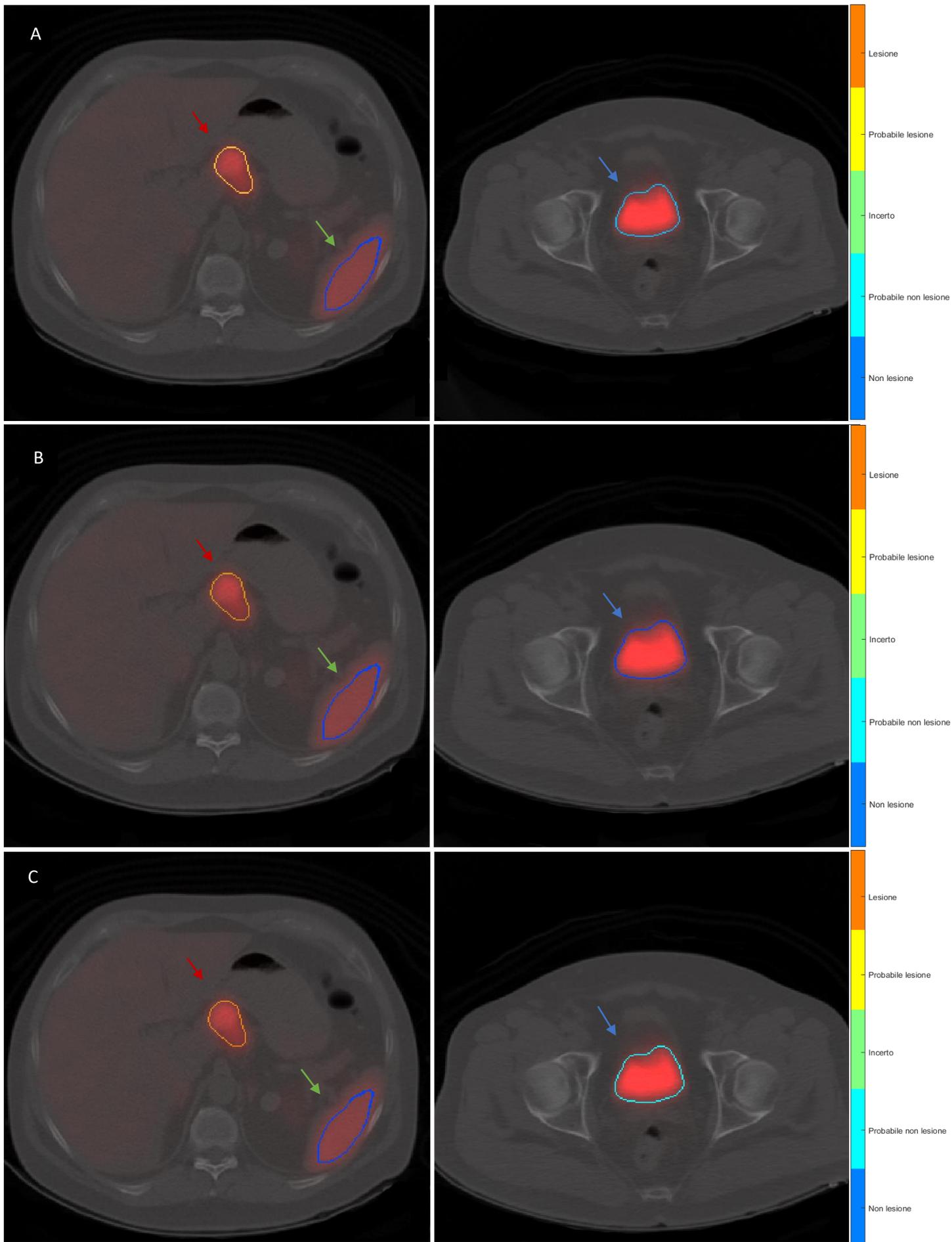


FIGURA 12 ESEMPIO DEI RISULTATI DI CLASSIFICAZIONE OTTENUTI CON I TRE CLASSIFICATORI: A) CLASSIFICATORE MANOVA B) SVM C) RETE NEURALE. PER OGNI CLASSIFICATORE SONO RIPORTATE LE STESSUE DUE FETTE DI UN PAZIENTE CON LESIONE PANCREATICA PRIMARIA: LA PRIMA FETTA CONTIENE LA LESIONE (FRECCIA ROSSA) ED UNO DEI RENI (FRECCIA VERDE), LA SECONDA CONTIENE LA VESICCA (FRECCIA BLU).

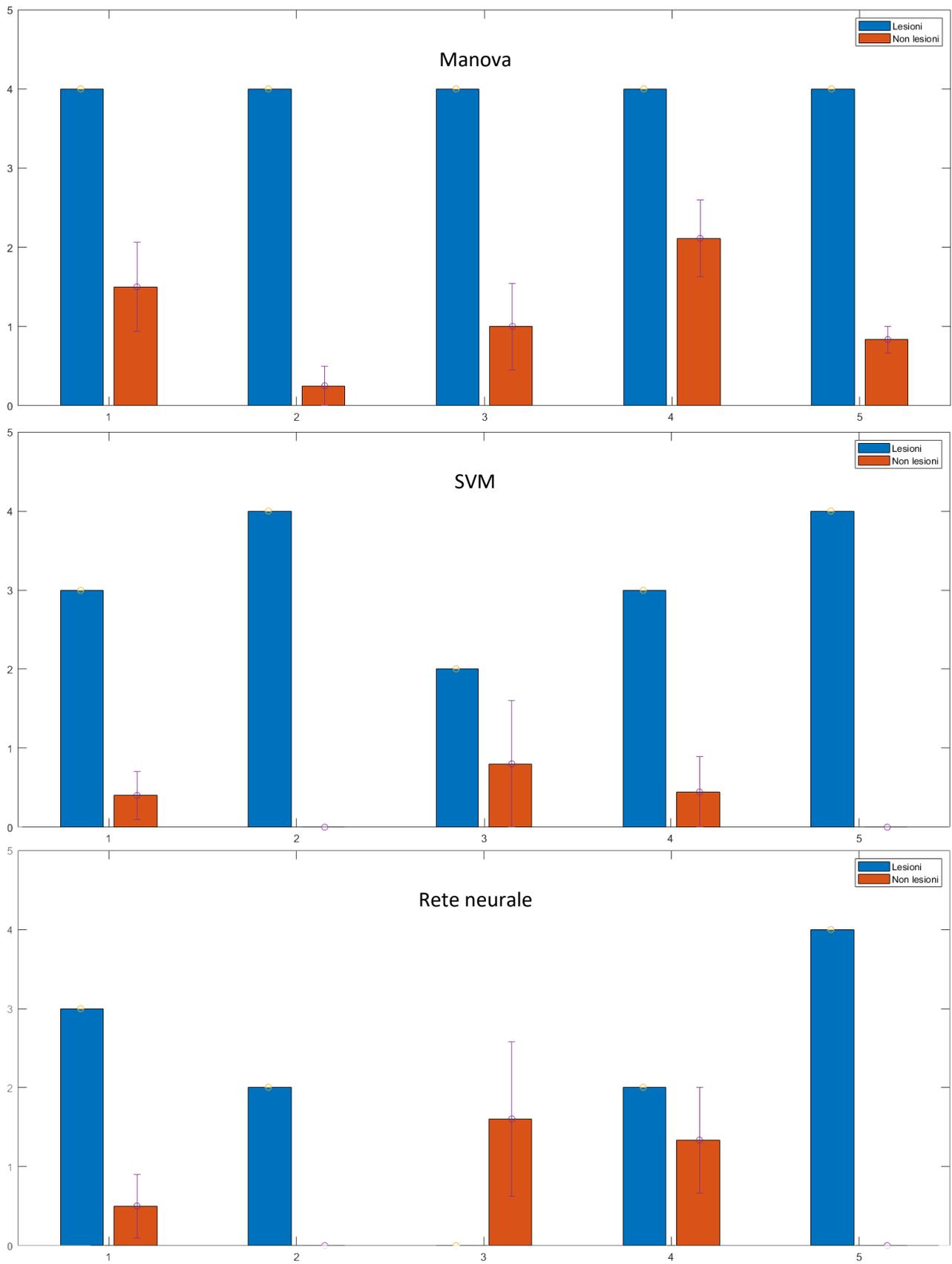


FIGURA 13 VALORI MEDI DEI GRADI ASSEGNATI PER LE DUE CLASSI PER I PAZIENTI DEL TEST SET UTILIZZANDO I TRE DIVERSI CLASSIFICATORI. SULLE ASCISSE SONO RIPORTATI I CINQUE PAZIENTI, SULLE ORDINATE I VALORI MEDI DEI GRADI DELLE DUE CLASSI E RELATIVE DEVIAZIONI STANDARD.

DISCUSSIONE

In questo lavoro è stato sviluppato un modello per la rilevazione automatica di tumori neuroendocrini all'interno di volumi ^{68}Ga -DOTATOC PET/CT. La rilevazione dei tumori neuroendocrini prevede allo stato dell'arte l'intervento più o meno esteso di un operatore esperto per una o più fasi del processo. Il medico sfoglia il volume del paziente fetta per fetta, osservando le due modalità separatamente o fuse, e tramite un software di supporto (es. LifeX) delinea i bordi delle porzioni di tessuto sospette. Tale approccio è impegnativo poiché richiede tempo, può essere complicato dal metodo con cui vengono visualizzati i volumi e dai parametri degli schermi utilizzati ed è inoltre risultato dell'esperienza dell'operatore e dei criteri che adotta nella segmentazione, pertanto soggetto ad elevata variabilità intra e inter operatore. L'automatizzazione è in parte raggiunta attraverso l'adozione di algoritmi di thresholding che consentono, tramite il confronto tra i valori di intensità dei voxel del volume PET e un valore soglia scelto con metodiche diverse a seconda del tipo di thresholding utilizzato, di delineare i margini delle porzioni di tessuto che mostrano una radioattività elevata. Le tecniche di thresholding sono considerate come una valida alternativa per via della facile implementazione ma la decisione di una soglia da applicare all'imaging PET è complessa e può necessitare di opportune calibrazioni dei parametri degli scanner. Il thresholding a soglia fissa è difficilmente generalizzabile a tutte le tipologie di neoplasie neuroendocrine, data l'eterogeneità che le caratterizza, mentre il thresholding con soglia adattiva necessita la calibrazione dei parametri tramite simulazioni con fantocci, presupponendo la conoscenza delle caratteristiche delle lesioni da simulare.

Il modello proposto prevede tra blocchi principali: il primo è il caricamento del volume PET e la delimitazione dei contorni delle porzioni di tessuto con valore di assorbimento al di sopra della soglia automatica calcolata sui valori SUV dell'immagine, ottenuto tramite algoritmo supervoxel; il secondo prevede la registrazione del volume CT e l'estrazione delle caratteristiche dei tessuti dalle due differenti modalità di imaging tramite un'analisi di tessitura multimodale; il terzo blocco prevede la classificazione delle porzioni di tessuto in base alle caratteristiche estratte, questo preceduto da un'analisi delle componenti principali al fine di ridurre e ripulire il dataset da eventuali collinearità tra le variabili.

La qualità delle segmentazioni ottenute è stata valutata tramite il coefficiente di Sørensen-Dice come confronto dei volumi delle segmentazioni manuali e automatiche. I risultati ottenuti per i pazienti su cui è stato validato il metodo, riportati in Tabella 2, dimostrano che l'efficacia dell'algoritmo supervoxel (SLIC) nel definire i margini delle lesioni è confrontabile con quella ottenuta tramite segmentazioni manuali. Un primo risultato notevole è l'abbattimento della variabilità della segmentazione, che costituisce uno dei punti deboli delle segmentazioni manuali. La segmentazione ottenuta dall'algoritmo SLIC, in quanto completamente automatica, è indipendente dall'operatore e i risultati sono consistenti a parità di parametri impostati. Tali parametri sono due: il numero di supervoxel N , parametro proprio dell'algoritmo che determina quanti punti iniziali vengono utilizzati e conseguentemente quanti supervoxel si desidera ottenere, ed una soglia, calcolata automaticamente sulla distribuzione dei valori SUV del volume PET come percentile. Tale soglia non è in realtà un parametro dell'algoritmo SLIC ma viene applicata in fase di preprocessing per ripulire il volume PET dalle porzioni di tessuto con valori di assorbimento non rilevanti allo studio dell'immagine. La soglia è determinata osservando la distribuzione dei valori SUV. Data l'elevata prevalenza di voxel con valori nulli o estremamente bassi di intensità, i voxel con valori elevati di intensità appaiono schiacciati verso uno degli estremi. La soglia è allora calcolata come i -esimo percentile della distribuzione e tutti i voxel con intensità inferiore vengono assimilati a sfondo. Per determinare quale percentile risulti la più efficace per segmentare adeguatamente tutte le lesioni presenti nel test set si è osservando l'andamento del valore Dice medio al variare del percentile scelto.

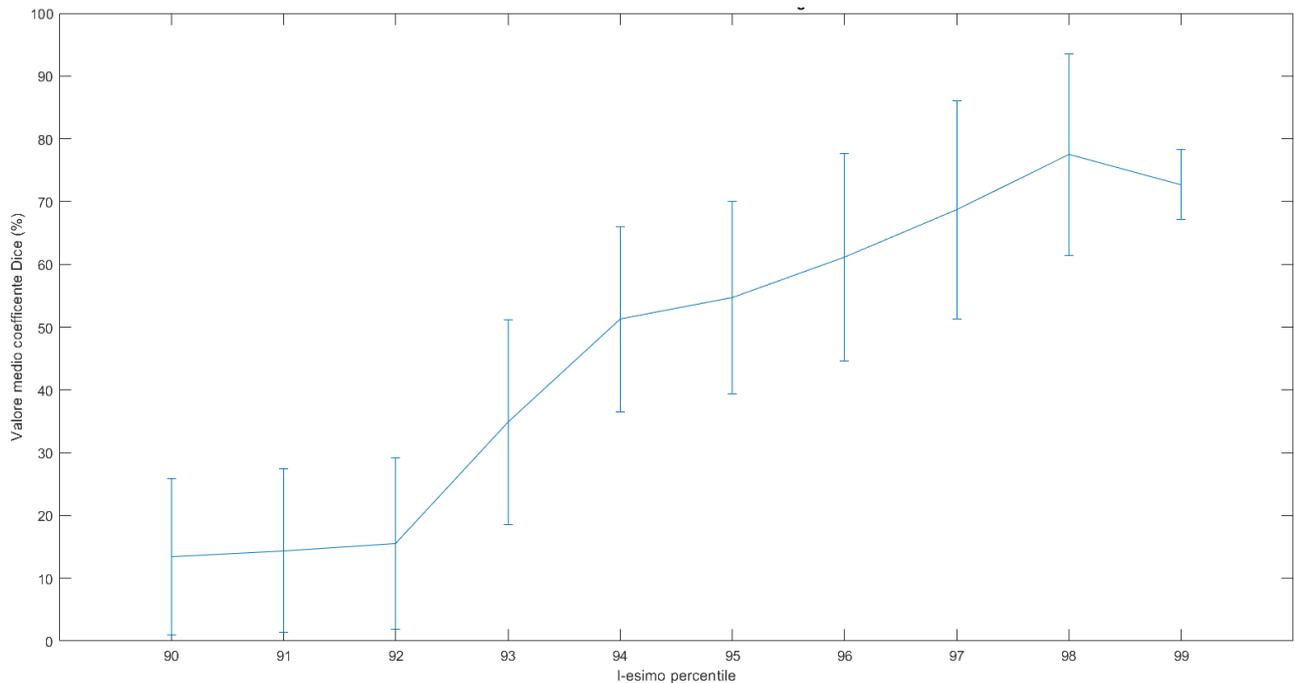


FIGURA 14 INFLUENZA DELLA SCELTA DELLA SOGLIA SUL VALORE MEDIO DEL COEFFICIENTE DICE NEI PAZIENTE DEL TEST SET

I migliori valori di coefficiente Dice sono stati osservati nelle segmentazioni precedute da preprocessing con soglia al 99-esimo percentile in quanto, sebbene mediamente le segmentazioni al 98-esimo percentili siano più simili a quelle manuali, nel caso di una lesione in particolare tale soglia portava a valori di Dice vicini allo 0.2, che si è ritenuto non accettabile. Inoltre, la variazione di Dice medio tra queste due soglie è circa del 5% a favore della soglia al 98. Tale variazione non è stata ritenuta significativa in quanto, sebbene il coefficiente Dice permetta di valutare la sovrapposizione volumetrica tra le maschere binarie e sia una delle metriche più utilizzate a tale scopo, bisogna tenere conto del volume ridotto dei tumori neuroendocrini che influenza i risultati ottenuti dalla metrica. Un discostamento dalle segmentazioni manuali non implica di per sé un fallimento dell'algorithm, all'interno di valori ragionevoli della metrica. La discordanza tra la segmentazione automatica e quella manuale, per quanto si è potuto osservare all'interno di questo lavoro, implica generalmente una segmentazione più aderente al contorno della lesione da parte dell'algorithm supervoxel, mentre le segmentazioni manuali tendono a mantenere una distanza maggiore dai contorni.

L'algorithm supervoxel ha mostrato buone performance con un valore medio della metrica Dice sui pazienti del test set $DSC = 81.48 \pm 9.92 \%$ e pertanto si ritiene che esso sia adeguato ad essere utilizzato per la segmentazione automatica di tumori neuroendocrini all'interno di volumi PET.

La registrazione del volume CT è stata ottenuta tramite trasformazione da un sistema di coordinate basate sull'immagine (ICS) ad un sistema di coordinate basate sul paziente (PCS) per entrambi il volume PET ed il volume CT. I parametri per eseguire questa trasformazione sono stati ottenuti dai file DICOM dei volumi.

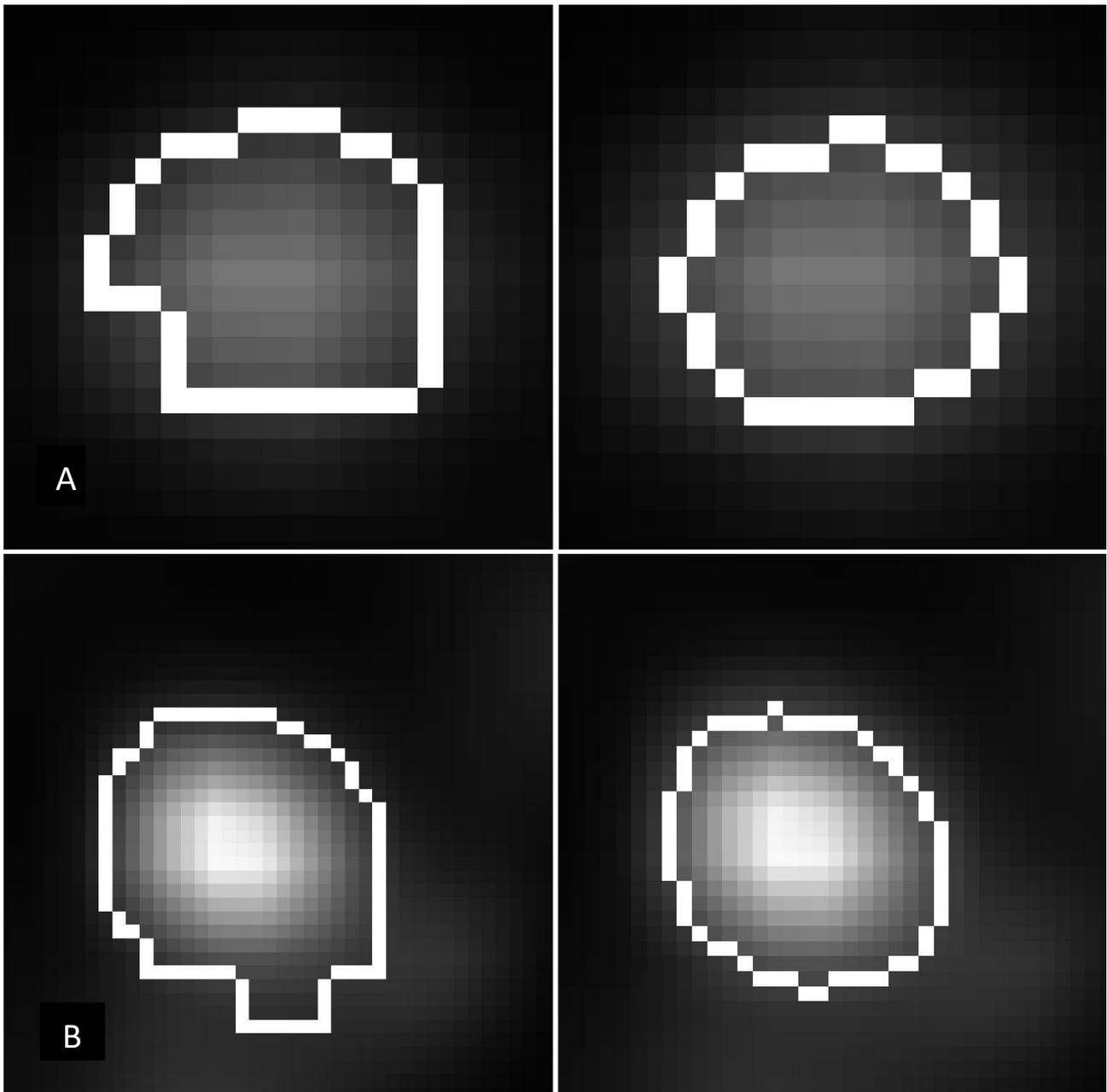


FIGURA 15 CONFRONTO DEI MARGINI OTTENUTI TRAMITE SEGMENTAZIONE MANUALE (SINISTRA) E ALGORITMO SUPERVOXEL (DESTRA) PER DUE DEI PAZIENTI DEL TEST SET

Registrati i volumi PET e CT è possibile procedere con l'analisi di tessitura multimodale sulle porzioni di tessuto individuate dalla segmentazione automatica. Lo scopo di tale analisi è quello di estrarre informazioni dalle caratteristiche di tessitura per entrambe le modalità di imaging. Data la complementarità delle informazioni fornite dai due metodi, l'idea è che entrambi possano contenere informazioni utili alla caratterizzazione delle lesioni. Questo viene confermato se si osserva la frequenza con cui ogni feature estratta viene utilizzata per la classificazione.

Per valutare tale frequenza si è contato quante volte ogni feature influenzasse la determinazione delle componenti principali utilizzate dal classificatore. A partire dal peso di ogni variabile sulla singola componente principale, osservabile dai valori contenuti nella matrice dei coefficienti restituita dall'analisi delle componenti principali, si è considerato come influente ogni variabile con peso maggiore o uguale all'80% del peso massimo per tale componente principale. I risultati in figura 14 mostrano l'influenza media delle variabili sui cinque pazienti. Si può osservare che, sebbene le maggiori frequenze appartengano a

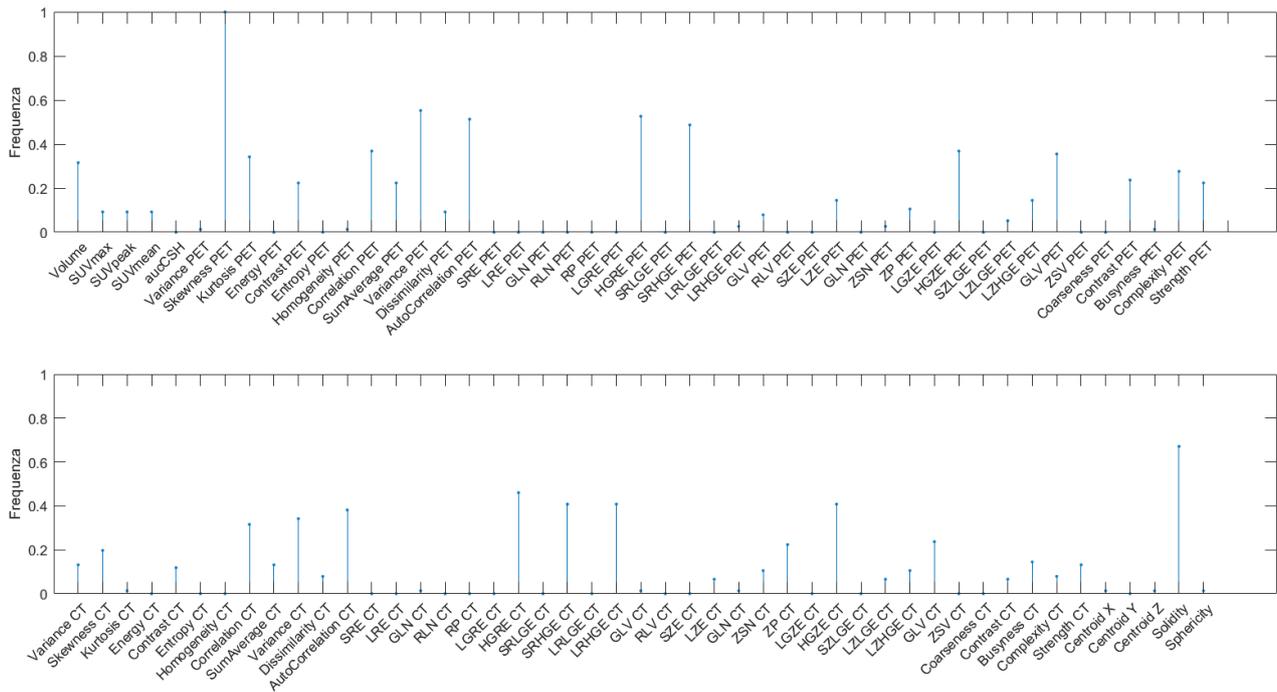


FIGURA 16 FREQUENZA DI UTILIZZO DELLE FEATURE ESTRATTE AI FINI DELLA CLASSIFICAZIONE

caratteristiche di tessitura calcolate sul volume PET, anche le texture estratte dal volume CT presentano dei picchi di influenza. In particolare modo si distinguono per frequenza la *Skewness* del volume PET, onnipresente nella creazione delle componenti principali, la *Varianza* e l'*Autocorrelazione* della PET, *High Gray-Level Run Emphasis* (HGRE) e *Short Run High Gray-Level Emphasis* (SRHGE) della PET e della CT, *Long Run Gray-Level Emphasis* (LRHGE) e *High Gray-Level Zone Emphasis* (HGZE) della CT, e la *Solidità*. Incidenza minore hanno caratteristiche generali (*Volume*), texture del volume PET (*Kurtosis*, *Correlazione*, *High Gray-Level Zone Emphasis*, *Gray-Level Variance* calcolata tramite *Gray-Level Size Zone Matrix*) e texture del volume CT (*Correlazione*, *Varianza*, *Autocorrelazione*). Il resto delle variabili ha rilevanze sensibilmente minori o più saltuarie nella classificazione.

Oltre alla *Solidità* e alla *Skewness* dei valori PET, entrambe presenti in almeno il 60% dei casi, il resto delle variabili con maggiore influenza è generalmente o una caratteristica globale, estratta su uno dei due volumi, o una caratteristica estratta tramite tre delle cinque famiglie di texture implementate: *Gray-Level Co-occurrence Matrix*, *Gray-Level Run-Length Matrix*, *Gray-Level Size Zone Matrix*. Le texture estratte tramite *Neighbourhood Gray-Tone Difference Matrix* presentano incidenza minore del 25% quando calcolate sul volume PET e minori del 15% quando valutate sul volume CT.

La PCA è un potente strumento per creare delle nuove variabili che siano la combinazione lineare delle variabili che compongono un insieme di dati. In questo lavoro essa viene applicata poiché il dataset è caratterizzato da un'elevata dimensionalità, con possibili collinearità tra le feature che lo costituiscono e l'analisi delle componenti principali permette di ridurre le dimensioni del dataset, conservando quanta più variabilità possibile, e ottenere delle nuove variabili scorrelate tra loro. Questo è importante dal momento che sono due fattori che influenzano la qualità della classificazione, poiché due gruppi di feature possono portare informazioni ridondanti o in contrasto tra loro. Inoltre, una delle assunzioni su cui basa la tecnica delle support vector machine, utilizzata per classificare gli oggetti rilevati, è che i dati in ingresso siano tra loro scorrelati. Il numero di componenti principali utilizzate per classificare i dati viene calcolato automaticamente con il criterio che almeno il 90% della variabilità totale venga preso in considerazione. Si è osservato all'interno di questo lavoro che tale criterio è generalmente soddisfatto dalle prime 8-9 componenti principali, con una conseguente riduzione delle dimensioni del dataset del circa 90%.

La classificazione è stata eseguita utilizzando tre differenti tipi di classificatori: il primo è un classificatore basato su un'analisi della varianza multivariata (MANOVA), il secondo su support vector machine (SVM), mentre il terzo modello è una rete neurale. I risultati sono stati valutati tramite il grado medio assegnato agli oggetti rilevati all'interno del volume. La classificazione peggiore è stata ottenuta tramite rete neurale, possibile causa di ciò è il numero limitato di esempi utilizzato in fase di addestramento della rete, fattore che insieme alla elevata eterogeneità delle lesioni neuroendocrine è stato osservato influire negativamente sulla qualità della classificazione degli algoritmi di learning[4][13]. Il classificatore basato sull'analisi della varianza multivariata mostra la maggiore sensibilità dei tre metodi: è sempre in grado di distinguere le lesioni dal resto degli oggetti e vi assegna sempre il grado massimo, denotando un'assegnazione certa della classe. Allo stesso tempo però risulta incerto su alcune porzioni di tessuto sano che vengono identificate come dubbie. Il grado assegnato a tali porzioni è comunque sempre inferiore a 2, considerate quindi incerte e non possibili lesioni. Questo fatto va in qualsiasi caso tenuto in considerazione in quanto è indice di una specificità minore rispetto al classificatore SVM. Il classificatore basato sulle support vector machine mostra infatti un grado medio per la classe delle non lesioni inferiore, sia se si osserva il singolo paziente che sul loro insieme. Questo indica una maggiore specificità. L'SVM pecca però di sensibilità, con un grado medio per le lesioni inferiore rispetto a quello ottenuto tramite classificatore Manova. Il classificatore SVM presenta performance diverse a seconda del paziente osservato. Questo potrebbe essere dovuto ad una maggiore specializzazione del modello SVM su un determinato insieme di dati e dunque una minore capacità di generalizzarsi e identificare lesioni con caratteristiche al di fuori di tale gruppo. Il classificatore MANOVA d'altro canto potrebbe risultare più generale ed essere quindi in grado di riconoscere lesioni con variazioni non incontrate durante l'addestramento. Questo spiegherebbe la minore specificità, dovuta a dei criteri meno stringenti per la classificazione che vanno ad includere porzioni non lesionali, oltre alla capacità di riconoscere ogni lesione incontrata.

Va poi fatta una considerazione sulla deviazione standard dei gradi medi assegnati dai tre classificatori. In tutti e tre i casi, la deviazione standard per la classe delle lesioni è 0, a prescindere dal grado medio assegnato. Questo indica che durante le 5 ripetizioni con training set differenti i criteri con cui vengono individuate le lesioni non variano, risultando sempre nello stesso grado assegnato a tali oggetti. La deviazione standard per la classe delle non lesioni è invece maggiore, anche se non elevata. Questo è probabilmente dovuto al fatto che all'interno del training set ciò che varia sono gli esempi di "non lesioni", selezionati dal più vasto dataset iniziale, mentre gli esempi di "lesioni" si conservano in tutte le iterazioni. Il ribilanciamento risulta comunque necessario, in quanto il dataset originale presenta una sproporzione di 10:1 a favore degli oggetti non lesionali. Questo influisce pesantemente sulla qualità dei classificatori SVM e Rete neurale, come si può osservare dal confronto dei gradi medi in figura 15.

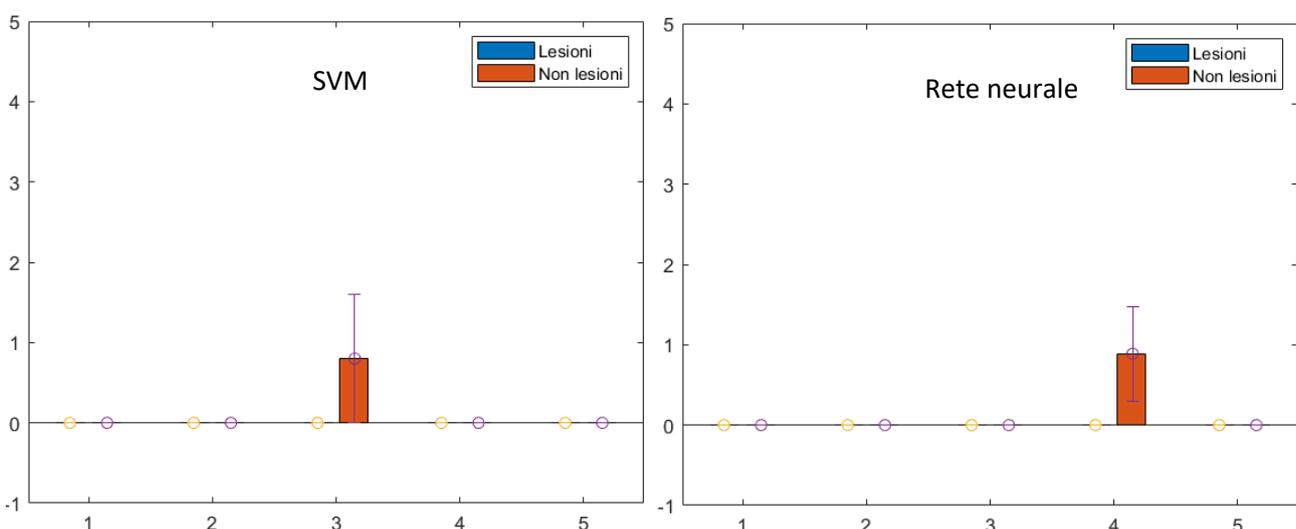


FIGURA 15 GRADO MEDIO ASSEGNATO AI PAZIENTI DEL TEST SET TRAMITE CLASSIFICATORI SVM E RETE NEURALE ADDESTRATI CON DATASET NON BILANCIATO

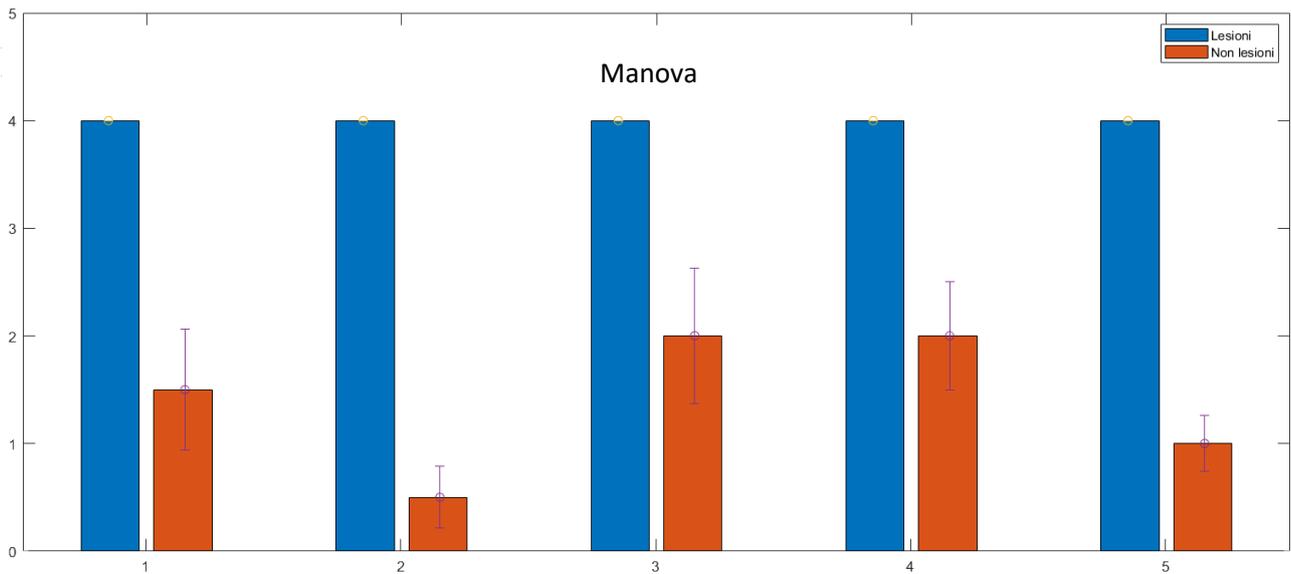


FIGURA 17 GRADO MEDIO ASSEGNATO AI PAZIENTI DEL TEST SET TRAMITE CLASSIFICATORE MANOVA ADDESTRATO CON DATASET NON BILANCIATO

L'effetto di un dataset non bilanciato sulla classificazione Manova è invece meno marcato e l'unica variazione è l'aumento del grado medio assegnato alle non lesioni. Tale incremento risulta comunque contenuto ed il grado delle non lesioni raramente supera il 2, non risultando quindi ancora classificate come lesioni. Questo dato rientra quindi nell'ipotesi che il classificatore Manova abbia criteri meno stringenti per la classificazione in grado di riconoscere lesioni con variazioni rispetto a quelle incontra durante l'addestramento.

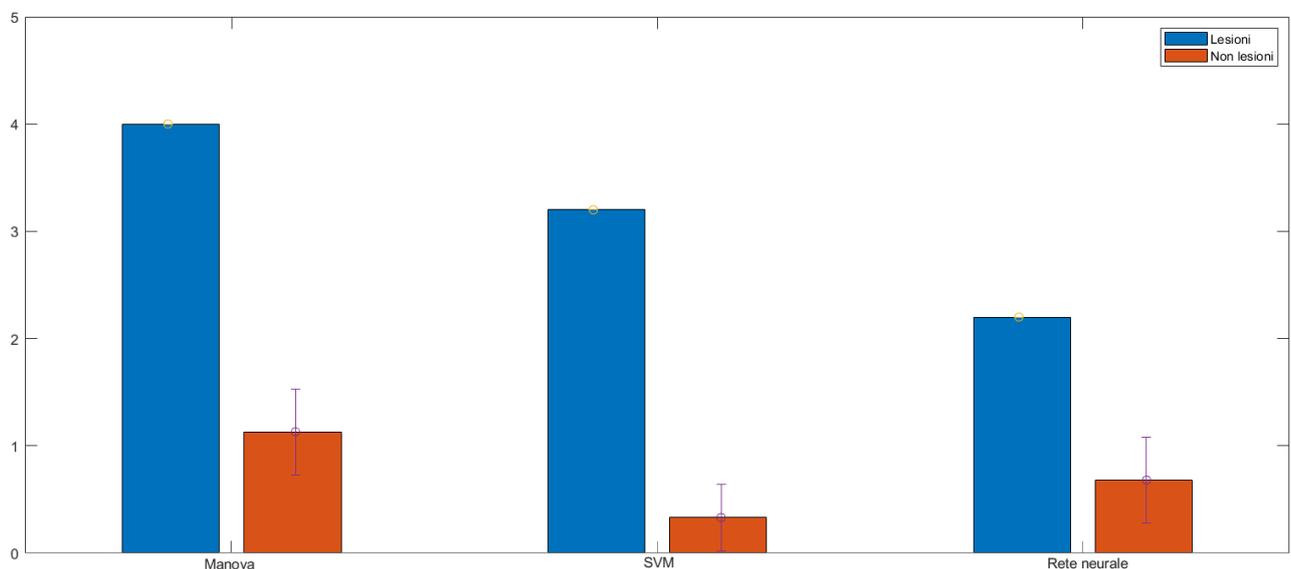


FIGURA 18 GRADI MEDI ASSEGNATI ALLE DUE CLASSI VALUTATI SUI CINQUE PAZIENTI DEL TEST SET

In generale il classificatore Manova mostra le performance migliori, con una maggiore sensibilità rispetto agli altri due classificatori e un grado medio per le non lesioni contenuto, anche quando confrontato con un dataset altamente sbilanciato. Le prestazioni dell'SVM risultano comunque soddisfacenti, a patto che venga addestrato su un insieme di dati bilanciato in cui le due classi siano rappresentate in maniera affine. La rete neurale mostra le prestazioni peggiori, probabilmente a causa del numero ridotto di esempi, e una

valutazione con un dataset più ampio sarebbe necessaria per stabilirne la praticabilità. I risultati inferiori ottenuti tramite SVM e rete neurale possono essere spiegati dalla maggiore propensione dei metodi di classificazione non lineari ad incorrere nell'over-fitting quando addestrate con un training set poco numeroso. Le reti neurali sono in particolar modo soggette a tale problematica, come viene evidenziato anche all'interno di questo lavoro. Nonostante la suddivisione del dataset in una porzione di training e una di test sia una tecnica generalmente adottata per prevenire il problema dell'over-fitting, essa può risultare poco efficace nel caso di dataset poco numerosi, per cui le reti neurali tendono a concentrarsi eccessivamente sugli esempi forniti durante l'addestramento perdendo in capacità di generalizzare. I metodi lineari sono al contrario meno soggetti a tale problematica e pertanto possono presentare performance migliori in tali situazioni. Per le considerazioni riportate la scelta finale ricade dunque sul classificatore Manova.

CONCLUSIONE

I tumori neuroendocrini sono delle neoplasie altamente eterogenee la cui valutazione diagnostica è complessa e in larga parte ancora dipendente dall'esperienza dell'oncologo. Sebbene siano stati adottati metodi automatici o semiautomatici per la delineazione delle lesioni, l'approccio ideale continua a essere la scansione visiva dei volumi PET/CT da parte del medico, procedura che richiede tempo ed è soggetta ad elevata variabilità inter e intra operatore. Il modello proposto in questo lavoro è un *proof-of-concept* per un software di supporto medico al fine di ridurre le tempistiche e la variabilità nella diagnosi dei tumori neuroendocrini. La segmentazione ottenuta tramite algoritmo supervoxel è coerente con le segmentazioni manuali ottenute da operatori esperti, come indicato dai valori del coefficiente di Sørensen-Dice, e non soggetta alla variabilità inter e intra operatore che limita il metodo manuale. La classificazione tramite classificatore basato sull'analisi della varianza multivariata (MANOVA), servendosi delle caratteristiche estratte dall'analisi di tessitura multimodale, è consistente nel distinguere le lesioni da altre porzioni di tessuti e organi con valori SUV elevati, automatizzando il processo decisionale del medico. Questo studio dimostra la fattibilità del metodo proposto che con una catena di algoritmi semplici e completamente automatici è in grado di individuare lesioni di natura neoplastica su tutti i tipi di tessuti generalmente interessati da tumori neuroendocrini. Data la natura ridotta del dataset su cui si è svolto lo studio è però azzardato dichiarare definitiva la funzionalità del metodo proposto e pertanto ulteriori studi su un numero maggiore di pazienti sarebbero auspicabili.

RIFERIMENTI

- [1] D. S. Klimstra, I. R. Modlin, D. Coppola, R. V. Lloyd, and S. Suster, "The pathologic classification of neuroendocrine tumors: A review of nomenclature, grading, and staging systems," *Pancreas*, vol. 39, no. 6, pp. 707–712, 2010, doi: 10.1097/MPA.0b013e3181ec124e.
- [2] J. Pollard, P. McNeely, and Y. Menda, "Nuclear Imaging of Neuroendocrine Tumors," *Surg. Oncol. Clin. N. Am.*, vol. 29, no. 2, pp. 209–221, 2020, doi: 10.1016/j.soc.2019.11.007.
- [3] A. Singh, J. J. Hines, and B. Friedman, "Multimodality Imaging of the Pancreatic Neuroendocrine Tumors," *Semin. Ultrasound, CT MRI*, vol. 40, no. 6, pp. 469–482, 2019, doi: 10.1053/j.sult.2019.04.005.
- [4] H. Zaidi and I. El Naqa, "PET-guided delineation of radiation therapy treatment volumes: A survey of image segmentation techniques," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 37, no. 11, pp. 2165–2187, 2010, doi: 10.1007/s00259-010-1423-3.
- [5] T. Baumann, C. Rottenburger, G. Nicolas, and D. Wild, "Gastroenteropancreatic neuroendocrine tumours (GEP-NET) - Imaging and staging," *Best Pract. Res. Clin. Endocrinol. Metab.*, vol. 30, no. 1, pp. 45–57, 2016, doi: 10.1016/j.beem.2016.01.003.
- [6] A. Sundin, "Novel Functional Imaging of Neuroendocrine Tumors," *Endocrinol. Metab. Clin. North Am.*, vol. 47, no. 3, pp. 505–523, 2018, doi: 10.1016/j.ecl.2018.04.003.
- [7] L. Evangelista, I. Ravelli, A. Bignotto, D. Cecchin, and P. Zucchetta, "Ga-68 DOTA-peptides and F-18 FDG PET/CT in patients with neuroendocrine tumor: A review," *Clin. Imaging*, vol. 67, no. April, pp. 113–116, 2020, doi: 10.1016/j.clinimag.2020.05.035.
- [8] S. H. Chen *et al.*, "68Ga-DOTATOC and 18F-FDG PET/CT for identifying the primary lesions of suspected and metastatic neuroendocrine tumors: A prospective study in Taiwan," *J. Formos. Med. Assoc.*, vol. 117, no. 6, pp. 480–487, 2018, doi: 10.1016/j.jfma.2017.07.007.
- [9] R. J. H. M. Steenbakkers *et al.*, "Reduction of observer variation using matched CT-PET for lung cancer delineation: A three-dimensional analysis," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 64, no. 2, pp. 435–448, 2006, doi: 10.1016/j.ijrobp.2005.06.034.
- [10] H. Li *et al.*, "A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours," *Med. Phys.*, vol. 35, no. 8, pp. 3711–3721, 2008, doi: 10.1118/1.2956713.
- [11] C. H. Kao *et al.*, "18F-FDG PET/CT-based gross tumor volume definition for radiotherapy in head and neck Cancer: A correlation study between suitable uptake value threshold and tumor parameters," *Radiat. Oncol.*, vol. 5, no. 1, pp. 1–8, 2010, doi: 10.1186/1748-717X-5-76.
- [12] E. C. Ford *et al.*, "Tumor delineation using PET in head and neck cancers: Threshold contouring and lesion volumes," *Med. Phys.*, vol. 33, no. 11, pp. 4280–4288, 2006, doi: 10.1118/1.2361076.
- [13] J. F. Daisne, M. Sibomana, A. Bol, T. Doumont, M. Lonnew, and V. Grégoire, "Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: Influence of reconstruction algorithms," *Radiother. Oncol.*, vol. 69, no. 3,

pp. 247–250, 2003, doi: 10.1016/S0167-8140(03)00270-6.

- [14] C. Y. Hsu, C. Y. Liu, and C. M. Chen, “Automatic segmentation of liver PET images,” *Comput. Med. Imaging Graph.*, vol. 32, no. 7, pp. 601–610, 2008, doi: 10.1016/j.compmedimag.2008.07.001.
- [15] M. Halt, C. C. Le Rest, A. Turzo, C. Roux, and D. Visvikis, “A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET,” *IEEE Trans. Med. Imaging*, vol. 28, no. 6, pp. 881–893, 2009, doi: 10.1109/TMI.2008.2012036.
- [16] I. Virgolini *et al.*, “Procedure guidelines for PET/CT tumour imaging with 68Ga-DOTA-conjugated peptides: 68Ga-DOTA-TOC, 68Ga-DOTA-NOC, 68Ga-DOTA-TATE,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 37, no. 10, pp. 2004–2010, 2010, doi: 10.1007/s00259-010-1512-3.
- [17] M. M. Dias, Vera Junn, Eunsung Mouradian, “PET/CT Standardized Uptake Values (SUVs) in Clinical Practice and Assessing Response to Therapy,” *Bone*, vol. 23, no. 1, pp. 1–7, 2008, doi: 10.1053/j.sult.2010.10.001.PET/CT.
- [18] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, 2012, doi: 10.1109/TPAMI.2012.120.
- [19] M. Tamajka and W. Beneřov, “Supervoxel algorithm for medical image processing,” *IEEE Int. Conf. Power, Control. Signals Instrum. Eng. ICPCSI 2017*, pp. 3121–3127, 2018, doi: 10.1109/ICPCSI.2017.8392300.
- [20] A. Lucchi, K. Smith, R. Achanta, G. Knott, and P. Fua, “Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features,” *IEEE Trans. Med. Imaging*, vol. 31, no. 2, pp. 474–486, 2012, doi: 10.1109/TMI.2011.2171705.
- [21] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, “SLIC Superpixels,” *EPFL Tech. Rep. 149300*, no. June, p. 15, 2010, doi: 10.1109/TPAMI.2012.120.
- [22] M. Vallires, C. R. Freeman, S. R. Skamene, and I. El Naqa, “Definition of texture features,” vol. V, pp. 1–9, 2015.
- [23] A. Ortiz, J. M. Grriz, J. Ramrez, and D. Salas-Gonzlez, “Improving MRI segmentation with probabilistic GHSOM and multiobjective optimization,” *Neurocomputing*, vol. 114, pp. 118–131, 2013, doi: 10.1016/j.neucom.2012.08.047.
- [24] C. Fernandez-lozano, J. A. Seoane, M. Gestal, T. R. Gaunt, and C. Campbell, “Texture Classification Using Kernel-Based Techniques,” pp. 427–434.
- [25] M. Asadur Rahman, M. Faisal Hossain, M. Hossain, and R. Ahmmed, “Employing PCA and t-statistical approach for feature extraction and classification of emotion from multichannel EEG signal,” *Egypt. Informatics J.*, vol. 21, no. 1, pp. 23–35, 2020, doi: 10.1016/j.eij.2019.10.002.
- [26] I. T. Jolliffe and J. Cadima, “Principal component analysis: A review and recent developments,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.
- [27] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. E. Mohamed, and H. Arshad, “State-of-the-art in artificial neural network applications: A survey,” *Heliyon*, vol. 4, no. 11, p. e00938, 2018, doi: 10.1016/j.heliyon.2018.e00938.

