



Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Biomedica

Orientamento e-Health

Tesi di Laurea Magistrale

**Algoritmi di intelligenza artificiale per la
predizione della risposta alla terapia nei tumori al
retto con immagini di risonanza magnetica**

Candidato:

Lorenzo Steno 253605

Relatori:

Prof. Rosati Samanta

Prof. Giannini Valentina

Anno accademico 2019-2020

ABSTRACT

Lo scopo di questo studio è l'elaborazione di un modello di classificazione ottenuto tramite algoritmi di machine learning ed immagini MRI forniti da tre diversi centri ospedalieri di Torino. In particolar modo i dati sono stati estratti da 90 pazienti ritenuti idonei allo studio con le relative immagini T2 e mappe ADC per la costruzione di matrici contenenti feature estratte dai parametri delle immagini. Le feature ricavate sono 42 per le immagini T2 e 45 per le immagini ADC.

Per la classificazione si è cercato di ottimizzare i parametri di un algoritmo genetico accostando sia un classificatore SVM sia un classificatore Bayesiano in delle prove distinte. Successivamente all'estrazione delle feature, alla loro normalizzazione e alla suddivisione in training e validation set sono stati posti dei parametri di controllo per l'algoritmo genetico mentre i coefficienti caratterizzanti del classificatore SVM sono stati gestiti come geni dello stesso. All'interno della costruzione del classificatore SVM e Bayesiano è stato applicato il metodo di cross validazione con k-fold di $K=5$. La feature delle soluzioni dell'algoritmo genetico è stata valutata secondo i parametri statistici derivanti dalla Confusion Matrix di paragone della classificazione ottenuta e quella reale. Si è scelta la soluzione con il minor valore di fitness come migliore di ogni ripetizione.

Tutte le soluzioni migliori dei classificatori SVM e Bayesiani sono state poi comparate con le classificazioni derivante dallo stesso modello applicato però sulle maschere automatiche. Per testare la robustezza del modello si è costruita una Confusion Matrix di paragone delle classificazioni automatiche e manuali. Si è considerata infine come modello migliore la prova con il maggior valore di accuratezza.

La conclusione dello studio ha portato ad un risultato promettente sia dal punto di vista di classificazione in vista della classe reale sia come comparazione tra classificazioni di diversi operatori. Feature di un grado superiore e un modello allenato sulle maschere automatiche potrebbero portare a risultati diversi ipoteticamente migliori.

INDICE

ABSTRACT.....	
1. INTRODUZIONE	1
1.1. II TUMORE AL RETTO E LA TERAPIA NEOADIUVANTE.....	1
1.2. DATI ED OBIETTIVO DELLA TESI	2
2. METODI.....	5
2.1. IMAGING E SEGMENTAZIONE.....	5
2.2. FEATURE EXTRACTION	7
2.3. PREPARAZIONE DEI DATI.....	9
2.3.1. NORMALIZZAZIONE.....	9
2.3.2. PARTIZIONE TRAINING E VALIDATION	10
2.3.3. CROSS VALIDATION (K-FOLD)	11
2.4. ALGORITMI GENETICI.....	12
2.4.1. CALCOLO DELLA FITNESS	14
2.5. CLASSIFICATORE SVM E NAIVE BAYES.....	15
2.6. SCELTA SOLUZIONE MIGLIORE DELL'ALGORITMO GENETICO	16
2.7. CONFRONTO MASCHERE MANUALI E AUTOMATICHE.....	18
3. RISULTATI	22
3.1. GA CON CLASSIFICATORI SVM.....	22
3.2. GA CON CLASSIFICATORI BAYESIANI.....	31
3.3. CONFRONTO MASCHERE AUTOMATICHE E MASCHERE MANUALI.....	35

4. CONCLUSIONI.....	41
BIBLIOGRAFIA	44

1. INTRODUZIONE

1.1. II TUMORE AL RETTO E LA TERAPIA NEOADIUVANTE

I carcinomi colon-retto (definiti anche come CRC) rappresentano in statistica un caso molto importante a livello europeo e mondiale. In Europa è la seconda tipologia di tumore più comune sia per uomini che per donne, con una maggiore incidenza per pazienti di sesso maschile piuttosto che femminile. In totalità nel 2018 sono stati calcolati all'incirca 500.000 carcinomi colon-rettali, corrispondenti al 12.8% sui casi totali, inoltre con una statistica del 12.2% ~ 12.6% è considerato la seconda causa di morte per pazienti sia di sesso femminile che maschile[1]. In Italia il CRC rappresenta l'11.6% dei casi presenti nel territorio stabilendosi come il secondo tumore più frequente nella popolazione, secondo al cancro alla mammella. L'incidenza dei carcinomi colon-retto è però in calo come indicato dai dati che coprono una finestra temporale dal 2008 al 2016; in particolar modo si ha una diminuzione del -4.1% annuo negli uomini tra 50 e 69 anni e -3.0% nelle donne della stessa fascia di età. Nonostante questo trend al ribasso e non stagnante il carcinoma colon-retto si presenta come la seconda causa di morte oncologica (10.6% - 10612 morti)[2]. In risposta a queste problematiche sono stati studiati diversi percorsi di trattamento e percorsi clinici per avere il miglior riscontro possibile, puntando ovviamente alla salute del paziente attraverso la rimozione del tumore e l'evitare eventuali ricorrenze della stessa patologia a distanza di anni. Il 30% dei tumori di questo tipo tendono a svilupparsi nel retto, ciò porta delle complicanze a livello chirurgico poiché lo scopo della rimozione è l'asportazione di ogni tipo di linfonodo. La terapia neoadiuvante è un processo utilizzato per minimizzare il rischio di ricorrenza del tumore ed include una varietà di chemioterapie e radioterapie somministrate singolarmente o in combinazione. La terapia neoadiuvante è un processo utilizzato per minimizzare il

rischio di ricorrenza del tumore ed include una varietà di chemioterapie e radioterapie somministrate singolarmente o in combinazione[3]. Questo tipo di trattamento tende a ridurre la tossicità, così come utilizzato per tumori rettali in stage II e III. Secondo le statistiche non tutti i pazienti raggiungono la così detta pCR (pathologic complete response – risposta patologica completa) ma solamente una percentuale che va dal 10% al 30% di chi si sottopone alla terapia[4]. Il grado di regressione del tumore (TRG) è stato quantificato in cinque gradi: TRG 1 (regressione completa) ha mostrato assenza di cancro residuo e fibrosi che si estende attraverso i diversi strati del tessuto del tumore; TRG 2 (risposta parziale) è stata caratterizzata dalla presenza di rare cellule tumorali residue sparse attraverso la fibrosi; TRG 3 era caratterizzato da un aumento del numero di cellule cancerose residue, ma la fibrosi predominava ancora; TRG 4 ha mostrato fibrosi residua di cancro eccessivo; e TRG 5 (completa non risposta) era caratterizzata dall'assenza di cambiamenti regressivi [5].

1.2. DATI ED OBIETTIVO DELLA TESI

Lo scopo di questo studio vuole essere l'ottimizzazione di un Algoritmo Genetico e di un classificatore ad esso associato per ottenere la migliore soluzione possibile alla classificazione di immagini di risonanza magnetica T2 e ADC. È particolarmente importante notare come i dati forniti riguardano tre centri piemontesi, ciò permette di poter identificare una soluzione non incentrata sullo studio di un solo ospedale ma spaziare su più dati convergenti in un solo dataset. All'interno dello studio sono stati analizzati una totalità di 90 pazienti e le loro relative classi. I pazienti di classe 0 non hanno risposto correttamente alla terapia mentre i pazienti di classe 1 hanno risposto in modo adeguato alla tecnica medica. I pazienti si presentano suddivisi nel seguente modo: 41 pazienti dal centro Candiolo con 16 pazienti di classe reale 1 e 25 pazienti di classe reale 0, 22 pazienti dal centro Mauriziano con 11 pazienti di classe reale 1 e 11 pazienti di classe reale 0, infine 27 pazienti dal centro Molinette con 11 pazienti di classe reale 1 e 16 pazienti di classe reale 0. I pazienti sono stati classificati come

responder (pR +), con TRG = 1 o 2 e non responder (pR-) con TRG \geq 3. Di ogni immagine di risonanza magnetica dei pazienti sono state scontornate in maschere le zone maligne attraverso un processo manuale da parte di un radiologo esperto e tramite anche una tecnica automatica derivante da un sistema deep learning U-Net di segmentazioni di cancro colonretto a partire da immagini MRI. Sono state così ottenute due serie di maschere applicate alle immagini ADC e T2 estraendo dunque rispettivamente 42 feature dalle immagini T2 e 45 feature dalle immagini ADC. In totale si ottengono due coppie di matrici, una per la segmentazione manuale ed una per la segmentazione automatica con le relative feature ADC e T2. A partire da questi dati sono stati affiancati dei classificatori all'algoritmo di feature selection, costituito da un algoritmo genetico, per ricercare le soluzioni migliori alla classificazione. Lo studio, infatti, si pone come obiettivo il classificare i pazienti come responsivi o meno alla terapia neoadiuvante. Questo potrebbe essere utile per evitare eventuali overtreatment sui pazienti oppure evitare eventuali disturbi derivanti dalla terapia che, in caso di risposta non positiva, potrebbero essere semplicemente dannosi per l'individuo. Può essere utile per il tentare di specializzare le terapie necessarie per la cura del paziente con maggiore precisione.

2. METODI

2.1. IMAGING E SEGMENTAZIONE

Gli esami RM sono stati eseguiti nei tre istituti seguenti utilizzando diversi scanner MR:

- Scanner da 1,5 T che utilizza una bobina di superficie phased-array a otto canali (HDx Signa Excite, GE HealthCare, Milwaukee, WI, USA), a Candiolo;
- Scanner da 1,5 T che utilizza una bobina body phased-array a 32 canali (Ingenia, Philips Medical Systems, Eindhoven, Paesi Bassi), a Mauriziano;
- Scanner da 1,5 T che utilizza una bobina body phased array a 32 canali (Achieva, versione 2.6, Philips Medical Systems, Eindhoven, Paesi Bassi), presso Molinette.

I pazienti sono stati posizionati in posizione supina e, salvo controindicazione, è stata somministrata un'iniezione da 20 mg di scopolamina butilbromuro per via endovenosa pochi minuti prima dell'inizio dell'esame o per via intramuscolare 20 minuti prima dell'inizio dell'esame, per ridurre al minimo gli artefatti da movimento indotti dalla peristalsi intestinale. Tutti gli esami includevano le sequenze di base per la segnalazione della risonanza magnetica pelvica per la stadiazione del cancro del retto, secondo le linee guida [6]. In particolare, il protocollo di tutti gli esami MRI comprendeva: 1) sequenza veloce spin-echo T2w acquisita sul piano assiale perpendicolare al diametro più lungo del tumore; 2) sequenza assiale EPI-SE pesata in diffusione (DWI). I parametri della sequenza T2w e DWI sono riportati, per ogni istituto,

rispettivamente nella Tabella 1 e nella Tabella 2. Inoltre, le mappe ADC sono state calcolate da DWI.

T2W sequence	Candiolo	Mauriziano	Molinette
Repetition/echo time (TR/TE)	7660/110 ms	3231/90 ms	5085 /100 ms
acquisition matrix	416 × 224	320 × 311	512 × 512
slice thickness	4 mm	3.5 mm	3 mm
pixel size	0.4297 × 0.4297 mm ²	0.47 × 0.47 mm ²	0.8 x 0.8 mm
field of view	220 mm x 220 mm	240 mm × 240 mm	250 mm x 250 mm
flip angle	90°	90°	90°

Tabella 1: Parametri di imaging T2

EPI-SE DWI	Candiolo	Mauriziano	Molinette
Repetition/echo time (TR/TE)	2000/87 ms	4011/91 ms	2694/68 ms
acquisition matrix	96 × 128	100 × 98	124 x 101
slice thickness	4 mm	3.5 mm	3 mm
pixel size	0.8594 × 0.8594 mm ²	1.88 × 1.88 mm ²	2.8 x 2.8 mm ²
field of view	220 mm x 220 mm	240 mm × 240 mm	345 mm x 288 mm
flip angle	90°	90°	90°
b-value max	800 s/mm ²	1000 s/mm ²	1000

Tabella 2: Parametri di imaging ADC

Gli esami MRI di ciascun paziente sono stati resi anonimi. La segmentazione manuale è stata progettata da un radiologo con più di due anni di esperienza nella valutazione dell'esame MRI. Le segmentazioni sono state supervisionate da radiologi con più di 10 anni di esperienza nella valutazione dell'imaging oncologico MRI. Il radiologo ha eseguito la segmentazione del tumore sulla sequenza T2w assiale, includendo tutte le sezioni al fine di ottenere un set di dati 3D. L'analisi è stata eseguita utilizzando l'applicativo di segmentazione open-source "3D Slicer" (v. 4.10.1, supportato da NA-MIC, NAC, BIRN, NCIGT e Slicer

Community). Le sequenze T2w assiali, coronali e sagittali e DWI (con mappa ADC) sono state visualizzate su un secondo monitor, per comprendere meglio l'estensione 3D dei tumori. La segmentazione è stata eseguita avendo cura di non includere il bordo esterno, escludendo così la possibile presenza di strie desmoplastiche, EMVI o linfonodi contigui alla lesione. Inoltre, sono state escluse anche sezioni di pareti ispessite dovute ad edema (non neoplastico). La maschera ottenuta è stata applicata sia su immagini T2w che DWI e le caratteristiche delle texture sono state estratte dai volumi segmentati.

2.2. FEATURE EXTRACTION

Diverse caratteristiche radiomiche sono state estratte dai voxel appartenenti alla maschera 3D segmentata in T2w e ADC: a) parametri del primo ordine, dove le caratteristiche sono statisticamente calcolate dai singoli pixel, tra cui dimensione, media, asimmetria, curtosi, deviazione standard, percentili; b) parametri di texture del secondo ordine, dove le misure considerano la relazione tra gruppi di pixel contigui, come quelli derivati dalla matrice di co-occorrenza del livello di grigio (GLCM) e dalla matrice di lunghezza di esecuzione del livello di grigio (GLRLM). L'elenco di tutte le caratteristiche estratte è riportato nella Tabella 3 [7].

Il GLCM è una tabulazione della frequenza con cui si verificano diverse combinazioni di valori di luminosità dei pixel (cioè livelli di grigio) tra voxel vicini in un'immagine. Al contrario, il GLRLM è l'insieme di pixel contigui aventi lo stesso livello di grigio: la lunghezza della corsa è il numero di livelli di grigio adiacenti in una particolare direzione, ed è calcolata contando il numero di volte che la corsa corrispondente si verifica nell'immagine [8].

Per estrarre i parametri della trama, abbiamo innanzitutto equalizzato l'istogramma riscaldando le intensità all'interno di ciascuna ROI tra il 1° e

il 99° percentile della ROI su 64 bin. L'utilizzo di 64 contenitori equamente suddivisi è stato un approccio comune per la quantificazione delle immagini nell'analisi radiomica e, allo stesso tempo, ha permesso di esplorare l'intera gamma di intensità del segnale del tumore [9]. Quindi, GLCM e GLRLM sono stati generati per ciascuna delle 13 direzioni di un'immagine 3D.

1° Ordine	2° Ordine	
	GLCM	GLRLM
ROI Volume (mm ³)	Joint Max	SRE
Skewness	Joint Average	LRE
Kurtosis	Joint Variance	LGRE
Intensity Kurtosis	Joint Entropy	HGRE
Intensity Variance	Diff Average	SRLGE
	Diff Variance	SRHGE
	Diff Entropy	LRLGE
	Sum Average	LRHGE
	Sum Variance	GLNU
	Sum Entropy	Normalised GLNU
	Angular Second Moment	RLNU
	Contrast	Normalised RLNU
	Dissimilarity	RP
	Inverse Difference	Grey Level Variance
	Normalised Inverse Difference	Run Length Variance
	Inverse Difference Moment	Run Entropy
	Normalised Inverse Difference Moment	
	Inverse Variance	
	Correlation	
	Autocorrelation	
	Cluster Tendency	
	Cluster Shade	
	Cluster Prominence	
	Information Measure Of Correlation1	
	Information Measure Of Correlation2	

Tabella 3: Feature estratte dalle immagini T2 e ADC

SRE: Short-Run Emphasis; LRE: Long-Run Emphasis; HGRE: High Gray-level Run Emphasis; LGRE: Low Gray-level Run Emphasis; SRLGE: Short-Run High Gray-level Emphasis; SRHGE: Short-Run High Gray-level Emphasis; LRLGE: Long-Run Low Gray-level Emphasis; LRHGE: Long-Run High Gray-level Emphasis; GLNU: Gray-level Non-Uniformity; RLNU: Run Length Non-Uniformity.

2.3. PREPARAZIONE DEI DATI

2.3.1. NORMALIZZAZIONE

Nella totalità di tutti i dati usati nello studio si ha un numero totale di pazienti uguale a 90 con 42 feature feature T2 e 45 feature ADC. Ancor prima del processo di Feature Selection, fondamentale per alleggerire il carico computazionale ed eliminare feature superflue, si è svolto un lavoro di normalizzazione per ri-scalare i dati secondo una sola tendenza. Questo tipo di processo è stato effettuato poiché un algoritmo di Machine Learning, avendo il compito di trovare un trend dei dati attraverso una comparazione di feature, può trovare evidenti difficoltà nell'utilizzare dei dati su diverse scale di grandezza. Secondo recenti studi le performance dei classificatori SVM, così come gli altri algoritmi di Machine Learning, possono essere migliorati attraverso un uso adeguato della normalizzazione dei dati[10]. Lo scopo della normalizzazione è quello di far visualizzare ogni punto dato con lo stesso fattore di equità, a livello di importanza, rispetto ad altri dati che originariamente sarebbero visualizzati con più enfasi a causa del loro ordine di grandezza. Il metodo che è stato utilizzato è il “*Min-Max Normalization*”, uno dei più comuni, attraverso la seguente formula.

$$\text{Valore Normalizzato} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Per ogni feature il valore minimo della colonna viene trasformato in 0 mentre il valore massimo in 1, su questa scala vengono quindi riassegnati tutti gli altri valori. Eseguendo questo processo abbiamo evitato che determinati valori di feature, se non feature stesse, potessero dominare sulla

classificazione rispetto ad altri valori con ordini di grandezza più piccoli.

2.3.2. PARTIZIONE TRAINING E VALIDATION

Uno dei fattori più importanti dei classificatori è non allenare gli algoritmi di machine learning sull'intero dataset. Lo scopo principale di questi algoritmi è quello di imparare, tramite dati, determinati trend ed eseguire specifiche azioni sui dati che vengono immessi. Se si vuole che l'algoritmo vada a generalizzare le soluzioni su qualsiasi altro tipo di dataset che si immette bisogna sacrificare parte dei propri dati di partenza costruendo due partizioni di "Training" e "Validation". Nel nostro caso si è deciso di suddividere i pazienti in due partizioni seguendo la loro appartenenza ad uno dei tre centri. In particolar modo si sono uniti i pazienti dei centri di Candiolo e Mauriziano da cui sono stati selezionati casualmente circa il 70% per comporre il Training Set. Il centro di Molinette è stato unito al 30% rimanente dei centri di Candiolo e Mauriziano uniti. Di seguito due tabelle che riassumono le due partizioni ottenute.

Numero pazienti in Training Set	
<i>Responder – Classe 0:</i>	25
<i>Non Responder – Classe 1:</i>	20
<i>Totale:</i>	45

Numero pazienti in Validation Set	
Responder – Classe 0:	24
Non Responder – Classe 1:	21
Totale:	45

2.3.3. CROSS VALIDATION (K-FOLD)

Il nostro metodo è di tipo “Supervised Learning” ciò vuol dire che è necessario effettuare uno split dei dati in Training e Test set, così da evitare che l’algoritmo vada ad allenarsi solo su sé stesso. Ciò vuol dire che il modello si va ad allenare secondo le feature e i dati del Training Set, queste soluzioni vengono poi testate nel test set così da poter evitare l’overfitting del modello. Il metodo per effettuare queste operazioni all’interno del nostro classificatore è la cross validation k-fold. La Cross validation è un metodo per la valutazione del modello, parte dei dati del training set sono infatti messi da parte durante la creazione del classificatore e, una volta ottenuto il modello, viene testato sulla parte dei dati restanti, così da simulare una compartizione di “nuovi dati” su cui far girare il modello. Il K-fold è uno dei metodi più semplici per implementare il modello “holdout” che è definito come il modo più naturale per effettuare la cross validazione. Vengono suddivisi i dati in k subset e viene ripetuto il metodo ‘holdout’ per k volte. Ogni volta, uno dei k sottoinsiemi viene utilizzato come set di test e gli altri k-1 vengono messi insieme per formare un insieme di addestramento. Quindi viene calcolato l’errore medio in tutte le k prove. Il vantaggio di questo metodo è che è meno importante il modo in cui i dati vengono suddivisi. Ogni dato deve trovarsi in un set di test esattamente una volta e in un set di addestramento k-1 volte. La varianza della stima risultante si riduce all’aumentare di k. Lo svantaggio di questo metodo è

che l'algoritmo di addestramento deve essere rieseguito da zero k volte, il che significa che sono necessarie k volte più calcoli per effettuare una valutazione. Il vantaggio di fare questo è che puoi scegliere indipendentemente la grandezza di ciascun set di test e il numero di prove su cui calcolare la media. All'interno del nostro studio il k scelto è 5.

2.4. ALGORITMI GENETICI

Gli algoritmi genetici sono dei sistemi di ricerca di tipo euristico che hanno lo scopo di emulare il normale processo di selezione naturale all'interno della realtà fisica. Sono dei metodi che vogliono introdurre nel concetto di feature selection il meccanismo di riproducibilità di soluzioni in più figli, la loro mutazione e la considerazione di una popolazione da cui vengono selezionati gli esemplari più forti per la riproduzione. Come specificato, questo tipo di algoritmi, vuole ricordare il meccanismo di selezione naturale partendo da una popolazione di partenza di soluzioni "genitori" che producono della prole, la generazione successiva, in cui saranno ereditate le caratteristiche dei genitori. Da questo semplice meccanismo si vuole fare in modo che le soluzioni pronte a riprodursi siano tra le migliori nella popolazione, a loro volta i figli saranno quindi ereditieri delle caratteristiche migliori dei genitori. Questo processo è reiterato più e più volte in delle iterazioni definite e che andranno sempre più a ricercare una soluzione migliore della precedente. Questo tipo di algoritmo però non consiste soltanto in una semplice riproduzione e selezione, va infatti specificato che esistono dei fattori randomici che aiutano le generazioni a differenziarsi dai genitori, così come avviene nella natura. Nell'uso degli Algoritmi Genetici nel Machine Learning si schematizza una soluzione come cromosoma, questo è composto da geni consistenti da un valore "0" oppure "1" che andranno ad indicare se quel gene è attivo all'interno del suo cromosoma. Nel nostro caso ogni gene è rappresentato da una feature. L'insieme dei cromosomi rappresenta invece una popolazione. Per stabilità la bontà di una soluzione si considera una funziona

di fitness che vuole quantificare in numero, detto fitness score, la capacità di una soluzione di competere con le altre poiché gli individui vengono selezionati attraverso questo fattore. È attraverso la selezione che si vanno a scegliere due genitori, in base al loro fitness score, e fatti riprodurre in soluzione figlie. Nella riproduzione però può intervenire un fattore randomico che può far avvenire il cross over tra le due soluzioni genitori; se l'evento avviene le soluzioni vengono spezzate in un punto di taglio, le metà di due soluzioni diverse vengono unite per creare le soluzioni figlie. Secondo un altro fattore probabilistico avviene l'evento della mutazione, ovvero il cambiamento del valore di alcuni geni all'interno delle soluzioni: ciò previene la prematura convergenza delle popolazioni. La convergenza è la condizione di stop dell'algoritmo, quando le soluzioni non tendono a migliorare dopo un determinato intervallo di ripetizioni l'algoritmo termina e si va a selezionare automaticamente la soluzione migliore dell'intero processo.

Feature:	1	2	3	-	-	-	87
Soluzione:	0	0	1	1

Esempio Soluzione di Algoritmo Genetico (0 – Feature disattivata; 1 – Feature attivata)

Nel nostro studio una soluzione dell'algoritmo è formata da tutte le feature estratte, sarà quindi rappresentata da una singola riga, ovvero una soluzione, con tante colonne tante quante sono le feature. Tramite l'algoritmo genetico vengono effettuate delle prove su queste soluzioni ottenendo la migliore per ogni ripetizione e attivando o disattivando determinate feature. L'intero processo è stato schematizzato come segue:



Come indicato in precedenza è necessaria l'inizializzazione dei parametri dell'algoritmo genetico. Il numero di individui della popolazione utilizzato è 400, il numero delle iterazioni massimo è 5000. Il numero di genitori è calcolato come l'80% del numero di individui. Sono state imposte 10 ripetizioni e il numero di geni è ovviamente uguale alle feature considerate nelle varie prove, particolarità aggiuntiva è nell'utilizzo del classificatore SVM che vede nella soluzione un'aggiunta di ulteriori 4 bit che vanno ad indicare i parametri del classificatore. La probabilità di Cross-Over è inserita con un valore di 0.9 mentre la probabilità di Cross-over è inizialmente 0.2 ma all'iterazione 200 viene automaticamente diminuita a 0.15 e all'iterazione 400 raggiunge definitivamente 0.1. Infine, è stata imposta una condizione di stop che vede fermare l'algoritmo genetico se in 250 iterazioni successive nella stessa ripetizione non v'è trovata nessuna nuova soluzione migliore.

2.4.1. CALCOLO DELLA FITNESS

La fitness di ogni soluzione del GA è stata calcolata come segue:

$$fitness = 1 - \left(\frac{|spec + sens|}{2} \right) + 0.3 * |spec - sens|$$

definendo *spec* e *sens* come specificità e sensibilità ottenuti dal classificatore SVM. La prima parte della formula tende a massimizzare l'accuratezza della classificazione mentre la seconda parte è una penalità introdotta per bilanciare le performance delle due classi. Secondo questo ragionamento vengono considerate soluzioni migliori quelle con un valore di fitness più basso rispetto alle altre.

2.5. CLASSIFICATORE SVM E NAIVE BAYES

Un classificatore Naive Bayes è costituito da un algoritmo che per trovare l'elemento di uscita (output) calcola la probabilità per ogni fattore (input), selezionando l'uscita che ha la probabilità più alta rispetto alle altre. Questo classificatore può essere adatto per lo studio in corso poiché la teoria che regge l'algoritmo assume che ogni feature deve essere indipendente dalle altre, ovvero che ogni feature contribuisca in modo unico all'interno della soluzione senza dar conto ad eventuali correlazioni tra feature. Un classificatore SVM vuole invece definire un iperpiano separatore tra le due classi così da attribuire ad ogni elemento la sua appartenenza ad un lato o l'altro dell'iperpiano. In una dimensione bidimensionale ogni paziente si trova in un piano bidimensionale, la loro distribuzione può essere alterata attraverso delle funzioni di Kernel che trasformano, non modificando le informazioni, la posizione di questi elementi così che la linea dell'iperpiano possa essere facilmente rappresentabile. Un altro parametro da tenere in considerazione è il parametro di regolarizzazione delle misclassificazioni, definito come BoxConstraint o penalty term all'interno dello studio. Questi due fattori sono stati caratterizzati nelle soluzioni dell'algoritmo genetico negli ultimi quattro bit. Nel caso del parametro BoxConstraint è stato variato durante le prove per evitare che l'algoritmo tenda immediatamente ad un C troppo alto, come si è evinto dalle prove.

BoxConstraint		Kernel	
Quartultimo Bit	Terzultimo Bit	Penultimo Bit	Ultimo Bit
0/1	0/1	0/1	0/1
0 - 0 $\Rightarrow C_1 = 1 \mid C_2 = 1$ 0 - 1 $\Rightarrow C_1 = 10 \mid C_2 = 5$ 1 - 0 $\Rightarrow C_1 = 50 \mid C_2 = 10$ 1 - 1 $\Rightarrow C_1 = 100 \mid C_2 = 25$		0 - 0 \Rightarrow Lineare 0 - 1 \Rightarrow Gaussiano 1 - 0 \Rightarrow Polynominal 2 1 - 1 \Rightarrow Polynominal 3	

Tabella 4: Caratteristiche del classificatore SVM usati

Il classificatore Bayes non ha avuto bisogno di particolari ritocchi nell'implementazione, non necessitando nessun tipo di parametro da preimpostare è stato scelto di utilizzare il modello multiclasse offerto dalla funzione Matlab

‘fitcnb’. Il metodo di cross validation, come è stato precedentemente spiegato, è stato utilizzato all’interno dei classificatori per testare le loro performance. Parte dei dati, infatti, sono stati considerati di training ed utilizzati per la creazione del modello, la restante parte è definita di test su cui sono state valutate le performance del classificatore con quella determinata soluzione.

2.6. SCELTA SOLUZIONE MIGLIORE DELL’ALGORITMO GENETICO

L’algoritmo genetico a fine di ogni ripetizione salva la soluzione migliore in una matrice che rappresenta nelle righe il numero di soluzioni uguale alle ripetizioni e come colonne la totalità delle feature ottenute. Affiancata alla soluzione di ogni ripetizione un valore di fitness viene aggiunto in un vettore che ha come righe il numero di soluzioni uguale alle ripetizioni dell’algoritmo. La soluzione migliore per ogni ripetizione è sì scelta dalla fitness migliore, ovvero la più bassa tra le soluzioni della prova, ma sono stati considerati dei fattori importanti per la valutazione. La fitness stessa viene calcolata alla successiva creazione di una Confusion Matrix del tipo:

		Classe Reale	
		0	1
Classe Predetta	0	TP	TN
	1	FN	TN

Esempio Confusion Matrix

Una matrice di confusione viene utilizzata nei modelli di classificazione in cui la risposta è categoriale, ovvero per classi, come nel nostro caso (classe 0 e classe 1).

Il modello prevede dunque un valore di risposta per ogni osservazione prevista andandolo a confrontare con il valore di risposta effettivo della realtà.

La matrice va quindi a valutare l'accuratezza del nostro modello fornendoci inoltre l'accesso a diverse metriche che possono aiutarci nella valutazione delle soluzioni e il confronto tra prove diverse. Di seguito l'elenco delle metriche usate nello studio.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

• **Accuratezza (ACC)** – Questo termine può essere uno dei più generici e va ad indicare quanto i miei valori di predizione siano nella totalità corretti. La sua valutazione può dipendere da molti fattori poiché l'accuratezza assume che gli errori positivi e negativi abbiano lo stesso peso nella valutazione, il che potrebbe non essere performante per il nostro caso.

$$TPR = \frac{TP}{TP + FN}$$

• **Sensitività o Recall (TPR)** – Questo termine indica quante volte il modello sceglie una classe positiva quando l'osservazione reale è di fatto positiva.

$$PPV = \frac{TP}{TP + FP}$$

• **Precision (PPV)** – Questo termine è utile per valutare quante volte il modello è corretto quando la predizione è una classe positiva.

$$NPV = \frac{TN}{TN + FN}$$

• **Negative Predictive Value (NPV)** - Questo termine è utile per valutare quante volte il modello è corretto quando la predizione è una classe negativa.

Affiancando questi valori alla fitness, comunque da loro dipendente, si è valutato i singoli modelli delle prove tentando di prediligere l'accuratezza per una classificazione migliore ma senza distogliere l'attenzione dagli altri parametri che possono contestualizzare il valore di accuratezza così da renderlo effettivamente un ottimo parametro da considerare per la scelta del modello migliore.

2.7. CONFRONTO MASCHERE MANUALI E AUTOMATICHE

Per lo studio sono state fornite delle maschere automatiche ottenute tramite un algoritmo di deep learning U-NET, questa opportunità permette di poter valutare le soluzioni delle prove testandone l'efficacia intra-operatore. Uno dei limiti degli algoritmi di machine learning è infatti la possibilità di poter variare i dati in ingresso provenienti inoltre da ulteriori operatori che hanno segmentato la stessa lesione in modo diverso. Attraverso il confronto con le maschere automatiche è possibile stabilire la robustezza del modello trovato e testarlo in un ambito inter-operatore. Per ogni prova effettuata è stata utilizzata la soluzione ottenuta con le feature estratte dalle maschere automatiche ottenendo di fatto una seconda classificazione comparabile con la prima proveniente dalla classificazione manuale. Di queste due classificazioni si son volute valutare le Confusion Matrix costruite sulla differenza di classificazione delle maschere automatiche e manuali ed inoltre sul loro confronto con la classificazione reale. Sono ottenute tre Confusion Matrix per prova. Si è cercato di approfondire i legami intrinseci delle maschere automatiche e quelle manuali ottenendo dei valori statistici sulle loro differenze e similarità.

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

$$DSC = \frac{2TP}{2TP + FP + FN}$$

Per dati in forma booleana

• **Dice Similarity Coefficient (DSC)** – È un indice di sovrapposizione spaziale, una metrica che può andare da 0 ad 1 dove il minimo rappresenta una sovrapposizione nulla tra le maschere, al contrario rappresenta una sovrapposizione combaciante.

A questi termini sono stati accostati i valori di Precision e Recall riguardanti le performance geometriche delle maschere. La precision è una valutazione sui falsi positivi, se questa è alta ne esistono sempre di meno indicando una segmentazione migliore. La recall invece indica la bontà della segmentazione sull'estensione del tumore, un'alta recall indica una buona percentuale di tumore nella propria maschera ma bisogna attenzionare questo valore poiché anche considerando l'intera immagine MRI si avrebbe una recall alta poiché si prenderebbe il 100% del tumore. Si è considerata la differenza percentuale tra le due maschere, inoltre è stata calcolata la differenza di ogni valore di feature delle maschere automatiche con il corrispondente del manuale ottenendo di fatto un valore percentuale di cui, per valutare la feature, è stata fatta la mediana. Di questo valore sono stati scelti dei threshold mobili da settare per effettuare nuove prove con i modelli migliori. Infine è stato considerato il valore di ICC.

$$ICC = \frac{RMS - EMS}{(RMS + (CMS + EMS)) * \frac{NumJudges}{NumTargets + (NumJudges - 1) * EMS}}$$

$$RMS = \frac{BSS}{NumTargets - 1}$$

$$BSS = \sum (meanTarget - meanTotal)^2 - NumJudges$$

$$ESS = WSS - RSS$$

$$EMS = \frac{ESS}{(numJudges - 1) * (numTargets - 1)}$$

$$RSS = \sum (meanJudges - meanTotal)^2 * NumTargets$$

$$CMS = \frac{RSS}{NumJudges - 1}$$

• **Intraclass correlation coefficient (ICC)** – È un indice che tende a voler rendere l'affidabilità sia del grado di correlazione che l'accordo tra le misurazioni. Di questo valore è stato scelto un threshold di 0.5 per considerare buone le feature con un valore maggiore del threshold. Il valore calcolato è un ICC di tipo Single Random Raters [11][12]

3. RISULTATI

Tutte le prove effettuate e i calcoli a contorno sono stati compilati con il software The Math Works, Inc. MATLAB. Versione R2020a e Microsoft Excel (2010) del pacchetto Microsoft 365 Apps for enterprise.

3.1. GA CON CLASSIFICATORI SVM

Una prima prova sul data set delle maschere manuali è stata fatta tenendo i dati separati secondo i loro centri e tenendo altresì separati i dataset delle immagini T2 e ADC. In questa prima prova sono stati ottenuti quindi un totale di tre modelli per le immagini T2 e tre modelli per le immagini ADC allenati, ognuno di questi, in uno dei tre centri. Il modello è rappresentato dalla soluzione dell’algoritmo con la feature minore e, quando queste uguale ad altre soluzioni, sono state scelte quelle con il minor numero di feature selezionate per agevolare il carico computazionale ed evitare il coinvolgimento di feature superflue. Questa prima prova è stata effettuata con dei valori di C per il classificatore SVM uguali a 1, 10, 50 e 100 rappresentati dal quart’ultimo e terz’ultimo bit della soluzione del GA.

<i>Feature T2</i>	<i>Candiolo</i>	<i>Mauriziano</i>	<i>Molinette</i>
<i>Skewness STAT</i>	0	0	1
<i>Kurtosis STAT</i>	1	1	0
<i>Intensity Kurtosis STAT</i>	0	0	1
<i>IntensityVariance STAT</i>	0	0	0
<i>Joint Max GLCM</i>	0	0	0
<i>Joint Average GLCM</i>	0	1	1
<i>JointVariance GLCM</i>	0	0	1
<i>Joint Entropy GLCM</i>	1	0	1
<i>Diff Average GLCM</i>	1	1	0
<i>Diff Variance GLCM</i>	0	0	1
<i>Diff Entropy GLCM</i>	0	1	0
<i>Sum Average GLCM</i>	0	0	0
<i>Sum Variance GLCM</i>	0	1	0

Sum Entropy GLCM	1	1	0
Angular Second Moment GLCM	0	0	1
Contrast GLCM	0	1	1
Dissimilarity GLCM	1	1	1
Inverse Difference GLCM	1	0	1
Normalised Inverse Difference GLCM	1	0	0
Inverse Difference Moment GLCM	0	1	0
Normalised Inverse Difference Moment GLCM	1	1	1
Inverse Variance GLCM	0	0	0
Correlation GLCM	1	0	1
Autocorrelation GLCM	0	1	1
Cluster Tendency GLCM	1	0	0
Cluster shade GLCM	0	0	0
Cluster Prominence GLCM	0	1	0
Information Measure Of Correlation1 GLCM	0	0	1
Information Measure Of Correlation2 GLCM	0	0	0
SRE GLRLM	0	1	1
LRE GLRLM	1	0	0
LGRE GLRLM	1	0	0
HGRE GLRLM	0	0	0
SRLGE GLRLM	0	0	0
SRHGE GLRLM	0	0	0
LRLGE GLRLM	0	0	1
LRHGE GLRLM	0	0	1
GLNU GLRLM	1	1	0
GLNU Normalised GLRLM	0	0	1
RLNU GLRLM	1	1	1
RLNU Normalised GLRLM	1	1	0
RP GLRLM	0	0	0
Totale Feature attivate	15	16	19
Fitness	0.243	0.236	0.136

Tabella 5: Risultati del GA con le tre migliori soluzioni monocentriche per le immagini T2.

Le celle grigie sono feature selezionate, le celle bianche sono feature non attivate

Feature ADC	Candiolo	Mauriziano	Molinette
Skewness STAT	1	1	0
Kurtosis STAT	1	0	0
Intensity Kurtosis STAT	1	0	1
IntensityVariance STAT	0	0	1
Joint Max GLCM	0	1	0
Joint Average GLCM	0	0	0
JointVariance GLCM	1	0	0
Joint Entropy GLCM	1	0	0

<i>Diff Average GLCM</i>	0	1	0
<i>Diff Variance GLCM</i>	0	0	0
<i>Diff Entropy GLCM</i>	0	0	0
<i>Sum Average GLCM</i>	0	1	0
<i>Sum Variance GLCM</i>	0	0	1
<i>Sum Entropy GLCM</i>	1	0	0
<i>Angular Second Moment GLCM</i>	0	0	0
<i>Contrast GLCM</i>	0	0	0
<i>Dissimilarity GLCM</i>	1	0	0
<i>Inverse Difference GLCM</i>	1	0	1
<i>Normalised Inverse Difference GLCM</i>	1	0	0
<i>Inverse Difference Moment GLCM</i>	1	0	0
<i>Normalised Inverse Difference Moment GLCM</i>	0	1	0
<i>Inverse Variance GLCM</i>	0	0	0
<i>Correlation GLCM</i>	0	1	1
<i>Autocorrelation GLCM</i>	0	0	0
<i>Cluster Tendency GLCM</i>	0	1	0
<i>Cluster shade GLCM</i>	1	0	1
<i>Cluster Prominence GLCM</i>	0	0	1
<i>Information Measure Of Correlation1 GLCM</i>	0	0	0
<i>Information Measure Of Correlation2 GLCM</i>	1	0	1
<i>SRE GLRLM</i>	1	0	0
<i>LRE GLRLM</i>	0	1	0
<i>LGRE GLRLM</i>	0	0	0
<i>HGRE GLRLM</i>	0	1	0
<i>SRLGE GLRLM</i>	0	0	1
<i>SRHGE GLRLM</i>	0	1	0
<i>LRLGE GLRLM</i>	0	0	0
<i>LRHGE GLRLM</i>	0	0	0
<i>GLNU GLRLM</i>	1	0	0
<i>GLNU Normalised GLRLM</i>	1	0	1
<i>RLNU GLRLM</i>	0	1	1
<i>RLNU Normalised GLRLM</i>	0	0	1
<i>RP GLRLM</i>	1	0	0
<i>Grey Level Variance GLRLM</i>	0	0	0
<i>Run Length Variance GLRLM</i>	0	0	1
<i>Run Entropy GLRLM</i>	1	1	0
<i>Totale Feature attivate</i>	17	12	13
<i>Fitness</i>	0.209	0.179	0.225

Tabella 6: Risultati del GA con le tre migliori soluzioni monocentriche per le immagini ADC. Le celle grigie sono feature selezionate, le celle bianche sono feature non attivate.

Queste sei prove presentano alcune particolarità che sono state successivamente testate per uno studio delle feature. Il database proveniente dal centro Mauriziano

delle immagini ADC presenta infatti soluzioni con lo stesso valore di fitness da cui è stata prelevata la soluzione migliore definendola come quella con il minor numero di feature selezionate, seguendo le specifiche precedentemente dettate nei metodi di valutazione delle soluzioni. Tra tutte queste soluzioni però si è notato come due feature in particolare sono presenti in tutte e dieci le soluzioni dell'algoritmo genetico, quindi è stata effettuata una prova testando il modello con solo e soltanto le due feature selezionate.

Feature utilizzate	Kernel	C	Acc	Sens	Spec	PPV	NPV	Fitness
<i>Joint Variance GLCM</i>	Lineare	50	0.696	0.818	0.583	0.643	0.778	0.370
<i>Inverse Difference Moment GLCM</i>	Lineare	100	0.652	0.818	0.500	0.600	0.750	0.436

Tabella 7: Soluzioni ottenute sul dataset del centro Mauriziano con immagini ADC sono le sole due feature selezionate in ogni soluzione dei primi dieci modelli proposti. Le due soluzioni variano nella selezione del Kernel e di C.

Tutte le soluzioni migliori, divise per il tipo di immagine, sono state poi testate con i restanti dataset provenienti dagli altri due centri così da poter avere una prima idea del funzionamento della soluzione su dati che non sono stati utilizzati per il training e per il testing. È stata quindi effettuata una validazione di cui riportiamo i risultati.

Soluzione	Applicato a	Acc	Sens	Spec	PPV	NPV	Fitness
Candiolo	Molinette	0.679	0.733	0.615	0.688	0.667	0.361
Candiolo	Mauriziano	0.565	0.545	0.583	0.545	0.583	0.447
Mauriziano	Molinette	0.643	0.667	0.615	0.667	0.615	0.374
Mauriziano	Candiolo	0.727	0.563	0.821	0.643	0.767	0.386
Molinette	Candiolo	0.727	0.563	0.821	0.643	0.767	0.386
Molinette	Mauriziano	0.609	0.636	0.583	0.583	0.636	0.406

Tabella 8: Soluzioni monocentriche di immagini T2 applicate ai due centri restanti

Soluzione	Applicato a	Acc	Sens	Spec	PPV	NPV	Fitness
Candiolo	MolINETTE	0.643	0.733	0.538	0.647	0.636	0.423
Candiolo	Mauriziano	0.522	0.545	0.500	0.500	0.545	0.491
Mauriziano	MolINETTE	0.571	0.600	0.538	0.600	0.538	0.449
Mauriziano	Candiolo	0.795	0.688	0.857	0.733	0.828	0.279
MolINETTE	Candiolo	0.795	0.688	0.857	0.733	0.828	0.279
MolINETTE	Mauriziano	0.609	0.636	0.583	0.583	0.636	0.406

Tabella 9: Soluzioni monocentriche di immagini ADC applicate ai due centri restanti

Successivamente sono state effettuate delle prove in cui sono stati creati due training set diversi, uno per ogni tipo di immagine, ADC e T2.

<i>Feature T2</i>	<i>Soluzione Training</i>
<i>Skewness STAT</i>	0
<i>Kurtosis STAT</i>	0
<i>Intensity Kurtosis STAT</i>	1
<i>IntensityVariance STAT</i>	1
<i>Joint Max GLCM</i>	1
<i>Joint Average GLCM</i>	1
<i>JointVariance GLCM</i>	0
<i>Joint Entropy GLCM</i>	0
<i>Diff Average GLCM</i>	0
<i>Diff Variance GLCM</i>	0
<i>Diff Entropy GLCM</i>	0
<i>Sum Average GLCM</i>	1
<i>Sum Variance GLCM</i>	0
<i>Sum Entropy GLCM</i>	1
<i>Angular Second Moment GLCM</i>	0
<i>Contrast GLCM</i>	0
<i>Dissimilarity GLCM</i>	1
<i>Inverse Difference GLCM</i>	0
<i>Normalised Inverse Difference GLCM</i>	0
<i>Inverse Difference Moment GLCM</i>	0
<i>Normalised Inverse Difference Moment GLCM</i>	0
<i>Inverse Variance GLCM</i>	0

<i>Correlation GLCM</i>	1
<i>Autocorrelation GLCM</i>	0
<i>Cluster Tendency GLCM</i>	0
<i>Cluster shade GLCM</i>	0
<i>Cluster Prominence GLCM</i>	0
<i>Information Measure Of Correlation1 GLCM</i>	0
<i>Information Measure Of Correlation2 GLCM</i>	0
<i>SRE GLRLM</i>	1
<i>LRE GLRLM</i>	0
<i>LGRE GLRLM</i>	0
<i>HGRE GLRLM</i>	1
<i>SRLGE GLRLM</i>	0
<i>SRHGE GLRLM</i>	1
<i>LRLGE GLRLM</i>	0
<i>LRHGE GLRLM</i>	1
<i>GLNU GLRLM</i>	1
<i>GLNU Normalised GLRLM</i>	1
<i>RLNU GLRLM</i>	0
<i>RLNU Normalised GLRLM</i>	0
<i>RP GLRLM</i>	0
<i>Totale Feature attivate</i>	14
<i>Fitness</i>	0.230

Tabella 10: Risultati del GA con la migliore soluzione policentrica del training set delle immagini T2. Le celle grigie sono feature selezionate, le celle bianche sono feature non attivate.

Kernel	C	Acc	Sens	Spec	PPV	NPV	Fitness
Lineare	5	0.583	0.182	0.923	0.667	0.571	0.670

Tabella 11: Performance della soluzione T2 sul Validation Set T2.

<i>Feature ADC</i>	<i>Soluzione Training</i>
<i>Skewness STAT</i>	1
<i>Kurtosis STAT</i>	1
<i>Intensity Kurtosis STAT</i>	1
<i>IntensityVariance STAT</i>	1
<i>Joint Max GLCM</i>	0
<i>Joint Average GLCM</i>	0
<i>JointVariance GLCM</i>	0
<i>Joint Entropy GLCM</i>	1

<i>Diff Average GLCM</i>	1
<i>Diff Variance GLCM</i>	0
<i>Diff Entropy GLCM</i>	0
<i>Sum Average GLCM</i>	1
<i>Sum Variance GLCM</i>	1
<i>Sum Entropy GLCM</i>	0
<i>Angular Second Moment GLCM</i>	0
<i>Contrast GLCM</i>	1
<i>Dissimilarity GLCM</i>	1
<i>Inverse Difference GLCM</i>	1
<i>Normalised Inverse Difference GLCM</i>	1
<i>Inverse Difference Moment GLCM</i>	1
<i>Normalised Inverse Difference Moment GLCM</i>	1
<i>Inverse Variance GLCM</i>	0
<i>Correlation GLCM</i>	0
<i>Autocorrelation GLCM</i>	0
<i>Cluster Tendency GLCM</i>	1
<i>Cluster shade GLCM</i>	1
<i>Cluster Prominence GLCM</i>	0
<i>Information Measure Of Correlation1 GLCM</i>	1
<i>Information Measure Of Correlation2 GLCM</i>	1
<i>SRE GLRLM</i>	0
<i>LRE GLRLM</i>	0
<i>LGRE GLRLM</i>	0
<i>HGRE GLRLM</i>	0
<i>SRLGE GLRLM</i>	0
<i>SRHGE GLRLM</i>	1
<i>LRLGE GLRLM</i>	0
<i>LRHGE GLRLM</i>	0
<i>GLNU GLRLM</i>	0
<i>GLNU Normalised GLRLM</i>	1
<i>RLNU GLRLM</i>	1
<i>RLNU Normalised GLRLM</i>	0
<i>RP GLRLM</i>	0
<i>Totale Feature attivate</i>	23
<i>Fitness</i>	0.222

Tabella 12: Risultati del GA con la migliore soluzione policentrica del training set delle immagini ADC. Le celle grigie sono feature selezionate, le celle bianche sono feature non attivate.

Kernel	C	Acc	Sens	Spec	PPV	NPV	Fitness
--------	---	-----	------	------	-----	-----	---------

Lineare	10	0.661	0.667	0.654	0.710	0.607	0.344
---------	----	-------	-------	-------	-------	-------	-------

Tabella 13: Performance della soluzione ADC sul Validation Set ADC.

Sono state infine fatte delle ultime prove sui classificatori SVM costruendo quattro dataset, tre monocentrici con le feature delle immagini ADC e T2 unite in una sola matrice, l'ultimo dataset costruito con la suddivisione di Training e Validation descritto nei metodi. Le soluzioni migliori per ogni prova sono presentate nella tabella sottostante.

<i>Feature T2+ADC</i>	Candiolo	Mauriziano	Molinette	Training
<i>ROI Volume(mm3)</i>	0	1	1	0
<i>Joint Max GLCM</i>	1	0	1	0
<i>JointAverage GLCM</i>	0	0	1	0
<i>Joint Variance GLCM</i>	1	0	0	0
<i>Joint Entropy GLCM</i>	0	1	0	1
<i>Diff Average GLCM</i>	0	0	1	1
<i>Diff Variance GLCM</i>	1	0	1	1
<i>Diff Entropy GLCM</i>	0	1	0	0
<i>Sum Average GLCM</i>	1	0	0	1
<i>Sum Variance GLCM</i>	1	0	0	0
<i>SumEntropy GLCM</i>	0	0	0	0
<i>Angular Second Moment GLCM</i>	0	0	0	0
<i>Contrast GLCM</i>	1	1	1	1
<i>Dissimilarity GLCM</i>	1	1	0	1
<i>Inverse Difference GLCM</i>	0	1	1	0
<i>Normalised Inverse Difference GLCM</i>	1	0	1	1
<i>Inverse Difference Moment GLCM</i>	1	1	0	1
<i>Normalised Inverse Difference Moment GLCM</i>	1	1	0	0
<i>Inverse Variance GLCM</i>	0	1	0	1
<i>Correlation GLCM</i>	1	0	0	0
<i>Autocorrelation GLCM</i>	1	1	1	1
<i>ClusterTendency GLCM</i>	1	1	0	0
<i>Cluster Shade GLCM</i>	1	0	0	1
<i>Cluster Prominence GLCM</i>	0	0	0	0
<i>Information Measure Of Correlation1 GLCM</i>	1	1	0	0
<i>Information Measure Of Correlation2 GLCM</i>	1	0	0	0
<i>SRE GLRLM</i>	1	0	0	1
<i>LRE GLRLM</i>	1	0	0	0

<i>LGRE GLRLM</i>	0	0	0	1
<i>HGRE GLRLM</i>	0	1	1	0
<i>SRLGE GLRLM</i>	1	0	0	1
<i>SRHGE GLRLM</i>	1	0	1	0
<i>LRLGE GLRLM</i>	1	0	0	1
<i>LRHGE GLRLM</i>	0	1	1	0
<i>GLNU GLRLM</i>	0	0	0	1
<i>GLNU norm GLRLM</i>	1	1	0	0
<i>RLNU GLRLM</i>	0	0	1	0
<i>RLNU norm GLRLM</i>	1	1	0	0
<i>RP GLRLM</i>	0	1	0	0
<i>GreylevelVariance GLRLM</i>	1	0	0	0
<i>RunlengthVariance GLRLM</i>	0	0	0	1
<i>RunEntropy GLRLM</i>	1	1	0	1
<i>Skewness STAT</i>	0	0	1	1
<i>Kurtosis STAT</i>	0	1	1	0
<i>Intensity Kurtosis STAT</i>	0	1	0	0
<i>IntensityVariance STAT</i>	1	0	1	0
<i>Joint Max GLCM</i>	1	1	1	0
<i>Joint Average GLCM</i>	0	0	0	0
<i>JointVariance GLCM</i>	1	1	1	0
<i>Joint Entropy GLCM</i>	1	0	1	0
<i>Diff Average GLCM</i>	1	0	1	0
<i>Diff Variance GLCM</i>	1	0	1	0
<i>Diff Entropy GLCM</i>	1	0	0	1
<i>Sum Average GLCM</i>	0	0	0	1
<i>Sum Variance GLCM</i>	1	0	0	0
<i>Sum Entropy GLCM</i>	1	1	0	0
<i>Angular Second Moment GLCM</i>	0	0	1	0
<i>Contrast GLCM</i>	0	1	1	0
<i>Dissimilarity GLCM</i>	1	1	0	0
<i>Inverse Difference GLCM</i>	0	0	1	0
<i>Normalised Inverse Difference GLCM</i>	0	0	0	1
<i>Inverse Difference Moment GLCM</i>	0	0	0	0
<i>Normalised Inverse Difference Moment GLCM</i>	0	0	0	0
<i>Inverse Variance GLCM</i>	1	1	1	1
<i>Correlation GLCM</i>	0	0	1	0
<i>Autocorrelation GLCM</i>	0	1	1	0
<i>Cluster Tendency GLCM</i>	0	1	1	0
<i>Cluster shade GLCM</i>	0	0	1	0
<i>Cluster Prominence GLCM</i>	1	1	0	1
<i>Information Measure Of Correlation1 GLCM</i>	1	0	0	0
<i>Information Measure Of Correlation2 GLCM</i>	0	1	0	1

SRE GLRLM	1	0	1	0
LRE GLRLM	1	0	0	1
LGRE GLRLM	0	0	1	0
HGRE GLRLM	0	0	0	0
SRLGE GLRLM	0	1	1	1
SRHGE GLRLM	1	1	1	0
LRLGE GLRLM	0	0	1	1
LRHGE GLRLM	0	1	0	0
GLNU GLRLM	0	0	1	1
GLNU Normalised GLRLM	1	0	0	0
RLNU GLRLM	0	0	0	0
RLNU Normalised GLRLM	0	0	1	1
RP GLRLM	1	0	0	0
Grey Level Variance GLRLM	1	1	1	0
Run Length Variance GLRLM	0	1	0	1
Run Entropy GLRLM	1	0	0	0
Totale Feature attivate	45	35	37	31
Fitness	0.336	0.053	0.218	0.315

Tabella 14: Risultati del GA con le quattro migliori soluzioni per le immagini T2 E ADC. Le celle grigie sono feature selezionate, le celle bianche sono feature non attivate. Le feature in verde sono appartenenti alle immagini ADC, le restanti alle immagini T2.

Acc	Sens	Spec	PPV	NPV
0.646	0.625	0.667	0.652	0.64

Tabella 15: Performance della soluzione Training T2+ADC sul Validation T2+ADC.

3.2. GA CON CLASSIFICATORI BAYESIANI

Le prove con il classificatore Bayesiano sono state solo due, si è considerato il modello proveniente dai dati T2+ADC del centro di Candiolo e il modello

ottenuto dai dati di Training T2+ADC. Questi due modelli sono stati scelti tra i quattro proposti poiché rappresentano le soluzioni con una maggiore accuratezza tra i modelli proposti. Partendo quindi dal dataset di Candiolo T2+ADC e quello del Training T2+ADC è stata effettuata una prova dell'algorithm genetico con il classificatore Bayesiano. I risultati sono qui di seguito tabellati.

<i>Feature T2+ADC</i>	<i>Candiolo</i>	<i>Training</i>
<i>ROI Volume(mm3)</i>	0	1
<i>Joint Max GLCM</i>	0	1
<i>JointAverage GLCM</i>	0	1
<i>Joint Variance GLCM</i>	1	0
<i>Joint Entropy GLCM</i>	0	0
<i>Diff Average GLCM</i>	0	1
<i>Diff Variance GLCM</i>	1	0
<i>Diff Entropy GLCM</i>	0	1
<i>Sum Average GLCM</i>	1	1
<i>Sum Variance GLCM</i>	0	0
<i>SumEntropy GLCM</i>	1	0
<i>Angular Second Moment GLCM</i>	1	1
<i>Contrast GLCM</i>	1	0
<i>Dissimilarity GLCM</i>	0	1
<i>Inverse Difference GLCM</i>	1	0
<i>Normalised Inverse Difference GLCM</i>	0	0
<i>Inverse Difference Moment GLCM</i>	0	0
<i>Normalised Inverse Difference Moment GLCM</i>	0	1
<i>Inverse Variance GLCM</i>	0	0
<i>Correlation GLCM</i>	0	0
<i>Autocorrelation GLCM</i>	0	0
<i>ClusterTendency GLCM</i>	0	0
<i>Cluster Shade GLCM</i>	1	1
<i>Cluster Prominence GLCM</i>	0	1
<i>Information Measure Of Correlation1 GLCM</i>	0	0
<i>Information Measure Of Correlation2 GLCM</i>	0	0
<i>SRE GLRLM</i>	0	0
<i>LRE GLRLM</i>	0	0
<i>LGRE GLRLM</i>	0	0
<i>HGRE GLRLM</i>	1	0
<i>SRLGE GLRLM</i>	0	0
<i>SRHGE GLRLM</i>	1	0
<i>LRLGE GLRLM</i>	0	0
<i>LRHGE GLRLM</i>	0	1

<i>GLNU GLRLM</i>	0	0
<i>GLNU norm GLRLM</i>	0	1
<i>RLNU GLRLM</i>	0	0
<i>RLNU norm GLRLM</i>	1	0
<i>RP GLRLM</i>	1	1
<i>GreylevelVariance GLRLM</i>	1	0
<i>RunlengthVariance GLRLM</i>	0	0
<i>RunEntropy GLRLM</i>	0	1
<i>Skewness STAT</i>	1	0
<i>Kurtosis STAT</i>	0	0
<i>Intensity Kurtosis STAT</i>	0	0
<i>IntensityVariance STAT</i>	0	1
<i>Joint Max GLCM</i>	0	1
<i>Joint Average GLCM</i>	0	1
<i>JointVariance GLCM</i>	0	1
<i>Joint Entropy GLCM</i>	0	0
<i>Diff Average GLCM</i>	0	1
<i>Diff Variance GLCM</i>	0	1
<i>Diff Entropy GLCM</i>	0	1
<i>Sum Average GLCM</i>	0	0
<i>Sum Variance GLCM</i>	0	0
<i>Sum Entropy GLCM</i>	1	1
<i>Angular Second Moment GLCM</i>	1	0
<i>Contrast GLCM</i>	0	1
<i>Dissimilarity GLCM</i>	0	1
<i>Inverse Difference GLCM</i>	0	0
<i>Normalised Inverse Difference GLCM</i>	0	0
<i>Inverse Difference Moment GLCM</i>	0	0
<i>Normalised Inverse Difference Moment GLCM</i>	1	0
<i>Inverse Variance GLCM</i>	0	0
<i>Correlation GLCM</i>	0	0
<i>Autocorrelation GLCM</i>	1	0
<i>Cluster Tendency GLCM</i>	0	1
<i>Cluster shade GLCM</i>	0	1
<i>Cluster Prominence GLCM</i>	0	1
<i>Information Measure Of Correlation1 GLCM</i>	1	0
<i>Information Measure Of Correlation2 GLCM</i>	0	1
<i>SRE GLRLM</i>	0	1
<i>LRE GLRLM</i>	0	0
<i>LGRE GLRLM</i>	0	1
<i>HGRE GLRLM</i>	1	1
<i>SRLGE GLRLM</i>	1	1
<i>SRHGE GLRLM</i>	0	0

LRLGE GLRLM	1	1
LRHGE GLRLM	1	1
GLNU GLRLM	0	0
GLNU Normalised GLRLM	0	0
RLNU GLRLM	0	0
RLNU Normalised GLRLM	0	1
RP GLRLM	1	1
Grey Level Variance GLRLM	0	1
Run Length Variance GLRLM	0	0
Run Entropy GLRLM	1	0
Totale Feature attivate	25	38
Fitness	0.307	0.219157

Tabella 16: Risultati del GA+Bayes con le due migliori soluzioni per le immagini T2 E ADC. Le celle grigie sono feature selezionate, le celle bianche sono feature non attivate. Le feature in verde sono appartenenti alle immagini ADC, le restanti alle immagini T2.

Tutti questi modelli sono stati usati per paragonare le classificazioni delle maschere manuali a quelle automatiche: lo scopo è quello di testare la possibilità del modello di adattarsi ad un cambio di operatore. Per questa comparazione è stata usata la soluzione di ogni modello sull'intero dataset proveniente dalle feature automatiche ottenendo quindi una classificazione per gli stessi pazienti. Questa classificazione è stata comparata alla stessa delle maschere automatiche, da questa parallela è stata ricavata una Confusion Matrix di cui sono riportati i valori indicativi per la valutazione delle performance dei modelli.

Dati	Classificatore	ACC	SENS	SPEC	PPV	NPV
Candiolo T2+ADC	Bayesiano	0.600	0.739	0.455	0.586	0.625
Training T2+ADC	Bayesiano	0.656	0.717	0.568	0.704	0.583
Candiolo T2+ADC	SVM	0.600	0.611	0.593	0.500	0.696
Mauriziano T2+ADC	SVM	0.656	0.717	0.568	0.704	0.583
Molinettes T2+ADC	SVM	0.622	0.732	0.531	0.566	0.703
Training T2+ADC	SVM	0.833	0.833	0.833	0.769	0.882

Tabella 17: Risultati della comparazione tra classificazione automatica e manuale.

Da questa comparazione si evince che il modello più robusto ottenuto tramite le comparazioni è quello ottenuto dai dati di Training delle immagini T2 e ADC

unite in una sola matrice e con un classificatore Bayesiano che ha come parametri un kernel Lineare e un valore di BoxConstraint C=10. Si sono fatte dunque le seguenti considerazioni su questo modello.

3.3. CONFRONTO MASCHERE AUTOMATICHE E MASCHERE MANUALI

Per verificare la qualità delle feature è stato calcolato il valore di ICC identificato come spiegato nei metodi, inoltre è stato valutato il valore percentuale di differenza tra i valori di feature per ogni paziente, di cui è stata ottenuta la mediana. Questi due dati sono utilizzati per tentare di migliorare il modello, eliminando le feature che possono risultare eccessivamente differenti e con un basso valore di ICC si dovrebbero registrare dei miglioramenti sulla comparazione delle due classificazioni. Per le differenze percentuali si è scelto di effettuare tre prove con tre threshold diversi ovvero il 10%, il 15% e il 20% di differenza media. Per il valore di ICC si è identificato come $ICC > 0.5$ un valore che identifica la feature come buona, eliminando le feature con un valore al di sotto di questo threshold. Di seguito le feature rimosse.

Feature Rimosse			
ICC < 0.5	Perc > 10	Perc > 15	Perc > 20
<i>Joint Variance GLCM</i>	ROI Volume(mm3)	ROI Volume(mm3)	ROI Volume(mm3)
<i>Sum Variance GLCM</i>	Joint Max GLCM	Cluster Shade GLCM	Cluster Shade GLCM
<i>ClusterTendency GLCM</i>	Diff Variance GLCM	Cluster Prominence GLCM	GLNU GLRLM
<i>Cluster Shade GLCM</i>	Contrast GLCM	GLNU GLRLM	RLNU GLRLM
<i>Cluster Prominence GLCM</i>	Autocorrelation GLCM	RLNU GLRLM	RunEntropy GLRLM
<i>GreylevelVariance GLRLM</i>	Cluster Shade GLCM	RunEntropy GLRLM	Skewness STAT
<i>Skewness STAT</i>	Cluster Prominence GLCM	Skewness STAT	Intensity Kurtosis STAT
<i>Kurtosis STAT</i>	Information Measure	Intensity Kurtosis STAT	IntensityVariance STAT

	Of Correlation1 GLCM		
<i>Intensity Kurtosis STAT</i>	LGRE GLRLM	IntensityVariance STAT	Cluster shade GLCM
<i>IntensityVariance STAT</i>	SRLGE GLRLM	Joint Max GLCM	LRLGE GLRLM
<i>JointVariance GLCM</i>	LRLGE GLRLM	Diff Variance GLCM	GLNU GLRLM
<i>Sum Variance GLCM</i>	GLNU GLRLM	Contrast GLCM	RLNU GLRLM
<i>Sum Entropy GLCM</i>	RLNU GLRLM	Cluster shade GLCM	
<i>Autocorrelation GLCM</i>	GreylevelVariance GLRLM	LGRE GLRLM	
<i>Cluster Tendency GLCM</i>	RunEntropy GLRLM	SRLGE GLRLM	
<i>Cluster shade GLCM</i>	Skewness STAT	LRLGE GLRLM	
<i>HGRE GLRLM</i>	Intensity Kurtosis STAT	GLNU GLRLM	
<i>LRHGE GLRLM</i>	IntensityVariance STAT	RLNU GLRLM	
<i>GLNU Normalised GLRLM</i>	Joint Max GLCM		
	Diff Variance GLCM		
	Angular Second Moment GLCM		
	Contrast GLCM		
	Autocorrelation GLCM		
	Cluster shade GLCM		

Tabella 18: Feature rimosse per ogni prova con i diversi valori di soglia della differenza percentuale e del valore di ICC.

Queste tre prove hanno portato a tre modelli differenti che sono stati successivamente testati sulle maschere automatiche per confrontare ancora una volta la classificazione. Seguono le performance ottenute dei tre modelli in confronto alle classi reali e il confronto dei modelli nelle due classificazioni, automatiche e manuali.

ICC	Classificatore	ACC	SENS	SPEC	PPV	NPV
< 0.5	> 10%	0.600	0.537	0.653	0.564	0.627
< 0.5	> 15%	0.556	0.439	0.653	0.514	0.582
< 0.5	> 20%	0.656	0.610	0.694	0.625	0.680

Tabella 19: Performance dei tre modelli in riferimento alle classi reali.

ICC < 0.5 Diff > 10%	Cl. Manuali	
Cl. Automatiche	32	19
	21	18

Tabella 20.a

ICC < 0.5 Diff > 10%	Cl. Manuali	
Cl. Automatiche	35.6%	21.1%
	23.3%	20.0%

Tabella 20.b

ICC < 0.5 Diff > 15%	Cl. Manuali	
Cl. Automatiche	29	26
	18	17

Tabella 20.c

ICC < 0.5 Diff > 15%	Cl. Manuali	
Cl. Automatiche	32.2%	28.9%
	20.0%	18.9%

Tabella 20.d

ICC < 0.5 Diff > 20%	Cl. Manuali	
Cl. Automatiche	34	16
	16	24

Tabella 20.e

ICC < 0.5 Diff > 20%	Cl. Manuali	
Cl. Automatiche	37.8%	17.8%
	17.8%	26.7%

Tabella 20.f

Tabella 20: Confusion Matrix dei tre modelli confrontate le classificazione manuali e automatiche. Le tabelle a,c, ed e contengono il numero di pazienti. Le tabelle b,d ed f le percentuali dei pazienti sul loro totale.

Per valutare al meglio queste nuove prove è stato compiuto uno studio all'interno delle feature selezionate e dei pazienti sbagliati tra le due classificazioni, sia tra loro che in riferimento alla classe reale. In prima istanza si sono statisticamente calcolate le variabili indicate nei metodi: precision, recall, DSC e la differenza percentuale di volume tra le maschere. Successivamente sono state fatte delle considerazioni su questi valori e sull'andamento degli errori alla ricerca di un collegamento tra l'intensità di questi valori e la quantità di errori segnalati nelle due classificazioni. In particolar modo ci si è concentrati sulla terza prova (differenza percentuale > 20% e ICC < 0.5) suddividendo i pazienti in quattro categorie secondo il loro valore di Precisione e Recall. Con un threshold di 0.75 sono stati definiti pazienti con alta precisione o recall chi presenta un valore maggiore del termine, di questi sono stati suddivisi in classificazioni uguali o diverse nel confronto tra automatiche e diverse.

Alta Precisione e Alta Recall		Alta Precisione e Bassa recall		Bassa Precisione e Alta Recall		Bassa Precisione e Bassa Recall	
Classificazione uguale	Classificazione diversa	Classificazione uguale	Classificazione diversa	Classificazione uguale	Classificazione diversa	Classificazione uguale	Classificazione diversa
62	19	5	25	11	1083	75	21
204	29	23	44	42	2028	203	37
212	231	31	53	1080		211	43
221	1072	56	99	1054		1134	47
228	1154	66	1138	2011		1171	2003
1081	17	73	6	2016		1151	2007
1120	1084	76	45	2032		2002	2013
1135	1121	80	64			2005	2026
1157		82	65			2008	2030
1159		208	93			2014	2039
1073		220	2031			2022	
1124		1153	2036			2041	
1130		35				2042	
1170		46				2045	
		230				2047	
		1136					
		2004					
		2015					
		2023					
		2029					
		2049					

Tabella 21: Pazienti classificati secondo la loro Recall e Precisione calcolata sulle loro maschere automatiche e manuali, suddivisi per errori di classificazione tra i risultati dello stesso modello.

Secondo questo trend statistica i pazienti con alta recall e alta precisione sono uguali al 63% nei casi, i pazienti con alta precisione e bassa recall sono uguali nel 63% dei casi, i pazienti con bassa precisione e alta recall sono uguali nel 77% dei casi, i pazienti con bassa precisione e bassa recall sono uguali nel 60% dei casi. Sono stati inoltre valutati dei pazienti interessanti con alta recall e alta precisione, venendo individuati come i pazienti con classificazione differente tra automatica e digitale e con un valore di differenza di volume tra maschere inferiore al 1%.

Pazienti interessanti	Errore sulla classificazione	Volume
19	Automatica	0.028280858
231	Manuale	0.003990
1072	Manuale	0.070133
1138	Automatica	0.010668
1154	Manuale	0.039196
2030	Manuale	0.097087379
2007	Manuale	0.030630491

Tabella 22: Pazienti "interessanti" con alta precisione e alta recall e differenza di volume tra maschere minore dell' < 10%.

Infine, per le feature selezionate dal modello sono stati costruiti dei boxplot a confronto per controllare, dei pazienti interessanti, il loro valore di feature con la corrispondenza dei boxplot ottenuti dalla matrice completa e le classi reali. Da quest'ultima valutazione sono state tratte le conclusioni finali sullo studio.

4. CONCLUSIONI

Il modello scelto come il migliore sembra essere rappresentato dalla combinazione dell'algoritmo genetico con un classificatore SVM poiché le performance presentate sono più incoraggianti rispetto ai classificatori Bayesiani. I dati più promettenti per la costruzione del modello sono la partizione di training formata dall'unione delle feature estratte dalle immagini T2 e immagini ADC. Dalla scelta di questa soluzione e a seguito delle successive prove si è potuto notare come al variare dei threshold e all'esclusione della feature teoricamente nocive per il modello le performance andavano a mostrare un trend discendente nel confronto tra la classificazione automatica e manuale, invece dimostrando un miglioramento delle performance rispetto alla classe reale. Ciò che si cerca è un modello robusto, che riesca a reggere un cambiamento di operatore, quindi potrebbe essere un'ottima scelta rimanere sulla scelta prima che con un'accuratezza di circa 83% sembra assicurare una coerenza di classificazioni anche ad un ipotetico scambio di operatori al processo di segmentazioni delle immagini. Inoltre, dallo studio delle feature si nota come le stesse non appaiono distinte al loro interno, cioè i valori sembrano appartenere ad un intervallo piuttosto ristretto il che potrebbe spiegare il trend discendente delle performance all'esclusione delle feature dannose. Nei boxplot si è potuto infatti osservare come fosse apparentemente labile l'intervallo a distinzione di una feature per la classificazione in caso 0 e caso 1, se non addirittura delle casistiche in cui alcuni intervalli sono interamente contenuti nell'altro. Ciò vuol dire che ad una minima variazione del valore la classificazione sarebbe cambiata drasticamente. A questo problema potrebbe ricondursi una soluzione riguardante la costruzione del modello sull'algoritmo genetico tentando di farlo allenare sulle maschere automatiche e non quelle manuali. Inoltre, si potrebbe valutare l'aggiunta di altre feature di ordine superiore per ottenere probabilmente più feature informative e decisive rispetto a quelle disponibili. Nell'osservazione delle maschere dei

pazienti interessanti si è potuto valutare come alcune maschere manuali contenessero delle zone bianche al loro interno, o come alcune maschere automatiche fossero più abbondanti rispetto alle corrispettive manuali. Si potrebbe ricercare in questi dettagli una motivazione per i risultati ottenuti nonostante i difetti trovati siano comunque di scala minore rispetto ad altri riportati, invece, corretti nella classificazione. I limiti di questo studio potrebbero essere rappresentati da alcune prove che possono essere confrontate con altri tentativi su ulteriori tipi di maschera, così come una nuova segmentazione avrebbe potuto fortificare le performance dei confronti dello stesso modello su dati provenienti da più operatori.

BIBLIOGRAFIA

- [1] Ferlay, J., et al. "Cancer Incidence and Mortality Patterns in Europe: Estimates for 40 Countries and 25 Major Cancers in 2018." *European Journal of Cancer*, vol. 103, Nov. 2018, pp. 356–387, 10.1016/j.ejca.2018.07.005.
- [2] "I Numeri Del Cancro in Italia." AIOM, www.aiom.it/i-numeri-del-cancro-in-italia/.
- [3] Feeney, Gerard, et al. "Neoadjuvant Radiotherapy for Rectal Cancer Management." *World Journal of Gastroenterology*, vol. 25, no. 33, 7 Sept. 2019, pp. 4850–4869, 10.3748/wjg.v25.i33.4850. Accessed 8 Nov. 2019.
- [4] De Campos-Lobato, Luiz Felipe, et al. "Pathologic Complete Response After Neoadjuvant Treatment for Rectal Cancer Decreases Distant Recurrence and Could Eradicate Local Recurrence." *Annals of Surgical Oncology*, vol. 18, no. 6, 5 Jan. 2011, pp. 1590–1598, 10.1245/s10434-010-1506-1. Accessed 10 Dec. 2019.
- [5] A. M. Mandard et al., «Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations», *Cancer*, vol. 73, n. 11, pagg. 2680–2686, giu. 1994.
- [6] V. Giannini et al., «Predicting locally advanced rectal cancer response to neoadjuvant therapy with 18F-FDG PET and MRI radiomics features», *Eur. J. Nucl. Med. Mol. Imaging*, vol. 46, n. 4, pagg. 878–888, apr. 2019.
- [7] K. Preetha e D. S. K. Jayanthi, «GLCM and GLRLM based Feature Extraction Technique in Mammogram Images», *Int. J. Eng. Technol.*, vol. 7, n. 2.21, pagg. 266–270, apr. 2018.
- [8] G. J. R. Cook, G. Azad, K. Owczarczyk, M. Siddique, e V. Goh, «Challenges and Promises of PET Radiomics», *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 102, n. 4, pagg. 1083–1089, 15 2018.
- [9] R. G. H. Beets-Tan et al., «Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting», *Eur. Radiol.*, vol. 28, n. 4, pagg. 1465–1475, apr. 2018.
- [10] Jo, Jun-Mo. "The effect of big data normalization preprocessing on machine learning performance." *The Journal of the Korean Institute of Electronics and Communication Sciences*, vol. 14, no. 3, The Electronics and Telecommunications Society, June 2019, pp. 547–552, doi:10.13067/JKIECS. 2019.14.3.547.
- [11] Shrout, Patrick E. and Fleiss, Joseph L. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 1979, 86, 420-3428.

[12] McGraw, Kenneth O. and Wong, S. P. (1996), Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30-46.