



**Politecnico  
di Torino**

**Politecnico di Torino**

Corso di Laurea Magistrale in Ingegneria Biomedica

A.a. 2023/2024

Sessione di Laurea Dicembre 2024

**Biomarcatore radiopatologico  
basato sull'intelligenza artificiale  
per la predizione della risposta  
patologica nei pazienti con cancro  
del retto**

Relatrice:  
Prof.ssa Samanta Rosati  
Corelatrice:  
Ing. Valentina Giannini

Candidata:  
Ilaria Oliva



# Sommario

Il cancro coloretale rappresenta uno dei tumori più diffusi a livello globale ed è una delle principali tipologie di cancro che causano un alto tasso di mortalità. Al fine di contrastarlo, le strategie terapeutiche che vengono maggiormente attuate sono la resezione chirurgica e la chemioterapia adiuvante. In particolare, quest'ultima non sempre è un beneficio per il paziente, soprattutto se risulta ad alto rischio di ricaduta. In relazione a ciò, la patomica sta emergendo come un campo promettente che può migliorare la diagnosi e la gestione dei pazienti, fornendo informazioni preziose sulle caratteristiche molecolari e cellulari del tumore. Sfruttando strumenti di intelligenza artificiale, come machine learning e deep learning, la patomica si occupa inoltre di analizzare immagini digitali di campioni di tessuto, come quelli ottenuti da biopsie o resezioni chirurgiche, ed estrarre pattern complessi che potrebbero essere correlati alle caratteristiche del tumore.

Lo scopo di questa tesi è quello di sviluppare un biomarcatore per predire la risposta patologica nei pazienti con cancro al retto, sfruttando tecniche di machine learning che ricevono come input immagini istologiche digitalizzate.

Nella prima fase del lavoro si avevano a disposizione 20 pazienti provenienti dal centro Niguarda, dalle cui immagini sono state estratte delle tiles di dimensioni 224x224 pixel ad alta risoluzione (0.5  $\mu\text{m}/\text{pixel}$ ). Successivamente sono state normalizzate con la color deconvolution di Macenko; quindi, utilizzate come dataset e suddivise in Training e Test set. Con lo scopo di ottenere le segmentazioni dei tessuti tumorali, sono stati implementati due modelli di rete convoluzionale, ResNet18 e ResNet50, e inoltre, tramite l'approccio del *transfer learning*, è stata sfruttata una rete VGG19 precedentemente allenata su un task simile. Dopo una fase di tuning dei parametri, le prestazioni migliori sono state ottenute dalla VGG19, con un'accuratezza nel Training set del 79% e nel Test set del 82%. Occorre precisare che il dataset risultava essere sbilanciato, e per questo è stata calcolata anche l'accuratezza bilanciata, pari al 74%.

Nella seconda fase del lavoro, sono stati aggiunti 116 nuovi pazienti, provenienti da diversi centri, tra cui l'ospedale Niguarda, l'ospedale Papa Giovanni XXIII, l'Istituto Oncologico Europeo e l'Istituto Oncologico Veneto. A partire da queste immagini sono state estratte delle tiles con le stesse caratteristiche delle precedenti, e queste sono state date in input alla rete VGG19, ottenendo le segmentazioni dei tessuti tumorali. Da queste sono state estratte 97 features tramite la libreria *pyradiomics*, e una volta aggregate le tiles per ciascun paziente, si sono ottenuti, tramite features selection, dei subset contenenti le feature più correlate alle due classificazioni: "responder" e "not responder" al trattamento oncologico. Con questi subset sono stati allenati diversi classificatori, tra cui: Naive Bayes, K-Nearest Neighbours (KNN), Support vector machine (SVM), Decision Tree e Random Forest. Tra i vari classificatori, le migliori prestazioni si sono ottenute dal Decision Tree, con accuratezza pari a 88% nel Training Set e 66% nel Test Set.

In conclusione, i risultati ottenuti hanno evidenziato alcune limitazioni nella capacità del classificatore di generalizzare su nuovi dati, suggerendo che un ampliamento del dataset

potrebbe migliorare le prestazioni riducendo l'overfitting. Tuttavia, questo studio ha posto le basi per lo sviluppo di un biomarcatore predittivo, aprendo nuove prospettive nell'uso dell'intelligenza artificiale per migliorare la predizione della risposta patologica nei pazienti con cancro del retto.

# Indice

Introduzione .....	1
Metodi e materiali .....	3
Pazienti .....	3
Pre-processing .....	3
Creazione del dataset per addestramento reti neurali .....	4
Convolutional Neural Network .....	4
Allenamento reti .....	7
Confronto prestazioni dei tre modelli .....	11
Estrazione delle feature .....	15
Creazione del dataset per feature selection .....	16
Feature selection e costruzione dei classificatori .....	18
RFE (Recursive Feature Elimination) .....	19
Minimum Redundancy - Maximum Relevance (mRmR) .....	19
Affinity Propagation .....	20
Correlazione assoluta tra features e target .....	21
Analisi delle componenti principali (PCA) .....	22
Naïve Bayes .....	22
K-Nearest Neighbours (KNN) .....	23
Support Vector Machine (SVM) .....	24
Decision Tree .....	25
Ensamble learning models .....	25
Random Forest .....	26
Stochastic Gradient Descent (SGD) Classifier .....	27
Risultati .....	27
Metodo di aggregazione: calcolo della media di ciascuna feature .....	27
Metodo di aggregazione: calcolo delle statistiche descrittive dell'istogramma per ciascuna feature: Media, Mediana, Skewness, Kurtosis, Percentili, range interquartili (IQR) ed Entropia .....	42
Prestazioni classificatori migliori nel Validation Set .....	54
Appendice .....	62
Bibliografia .....	97

# Introduzione

Il tumore del colon retto (CRC) rappresenta ancora oggi una delle neoplasie più importanti, con indici di mortalità e progressione metastatica della malattia legati alle caratteristiche biologiche e strutturali del tumore primario. Detiene il terzo posto in termini di incidenza (10.0 % rispetto a tutti i siti tumorali) preceduto dal tumore alla mammella (11.7 %) e dal cancro ai polmoni (11.4 %). Nell'ultimo quinquennio si è registrato un calo della mortalità in Italia di circa il 10%. A ciò hanno dato un contributo notevole gli interventi di medicina predittiva e preventiva che si avvalgono di una serie di strumenti diagnostici e dispositivi chirurgici esplorativi (come l'endoscopia) che consentono diagnosi molto precoci e interventi ambulatoriali sulle lesioni displasiche e poliposiche, evitando la trasformazione neoplastica e l'invasività della massa tumorale [1]. Tuttavia, è importante notare che, a causa della pandemia di COVID-19, si è verificata una riduzione degli screening oncologici nel 2020, con conseguente diminuzione delle diagnosi precoci. Questo ha portato a un aumento dei casi diagnosticati in stadi più avanzati, influenzando negativamente le prospettive di sopravvivenza [2].

Quasi il 30% dei CRC sono localizzati nel retto, compreso un sottogruppo di tumori rettali localmente avanzati tumori (LARC). Il cancro del retto, noto anche come adenocarcinoma del retto, è un tumore maligno che si sviluppa dalle cellule ghiandolari presenti nella mucosa del retto. Quando il tumore penetra nella sottomucosa, si parla di carcinoma invasivo, caratterizzato da una crescita locale e dalla capacità di diffondersi attraverso i vasi linfatici o sanguigni, con metastasi che interessano comunemente il fegato e i polmoni. L'eziologia di questo tumore è complessa e coinvolge una serie di alterazioni genetiche delle cellule della mucosa rettale, che gradualmente portano alla trasformazione della mucosa sana in adenoma e, successivamente, in adenocarcinoma. I principali fattori di rischio per lo sviluppo del cancro del retto sono:

- l'eccessivo consumo di carni rosse e di insaccati, farine e zuccheri raffinati
- il sovrappeso e anche la scarsa attività fisica
- il consumo eccessivo di alcool
- il fumo
- la presenza di malattie infiammatorie croniche intestinali, come la malattia di Crohn e rettocolite ulcerosa.

Al contrario, alcuni fattori di protezione dal rischio di insorgenza sono:

- il consumo di frutta e verdure e carboidrati non raffinati
- il calcio e la vitamina D

Esistono anche delle sindromi in cui sono state identificate mutazioni genetiche, che possono essere trasmesse in maniera ereditaria, come la sindrome di Lynch [3] o la poliposi adenomatosa familiare (FAP)

Fino al 2017, il trattamento del cancro del retto prevedeva principalmente l'uso di chemio-radioterapia seguita da chirurgia, con una fase di chemioterapia adiuvante post-operatoria per prevenire le recidive. Tuttavia, studi recenti hanno evidenziato che il maggiore beneficio terapeutico si ottiene intervenendo nella fase preoperatoria. Per questo, si sta sempre più

puntando a intensificare il trattamento chemioterapico prima dell'intervento chirurgico e, ove possibile, a ridurre la componente di radioterapia. Questo approccio, noto come **total neoadjuvant treatment (TNT)**, concentra l'intero percorso oncologico nella fase preoperatoria.

L'obiettivo principale della terapia preoperatoria è quello di ridurre lo stadio o le dimensioni del tumore, migliorando il controllo della malattia a livello locale. Tuttavia, la possibilità di metastasi a distanza di tempo rimane un problema. Per affrontare questo rischio di recidiva, studi recenti hanno valutato l'efficacia di una chemioterapia sistemica precoce, introdotta già nelle prime fasi del trattamento. I risultati di recenti studi prospettici randomizzati e controllati (RCT) hanno mostrato che il TNT può migliorare la **sopravvivenza libera da malattia (DFS)**, aumentare i **tassi di risposta patologica completa (pCR)** e i tassi di completamento della chemioterapia, dimostrando così un impatto positivo su diversi indicatori di esito del trattamento.

Uno dei principali studi in questo ambito è il NO-CUT Trial, presentato al Congresso ESMO 2024 a Barcellona, che valuta la TNT seguita da un approccio di Non-Operative Management (NOM) per il carcinoma del retto medio-basso localmente avanzato. I risultati mostrano che un quarto dei pazienti ha raggiunto una remissione clinica completa, mantenuta nel tempo senza ricorrere a chirurgia e colostomia, con una sopravvivenza libera da metastasi del 97% a 30 mesi. Lo studio si distingue anche per l'impiego di biomarcatori multiomici, analisi genomiche e trascrittomiche, con lo scopo di identificare in anticipo i pazienti che possono evitare la chirurgia o che potrebbero trarre beneficio da terapie alternative.

In questo contesto, lo studio NO-CUT apre nuove prospettive nell'uso della radiopatologia, che combina l'analisi radiologica e patologica delle immagini per studiare le caratteristiche tumorali, permettendo di estrarre specifiche *features* da immagini diagnostiche potenzialmente utili nel predire la risposta clinica e migliorare la personalizzazione dei trattamenti [4].

In merito a ciò, ci sono diversi studi che dimostrano l'efficacia di queste features come biomarcatori. Ad esempio, Chiloiro et Al., sfruttando la risonanza magnetica (MRI) pre e post chemioterapia neoadiuvante (nCRT), hanno indagato sulla correlazione tra la variazione nel tempo delle features radiomiche (radiomica delta) e il tasso di metastasi a distanza di 2 anni nei pazienti LARC, riuscendo così a stimare il comportamento della malattia nella LARC e prevedendo il rischio di recidiva sistemica precoce [5].

Rosati, Giannini et al. hanno, invece, dimostrato che l'ottimizzazione simultanea di un sottoinsieme di features e dei parametri dei classificatori su diversi set di dati di imaging (T2w, DWI e PET) potrebbe migliorare la performance nella predizione della terapia prima del trattamento [6].

Alrahawy et Al. hanno studiato e utilizzato la Texture Analysis, parte del campo radiomico che quantifica le caratteristiche dell'immagine come la misurazione statistica della distribuzione dei pixel in scala di grigio, con lo scopo di riuscire a predire importanti esiti clinici nel cancro del retto, come la risposta al trattamento dopo la chemioradioterapia neoadiuvante, la sopravvivenza, la recidiva del tumore e la diffusione a distanza. [7]

Lo studio di Kim et Al. hanno, invece, valutato l'associazione delle texture features di tomografie computerizzate (TC) di carcinoma rettale localmente avanzato, pre-trattamento potenziato, con la risposta alla CRT neoadiuvante, valutando anche il ruolo delle texture features come fattori indipendenti per la previsione della sopravvivenza libera dalla malattia (DFS). Il gruppo dei responder al trattamento ha mostrato texture features omogenee, rispetto al gruppo dei non-responder, e il gruppo di pazienti con texture features omogenee ha dimostrato una DFS più elevata. Questi risultati mostrano il potenziale delle texture features della TC come biomarcatore di imaging non invasivo in grado di predire la risposta al trattamento e la prognosi del paziente [8].

Sulla base degli studi precedenti, lo scopo di questo lavoro di tesi è quello di sviluppare un biomarcatore per predire la risposta patologica alla terapia totale neoadiuvante (TNT) nei pazienti con cancro al retto, sfruttando tecniche di machine learning che ricevono come input le features radiomiche estratte da immagini istologiche.

## Metodi e materiali

In questa prima fase del lavoro di tesi lo scopo era quello di sfruttare la capacità delle deep convolutional neural network (CNN) [9] di classificare le componenti dei tessuti delle immagini istologiche per ottenere le segmentazioni tumorali anche dalle immagini dei pazienti raccolti nella seconda fase, che risultavano sprovvisti di una segmentazione manuale.

### Pazienti

I pazienti a disposizione nella prima fase del lavoro di tesi erano 20, provenienti tutti dall' ASST Grande Ospedale Metropolitano Niguarda. Facenti parte dello studio NO-CUT [10] citato precedentemente, sono soggetti a cui è stato diagnosticato un carcinoma del retto medio-basso localmente avanzato e somministrato il TNT seguito da un approccio di Non-Operative Management (NOM). Di ciascun paziente si aveva a disposizione un'immagine istologica digitalizzata della sezione di tessuto prelevato dalla biopsia, colorata con l'ematossilina-eosina (H&E), che tinge i nuclei in blu/violetto e il citoplasma in rosa, per evidenziare le strutture cellulari e tissutali. In seguito, una patologa ha segmentato manualmente le lesioni presenti tramite QuPath [11], un software open source per l'analisi delle bioimmagini. I dati sono stati completamente resi anonimi, assegnando un identificativo unico casuale ad ogni paziente in modo da garantirne la protezione della privacy.

### Pre-processing

Da ogni immagine WSI (whole slide image) sono state estratte delle tiles di dimensione 224x224 pixel e risoluzione 0.5466  $\mu\text{m}/\text{pixel}$ , tramite QuPath. Successivamente, per ogni tiles è stata effettuata la Color Normalization con il metodo "Macenko" [12].



## Creazione del dataset per addestramento reti neurali

Il dataset completo era composto da 15360 tiles tumorali e 7729 tiles non tumorali. È stato suddiviso per il 70% in Training set e 30% in Test set per allenare e testare il modello neurale nella classificazione delle tiles in “tumore” e “non tumore”. Si sono ottenuti così due dataset composti da:

- 10752 tiles tumorali e 5410 tiles non tumorali, per un totale di 16162 tiles nel Training set
- 4608 tiles tumorali e 2319 tiles non tumorali, per un totale di 6927 tiles nel Test set

Poiché il dataset risulta essere fortemente sbilanciato, sono state applicate in fase di allenamento delle tecniche di Data Augmentation [13].

## Convolutional Neural Network

Le reti neurali sono un insieme di algoritmi atti a riconoscere delle relazioni fra i dati presenti in un dataset di addestramento. Il funzionamento di questi algoritmi si ispira a come i neuroni del cervello umano elaborano le informazioni [14]. Sono costituite da un input layer, una serie di hidden layer e da un output layer. Ogni neurone è connesso ai neuroni del layer precedente e del layer successivo e ogni connessione è associata ad un valore numerico detto “peso”.

Ogni input di addestramento viene passato attraverso la rete neurale in un processo chiamato “Forward Propagation”, ovvero layer per layer, fino ai neuroni di output. Una volta che il modello ha prodotto un output, questo, detto “output previsto”, viene confrontato con l'output target dato in un processo chiamato “Back Propagation”. Questo processo si occupa di calcolare come deve essere modificato ogni peso della rete per ridurre l'errore o **loss**, ovvero, la differenza tra l'output previsto e l'output target. L'obiettivo è ridurre al minimo questo errore, poiché minori sono le differenze, più accurato è il modello.

L'output di ciascun neurone è rappresentato dalla seguente equazione [eq. 1]:

$$h_i = \sigma\left(\sum_{j=1}^N V_{ij}x_j + T_i^{hid}\right)$$

*Equazione 1 Funzione di attivazione*

Dove  $\sigma$  è la funzione di attivazione,  $N$  il numero di neuroni di ingresso,  $V_{ij}$  i pesi,  $x_j$  i dati in ingresso all' $i$ -esimo neurone e  $T_i^{hid}$  i termini di soglia o bias che consentono al neurone di regolare il livello di attivazione anche in assenza di input.

Una **rete neurale convoluzionale (CNN)** è una tipologia di algoritmo di apprendimento profondo [15] adatto al riconoscimento ed elaborazione delle immagini; in particolare, riesce a catturare modelli gerarchici e dipendenze spaziali all'interno delle immagini.

I principali componenti di una rete convoluzionale sono:

- Livelli convoluzionali: applicano operazioni convoluzionali alle immagini di input, utilizzando filtri (o Kernel) per rilevare delle caratteristiche, come i bordi, le texture e pattern più complessi. Questo tipo di operazioni viene effettuato per mantenere le relazioni spaziali tra i pixel.
- Livelli di pooling: questi livelli si occupano di sottodimensionare le dimensioni spaziali dell'input (essendo generalmente immagini, si avrebbe un input a tre dimensioni) per ridurre la quantità di parametri e di calcolo nella rete.
- Funzioni di attivazione: le funzioni di attivazione non lineari, come la Rectified Linear Unit (ReLU), aggiungono non linearità al modello, permettendogli di apprendere relazioni più complesse all'interno dei dati.
- Livelli completamente connessi: sono responsabili delle previsioni, utilizzando le caratteristiche di alto livello apprese dai livelli precedenti. In questi livelli, ogni neurone è collegato a tutti i neuroni del livello successivo.

Le CNN vengono addestrate tramite apprendimento supervisionato, ovvero con un set di immagini di addestramento già etichettate. In questo modo, la rete impara a collegare le immagini di input alle loro etichette corrette.

Le CNN confrontate in questo studio sono le ResNet18, ResNet50 e VGG19:

- Le ResNet (Reti Residue) [16] introducono l'utilizzo dei "blocchi residui" che consentono ai gradienti di propagarsi meglio nelle reti profonde ed evitare il problema del vanishing gradient: le derivate utilizzate per aggiornare i pesi che diventano sempre più piccole man mano che si propagano all'indietro attraverso la rete [17]. Ogni blocco residuo contiene delle connessioni di salto (skip connections) che permettono ai dati di saltare alcuni strati, consentendo al modello di apprendere solo le differenze, o i "residui", tra input e output del blocco. Ciò dovrebbe evitare il peggioramento delle prestazioni con l'aumento della profondità. La principale differenza tra la ResNet18 e la ResNet50 è il numero di layers allenabili, 18 nel primo caso e 50 nel secondo caso.
- La ResNet18 (figura 1) utilizza blocchi residui standard (basic block), composti ciascuno da due convoluzioni 3x3 con differenti numeri di neuroni per ciascun blocco. Questi blocchi sono più semplici e meno costosi in termini di calcolo.

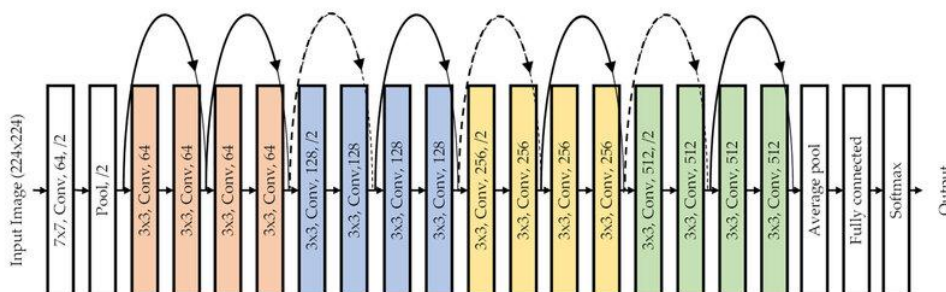


Figura 1 ResNet18 architecture

- La ResNet50 (figura 2) invece è caratterizzata da blocchi residui più complessi, chiamati "bottleneck blocks". Ogni blocco "bottleneck" è composto da tre convoluzioni: una

convoluzione 1x1 per ridurre il numero di canali, una convoluzione 3x3, e infine un'altra convoluzione 1x1 per riportare il numero di canali alla dimensione originale. Questo schema riduce il numero di parametri mantenendo la capacità di apprendimento, offrendo così prestazioni migliori su compiti più complessi.

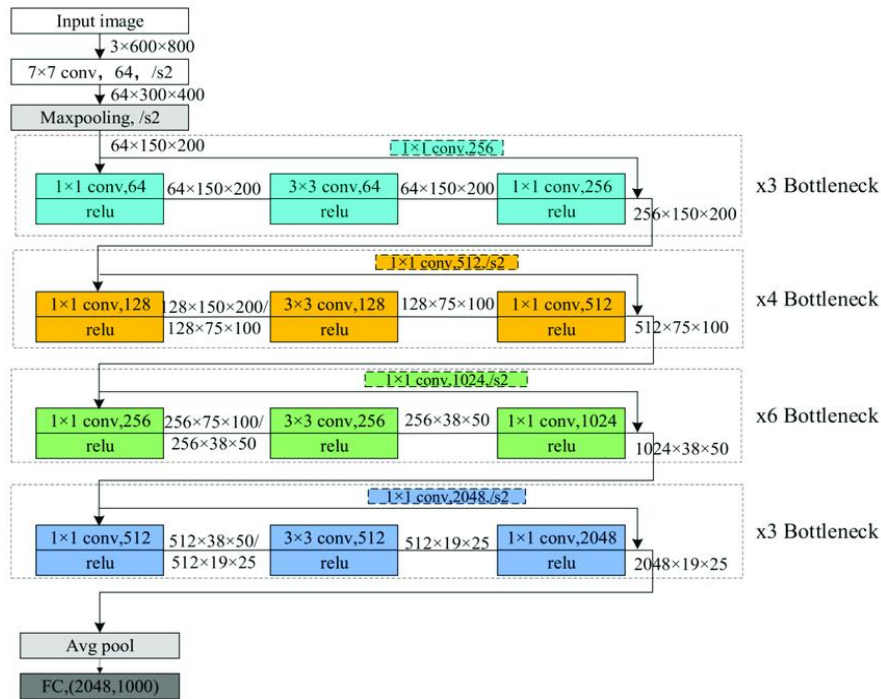


Figura 2 ResNet50 architecture

- La VGG19 è una rete neurale convoluzionale profonda che fa parte di una famiglia di modelli chiamati VGGNet [18]. Anche questa come le precedenti presenta una struttura a blocchi: presenta 19 layers allenabili, di cui 16 sono layers convoluzionali e 3 sono layers completamente connessi. Ogni blocco convoluzionale è composto da strati convoluzionali con kernel di dimensioni 3x3, seguiti da layers di pooling 2x2. Questa struttura a blocchi permette alla rete di avere una grande capacità di apprendimento senza un numero eccessivo di parametri, in quanto, i kernel di dimensioni 3x3 e stride pari a 1 dei layers convoluzionali permettono alla rete di approfondire la comprensione spaziale senza perdere dettagli, mentre i layer di pooling riducono la dimensionalità (figura 3).

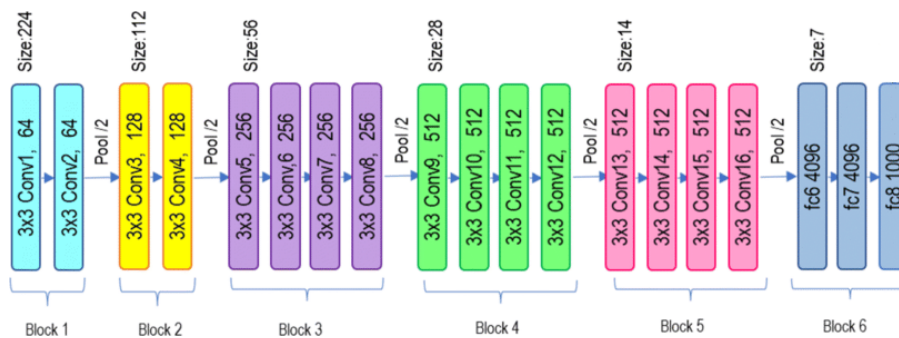


Figura 3 VGG19 architecture

## Allenamento reti

- ResNet18

Le fasi di costruzione, allenamento e testing dei modelli sono state effettuate sfruttando la piattaforma “Colab” di Google, che ha fornito le risorse necessarie all’allenamento di reti neurali, per mezzo dell’API “Keras” [19].

Come primo passo, dal 30% del dataset di training è stato ricavato un subset utilizzato nella fase di validazione (Validation set) da utilizzare al termine di ogni epoca per calcolare la “Validation Loss” e le metriche di validazione. Poiché, come detto in precedenza, il dataset risultava essere sbilanciato, sono state applicate delle tecniche di Data Augmentation ai dati del Training Set:

- Random Flip orizzontale e verticale
- Random Rotation 0.2

Successivamente, i dati del Training e del Validation set sono stati normalizzati tra 0 e 1, in modo da stabilizzare l’addestramento e velocizzarne la convergenza. Quindi, è stata definita la funzione di Early Stopping, che monitora la validation loss e interrompe l’allenamento quando questa smette di migliorare per un numero definito di epoche consecutive. A questo punto è stato implementato manualmente il modello di ResNet18, in quanto non presente in Keras

Sono state effettuate diverse prove con lo scopo di ottimizzare gli iperparametri: gli iperparametri sono quei valori preimpostati che non vengono appresi dal modello durante l’addestramento, come il “learning rate”, la funzione di ottimizzazione, la profondità della rete, il numero di neuroni in ciascun layer, il numero di epoche, ovvero il numero di cicli completi attraverso tutto il dataset, e il batch size, il numero di campioni che vengono dati in input al modello prima di aggiornare i pesi.

Le prestazioni migliori si sono ottenute con le seguenti combinazioni di iperparametri (tab. 1):

Optimizer	Learning rate	Funzione di Loss	Batch size	Epochs
Adam	0.0001	“Binary_crossentropy”	64	100

*Tabella 1 Migliore combinazione iperparametri ResNet18i*

La funzione di ottimizzazione “Adam” aggiorna il learning rate in maniera dinamica, per ciascun parametro, durante l’allenamento. Questa funzione calcola due momenti esponenzialmente decrescenti: la media dei gradienti e la varianza non centrata dei gradienti. I gradienti forniscono le informazioni necessarie per aggiustare i parametri del modello e ridurre l’errore. Grazie alla combinazione della media e della varianza non centrata dei gradienti, l’ottimizzatore Adam può regolare dinamicamente il tasso di apprendimento in base alla magnitudine e alla varianza dei gradienti. Questo rende Adam

capace di adattarsi automaticamente ai cambiamenti della distribuzione dei gradienti, rendendolo più efficace e stabile, specialmente in problemi complessi [20].

La funzione di Loss utilizzata è la Binary Cross-Entropy, tipicamente utilizzata nei problemi di classificazione binaria. La formula è: [eq. 2]

$$\text{Loss} = -[y * \log(p(i)) + (1 - y) * \log(1 - p(i))]$$

*Equazione 2 Loss Binary Cross-Entropy*

dove:

- $y$  è il valore reale (0 o 1),
- $p(i)$  è la probabilità prevista per la classe 1.

Questa funzione di loss penalizza fortemente il modello quando prevede una probabilità lontana dal valore reale, migliorando la sua accuratezza [21].

Inoltre, per migliorare le prestazioni del modello ResNet-18, sono state apportate modifiche ai layer finali successivi al layer di Global Average Pooling. In particolare, sono stati aggiunti:

- un layer Dense con 64 unità e attivazione “ReLU”
- un layer di Dropout con una probabilità di 0.30 per ridurre l'overfitting. Consiste nel selezionare e disattivare in modo casuale alcuni neuroni durante la fase di addestramento.
- un layer Dense con attivazione “Sigmoid”, per ottenere la classificazione binaria.

La funzione di attivazione ReLU (Rectified Linear Unit) “rettifica” i valori negativi a zero, mantenendo invariati i valori positivi [eq.3].

$$f(z) = \begin{cases} 0, & z < 0 \\ z, & z \geq 0 \end{cases}$$

*Equazione 3 Funzione di attivazione ReLU*

La funzione di attivazione Sigmoid, invece, mappa qualsiasi input in un intervallo compreso tra 0 e 1, ottimale per le previsioni binarie [22], [eq.4].

$$\phi(z) = \frac{1}{1 + e^{-z}}$$

*Equazione 4 Funzione di attivazione Sigmoid*

#### • ResNet50

Così come per la ResNet18, anche per allenare la ResNet50 è stato ricavato dal 30% del Training set un subset utilizzato nella fase di validazione. Sono state applicate le stesse modifiche di Data Augmentation per cercare di ridurre lo sbilanciamento del dataset.

Anche in questo caso i dati del Training e del Validation set sono stati normalizzati tra 0 e 1, e allo stesso modo è stata definita la funzione di Early Stopping. Poiché per questo modello di rete esiste già la funzione in Keras che lo implementa [23], non è stato necessario farlo manualmente. La ResNet50, essendo più profonda e quindi con più layers, richiede una maggiore quantità di risorse sia computazionali che temporali per poterla allenare: si introduce quindi la tecnica del “Transfer Learning” [24]. Il transfer learning è un approccio nell'ambito del machine learning e dell'intelligenza artificiale in cui un modello di apprendimento, allenato su un determinato compito, viene riutilizzato come punto di partenza per un altro compito, simile o meno, riducendo così le risorse necessarie e migliorandone le prestazioni.

Inizialmente, si allena un modello di machine learning su un compito specifico utilizzando un ampio set di dati, magari più semplici da reperire. Vengono quindi conservate le conoscenze apprese durante l'allenamento, come ad esempio caratteristiche delle immagini, delle strutture dei dati o delle relazioni tra le etichette. Il modello preaddestrato viene successivamente allenato su un nuovo compito specifico o su un set di dati più piccolo relativo a un compito correlato. Durante questa fase, il modello adatta le sue conoscenze preesistenti per il nuovo compito, regolando i pesi e i parametri in base alle nuove informazioni.

In questo caso, la ResNet50 è stata inizializzata con i pesi pre-addestrati su ImageNet [25], un dataset composto da 14 milioni di immagini e 1000 classi. Sono stati rimossi gli ultimi layer dense (fully connected) e sono stati frozen i rimanenti layer convoluzionali. Sono stati aggiunti i seguenti layers:

- un layer di Global Average Pooling
- un layer Dense con 512 unità, attivazione “ReLU” e Regularizer di tipo L2
- un layer di Dropout con una probabilità di 0.80
- un layer Dense con 256 unità, attivazione “ReLU” e Regularizer di tipo L2
- un layer di Dropout con una probabilità di 0.80
- un layer Dense con attivazione “Sigmoid”, per ottenere la classificazione binaria.

Per ridurre l’overfitting in questo modello è stata introdotta, oltre al layer di Dropout, la Regularizzazione: è una tecnica che aggiunge un termine di penalità alla funzione di loss, riducendo i coefficienti grandi e limitandone l'entità, impedendo così che un modello diventi eccessivamente complesso e memorizzi i dati di addestramento invece di apprendere i modelli sottostanti [26]. La tecnica di regularizzazione utilizzata in questo modello è la L2, detta anche “Regression Ridge”: favorisce pesi più piccoli e distribuiti più uniformemente, aggiungendo una penalità basata sul quadrato dei coefficienti [eq. 5].

$$Loss_{reg} = Loss_{original} + \lambda \sum_{i=1}^n \omega_i^2$$

*Equazione 5 Regularization L2*

Dove  $Loss_{original}$  è la funzione di perdita originale,  $\lambda$  il parametro di regularizzazione, che in questo modello è stato impostato a 0.001, e  $\omega_i$  i pesi del modello. Più è grande  $\lambda$ , maggiore sarà la penalizzazione per i pesi e il modello è meno incline all’overfitting.

Le prestazioni migliori si sono ottenute con le seguenti combinazioni di iperparametri (tab. 2):

Optimizer	Learning rate	Funzione di Loss	Batch size	Epochs
SGD	0.0001	“Binary crossentropy”	32	200

*Tabella 2 Migliore combinazione iperparametri ResNet50*

La funzione di ottimizzazione “SGD”, ovvero “Stochastic Gradient Descent”, è una variante del Gradient Descent [27], che invece di utilizzare il dataset intero, seleziona un piccolo batch per calcolare il gradiente della funzione di loss e aggiornare i parametri del modello. Questo procedimento permette di introdurre casualità nel processo di ottimizzazione, da cui il nome “Stocastico”. Inoltre, utilizzando piccoli batch, gli aggiornamenti sono più frequenti e meno costosi a livello computazionale. I parametri vengono aggiornati in direzione opposta al gradiente, per minimizzare la perdita.

#### • VGG19

La rete VGG19 che è stata utilizzata sfruttando la tecnica del Transfer Learning, è stata implementata da Kather JN et al. nel loro studio [28], in cui dimostrano che un modello CNN può valutare il microambiente tumorale dell’uomo e prevedere la prognosi direttamente dalle immagini istologiche. Questo modello è stato allenato su un dataset composto da immagini istologiche che comprendevano i seguenti tessuti: tessuto adiposo, detriti, linfociti, muco, muscolo liscio, mucosa normale del colon, stroma associato al cancro, sfondo ed epitelio CRC, ed effettuava una classificazione in 9 classi. Quindi, per poterla utilizzare in questo studio in cui viene effettuata una classificazione binaria, sono stati modificati gli ultimi gli ultimi layer dense (fully connected) e sono stati freezeati i rimanenti layer convoluzionali:

- un layer di Dropout con una probabilità di 0.80
- un layer Dense con 512 unità, attivazione “ReLU” e Regularizer di tipo L2, con  $\lambda=0.01$
- un layer di Dropout con una probabilità di 0.80
- un layer Dense con attivazione “Sigmoid”, per ottenere la classificazione binaria.

In questo modello di VGG19, oltre alla struttura classica dei 16 layer convoluzionali e 3 fully connected, sono presenti dei layer non trainabili:

- i layer di Padding, che si occupano di mantenere le dimensioni degli input durante le convoluzioni;
- un layer di Normalizzazione, ad inizio modello, che sottrae la media dai dati di input e divide per la deviazione standard, calcolate lungo le dimensioni spaziali e il canale colore;
- un layer di Reshape che converte le immagini da 3 dimensioni ad un vettore ad una dimensione, prima dei layer fully connected.

Anche per allenare la VGG19 è stato ricavato dal 30% del Training set un subset utilizzato nella fase di validazione. Sono state applicate le stesse modifiche di Data Augmentation per



cercare di ridurre lo sbilanciamento del dataset. La funzione di Early Stopping è stata definita e allo stesso modo delle precedenti.

Le prestazioni migliori si sono ottenute con le seguenti combinazioni di iperparametri (Tab 3)

Optimizer	Learning rate	Funzione di Loss	Batch size	Epochs
Adam	0.0001	“Binary crossentropy”	64	100

Tabella 3 Migliore combinazione iperparametri VGG19

## Confronto prestazioni dei tre modelli

Come metrica di valutazione per le performance ottenute dai modelli di rete neurale, si è calcolata l'accuracy [29]: calcola il rapporto fra i corretti classificati, sia positivi che negativi, e il totale delle classificazioni [eq.6]

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

Equazione 6 Accuracy

Dove:

- TP sono i True Positive, ovvero i “veri positivi”,
- TN i True Negative, i “veri negativi”
- FP i False Positive, ovvero un elemento negativo classificato erroneamente come positivo
- FN i False Negative, ovvero un elemento positivo classificato erroneamente come negativo

Nel modello ResNet18 si sono ottenute le seguenti prestazioni (fig. 4) (Tab 4):

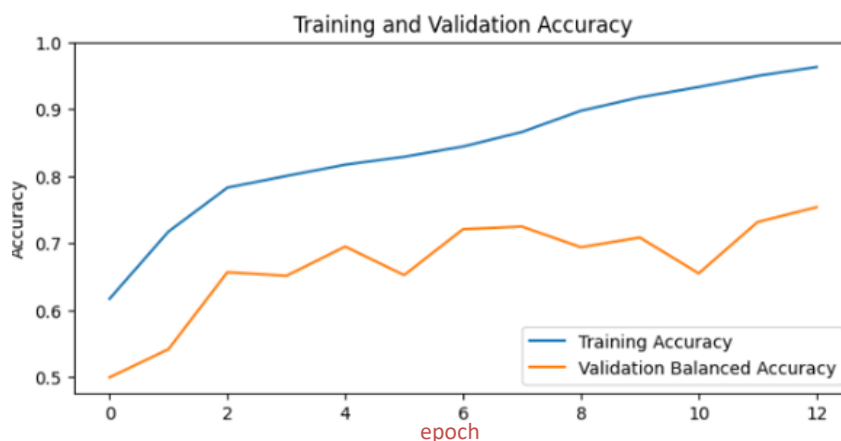


Figura 4 Andamento Prestazioni ResNet18

	Accuracy	Loss
Training	99%	0.02
Validation	78%	1.24

Tabella 4 Prestazioni della ResNet18



Poiché il dataset, nonostante il Data Augmentation applicato in fase di allenamento, risulta ancora essere sbilanciato, sono state calcolate anche la Balanced Accuracy e F1-Score nella fase di Validation (Tab. 5).

Validation Balanced Accuracy	Validation F1-score
74%	75%

Tabella 5 Balanced accuracy e F1 score della ResNet18

La Balanced Accuracy [30] calcola la media della somma tra la Sensitivity e Specificity [eq.7]:

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2}$$

Equazione 7 Balanced Accuracy

Dove la Sensitivity misura la percentuale dei positivi correttamente identificati rispetto a tutti gli esempi positivi reali [eq.8]:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

Equazione 8 Sensitivity

E la Specificity misura la percentuale dei negativi correttamente identificati rispetto a tutti gli esempi negativi reali [eq. 9]:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Equazione 9 Specificity

La metrica F1-score [eq. 10] bilancia “Precision” e “Recall” dove la Precision [eq. 11] calcola la percentuale dei positivi corretti rispetto a tutte le predizioni positive fatte dal modello, mentre la Recall [eq. 12] calcola la percentuale dei positivi correttamente identificati, come la Sensitivity.

$$F1 - score = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Equazione 10 F1 score

$$\text{Precision} = \frac{TP}{TP + FP}$$

Equazione 11 Precision

$$\text{Recall} = \frac{TP}{TP + FN}$$

Equazione 12 Recall

Nonostante nella fase di Training le prestazioni sembrano buone, nella fase di Testing, si è ottenuta una Balanced Accuracy del 50%. Come si può notare dall'andamento della funzione di loss nella fase di validazione (fig. 5), questo è altalenante, indice di scarsa generalizzabilità e che sta memorizzando i dati del training. Anche dai valori di loss finale (0.02 nel training e 1.24 nel validation) si può dedurre questa condizione, chiamata overfitting [31]. Poiché il dataset risulta essere fortemente sbilanciato, probabilmente il modello ha appreso meglio a identificare una classe rispetto ad un'altra.

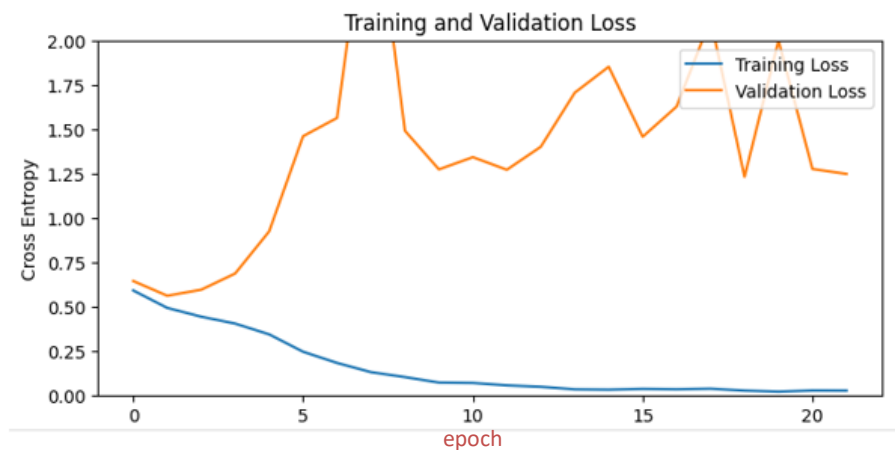


Figura 5 Andamento funzione di loss ResNet18

Anche nel modello ResNet50 si sono ottenuti dei risultati simili (fig. 6) (Tab. 6 e Tab. 7):

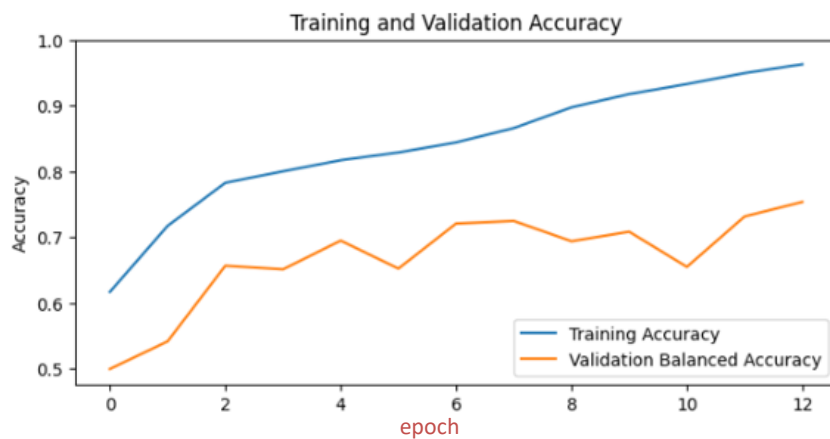


Figura 6 Andamento Prestazioni ResNet50

	Accuracy	Loss
Training	95%	0.51
Validation	78%	1.30

Tabella 6 Prestazioni della ResNet50

Validation Balanced Accuracy	Validation F1-score	Test Balanced Accuracy
75%	75%	50%

Tabella 7 Balanced Accuracy e F1 score della ResNet50

Così come per la ResNet18, anche per la ResNet50, sebbene in maniera meno pronunciata, il modello si adatta troppo ai dati di training, perdendo di generalizzazione. La funzione di loss (fig. 7) presenta oscillazioni, anche se meno altalenante rispetto alla ResNet18, e resta comunque alta.

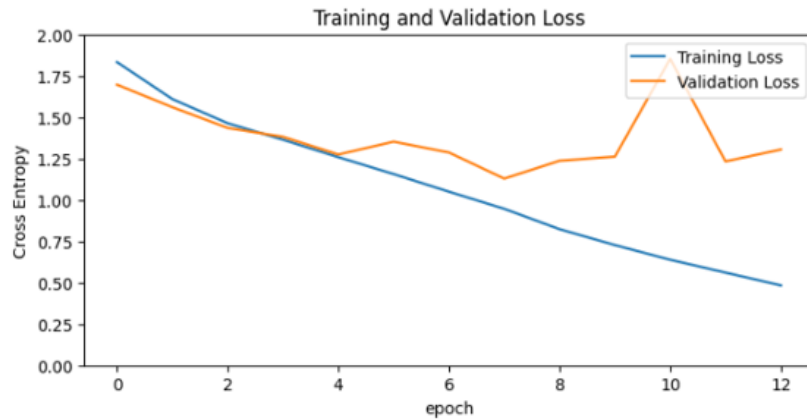


Figura 7 Andamento funzione di loss ResNet50

Il modello VGG19 pre-allenato, invece, ha dato dei risultati più performanti (fig. 8) (Tab. 8 e 9)

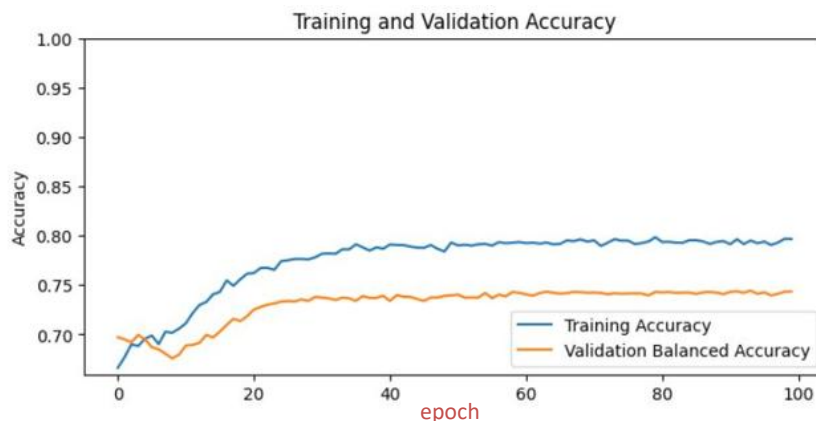


Figura 8 Andamento prestazioni VGG19

	Accuracy	Loss
Training	79%	0.57
Validation	81%	0.52

Tabella 8 Prestazioni della VGG19

Validation Balanced Accuracy	Validation F1-score	Test Balanced Accuracy
75%	75%	82%

Tabella 9 Balanced Accuracy e F1 score

L'accuracy del training non è eccessivamente alta, ciò implica che il modello non sta andando in overfitting. Anche le metriche di validazione sono abbastanza allineate con quelle di training,

indicando una buona generalizzazione. L'andamento della funzione di loss (fig. 9) decresce stabilmente, senza segni di instabilità, sia durante il training che il validation.

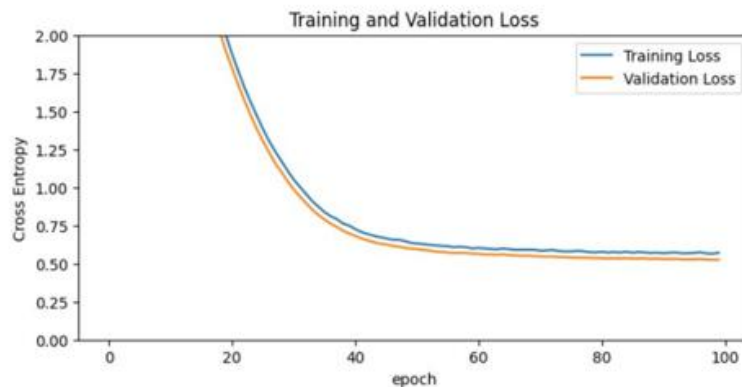


Figura 9 Andamento funzione di loss VGG19

Il fatto che la Balanced Accuracy (74%) è leggermente più bassa rispetto alla Validation Accuracy (81%) suggerisce che questo modello potrebbe avere difficoltà con classi meno rappresentate.

## Estrazione delle feature

Nella seconda fase del lavoro, sono stati aggiunti 116 nuovi pazienti, provenienti da diversi centri, tra cui l'ospedale Niguarda, l'ospedale Papa Giovanni XXIII, l'Istituto Oncologico Europeo e l'Istituto Oncologico Veneto. Dopo una prima fase di pulitura delle immagini, che risultavano macchiate da pennarelli, sono state estratte delle tiles con le stesse dimensioni di quelle della prima fase (224x224 pixel) e stessa risoluzione (0.5  $\mu\text{m}/\text{pixel}$ ). Dopo aver effettuato la color deconvolution sono state date in input alla rete VGG19, con lo scopo di segmentare automaticamente i tessuti tumorali, in assenza di quelle manuali. Si è deciso di ricavare le feature da queste tiles in quanto le WSI istologiche contenevano aree eterogenee (tessuto sano, sfondo, vari stadi tumorali e artefatti) e dividerle in tiles di piccole dimensioni avrebbe permesso di ottenere porzioni più omogenee, riducendo così il rumore dovuto alla variabilità e ottenendo features più rappresentative.

Si sono quindi ottenute 93 features tramite la libreria *pyradiomics* [32], un pacchetto open-source scritto in linguaggio python che permette, appunto, l'estrazione di features radiomiche da immagini mediche.

Le feature sono suddivise nel seguente modo:

- 18 features **First order**: 10th percentile, 90th percentile, energy, entropy, interquartile range, kurtosis, minimum, maximum, mean, median, range, mean absolute deviation, robust mean absolute deviation, root mean squared, skewness, total energy, variance e uniformity
- 24 features **Gray Level Cooccurrence Matrix (GLCM)**: autocorrelation, cluster prominence, cluster shade, cluster tendency, contrast, correlation, difference average, difference entropy, difference variance, inverse difference (Id), inverse difference

moment (Idm), inverse difference moment normalized (Idmn), inverse difference normalized (Idn), informational measure of correlation (Imc1), informational measure of correlation (Imc2), inverse variance, joint average, joint energy, joint entropy, maximal correlation coefficient (MCC), maximum probability, sum average, sum entropy e sum squares

- 16 features **Gray Level Run Length Matrix (GLRLM)**: gray level non uniformity, gray level non uniformity normalized, gray level variance, high gray level run emphasis, long run emphasis, long run high gray level emphasis, long run low gray level emphasis, low gray level run emphasis, run entropy, run length non uniformity, run length non uniformity normalized, run percentage, run variance, short run emphasis, short run high gray level emphasis e short run low gray level emphasis;
- 16 features **Gray Level Size Zone Matrix (GLSZM)**: gray level non uniformity, gray level non uniformity normalized, gray level variance, high gray level zone emphasis, large area emphasis, large area high gray level emphasis, large area low gray level emphasis, low gray level zone emphasis, size zone non uniformity, size zone non uniformity normalized, small area emphasis, small area high gray level emphasis, small area low gray level emphasis, zone entropy, zone percentage e zone variance;
- 5 features **Neighbouring Gray Tone Difference Matrix (NGTDM)**: Busyness, Coarseness, Complexity, Contrast e Strength
- 14 features **Gray Level Dependence Matrix (GLDM)**: dependence entropy, dependence non uniformity, dependence non uniformity normalized, dependence variance, gray level non uniformity, gray level variance, high gray level emphasis, large dependence emphasis, large dependence high gray level emphasis, large dependence low gray level emphasis, low gray level emphasis, small dependence emphasis, small dependence high gray level emphasis e small dependence low gray level emphasis.

Per estrarre le feature, le immagini sono state discretizzate utilizzando un numero fisso di bins ( $n=32$ ), sono state create le maschere delle tiles tumorali che sono state risegmentate tra il 1° e il 99° percentile per rimuovere gli outlier, ed è stato forzato il calcolo delle feature su ogni singola tiles (force2D).

## Creazione del dataset per feature selection

Prima di procedere alla fase di selezione delle feature, è stato necessario aggregare le tiles tumorali per ottenere una rappresentazione globale del paziente. Sono stati proposti due metodi di aggregazione che si basano su statistiche robuste, per mantenere la variabilità e la distribuzione delle feature a livello di paziente:

- Calcolo della media per ciascuna feature, ottenendo 137 pazienti e 93 feature
- Calcolo delle statistiche descrittive dell'istogramma per ciascuna feature: Media, Mediana, Skewness, Kurtosis, Percentili, range interquartili (IQR) ed Entropia; in totale si sono ottenute 744 features per 137 pazienti

La normalizzazione è stata effettuata precedentemente alla fase di aggregazione, calcolando i massimi e i minimi delle feature con il metodo del “Min-Max Scaler” [33], tra le tiles di ciascun paziente, per garantire una rappresentazione uniforme delle caratteristiche intra-paziente.

Si sono quindi ottenuti 137 pazienti, di cui 66 appartenenti alla classe 0 (Not responder alla terapia) e 71 appartenenti alla classe 1 (Responder alla terapia), con la seguente struttura riportata in tabella (Tab. 10):

Paziente ID	93 features					Label

Tabella 10 Struttura dataset

Il dataset così ottenuto è stato suddiviso un 80% Construction set e un 20% Validation set, rispettando un bilanciamento delle classi:

- Construction set: 53 soggetti not responder (classe 0) e 56 soggetti responder (classe 1)
- Validation set: 13 soggetti not responder e 15 soggetti responder.

Il Construction set a sua volta è stato suddiviso con proporzione 75-25% per la creazione del Training e Test set:

- Training set: 39 soggetti not responder e 42 soggetti responder
- Test set: 14 soggetti not responder e 14 soggetti responder.

I pazienti in questo nuovo dataset, come mostrato in figura (fig. 10), risultano più bilanciati rispetto al dataset della prima fase.

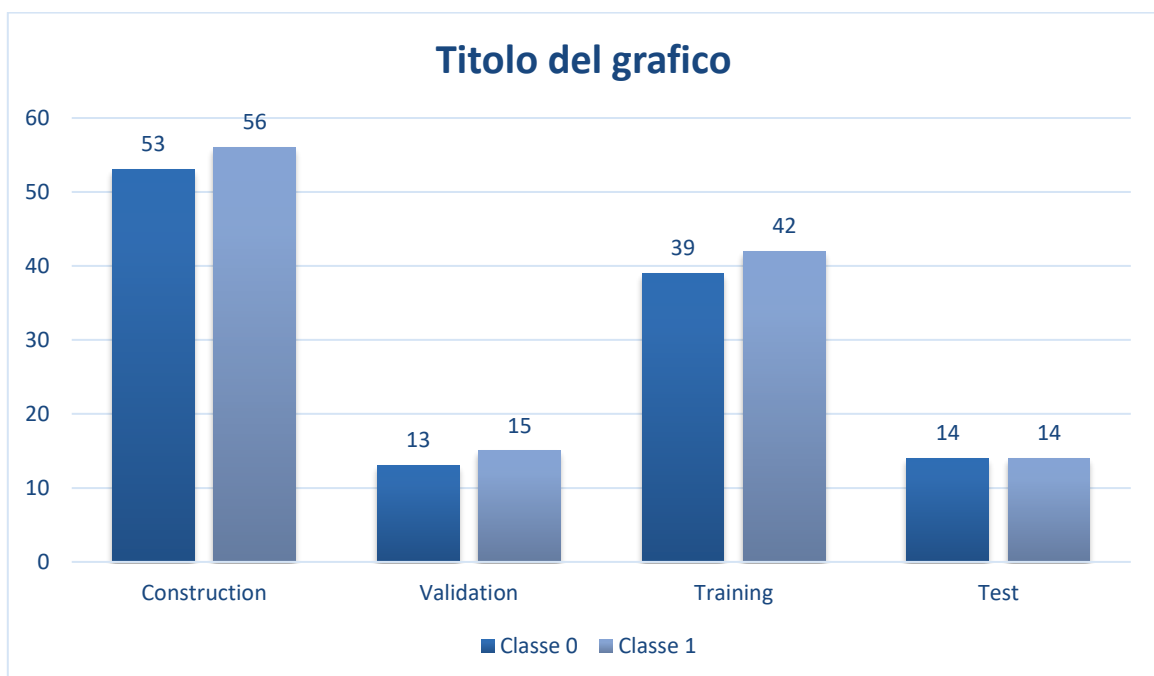


Figura 10 Suddivisione dataset

## Feature selection e costruzione dei classificatori

La feature selection è un processo molto importante per i problemi di classificazione nel machine learning, in quanto l'elevata dimensionalità dei dataset, spesso data da feature ridondanti, porta il modello incontro alla “curse of dimensionality” [34]:

- Maggiore complessità computazionale
- Tempi di addestramento più lunghi
- Aumento rischio di overfitting
- Riduzione capacità di generalizzazione.

La feature selection si occupa quindi di selezionare un subset di feature che contribuisca maggiormente nella predizione, riducendo in questo modo la dimensionalità del dataset e migliorando la prestazione dei classificatori [35].

I metodi di feature selection si suddividono [36] in:

- Filter: la selezione delle feature avviene in maniera indipendente dall'algoritmo di machine learning che viene utilizzato e si basa sulle caratteristiche delle feature stesse. Sono metodi rapidi ed economici per quanto riguarda il costo computazionale, però ignorano totalmente gli effetti che il subset di feature potrebbe avere sulle prestazioni del classificatore.
- Wrapper: questi metodi utilizzano un classificatore che valuta la qualità di un subset di feature in base alle prestazioni del classificatore stesso. È un procedimento che si ripete, suddividendo il dataset in più subset di feature, finché non si identifica quello che permette al classificatore di ottenere delle buone prestazioni. Nonostante si ottenga un'accuratezza predittiva migliore rispetto alle feature selezionate con i metodi Filter, risultano molto più costosi computazionalmente e comunque influenzati dal classificatore utilizzato per selezionarle: non è detto che risultino ottimali anche per altri classificatori.
- Embedded: la feature selection avviene durante l'esecuzione dell'algoritmo di apprendimento, sfruttando così il vantaggio dei metodi Wrapper, ovvero l'interazione delle feature con il classificatore, e il vantaggio dei metodi Filter, una computazione meno costosa.

I metodi di feature selection che si sono impiegati in questo studio sono stati:

- RFE (Recursive Feature Elimination)
- Minimum Redundancy - Maximum Relevance (mRmR)
- Affinity propagation
- Correlazione assoluta tra features e target
- Analisi delle componenti principali (PCA)

## RFE (Recursive Feature Elimination)

Il metodo RFE è un metodo Wrapper di feature selection: utilizza un modello di machine learning (in questo caso si utilizza un classificatore di tipo albero decisionale, il “Random Forest”) per stimare la stabilità di ogni feature [37]. Questo algoritmo rimuove in maniera ricorsiva le feature con il valore di ‘importanza’ più basso, finché non si ottiene il dataset composto da quelle più rilevanti. Per importanza delle feature si intende la misura in cui ciascuna feature influenza le previsioni di un modello. Nel caso del Random Forest l’importanza delle feature viene calcolata basandosi sulla riduzione dell’impurità, o sulla diminuzione dell’errore quando una feature è utilizzata per suddividere i nodi. L’impurità, quindi, misura l’omogeneità di un gruppo di elementi rispetto le due classi. Viene calcolata tramite l’indice di Gini [eq. 13]:

$$Gini\ Index = 1 - \sum_{i=1}^C p_i^2$$

*Equazione 13 Gini Index*

Dove C è il numero delle classi e  $p_i$  la proporzione delle feature appartenenti alla classe i-esima nel nodo corrente [38]. Più il valore dell’indice è vicino allo zero, più campioni appartengono alla stessa classe.

## Minimum Redundancy - Maximum Relevance (mRmR)

Il metodo Minimum Redundancy - Maximum Relevance (mRmR) seleziona le feature che hanno un’alta correlazione con la classe, quindi un’elevata rilevanza, e una bassa correlazione tra di loro, rimuovendo quelle altamente ridondanti [41]. Come misura di rilevanza è stata utilizzata la **mutua informazione** (Mutual Information) [42], mentre per la misura di ridondanza è stata utilizzata la mutua informazione media. La mutua informazione di due feature è la misura della dipendenza reciproca tra le due feature stesse, ovvero misura la quantità di ‘informazioni’ su una variabile ottenute osservando l’altra variabile. È definita con la seguente equazione [eq.14]:

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

*Equazione 14 Mutual Information*

Dove:

- $p(x,y)$  è la probabilità congiunta di  $X=x$  e  $Y=y$
- $p(x)$  è la probabilità marginale di  $X=x$ , dove la probabilità marginale sta per il valore associato alla probabilità di un singolo evento o di una singola variabile, ignorando ogni altra variabile o condizione nel sistema
- $p(y)$  è la probabilità marginale di  $Y=y$



La Mutual Information si calcola per valori discreti, di conseguenza è stato necessario discretizzare le feature. Per fare ciò si è utilizzato l’algoritmo di discretizzazione ChiMerge.

L’algoritmo ChiMerge, introdotto da Kerber [43], è un metodo supervisionato che suddivide il range dei valori continui in intervalli, basato su un approccio statistico che utilizza il test del chi-quadrato ( $\chi^2$ ) per determinare i confini degli intervalli. L’equazione per calcolare  $\chi^2$  è [eq.15]:

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(A_{ij} - E_{ij})^2}{E_{ij}}$$

*Equazione 15 Formula chi-quadrato*

Dove:

- $m = 2$ , gli intervalli che vengono confrontati
- $k$  = numero delle classi
- $A_{ij}$  = numero di elementi nell’intervallo  $i$ -esimo,  $j$ -esima classe
- $R_i$  = numero di elementi nell’intervallo  $i$ -esimo =  $\sum_{j=1}^k A_{ij}$
- $C_j$  = numero di elementi nella  $j$ -esima classe =  $\sum_{i=1}^m A_{ij}$
- $N$  = numero totale di elementi =  $\sum_{j=1}^k C_j$
- $E_{ij}$  = frequenza attesa di  $A_{ij} = (R_i \cdot C_j) / N$

L’algoritmo valuta una variabile alla volta, calcolando  $\chi^2$  per ciascuna coppia di intervalli adiacenti: lo scopo è verificare se la distribuzione delle classi degli elementi nei due intervalli sia indipendente dai valori assunti dalla variabile per quegli stessi elementi. Gli intervalli che presentano il valore minimo di  $\chi^2$  vengono fusi.

## Affinity Propagation

L’affinity propagation, più che un metodo di feature selection, è un algoritmo di clustering che si occupa di raggruppare i dati simili tra loro. A differenza dei tradizionali non è necessario specificare il numero di cluster sin dall’inizio. Si basa sullo scambio di “messaggi” tra i data points per identificare i centri dei cluster, ovvero dei campioni del dataset che rappresentano il più possibile i dati del dataset, e l’assegnazione al cluster degli altri dati che “somigliano” suddetto centro. Per determinare la similarità tra i punti dati viene calcolato il “punteggio di similarità”, ovvero la distanza euclidea negativa tra i data points. Mettendo insieme i punteggi di similarità calcolati per ogni coppia di dati si ottiene la matrice di similarità  $S$  [eq.16]:

$$S(i, k) = - || x(i) - x(k) ||^2$$

*Equazione 16 Matrice di similarità*

Dove  $i$  e  $k$  corrispondono alla coppia di data points.

I “messaggi” scambiati tra i data points riguardano, invece, la “Responsabilità” e la “Disponibilità”:

- La Responsabilità (R) è la matrice che indica quanto ogni data point è adatto ad essere l'esemplare, ovvero il centro del cluster, per un altro data point [eq.17].
- La Disponibilità (A) è la matrice che indica quanto ogni data point è disponibile ad essere l'esemplare per un altro data point. Questo valore tiene conto delle responsabilità di tutti gli altri punti rispetto al k-esimo punto, per stabilire se k può essere o meno un esemplare affidabile [eq.18].

$$R(i, k) = S(i, k) - \max(A(i, k') + S(i, k')) \forall k' \neq k$$

*Equazione 17 Matrice di responsabilità*

$$A(i, k) = \min\left(0, R(k, k) + \sum \max(0, R(i', k))\right) \forall i' \neq i, k$$

$$A(k, k) = \sum \max(0, R(i', k)) \forall i' \neq i, k$$

*Equazione 18 Matrice di disponibilità*

Dove:

- $A(i, k)$  calcola il valore di disponibilità per uno specifico data point  $i$  e un potenziale esemplare  $k$ .
- $R(k, k)$  rappresenta la responsabilità dell'ipotetico esemplare  $k$  verso se stesso
- $\max(0, R(i', k)) \forall i' \neq i, k$  appresenta la "responsabilità" degli altri data point  $i'$  verso l'ipotetico esemplare  $k$ .

Le equazioni vengono applicate iterativamente fino alla convergenza, ovvero fino a quando i valori nelle matrici non si stabilizzano [43]. In questo studio, l'Affinity Propagation è stato usato per fare il clustering delle feature, con lo scopo di ridurre la dimensionalità ed effettuare una selezione delle feature.

## Correlazione assoluta tra features e target

In questo metodo di feature selection è stata calcolata la correlazione assoluta tra ciascuna feature del dataset e la variabile target. Essendo feature continue, è stata calcolata la correlazione di Pearson [eq. 19]:

$$r = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y}$$

*Equazione 19 Correlazione di Pearson*

Dove:

- $Cov(X, Y)$  calcola la covarianza tra  $X$  e  $Y$
- $\sigma_X, \sigma_Y$  rappresentano la deviazione standard di  $X$  e  $Y$

$R$  può assumere valori compresi tra -1 e 1, che indicano rispettivamente che la relazione è linearmente negativa perfetta o linearmente positiva perfetta.

## Analisi delle componenti principali (PCA)

L'analisi delle componenti principali [44], più che una tecnica di features selection, è una tecnica di riduzione della dimensionalità nell'ambito della statistica multivariata. Dalle feature originarie si sono ottenute delle feature "artificiali" che permettono di ridurre lo spazio delle dimensioni, minimizzando la perdita d'informazione. Queste feature artificiali, chiamate **componenti principali**, rappresentano combinazioni lineari delle features originali nello spazio dei dati che massimizzano la varianza. Il primo step consiste nel calcolare la matrice di covarianza o di correlazione tra le variabili originali. Da questa matrice si ottengono gli autovalori, che indicano quanta variabilità viene spiegata da ogni componente, e gli autovettori che indicano la direzione delle componenti principali nel nuovo spazio. Questi autovettori forniscono i pesi, che sono calcolati considerando che le componenti principali debbano essere ortogonali, quindi non correlate tra di loro, e in modo tale che la prima componente che viene costruita contenga la massima variabilità possibile, la seconda componente contenga la massima variabilità residua e così via. In questo studio si sono mantenute le componenti che spiegavano il 90% della variabilità totale.

Per valutare i nuovi dataset ottenuti dalle feature selection sopracitate, sono stati utilizzati i seguenti classificatori:

- Naïve Bayes
- K-Nearest Neighbours (KNN)
- Support Vector Machine (SVM)
- Decision Tree (DT)
- Bagging Decision Tree
- Boosting Decision Tree (AdaBoost e Gradient Boosting)
- Random Forest
- Stochastic Gradient Descent (SGD) Classifier

## Naïve Bayes

Il classificatore Naïve Bayes è un algoritmo di apprendimento supervisionato basato sul teorema di Bayes. Il teorema di Bayes [45] viene utilizzato per determinare la probabilità condizionata di un evento che si è già verificato [eq. 21]:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

*Equazione 20 Teorema di Bayes*

Dove:

- $P(A)$  e  $P(B)$  sono le probabilità degli eventi A e B
- $P(A|B)$  è la probabilità dell'evento A quando si verifica l'evento B
- $P(B|A)$  è la probabilità dell'evento B quando si verifica A

Nel caso della classificazione binaria, il Naïve Bayes predice la probabilità che quel determinato campione appartenga o meno a quella determinata classe. È chiamato "naïve" (ingenuo) perché assume che tutte le feature siano indipendenti tra loro, un'ipotesi spesso non vera ma che rende l'algoritmo semplice ed efficace, anche in piccoli dataset.

Esistono diverse varianti di Naïve Bayes, a seconda del tipo di dati utilizzati nel dataset. In questo studio si utilizzano il Gaussian Naïve Bayes, adatto per dati numerici continui con distribuzione normale e il Multinomial Naïve Bayes, utilizzato per dati discreti.

## K-Nearest Neighbours (KNN)

Il classificatore K-Nearest Neighbors (KNN) è un algoritmo di apprendimento supervisionato che si basa sul principio che dati simili tendono ad essere vicini nello spazio delle feature [46]. L'algoritmo, quindi, assume che i punti più vicini a un'osservazione siano quelli che probabilmente condividono la stessa classe o un valore simile. Per determinare quali sono i «k» vicini più vicini, il KNN calcola la distanza tra il punto da classificare e tutti gli altri punti nel dataset e li ordina in base alla distanza calcolata.

Le distanze più comuni sono:

- Distanza Euclidea: Misura la distanza "lineare" nello spazio [eq. 22].

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

*Equazione 21 Distanza euclidea*

- Distanza Manhattan: Misura la distanza in termini assoluti [eq. 23].

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |x_i - y_i|$$

*Equazione 22 Distanza Manhattan*

- Distanza Minkowski: Una generalizzazione delle due precedenti, in cui il parametro p permette l'utilizzo di altre metriche di distanza [eq. 24].

$$d(\mathbf{x}, \mathbf{y}) = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

*Equazione 23 Distanza Minkowski*

La classificazione con questo modello dipende fortemente dalla scelta del valore k, in quanto, se il valore è troppo piccolo, renderà il modello più sensibile al rumore e quindi diminuirà l'accuratezza; se troppo grande si introdurrebbe bias dovuto all'alto numero di punti. Con questo metodo, la classe viene assegnata tenendo conto della classificazione più frequentemente rappresentata tra i k vicini più vicini, usando il Majority Voting [47].

Per ottimizzare la ricerca della combinazione di iperparametri che diano le prestazioni migliori, è stata utilizzata la funzione GridSearchCV, fornita dalla libreria Scikit-learn. Da un “dizionario” contenente i parametri del modello che si vogliono ottimizzare e i loro possibili valori, la funzione GridSearchCV effettua una ricerca esaustiva della migliore combinazione, eseguendo una validazione incrociata che permette di ottenere le migliori prestazioni.

## Support Vector Machine (SVM)

Il Support Vector Machine è un algoritmo di apprendimento supervisionato che classifica i dati trovando una linea o un iperpiano ottimale che massimizza la distanza tra ciascuna classe in uno spazio N-dimensionale. Il numero di dimensioni è dato dal numero di feature nei dati di input. Poiché esistono diversi iperpiani in grado di separare le classi, l’algoritmo cerca di massimizzare il margine tra i punti, garantendo così il miglior confine decisionale tra le classi. Se i dati non sono linearmente separabili, vengono utilizzate funzioni Kernel per proiettare i dati in uno spazio multidimensionale, dove diventa possibile la separazione lineare. I Kernel calcolano i prodotti scalari tra i vettori nello spazio trasformato, permettendo di lavorare in spazi a dimensione superiore in modo efficiente [48].

Tipi di Kernel utilizzati:

- Lineare: per dati linearmente separabili [eq. 25]

$$K(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$$

*Equazione 24 Kernel lineare*

- Polinomiale: per relazioni più complesse [eq. 26]

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + c)^d$$

*Equazione 25 Kernel Polinomiale*

- Radial Basis Function (RBF): per separare dati non linearmente separabili [eq. 27]

$$K(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x} - \mathbf{y}\|^2}$$

*Equazione 26 Kernel Radial Basis Function*

- Sigmoid: riprende il funzionamento di una funzione di attivazione nelle reti neurali [eq.28].

$$K(\mathbf{x}, \mathbf{y}) = \tanh(\mathbf{x} \cdot \mathbf{y} + c)$$

*Equazione 27 Kernel Sigmoid*

L’efficienza in spazi ad alta dimensionalità lo rende computazionalmente costoso, ma robusto all’overfitting per valori dei parametri ben selezionati. Sfruttando la funzione GridSearchCV sono state testate le combinazioni di Kernel con gli iperparametri, per ottenere le prestazioni migliori.

## Decision Tree

Il Decision Tree è un algoritmo di apprendimento supervisionato, caratterizzato da una struttura ad albero [49]: il dataset, detto anche “nodo radice”, viene suddiviso ripetutamente in sottogruppi detti «nodi decisionali», in base ad una regola decisionale: in questo caso è la feature e il valore di soglia che minimizzano l’impurità, ovvero, la probabilità che un punto scelto a caso sia classificato in modo errato; i “rami” rappresentano il risultato di queste decisioni, mentre le “foglie” rappresentano i risultati finali o le decisioni (fig. 11). La costruzione dell’albero termina quando tutti i dati in un nodo sono «puri», cioè, appartengono alla stessa classe, o secondo altri criteri di stop. Anche in questo caso, l’impurità può essere calcolata attraverso lo *Gini Index* di cui si è discusso sopra o tramite l’equazione dell’Entropia [eq. 29].

$$Entropia = - \sum_{i=1}^c p_i \log_2(p_i)$$

Equazione 28 Entropia

Anche in questo caso è stata sfruttata la funzione GridSearchCV per provare le diverse combinazioni di profondità e di numero di split.

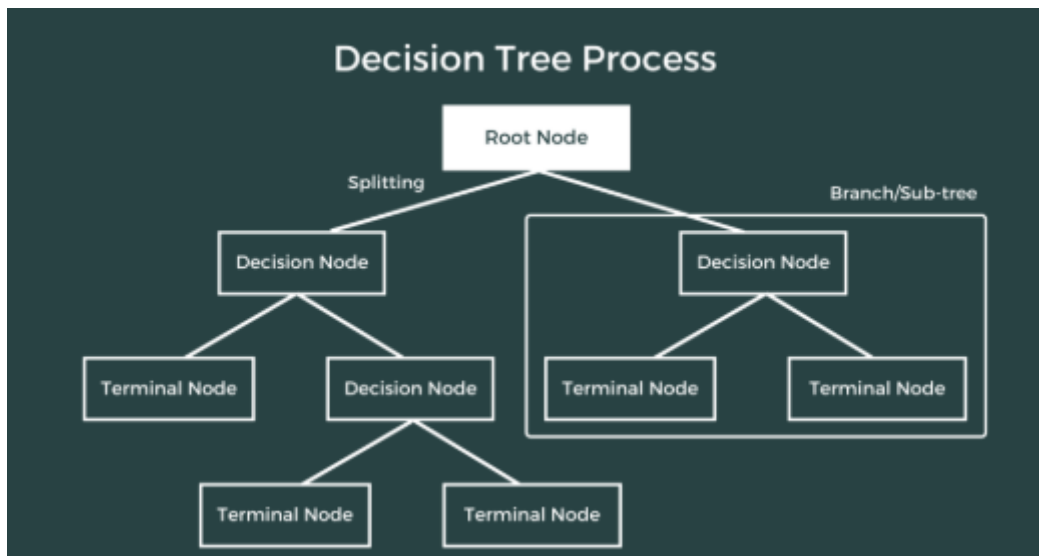


Figura 11 Struttura del Decision Tree

## Ensamble learning models

Gli ensemble learning models, o modelli di apprendimento d’insieme, sono gruppi di modelli, uguali o diversi, che vengono allenati sullo stesso dataset, per poi combinarne le predizioni. Lo scopo è quello di ottenere migliori prestazioni rispetto al singolo modello [50]. Gli ensemble learning models utilizzati in questo studio sono:

- Bagging Decision Tree
- Boosting Decision Tree

Il Bagging Decision Tree combina più alberi decisionali per migliorare la robustezza e la precisione del modello. In particolare, vengono combinate le predizioni ottenute dai modelli su subset campionati con sostituzione casuale (bootstrap). La predizione finale è la media delle predizioni di tutti gli alberi. Si concentra principalmente sulla riduzione dell'overfitting.

Il Boosting Decision Tree è una tecnica che combina più alberi decisionali deboli, ovvero poco profondi e semplici, per migliorare la robustezza e la precisione del modello (riduzione del bias). In questa tecnica gli alberi non sono indipendenti ma costruiti in sequenza, e ciascun albero cerca di correggere gli errori commessi negli alberi precedenti, assegnando un peso maggiore agli esempi più difficili, in modo da venire attenzionati nel modello successivo.

In questo studio sono stati utilizzati l'algoritmo AdaBoost (Adaptive Boosting) e il Gradient Boosting. La differenza principale tra questi due classificatori è che nel primo metodo viene aumentato il peso degli esempi che vengono classificati male, mentre nel secondo metodo non vengono usati pesi ma si minimizza una funzione di perdita.

## Random Forest

Il Random Forest è un algoritmo di apprendimento particolare del metodo «Bagging» che, oltre a combinare l'utilizzo di più alberi decisionali (fig. 12), costruiti utilizzando sottoinsiemi casuali del dataset, seleziona casualmente sottoinsiemi di features a ogni nodo, aumentando la decorrelazione tra gli alberi. In questo modo si aumenta la robustezza e si riduce l'overfitting. Nella fase di previsione, l'algoritmo aggrega i risultati tramite «votazione», ovvero basandosi sulla frequenza di quella determinata previsione [51].

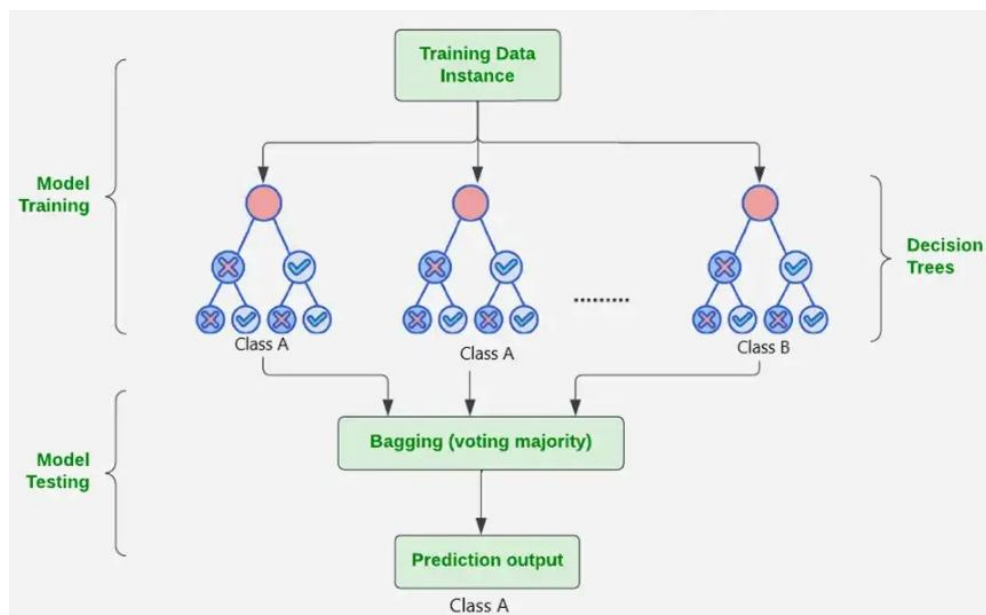


Figura 12 Struttura Random Forest

## Stochastic Gradient Descent (SGD) Classifier

Il classificatore Stochastic Gradient Descent (SGD) costruisce un confine lineare per separare le classi, ovvero assume che la relazione tra le feature e la variabile target possa essere approssimata da una linea retta o da un iperpiano. Il classificatore è definito “stocastico” perché nella fase di training vengono utilizzati mini-batch, o sottoinsiemi casuali dei dati, per introdurre casualità e per calcolare il gradiente della funzione di loss e aggiornare i parametri del modello ad ogni iterazione. L’uso di mini-batch permette di velocizzare l’algoritmo, in quanto, la funzione di loss non viene calcolata su tutto il dataset ma appunto su piccoli subset e l’introduzione della casualità permette al modello di sfuggire ai “minimi locali”: ovvero, quei punti in cui la funzione sembra minima ma non a livello globale [27].

## Risultati

Dopo aver discusso i metodi di feature selection e i classificatori utilizzati, si mostrano in questo paragrafo i risultati ottenuti per entrambi i metodi di aggregazione citati prima.

### Metodo di aggregazione: calcolo della media di ciascuna feature

#### 1. Feature selection: RFE (Recursive Feature Elimination)

Per questo metodo, come accennato nel paragrafo precedente, si è utilizzato come parametro *estimator* il classificatore Random Forest. Sono state effettuate 100 iterazioni, in cui i dati venivano ricampionati tramite il Bootstrap Sampling [39], per creare un dataset leggermente diverso ad ogni iterazione. L’algoritmo RFE ha selezionato iterativamente le 10 feature più rilevanti, valutandole tramite il Random Forest. Al termine delle iterazioni, si è calcolata la frequenza di selezione di ogni feature, mediando il numero di iterazioni in cui una feature è stata selezionata rispetto al numero totale delle iterazioni. Lo scopo era quello di individuare le feature rilevanti in modo robusto e consistente; perciò, sono state valutate le performance del modello Random forest con il dataset risultante, tramite una cross-validation [40], ovvero suddividendo i dati in più *fold* e utilizzando un subset come set di validazione e il resto come set di allenamento. Nella figura sottostante (fig.13) è possibile vedere la frequenza di selezione di ciascuna feature.



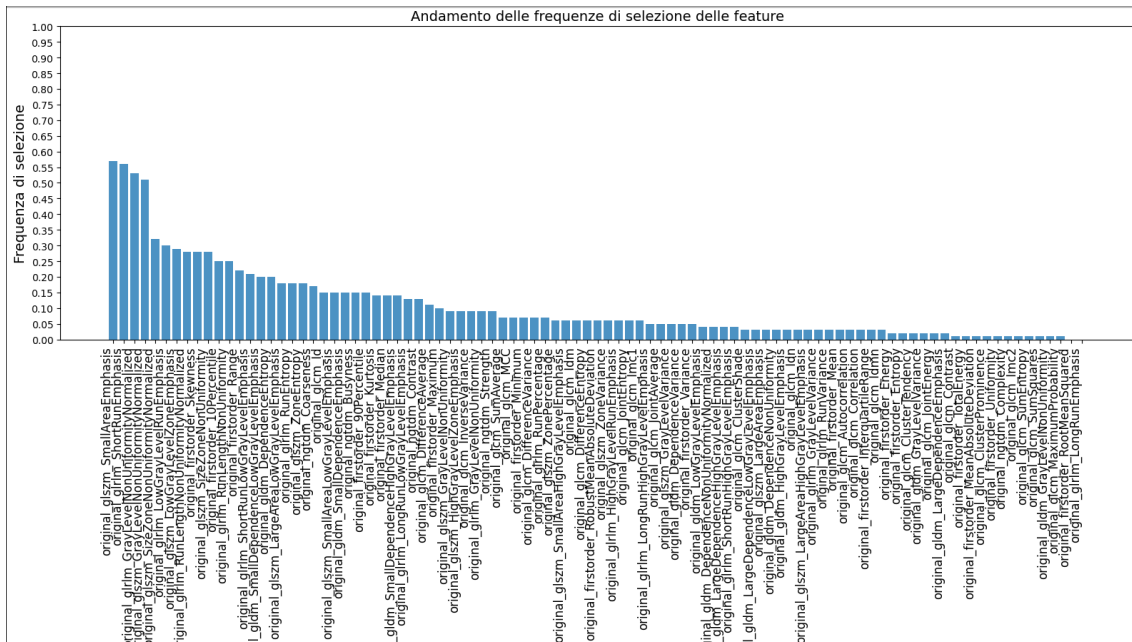


Figura 13 Frequenze di selezione delle feature

Dal grafico delle frequenze di selezione si può osservare un “gradino” intorno alla soglia 0.4, che è stata quindi selezionata per individuare le feature con maggiore consistenza. Si è ottenuto un dataset di quattro feature:

- original\_glszm\_SmallAreaEmphasis, Selection Frequency: 0.57
- original\_glrIm\_ShortRunEmphasis, Selection Frequency: 0.56
- original\_glszm\_GrayLevelNonUniformityNormalized, Selection Frequency: 0.53
- original\_glszm\_GrayLevelNonUniformity, Selection Frequency: 0.51

L’accuratezza è stata calcolata su 5 fold per rendere la stima delle prestazioni più stabile e meno soggetta alla casualità: [0.71 0.46 0.51 0.55 0.62] la cui media è pari al 57%.

In figura (fig.14) si possono osservare le prestazioni ottenute:

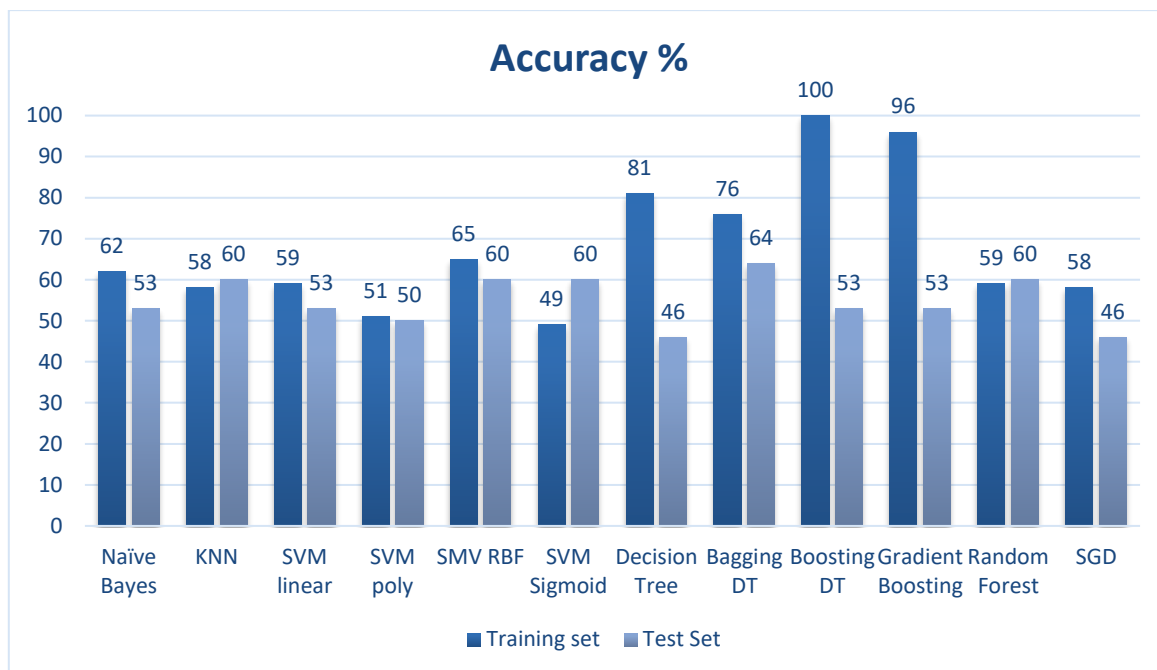


Figura 14 RFE 0.4

## Tuning dei parametri dei classificatori

Classificatore	Parametri Tuning
<i>Naïve Bayes</i>	Default
<i>K-Nearest Neighbours (KNN)</i>	algorithm: auto, n_neighbors: 13, p: 5, weights: distance
<i>SVM Kernel linear</i>	C: 10, gamma:1
<i>SVM Kernel polynomial</i>	C: 0.1, gamma:1
<i>SVM Kernel RBF</i>	C: 100, gamma:1
<i>SVM Kernel sigmoid</i>	C: 10, gamma:1
<i>Decision Tree (DT)</i>	Min_samples_split: 10, Max_depth:6
<i>Bagging Decision Tree</i>	N_estimators: 50, Max_features: 1, Max_samples: 0.5, Max_depth: 8, Min_samples_split:10
<i>Boosting Decision Tree (AdaBoost)</i>	N_estimators: 200, Max_depth:2, Learning_rate: 0.5
<i>Gradient Boosting</i>	N_estimators: 20
<i>Random Forest</i>	N_estimators: 50, Max_depth:None, min_samples_leaf: 1, min_samples_split:10, Max_features: sqrt
<i>Stochastic Gradient Descent (SGD)</i>	Default

Tabella 11 Tuning parametri RFE 0.4

Nella tabella a seguire (tab. 12) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

	Training Set	Test Set
Bagging DT		
Sensitivity %	79	57
Specificity %	74	71

PPV (Positive Predictive Value) %	77	67
NPV (Negative Predictive Value) %	76	62

Tabella 12 Metriche migliore classificatore

È stata valutata anche la soglia superiore a 0.25, ottenendo un dataset di 10 feature:

- original\_glszm\_SmallAreaEmphasis, Selection Frequency: 0.57
- original\_glrlm\_ShortRunEmphasis, Selection Frequency: 0.56
- original\_glrlm\_GrayLevelNonUniformityNormalized, Selection Frequency: 0.53
- original\_glszm\_GrayLevelNonUniformityNormalized, Selection Frequency: 0.51
- original\_glszm\_SizeZoneNonUniformityNormalized, Selection Frequency: 0.32
- original\_glrlm\_LowGrayLevelRunEmphasis, Selection Frequency: 0.3
- original\_glszm\_LowGrayLevelZoneEmphasis, Selection Frequency: 0.29
- original\_glrlm\_RunLengthNonUniformityNormalized, Selection Frequency: 0.28
- original\_firstorder\_Skewness, Selection Frequency: 0.28
- original\_glszm\_SizeZoneNonUniformity, Selection Frequency: 0.28

L'accuratezza ottenuta sui cinque fold è stata: [0.64 0.64 0.62 0.66 0.55], la cui media è pari al 62%. Nella figura (fig.15) sottostante si possono osservare le prestazioni ottenute:

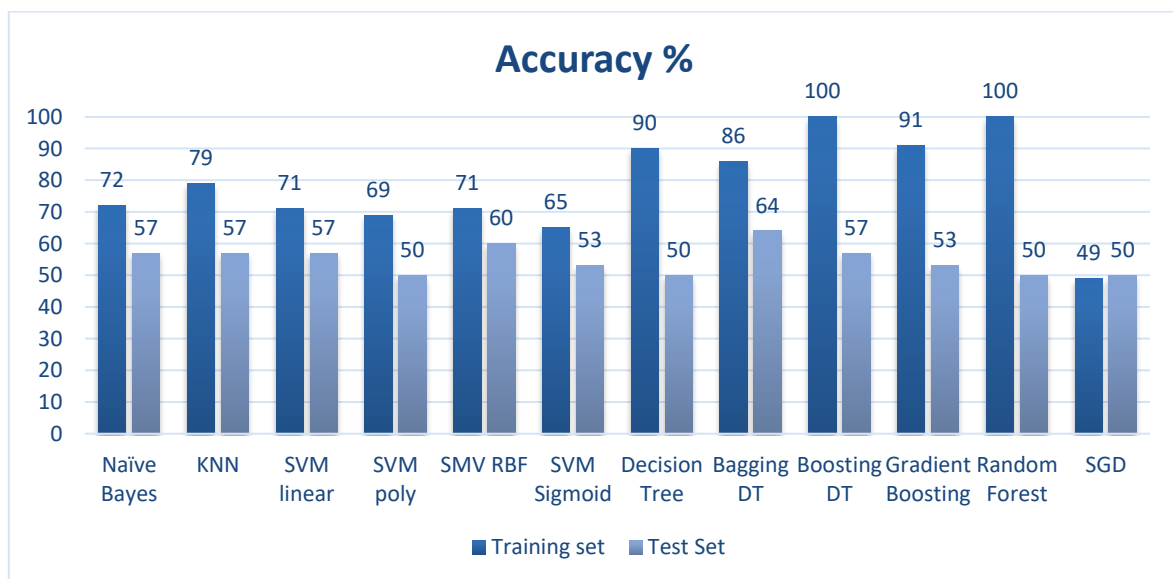


Figura 15 RFE 0.25

## Tuning dei parametri dei classificatori

Classificatore	Parametri Tuning
Naïve Bayes	Default
K-Nearest Neighbours (KNN)	algorithm: brute, leaf_size: 30, metric: minkowski, metric_params: None, n_jobs: -1, n_neighbors: 5, p: 2, weights: uniform
SVM Kernel linear	C: 1000, gamma: 1

<i>SVM Kernel polynomial</i>	C: 1, gamma: 1
<i>SVM Kernel RBF</i>	C: 1000, gamma:0.1
<i>SVM Kernel sigmoid</i>	C: 100, gamma: 0.1
<i>Decision Tree (DT)</i>	Min samples split: 5, Max depth:18
<i>Bagging Decision Tree</i>	N_estimators: 50, Max_features: 1, Max_samples: 0.5, Max_depth: 8, Min samples split:10
<i>Boosting Decision Tree (AdaBoost)</i>	N_estimators: 200, Max_depth:2, Learning rate: 0.5
<i>Gradient Boosting</i>	N_estimators: 20
<i>Random Forest</i>	N_estimators: 100, Max_depth:20 min_samples leaf: 5, min_samples_split:2 Max_features: sqrt
<i>Stochastic Gradient Descent (SGD)</i>	Default

Tabella 13 Tuning parametri RFE 0.25

Nella tabella a seguire (tab. 14) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

Bagging DT	Training Set	Test Set
Sensitivity %	86	64
Specificity %	87	64
PPV (Positive Predictive Value) %	88	64
NPV (Negative Predictive Value) %	85	64

Tabella 14 Metriche migliore classificatore

L'aumento del numero di feature ha portato ad un leggero miglioramento nelle prestazioni: in quasi tutti i modelli si può notare un aumento di accuratezza nel training, anche se ciò potrebbe essere dovuto ad overfitting, in quanto i valori di accuratezza nel test set rimangono circa identici, o presentano dei miglioramenti marginali. Ciò suggerisce che le feature aggiuntive potrebbero portare informazioni poco rilevanti

## 2. Feature selection: minimum Redundancy - maximum Relevance (mRmR)

Per questo metodo di feature selection è stato necessario discretizzare le feature, come detto in precedenza. Si è utilizzato l'algoritmo ChiMerge, la cui condizione di uscita è stata implementata in modo tale che si interrompa quando il numero di intervalli è pari a max\_intervals, impostato a priori. In questo studio il numero di max\_intervals è stato posto a 10, per catturare più dettagli e differenze nella distribuzione della feature.

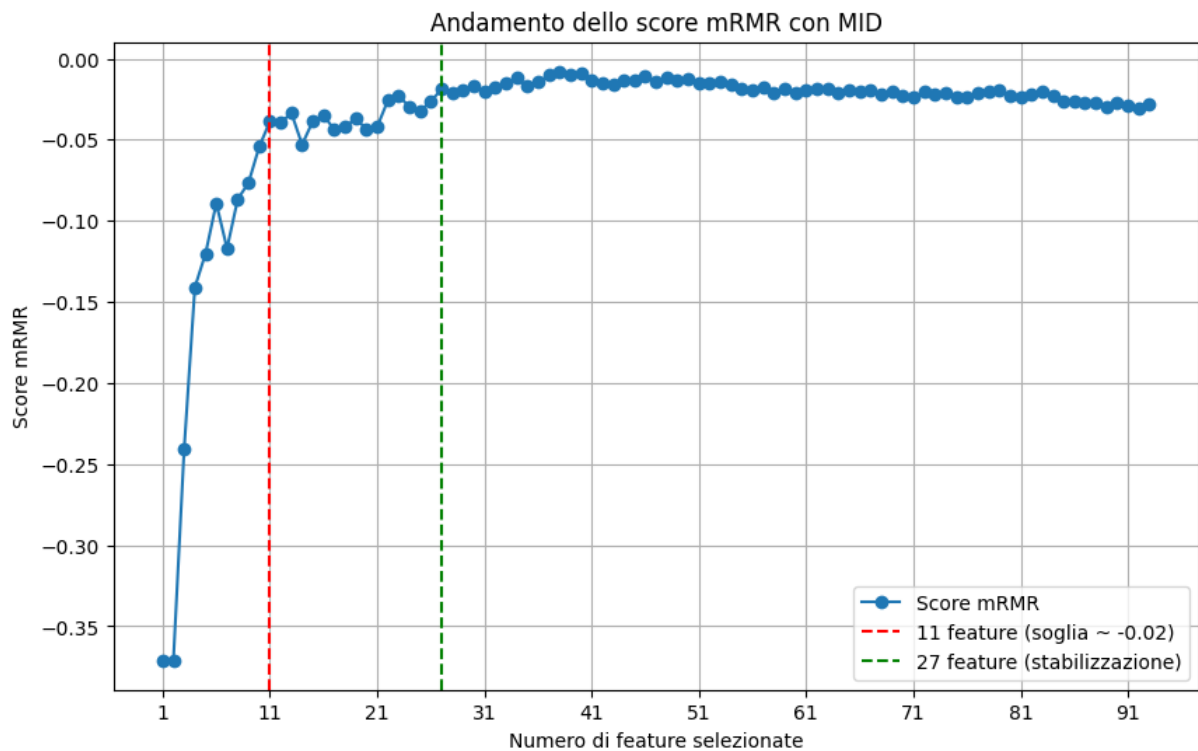


Figura 16 Andamento dello score mRMR con MID

Nella figura soprastante (fig.16) si può osservare l'andamento dello score MID (Mutual Information Difference), ovvero la differenza tra la rilevanza e la ridondanza calcolate dall'algoritmo mRMR per aggiungere ad ogni iterazione la feature più rilevante rispetto al target e meno ridondante tra quelle già selezionate. Dopo 27 feature si osserva una stabilizzazione dello score: ciò suggerisce che le feature successivamente aggiunte siano meno rilevanti rispetto al target o che introducano troppa ridondanza con le feature già selezionate. Il fatto che lo score rimanga negativo può essere dovuto al fatto che le feature del dataset sono molto correlate tra loro e quindi la ridondanza supera, anche se di poco, la rilevanza nel calcolo dello score MID. Si è voluto confrontare il dataset delle feature selezionate nella prima fase, quando principalmente sono presenti quelle altamente rilevanti, con il dataset di feature prima che l'andamento dello score si stabilizzi.

Nel primo dataset le feature sono undici e sono le seguenti:

- original\_firstorder\_90Percentile
- original\_firstorder\_Maximum
- original\_glcm\_ClusterShade
- original\_glcm\_DifferenceAverage
- original\_glrIm\_ShortRunLowGrayLevelEmphasis
- original\_glszm\_GrayLevelNonUniformityNormalized
- original\_glszm\_LargeAreaLowGrayLevelEmphasis
- original\_gldm\_DependenceEntropy
- original\_gldm\_DependenceVariance
- original\_gldm\_HighGrayLevelEmphasis

- original\_gldm\_SmallDependenceHighGrayLevelEmphasis

Nella figura (fig.17) si possono osservare le prestazioni ottenute:

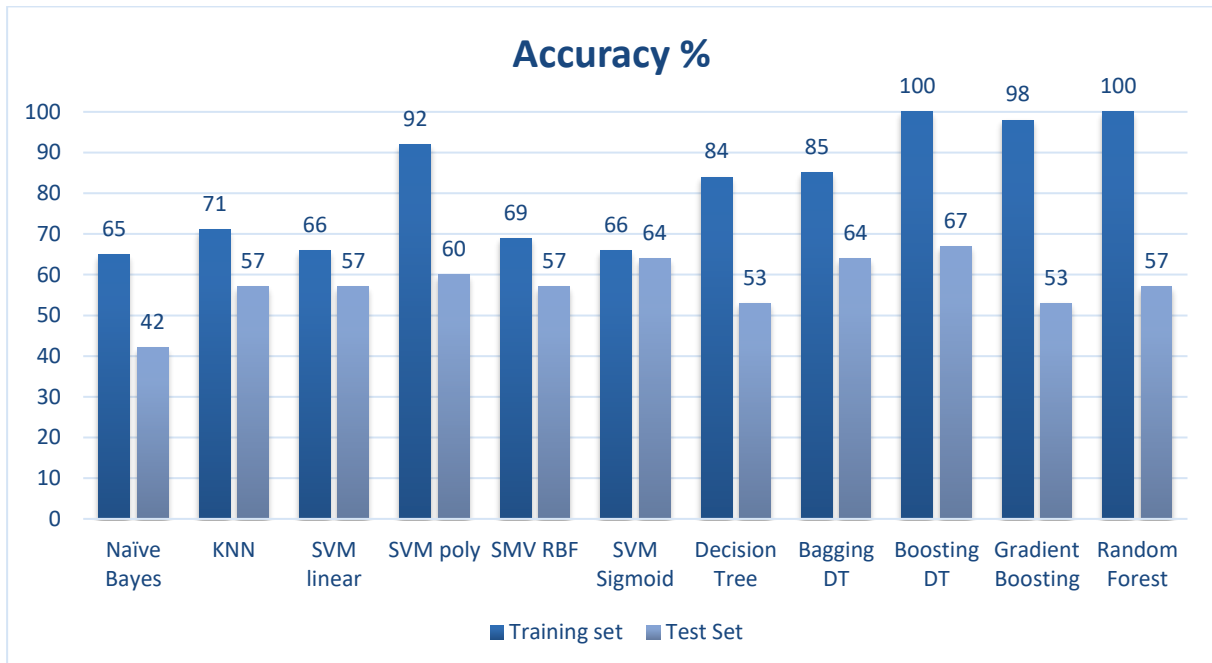


Figura 17 mrmr 11 feature

## Tuning dei parametri dei classificatori

Classificatore	Parametri Tuning
Naïve Bayes	Default
<i>K-Nearest Neighbours (KNN)</i>	algorithm: brute, leaf_size: 30, metric: minkowski, metric_params: None, n_jobs: -1, n_neighbors: 5, p: 2, weights: uniform
<i>SVM Kernel linear</i>	C: 0.1, gamma: 1
<i>SVM Kernel polynomial</i>	C: 1, gamma: 0.01
<i>SVM Kernel RBF</i>	C: 10, gamma: 0.001
<i>SVM Kernel sigmoid</i>	C: 10, gamma: 0.001
<i>Decision Tree (DT)</i>	Min_samples_split: 10, Max_depth: 6
<i>Bagging Decision Tree</i>	N_estimators: 50, Max_features: 1, Max_samples: 0.5, Max_depth: 8, Min_samples_split: 10
<i>Boosting Decision Tree (AdaBoost)</i>	N_estimators: 100, Max_depth: 2, Learning_rate: 0.5
<i>Gradient Boosting</i>	N_estimators: 20
<i>Random Forest</i>	N_estimators: 100, Max_depth: 20, min_samples_leaf: 5, min_samples_split: 2, Max_features: sqrt

Tabella 15 Tuning dei parametri mrmr 10 feature 5 intervalli

Nella tabella a seguire (tab. 16) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

Bagging DT	Training Set	Test Set
Sensitivity %	86	50
Specificity %	85	79
PPV (Positive Predictive Value) %	86	70
NPV (Negative Predictive Value) %	85	61

*Tabella 16 Metriche migliore classificatore*

Nel secondo dataset le feature sono ventisette e sono le seguenti:

- original\_firstorder\_10Percentile
- original\_firstorder\_90Percentile
- original\_firstorder\_Maximum
- original\_firstorder\_Minimum
- original\_firstorder\_RobustMeanAbsoluteDeviation
- original\_glcm\_Autocorrelation
- original\_glcm\_ClusterShade
- original\_glcm\_DifferenceAverage
- original\_glcm\_Idmn
- original\_glcm\_MCC
- original\_glcm\_SumSquares
- original\_glrIm\_GrayLevelNonUniformityNormalized
- original\_glrIm\_GrayLevelVariance
- original\_glrIm\_HighGrayLevelRunEmphasis
- original\_glrIm\_RunVariance
- original\_glrIm\_ShortRunLowGrayLevelEmphasis
- original\_glszm\_GrayLevelNonUniformity
- original\_glszm\_GrayLevelNonUniformityNormalized
- original\_glszm\_LargeAreaLowGrayLevelEmphasis
- original\_glszm\_SizeZoneNonUniformity
- original\_glszm\_SmallAreaEmphasis
- original\_ngtdm\_Coarseness
- original\_gldm\_DependenceEntropy
- original\_gldm\_DependenceNonUniformityNormalized
- original\_gldm\_DependenceVariance
- original\_gldm\_HighGrayLevelEmphasis
- original\_gldm\_SmallDependenceHighGrayLevelEmphasis

Nella figura (fig.18) si possono osservare le prestazioni ottenute:

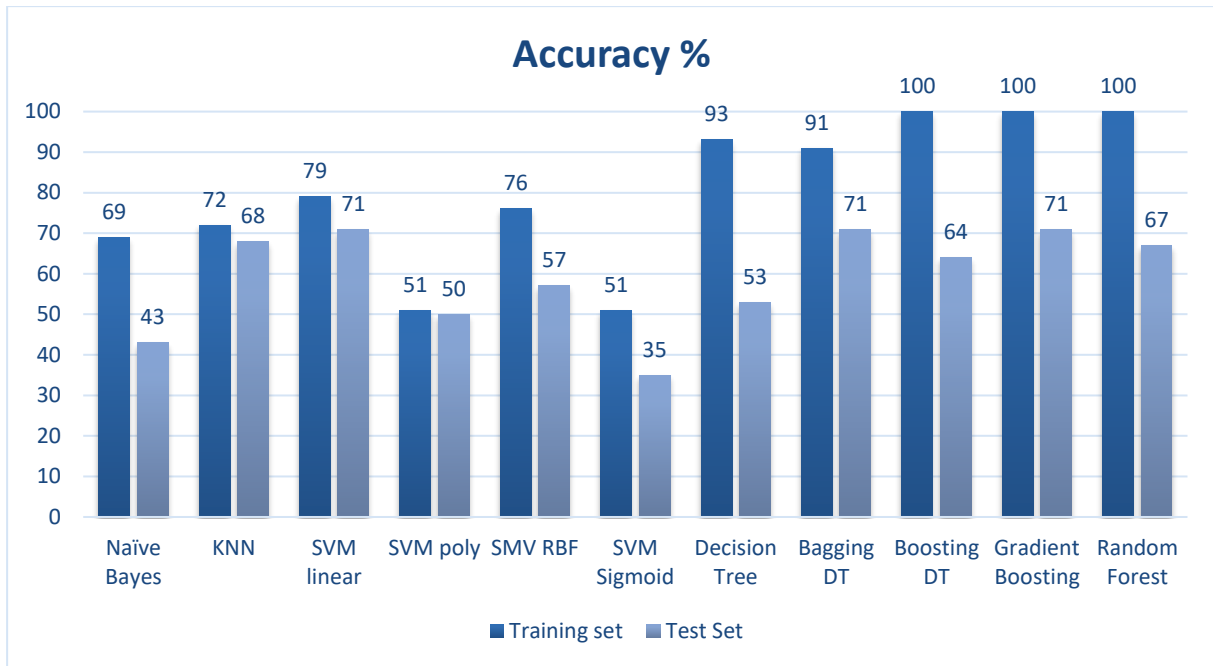


Figura 18 Tuning dei parametri mrmr 27 feature

## Tuning dei parametri dei classificatori

Classificatore	Parametri Tuning
<i>Naïve Bayes</i>	Default
<i>K-Nearest Neighbours (KNN)</i>	algorithm: brute, leaf_size: 30, metric: minkowski, metric_params: None, n_jobs: -1, n_neighbors: 5, p: 2, weights: uniform
<i>SVM Kernel linear</i>	C: 100, gamma: 1
<i>SVM Kernel polynomial</i>	C: 0.1, gamma: 0.001
<i>SVM Kernel RBF</i>	C: 100, gamma: 0.0001
<i>SVM Kernel sigmoid</i>	C: 100, gamma: 0.01
<i>Decision Tree (DT)</i>	Min_samples_split: 18, Max_depth: 5
<i>Bagging Decision Tree</i>	N_estimators: 50, Max_features: 1, Max_samples: 0.5, Max_depth: 8, Min_samples_split: 10
<i>Boosting Decision Tree (AdaBoost)</i>	N_estimators: 100, Max_depth: 2, Learning rate: 0.5
<i>Gradient Boosting</i>	N_estimators: 100
<i>Random Forest</i>	N_estimators: 100, Max_depth: 20, min_samples_leaf: 5, min_samples_split: 2, Max_features: sqrt

Tabella 17 Tuning dei parametri mrmr 27 feature

Nella tabella a seguire (tab. 18) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:



SVM linear	Training Set	Test Set
Sensitivity %	83	64
Specificity %	74	79
PPV (Positive Predictive Value) %	78	75
NPV (Negative Predictive Value) %	81	69

*Tabella 18 Metriche migliore classificatore*

Essendo necessari i dati discretizzati in questo metodo di feature selection, si è deciso di non utilizzare il classificatore SGD in quanto ottimizzato prevalentemente per le feature continue. L'aggiunta di feature sembra migliorare la capacità di generalizzazione per alcuni modelli, mentre per altri, come Bayes e SVM con Kernel Sigmoid, le prestazioni sul set restano basse, e in altri ancora restano costanti, indicando un leggero overfitting sui dati di training.

### 3. Feature selection: Affinity Propagation

In questo studio, come già detto in precedenza, l'Affinity Propagation è stato usato per ridurre la dimensionalità del dataset e per individuare le feature più rappresentative: l'algoritmo ha raggruppato le feature più simili in cluster, ed individuato un esemplare per ciascuno di essi. La matrice di similarità utilizzata è quella di default, che calcola la distanza Euclidea. Ogni esemplare è stato quindi inserito nel nuovo dataset di feature.

Si sono ottenuti otto cluster, i cui esemplari sono le seguenti feature:

- Cluster 1: original\_firstorder\_TotalEnergy
- Cluster 2: original\_glcm\_ClusterTendency
- Cluster 3: original\_glcm\_Correlation
- Cluster 4: original\_glcm\_JointEnergy
- Cluster 5: original\_glszm\_SmallAreaHighGrayLevelEmphasis
- Cluster 6: original\_ngtdm\_Complexity
- Cluster 7: original\_gldm\_LargeDependenceEmphasis
- Cluster 8: original\_gldm\_SmallDependenceEmphasis

Nella figura (fig.19) si possono osservare le prestazioni ottenute:

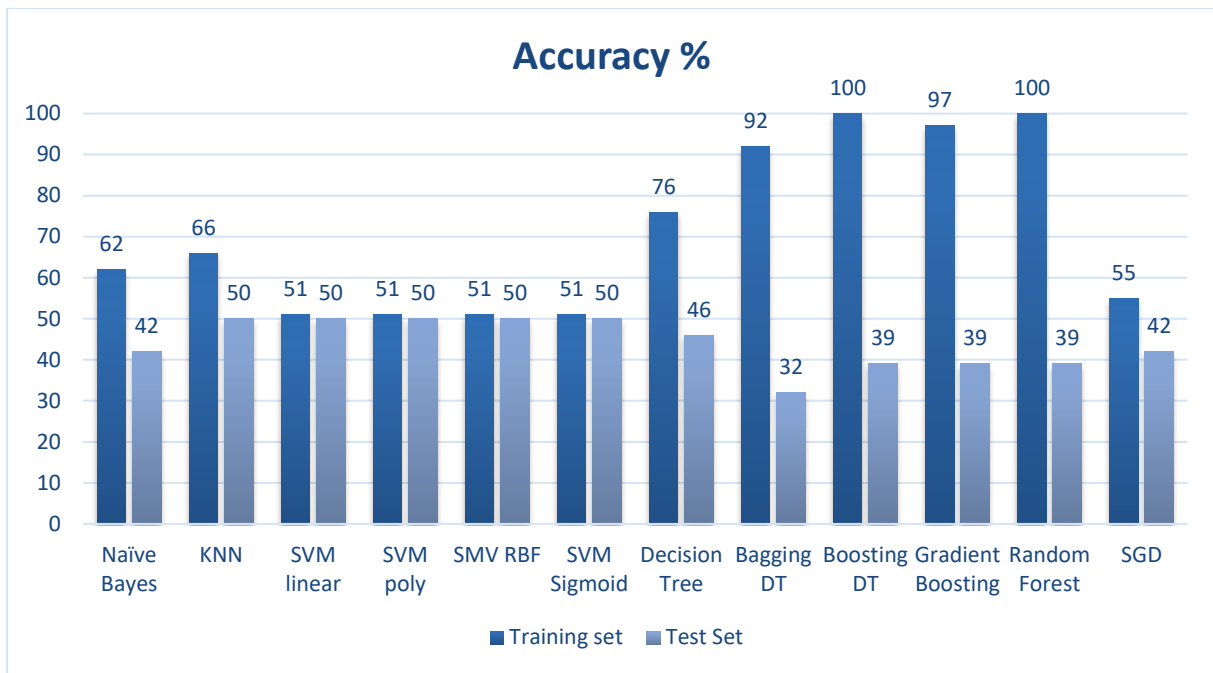


Figura 19 Affinity Propagation 8 cluster

## Tuning dei parametri dei classificatori

Classificatore	Parametri Tuning
<i>Naïve Bayes</i>	Default
<i>K-Nearest Neighbours (KNN)</i>	algorithm: brute, leaf_size: 30, metric: minkowski, metric_params: None, n_jobs: -1, n_neighbors: 5, p: 2, weights: uniform
<i>SVM Kernel linear</i>	C:0.1, gamma:1
<i>SVM Kernel polynomial</i>	C: 0.1, gamma: 0.1
<i>SVM Kernel RBF</i>	C: 0.1, gamma: 1
<i>SVM Kernel sigmoid</i>	C: 0.1, gamma: 1
<i>Decision Tree (DT)</i>	Min_samples_split: 10, Max_depth:6
<i>Bagging Decision Tree</i>	N_estimators: 50, Max_features: 1, Max_samples: 0.5, Max_depth: 8, Min samples split:10
<i>Boosting Decision Tree (AdaBoost)</i>	N_estimators: 100, Max_depth:2, Learning rate: 0.5
<i>Gradient Boosting</i>	N_estimators: 100
<i>Random Forest</i>	N_estimators: 50, Max_depth: None, min_samples_leaf: 1, min_samples_split:10, Max_features: sqrt
<i>Stochastic Gradient Descent (SGD)</i>	Default

Tabella 19 Tuning dei parametri Affinity Propagation 8 cluster

Nella tabella a seguire (tab. 20) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

KNN	Training Set	Test Set
Sensitivity %	64	71
Specificity %	69	29
PPV (Positive Predictive Value) %	69	50
NPV (Negative Predictive Value) %	64	50

Tabella 20 Metriche migliore classificatore

Dal grafico delle prestazioni è possibile notare che i risultati dell'accuracy sul test set presentano valori molto bassi, simili al caso casuale, dimostrando che il dataset ottenuto dalla clusterizzazione non contiene feature abbastanza informative.

#### 4. Feature selection: Correlazione assoluta tra features e target

È stata calcolata la correlazione assoluta tra ciascuna feature e la variabile target, e provando ad aggiungere di volta in volta le feature più correlate, secondo la correlazione di Pearson, sono state calcolate le prestazioni in accuratezza su training e test set per i tre classificatori lineari utilizzati nello studio:

- Naive Bayes
- SVM lineare
- SGD Classifier

Nella figura sottostante (fig. 20) è possibile visualizzare l'andamento dell'accuratezza nel test set al variare del numero delle feature, ordinate in modo decrescente secondo il valore di correlazione di Pearson.

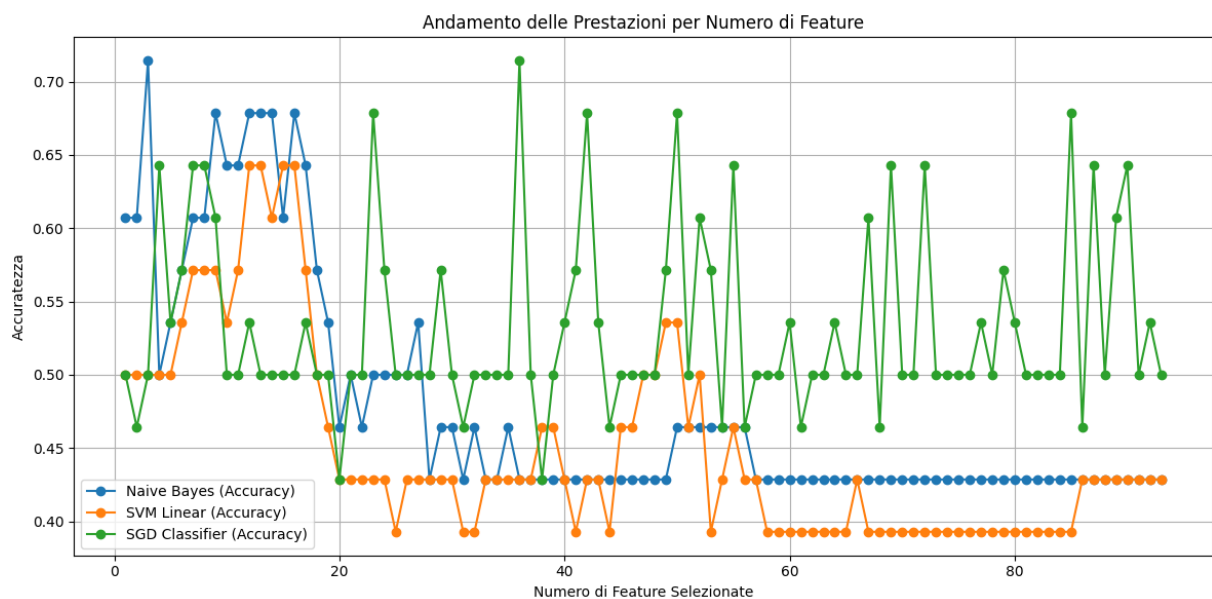


Figura 20 Performance di accuratezza in funzione delle feature aggiunte con Pearson

Nel caso del classificatore Naïve Bayes, la migliore accuratezza si è ottenuta con tre feature:

- original\_glrIm\_GrayLevelNonUniformityNormalized,
- original\_glszm\_GrayLevelNonUniformityNormalized,
- original\_glrIm\_LowGrayLevelRunEmphasis

Nel caso del classificatore SVM Lineare, si sono ottenute le migliori prestazioni con un dataset di 12 feature:

- original\_glrIm\_GrayLevelNonUniformityNormalized,
- original\_glszm\_GrayLevelNonUniformityNormalized,
- original\_glrIm\_LowGrayLevelRunEmphasis,
- original\_glszm\_LowGrayLevelZoneEmphasis,
- original\_glszm\_SmallAreaLowGrayLevelEmphasis,
- original\_glrIm\_ShortRunLowGrayLevelEmphasis,
- original\_firstorder\_Skewness,
- original\_firstorder\_Kurtosis,
- original\_gldm\_DependenceEntropy,
- original\_glrIm\_GrayLevelNonUniformity,
- original\_glszm\_GrayLevelNonUniformity,
- original\_firstorder\_10Percentile

Nel caso del classificatore SGD, si sono ottenute le migliori prestazioni con un dataset di 36 feature, visualizzabili in Appendice.

Nella figura (fig.21) si possono osservare le prestazioni ottenute:

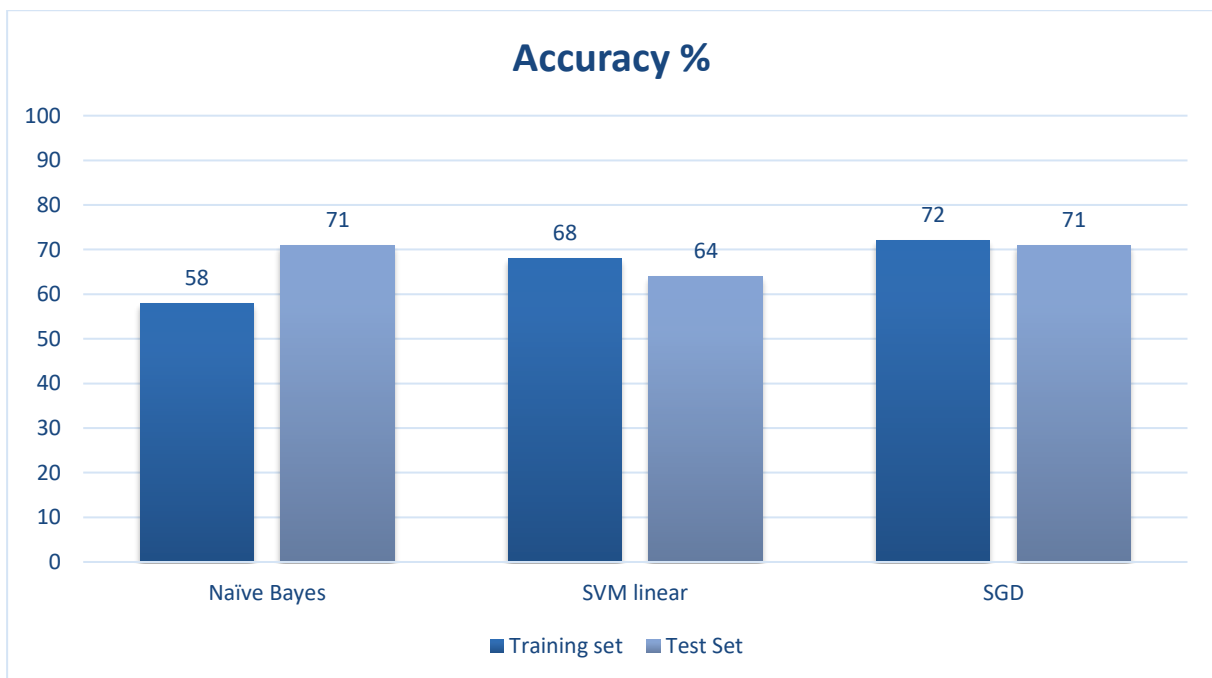


Figura 21 Correlazione assoluta misurata con Pearson

In questo metodo di feature selection non è stato effettuato tuning dei parametri ma sono stati utilizzati i parametri di default, in particolare nel caso del Naïve Bayes si è utilizzata la variante “Gaussian” e come Kernel del SVM il lineare.

Nella tabella a seguire (tab. 21) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

SGD	Training Set	Test Set
Sensitivity %	74	64
Specificity %	72	79
PPV (Positive Predictive Value) %	74	75
NPV (Negative Predictive Value) %	72	69

*Tabella 21 Metriche migliore classificatore*

Osservando il grafico in figura 20, si può notare che il Naïve Bayes non necessita di un numero elevato di feature per ottenere alte prestazioni. Le prestazioni più elevate nel test set potrebbero essere dovute ad una distribuzione più semplice dei campioni che ne fanno parte.

Nel caso del SVM lineare si presentano delle prestazioni non eccessivamente elevate per poi crollare, all’aumentare delle feature, intorno un’accuratezza del 40-45%, indicando che le feature successive non portano informazioni significative.

L’andamento dell’accuratezza del classificatore SGD, invece, presenta forti oscillazioni, indicando che il modello è molto sensibile al set di feature selezionate.

##### 5. Feature selection: Analisi delle componenti principali (PCA)

Prima di applicare questo metodo di feature selection, sono state standardizzate le feature, in quanto la PCA si basa sulla varianza delle feature e quelle con valori più grandi potrebbero falsare il calcolo delle componenti principali. In questo studio si sono mantenute le componenti che spiegavano il 90% della variabilità totale (fig. 22) e se ne sono ottenute 13. Poiché le componenti principali sono combinazioni lineari delle feature originali, sono stati utilizzati solo i classificatori lineari, più adatti alla semplificazione dello spazio e con meno rischio di overfitting.

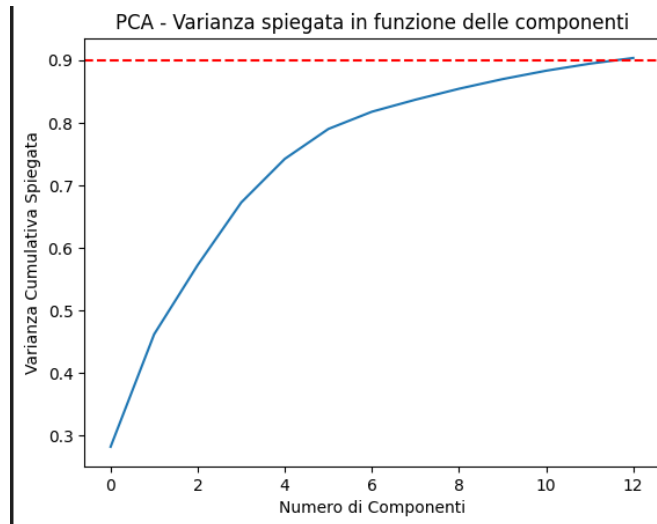


Figura 22 Varianza spiegata

Di seguito (fig. 23) si possono visionare i risultati ottenuti:

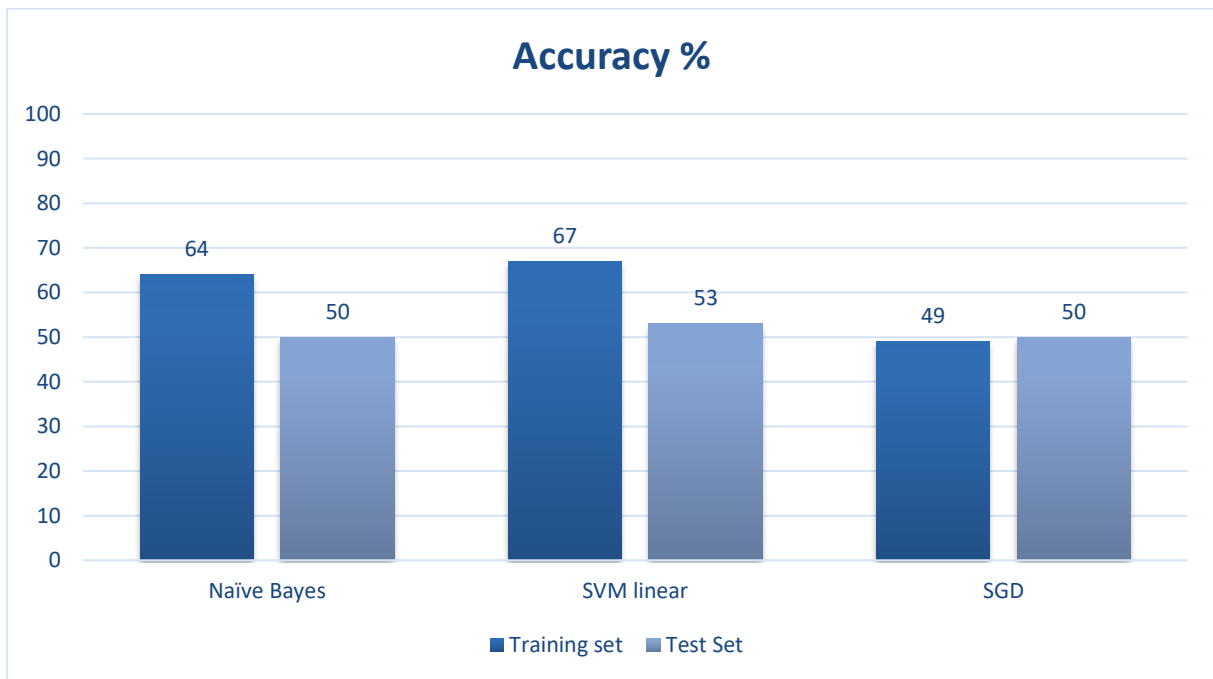


Figura 23 PCA

### Tuning dei parametri dei classificatori

Classificatore	Parametri Tuning
Naïve Bayes	Default
SVM Kernel linear	C: 1, gamma:1
Stochastic Gradient Descent (SGD)	Default

Tabella 22 Tuning dei parametri PCA

Nella tabella a seguire (tab. 23) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

SVM linear	Training Set	Test Set
Sensitivity %	71	43
Specificity %	64	64
PPV (Positive Predictive Value) %	68	55
NPV (Negative Predictive Value) %	68	53

*Tabella 23 Metriche migliore classificatore*

In generale i tre modelli non presentano prestazioni soddisfacenti: nei primi due casi si ha un chiaro esempio di overfitting; invece, il SGD sembra non aver appreso con le componenti che sono state ottenute.

## **Metodo di aggregazione: calcolo delle statistiche descrittive dell'istogramma per ciascuna feature: Media, Mediana, Skewness, Kurtosis, Percentili, range interquartili (IQR) ed Entropia**

### 1. RFE (Recursive Feature Elimination)

In questo metodo di aggregazione si sono ottenute frequenze di selezione molto basse (la maggior parte delle feature hanno una frequenza al di sotto dello 0.1, nella sezione Appendice è possibile vedere la frequenza di selezione di ciascuna feature), probabilmente dovuto al numero elevato di feature di partenza (744). Anche in questo caso sono state valutate due soglie: nel primo caso sono state selezionate cinque feature con frequenza di selezione superiore o uguale allo 0.19:

- original\_firstorder\_Skewness\_skewness, Selection Frequency: 0.52
- original\_glszm\_SmallAreaEmphasis\_25th\_percentile, Selection Frequency: 0.26
- original\_firstorder\_Minimum\_kurtosis, Selection Frequency: 0.25
- original\_glszm\_GrayLevelVariance\_iqr, Selection Frequency: 0.19
- original\_glrmlm\_GrayLevelNonUniformity\_75th\_percentile, Selection Frequency: 0.19

L'accuratezza ottenuta sui cinque fold è stata: [0.71 0.75 0.55 0.62 0.66] la cui media è pari al 66%.

In figura (fig. 24) si possono osservare le prestazioni ottenute:

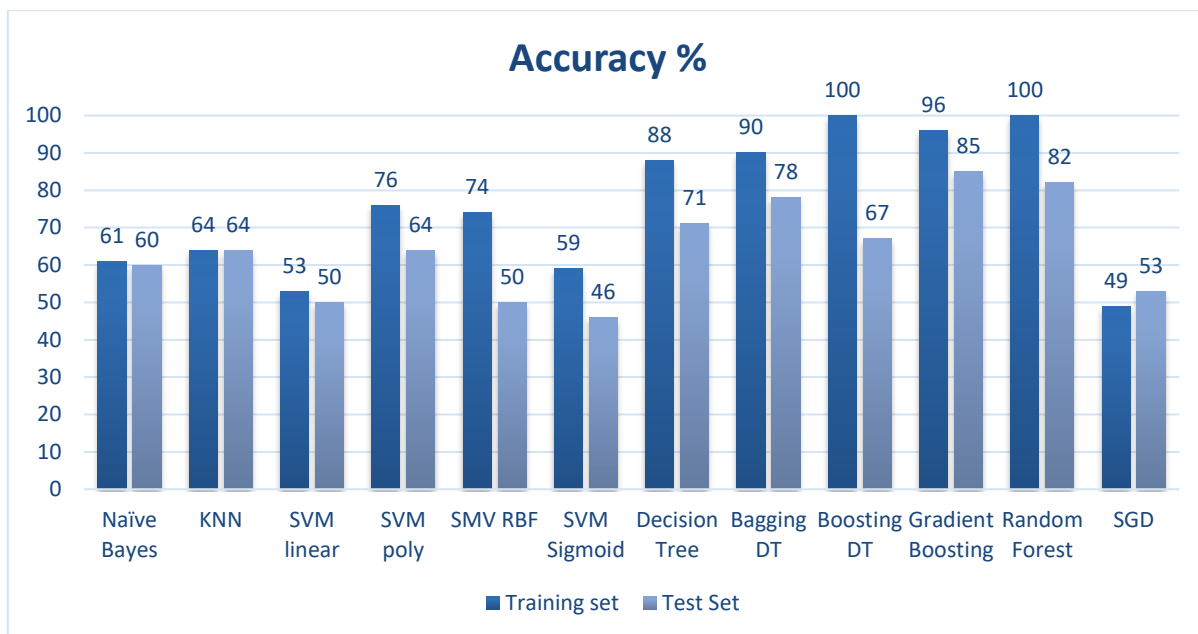


Figura 24 RFE 0.19

## Tuning dei parametri dei classificatori

Classificatore	Parametri Tuning
<i>Naïve Bayes</i>	Default
<i>K-Nearest Neighbours (KNN)</i>	algorithm: auto, leaf_size: 30, metric: minkowski, metric_params: None, n_jobs: -1, n_neighbors: 15 p: 3, weights: uniform
<i>SVM Kernel linear</i>	C: 0.1, gamma: 1
<i>SVM Kernel polynomial</i>	C: 100, gamma: 1
<i>SVM Kernel RBF</i>	C: 1, gamma: 1
<i>SVM Kernel sigmoid</i>	C: 100, gamma: 0.1
<i>Decision Tree (DT)</i>	Min_samples_split: 10, Max_depth: 6
<i>Bagging Decision Tree</i>	N_estimators: 50, Max_features: 1, Max_samples: 0.5, Max_depth: 8, Min_samples_split: 10
<i>Boosting Decision Tree (AdaBoost)</i>	N_estimators: 10, Max_depth: 2, Learning_rate: 0.5
<i>Gradient Boosting</i>	N_estimators: 20
<i>Random Forest</i>	N_estimators: 100, Max_depth: 20
<i>Stochastic Gradient Descent (SGD)</i>	Default

Tabella 24 Tuning dei parametri RFE 0.19

Nella tabella a seguire (tab. 25) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

	Training Set	Test Set
Gradient Boosting		
Sensitivity %	100	79
Specificity %	92	93
PPV (Positive Predictive Value) %	93	92



NPV (Negative Predictive Value) %	100	81
-----------------------------------	-----	----

Tabella 25 Metriche migliore classificatore

È stata valutata anche la soglia superiore a 0.14, ottenendo un dataset di 13 feature:

- original\_firstorder\_Skewness\_skewness, Selection Frequency: 0.52
- original\_glszm\_SmallAreaEmphasis\_25th\_percentile, Selection Frequency: 0.26
- original\_firstorder\_Minimum\_kurtosis, Selection Frequency: 0.25
- original\_glszm\_GrayLevelVariance\_iqr, Selection Frequency: 0.19
- original\_glrlm\_GrayLevelNonUniformity\_75th\_percentile, Selection Frequency: 0.19
- original\_glszm\_GrayLevelNonUniformity\_75th\_percentile, Selection Frequency: 0.17
- original\_glrlm\_GrayLevelNonUniformityNormalized\_25th\_percentile, Selection Frequency: 0.15
- original\_firstorder\_Minimum\_iqr, Selection Frequency: 0.14
- original\_glszm\_SmallAreaEmphasis\_mean, Selection Frequency: 0.14
- original\_glrlm\_GrayLevelVariance\_iqr, Selection Frequency: 0.14
- original\_glrlm\_GrayLevelNonUniformity\_kurtosis, Selection Frequency: 0.14
- original\_glrlm\_GrayLevelNonUniformityNormalized\_kurtosis, Selection Frequency: 0.14
- original\_glszm\_GrayLevelNonUniformityNormalized\_25th\_percentile, Selection Frequency: 0.14

L'accuratezza ottenuta sui cinque fold è stata: [0.71 0.64 0.62 0.59 0.66], la cui media è pari al 65%.

In figura (fig. 25) si possono osservare le prestazioni ottenute:

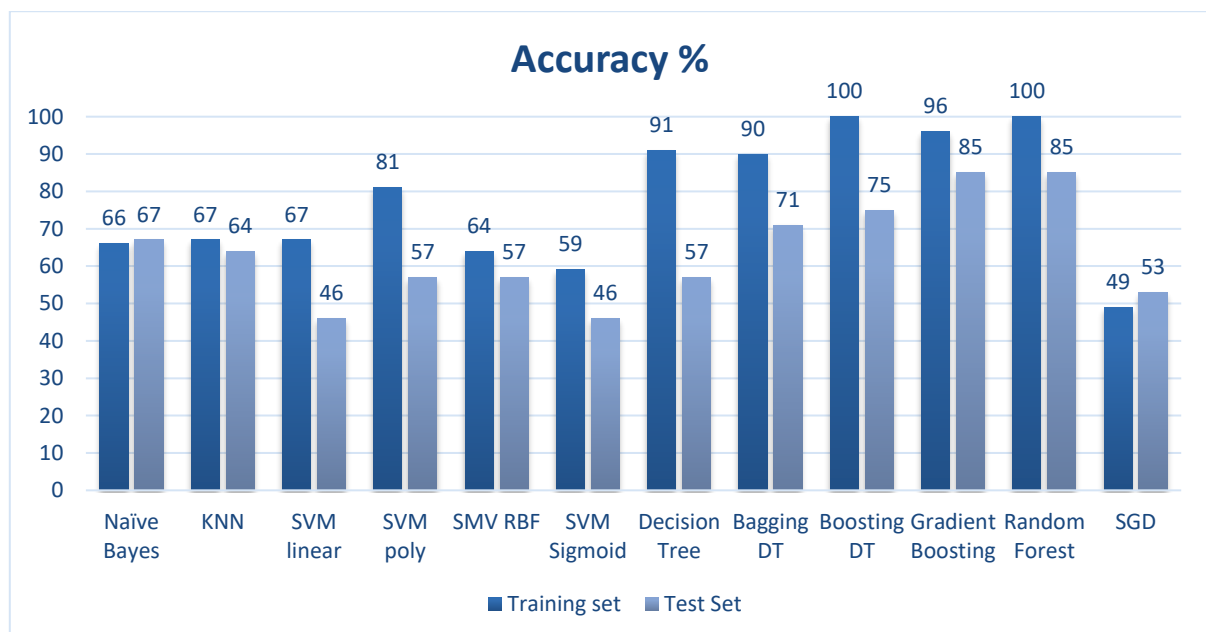


Figura 25 RFE 0.14

## Tuning dei parametri dei classificatori

<i>Classificatore</i>	<b>Parametri Tuning</b>
<i>Naïve Bayes</i>	Default
<i>K-Nearest Neighbours (KNN)</i>	algorithm: auto, leaf_size: 30, metric: minkowski, metric_params: None, n_jobs: -1, n_neighbors: 9, p: 5, weights: uniform
<i>SVM Kernel linear</i>	C: 1000, gamma: 1
<i>SVM Kernel polynomial</i>	C: 10, gamma: 0.1
<i>SVM Kernel RBF</i>	C: 100, gamma: 0.001
<i>SVM Kernel sigmoid</i>	C: 1000, gamma: 0.01
<i>Decision Tree (DT)</i>	Min_samples_split: 10, Max_depth:6
<i>Bagging Decision Tree</i>	N_estimators: 50, Max_features: 1, Max_samples: 0.5, Max_depth: 8, Min samples split:10
<i>Boosting Decision Tree (AdaBoost)</i>	N_estimators: 10, Max_depth:2, Learning_rate: 0.5
<i>Gradient Boosting</i>	N_estimators: 20
<i>Random Forest</i>	N_estimators:100, Max_depth: 20
<i>Stochastic Gradient Descent (SGD)</i>	Default

Tabella 26 Tuning dei parametri RFE 0.14

Nella tabella a seguire (tab. 27) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

Gradient Boosting	Training Set	Test Set
Sensitivity %	100	79
Specificity %	92	93
PPV (Positive Predictive Value) %	93	92
NPV (Negative Predictive Value) %	100	81

Tabella 27 Metriche migliore classificatore

Osservando le prestazioni dei due dataset, si può notare che l'accuratezza sul test set è migliorata per una buona parte dei classificatori, mentre per gli ensemble models è rimasta la stessa: ciò significa che l'aggiunta delle feature con soglia maggiore o uguale a 0.14 ha aumentato la capacità predittiva, migliorando la separazione delle classi.

## 2. Feature selection: minimum Redundancy - maximum Relevance (mRmR)

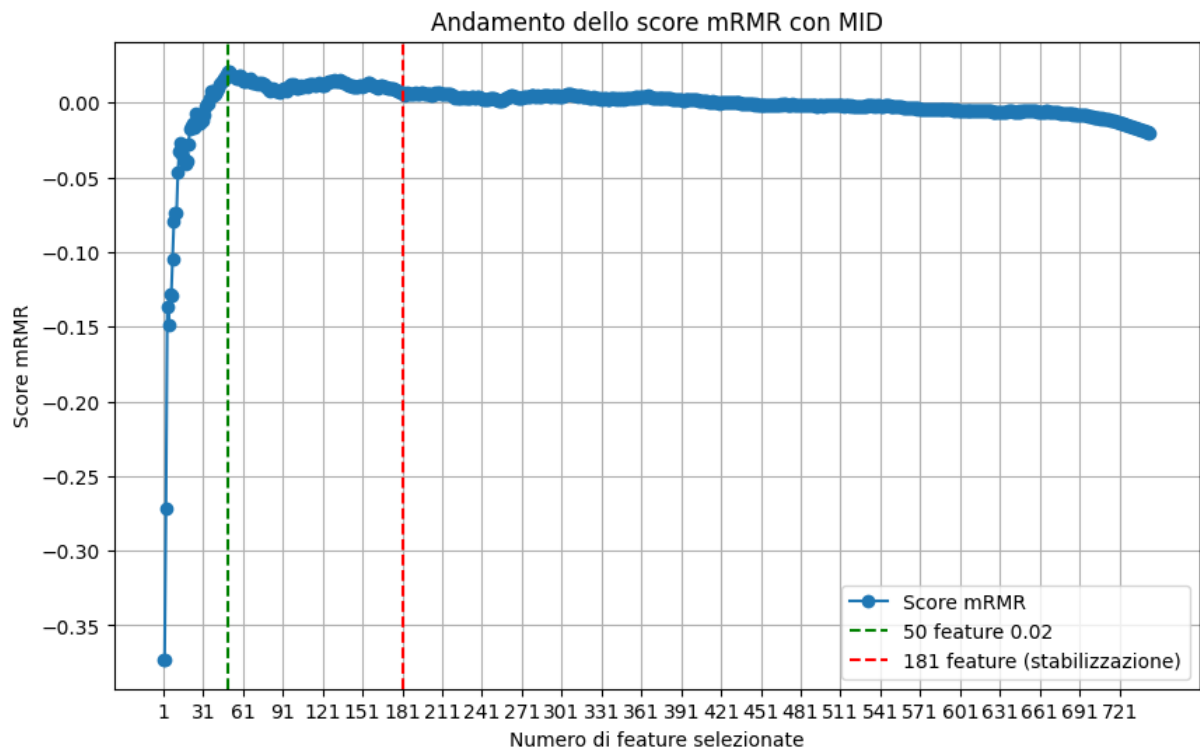


Figura 26 Andamento dello score mRMR con MID

Nella figura soprastante (fig. 26) si può osservare l'andamento dello score MID ovvero la differenza tra la rilevanza e la ridondanza calcolate dall'algoritmo mRMR per aggiungere ad ogni iterazione la feature più rilevante rispetto al target e meno ridondante tra quelle già selezionate. Dopo 181 feature si osserva una stabilizzazione dello score: ciò suggerisce che le feature successivamente aggiunte siano meno rilevanti rispetto al target o che introducano troppa ridondanza con le feature già selezionate. Si è voluto confrontare anche in questo caso il dataset delle feature selezionate nella prima fase, quando principalmente sono presenti quelle altamente rilevanti, con il dataset di feature prima che l'andamento dello score si stabilizzi. Si può notare che lo score di queste due soglie è positivo, ad indicare che queste feature sono molto rilevanti e che la loro ridondanza risulta leggermente meno impattante.

Nel primo dataset le feature sono cinquanta e sono visionabili nell'Appendice.

In figura (fig. 27) è possibile visionare le prestazioni dei classificatori sul primo dataset.

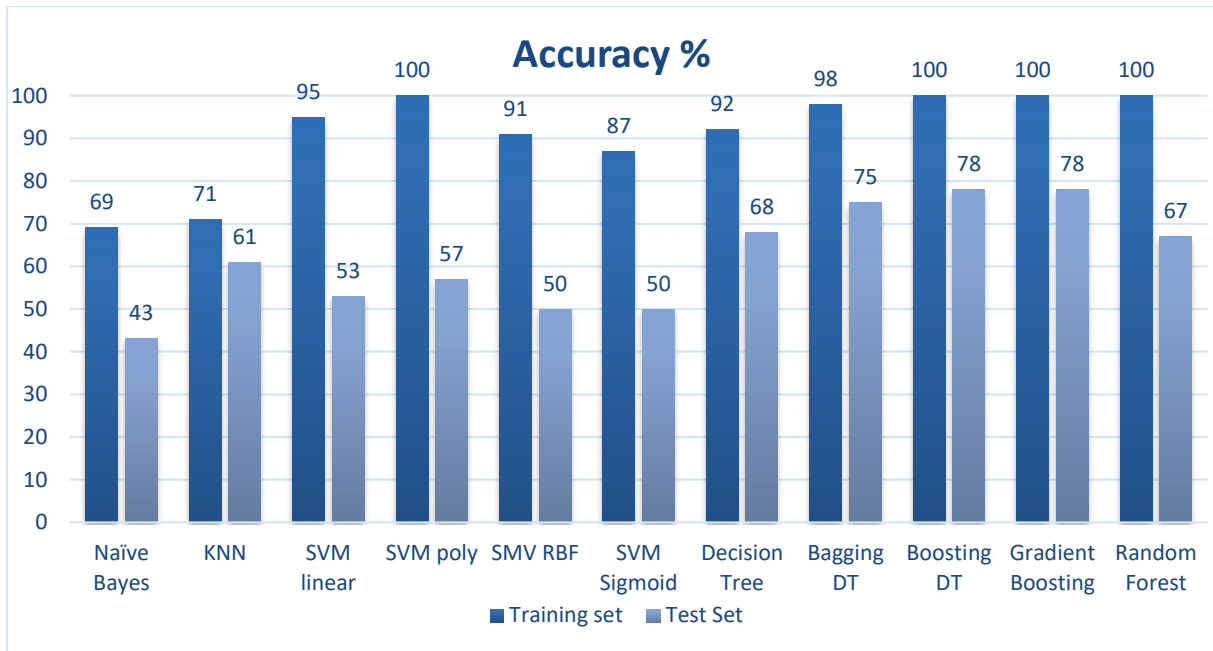


Figura 27 mrmr 50 feature

## Tuning dei parametri dei classificatori

Classificatore	Parametri Tuning
<i>Naïve Bayes</i>	Default
<i>K-Nearest Neighbours (KNN)</i>	algorithm: auto, leaf_size: 30, metric: minkowski, metric_params: None, n_jobs: -1, n_neighbors: 4 p: 5, weights: uniform
<i>SVM Kernel linear</i>	C: 0.1, gamma: 1
<i>SVM Kernel polynomial</i>	C: 0.1, gamma: 1
<i>SVM Kernel RBF</i>	C: 100, gamma: 0.0001
<i>SVM Kernel sigmoid</i>	C: 100, gamma: 0.0001
<i>Decision Tree (DT)</i>	Min_samples_split: 5, Max_depth: 18
<i>Bagging Decision Tree</i>	N_estimators: 50, Max_features: 1, Max_samples: 0.5, Max_depth: 8, Min_samples_split: 10
<i>Boosting Decision Tree (AdaBoost)</i>	N_estimators: 10, Max_depth: 2, Learning_rate: 0.5
<i>Gradient Boosting</i>	N_estimators: 20
<i>Random Forest</i>	N_estimators: 100, Max_depth: 20, min_samples_leaf: 4, min_samples_split: 2, Max_features: sqrt

Tabella 28 Tuning dei parametri mrmr 50 feature

Nella tabella a seguire (tab. 28) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

Bagging DT	Training Set	Test Set
Sensitivity %	100	71
Specificity %	97	79
PPV (Positive Predictive Value) %	98	77
NPV (Negative Predictive Value) %	100	73

Figura 28 Metriche migliore classificatore

Nel secondo dataset le feature sono 181 e sono leggibili nella sezione Appendice.

In figura (fig. 29) è possibile visionare le prestazioni dei classificatori sul secondo dataset.

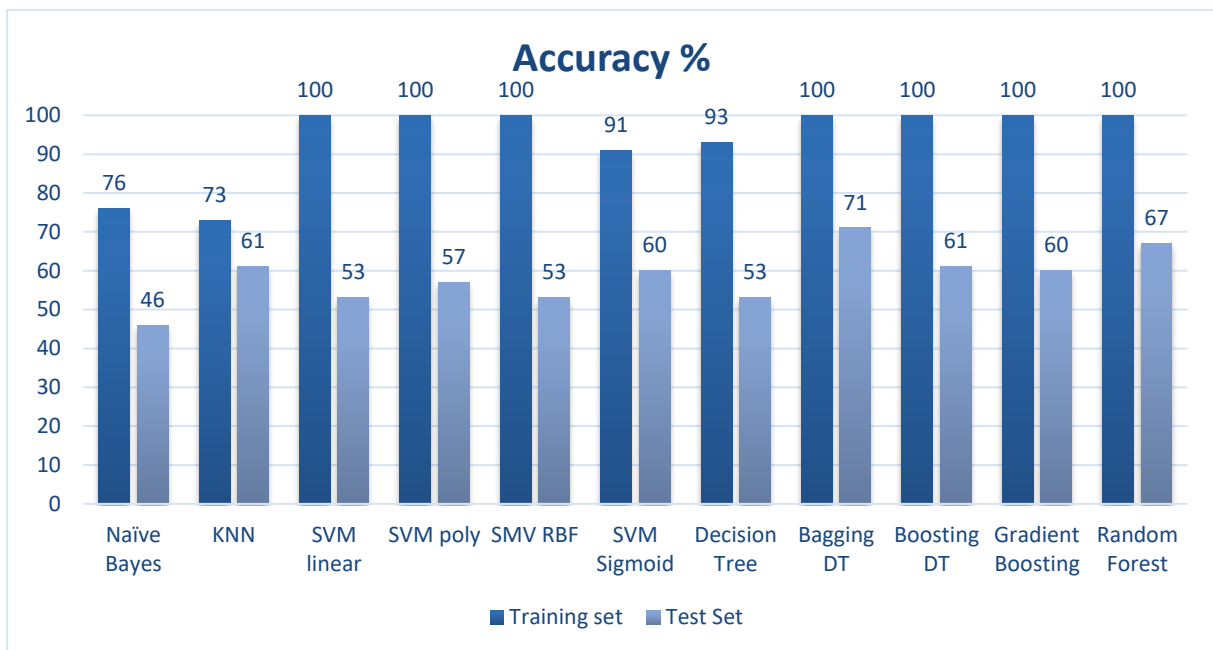


Figura 29 mrmr 181 feature

## Tuning dei parametri dei classificatori

Classificatore	Parametri Tuning
<i>Naïve Bayes</i>	Default
<i>K-Nearest Neighbours (KNN)</i>	algorithm: brute, leaf_size: 30, metric: minkowski, metric_params: None, n_jobs: -1, n_neighbors: 5, p: 2, weights: uniform
<i>SVM Kernel linear</i>	C: 0.1, gamma: 1
<i>SVM Kernel polynomial</i>	C: 0.1, gamma: 1
<i>SVM Kernel RBF</i>	C: 100, gamma: 0.0001
<i>SVM Kernel sigmoid</i>	C: 100, gamma: 0.0001
<i>Decision Tree (DT)</i>	Min_samples_split: 5, Max_depth: 18
<i>Bagging Decision Tree</i>	N_estimators: 50, Max_features: 1, Max_samples: 0.5, Max_depth: 8, Min samples split: 10
<i>Boosting Decision Tree (AdaBoost)</i>	N_estimators: 10, Max_depth: 2, Learning rate: 0.5

*Gradient Boosting  
Random Forest*

N_estimators: 100
N_estimators:100, Max_depth: 200 min_samples_leaf: 5, min_samples_split:2 Max_features: sqrt

Tabella 29 Tuning dei parametri mrmr 181

Nella tabella a seguire (tab. 30) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

Bagging DT	Training Set	Test Set
Sensitivity %	100	57
Specificity %	100	86
PPV (Positive Predictive Value) %	100	80
NPV (Negative Predictive Value) %	100	67

Tabella 30 Metriche migliore classificatore

Le feature aggiuntive si sono rilevate ridondanti, in quanto, hanno portato all'overfitting la maggioranza dei classificatori, e in alcuni casi si è anche visto il peggioramento nel test set, mostrando appunto una perdita nella capacità di generalizzare.

### 3. Feature selection: Affinity Propagation

In questo dataset l’Affinity Propagation ha raggruppato le 744 feature in 142 cluster, ottenendo quindi 142 feature. È possibile visionare l’elenco delle feature selezionate in Appendice.

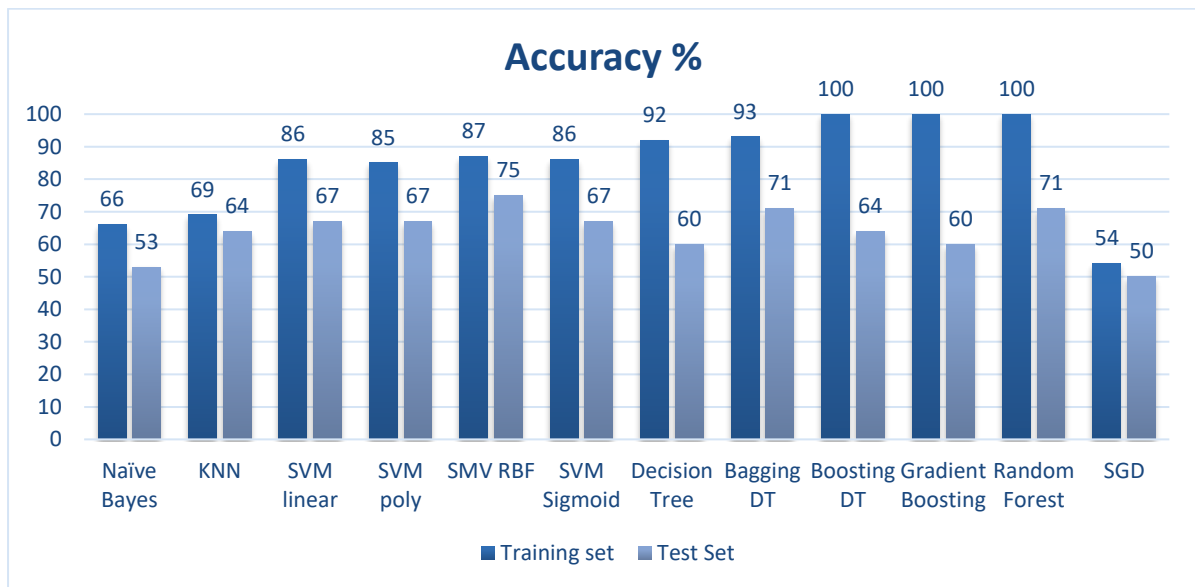


Figura 30 Affinity Propagation 142 cluster

## Tuning dei parametri dei classificatori

<i>Classificatore</i>	<b>Parametri Tuning</b>
<i>Naïve Bayes</i>	Default
<i>K-Nearest Neighbours (KNN)</i>	algorithm: brute, leaf_size: 30, metric: minkowski, metric_params: None, n_jobs: -1, n_neighbors: 5, p: 2, weights: uniform
<i>SVM Kernel linear</i>	C:1, gamma:1
<i>SVM Kernel polynomial</i>	C: 0.1, gamma: 0.1
<i>SVM Kernel RBF</i>	C: 100, gamma: 0.01
<i>SVM Kernel sigmoid</i>	C: 1000, gamma: 0.001
<i>Decision Tree (DT)</i>	Min_samples_split: 10, Max_depth:6
<i>Bagging Decision Tree</i>	N_estimators: 50, Max_features: 1, Max_samples: 0.5, Max_depth: 8, Min_samples_split:10
<i>Boosting Decision Tree (AdaBoost)</i>	N_estimators: 100, Max_depth:2, Learning_rate: 0.5
<i>Gradient Boosting</i>	N_estimators: 100
<i>Random Forest</i>	N_estimators: 50, Max_depth: None, min_samples_leaf: 1, min_samples_split:10, Max_features: sqrt
<i>Stochastic Gradient Descent (SGD)</i>	Default

Tabella 31 Tuning dei parametri Affinity propagation 142 cluster

Nella tabella a seguire (tab. 32) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

SVM RBF	Training Set	Test Set
Sensitivity %	52	36
Specificity %	51	57
PPV (Positive Predictive Value) %	54	45
NPV (Negative Predictive Value) %	50	47

Tabella 32 Metriche migliore classificatore

In questo dataset si sono ottenute prestazioni migliori rispetto alle feature selezionate nel dataset aggregato secondo la media di ciascuna feature: l'accuracy sul test set è aumentata per tutti i classificatori, ciò significa che le feature ottenute da questo clustering risultano essere, oltre che in maggiore quantità, più informative.

#### 4. Feature selection: Correlazione assoluta tra features e target

Anche in questo dataset è stata calcolata la correlazione di Pearson tra ciascuna feature e la variabile target, e aggiungendo iterativamente le feature più correlate, sono state calcolate le prestazioni su training e test set per i tre classificatori lineari utilizzati nello studio.

Nel caso del classificatore Naïve Bayes, la migliore accuratezza si è ottenuta con 18 feature:

- original\_glrlm\_GrayLevelNonUniformityNormalized\_median,
- original\_glszm\_GrayLevelNonUniformityNormalized\_median,
- original\_glrlm\_GrayLevelNonUniformity\_skewness,
- original\_glszm\_GrayLevelNonUniformity\_skewness,
- original\_glrlm\_GrayLevelNonUniformityNormalized\_mean,
- original\_glszm\_GrayLevelNonUniformityNormalized\_mean,
- original\_glszm\_GrayLevelNonUniformityNormalized\_75th\_percentile,
- original\_glrlm\_GrayLevelNonUniformityNormalized\_75th\_percentile,
- original\_glszm\_GrayLevelNonUniformityNormalized\_25th\_percentile,
- original\_glrlm\_GrayLevelNonUniformityNormalized\_25th\_percentile,
- original\_glrlm\_GrayLevelNonUniformity\_75th\_percentile,
- original\_firstorder\_Minimum\_iqr,
- original\_gldm\_DependenceNonUniformity\_kurtosis,
- original\_glszm\_GrayLevelNonUniformity\_75th\_percentile,
- original\_firstorder\_Skewness\_median,
- original\_glszm\_SmallAreaLowGrayLevelEmphasis\_75th\_percentile,
- original\_glrlm\_ShortRunLowGrayLevelEmphasis\_75th\_percentile,
- original\_glrlm\_LowGrayLevelRunEmphasis\_mean

Nel caso del classificatore SVM Lineare, si sono ottenute le migliori prestazioni con un dataset di cinque feature:

- original\_glrlm\_GrayLevelNonUniformityNormalized\_median,
- original\_glszm\_GrayLevelNonUniformityNormalized\_median,
- original\_glrlm\_GrayLevelNonUniformity\_skewness,
- original\_glszm\_GrayLevelNonUniformity\_skewness,
- original\_glrlm\_GrayLevelNonUniformityNormalized\_mean

Nel caso del classificatore SGD, si sono ottenute le migliori prestazioni con un dataset di 126 feature, visualizzabili in Appendice.

Nella figura sottostante (fig. 31) è possibile visualizzare l'andamento dell'accuratezza nel test set al variare del numero delle feature, ordinate in modo decrescente secondo il valore di correlazione di Pearson.



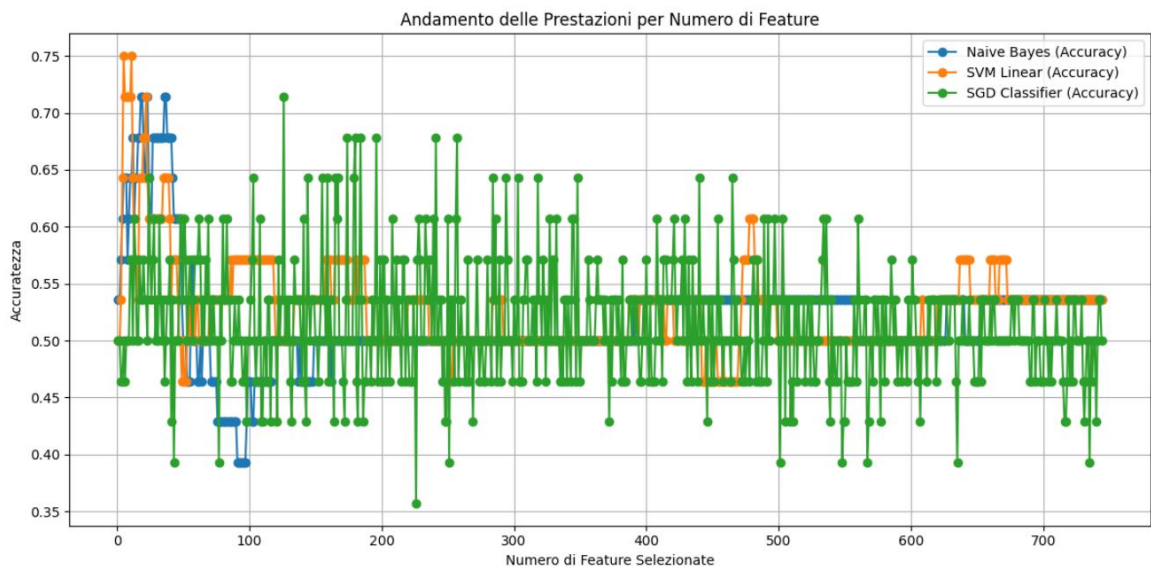


Figura 31 Performance in funzione delle feature aggiunte con Pearson

Nella figura (fig. 32) si possono osservare le prestazioni ottenute:

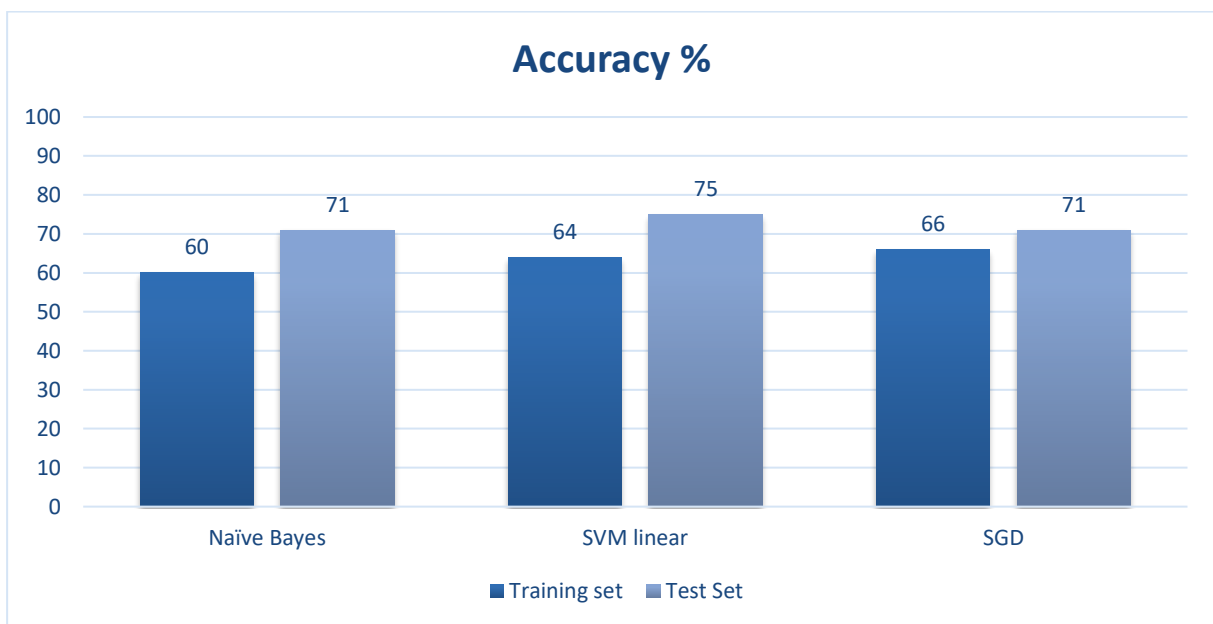


Figura 32 Correlazione assoluta misurata con Pearson

Nella tabella a seguire (tab. 33) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

SVM RBF	Training Set	Test Set
Sensitivity %	95	86
Specificity %	36	57
PPV (Positive Predictive Value) %	62	67
NPV (Negative Predictive Value) %	88	80

Tabella 33 Metriche migliore classificatore

Dall'andamento delle prestazioni, si osserva che il Naïve Bayes si stabilizza rapidamente con un piccolo numero di feature e rimane invariato per il resto del test set: questo indica che le prime feature selezionate contengono la maggior parte dell'informazione utile.

Il SVM lineare invece raggiunge prestazioni migliori con meno feature e raggiunge prestazioni maggiori sul test set.

Il SGD, invece, presenta sempre forti oscillazioni, probabilmente dovuto alla sua natura stocastica, che dipende dal learning rate e dal numero di iterazioni. Nessuno dei modelli presenta overfitting,

### 5. Feature selection: Analisi delle componenti principali (PCA)

Anche in questo dataset si sono mantenute le componenti che spiegavano il 90% della variabilità totale (fig. 33): se ne sono ottenute 20.

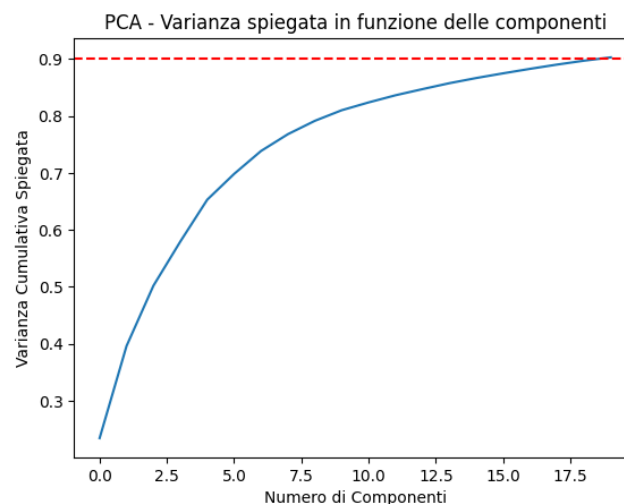


Figura 33 Varianza spiegata

Di seguito (fig. 34) si possono visionare i risultati ottenuti:

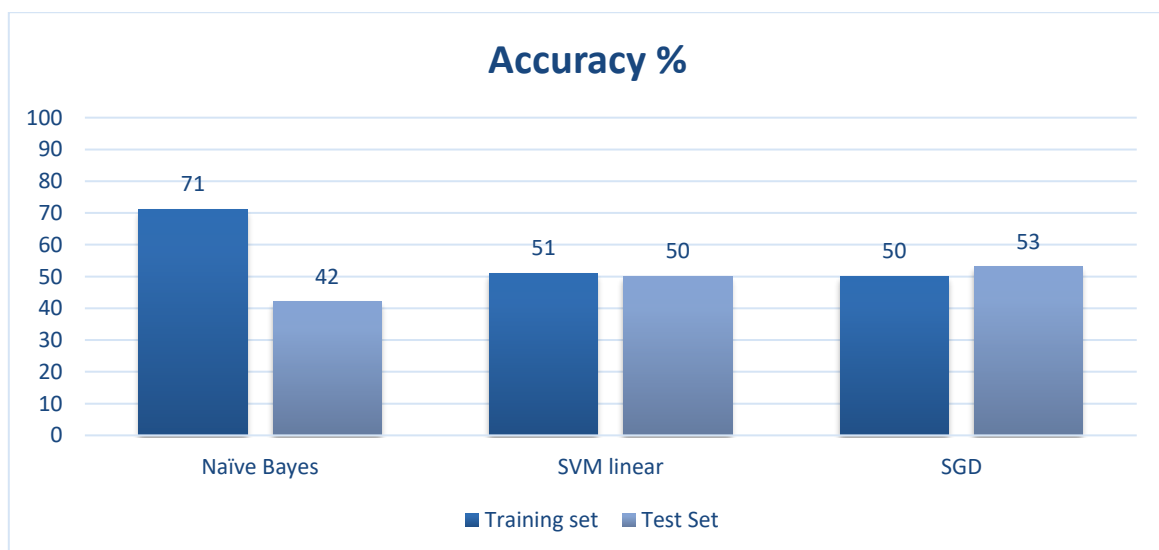


Figura 34 PCA

## Tuning dei parametri dei classificatori

<i>Classificatore</i>	<b>Parametri Tuning</b>
<i>Naïve Bayes</i>	Default
<i>SVM Kernel linear</i>	C: 0.1, gamma:1
<i>Stochastic Gradient Descent (SGD)</i>	Default

Tabella 34 Tuning dei parametri PCA

Nella tabella a seguire (tab. 35) è possibile prendere visione delle metriche valutate sul Training e Test set del classificatore che ottiene le migliori prestazioni:

SVM RBF	Training Set	Test Set
Sensitivity %	100	100
Specificity %	0	0
PPV (Positive Predictive Value) %	52	50
NPV (Negative Predictive Value) %	0	0

Tabella 35 Metriche migliore classificatore

Anche in questo dataset i tre modelli non presentano prestazioni soddisfacenti

## Prestazioni classificatori migliori nel Validation Set

Dai vari metodi di feature selection di entrambi i metodi di aggregazione si è voluto mettere a confronto le prestazioni dei migliori classificatori nel Validation Set (fig. 35)

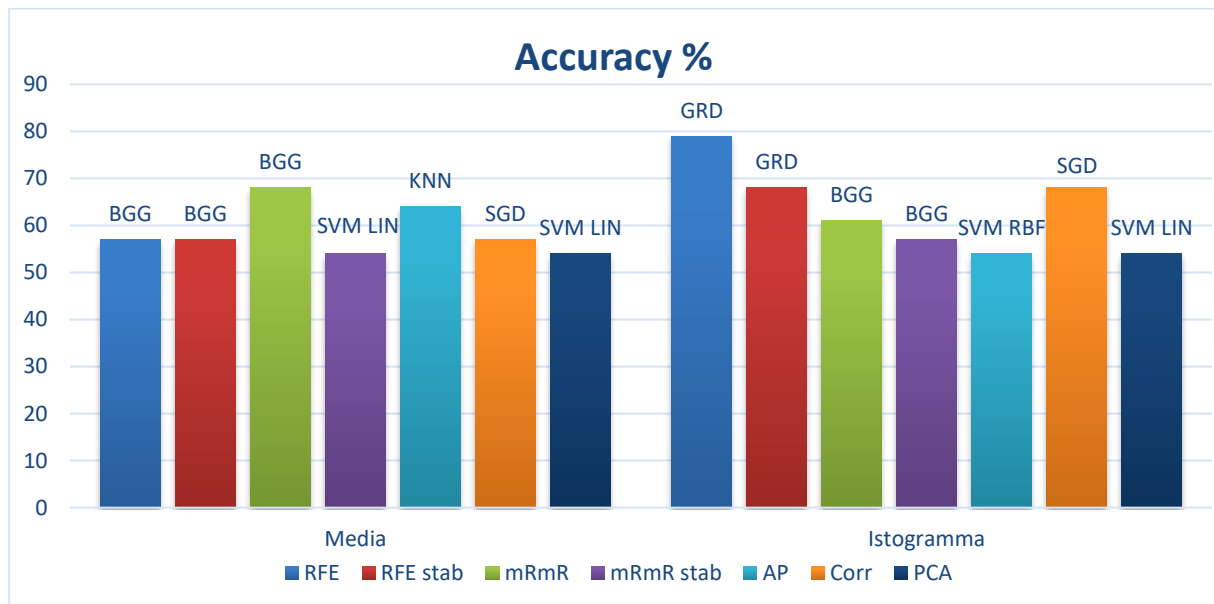


Figura 35 Prestazioni migliori classificatori

I classificatori presenti nel grafico sono stati valutati anche con sensitivity, specificity, confusion matrix e la curva ROC. Nelle tabelle di seguito (tab. 36) è possibile visualizzare le suddette metriche, per i classificatori utilizzati nei dataset ottenuti dalla media delle feature:

Classificatore	Sensitivity %	Specificity %	PPV %	NPV %	AUC (ROC) %
Bagging DT	67	46	59	55	68
Bagging DT	53	62	62	53	67
Bagging DT	53	85	80	61	71
SVM Lineare	47	62	58	50	31
KNN	67	62	67	62	54
SGD	86	29	55	67	64
SVM Lineare	47	62	55	67	66

Tabella 36 Metriche classificatori metodo di aggregazione media

True	Predicted	
	Classe 0	Classe 1
	Classe 0	6
Classe 1	5	10

Tabella 37 RFE 0.4 Bagging DT

True	Predicted	
	Classe 0	Classe 1
	Classe 0	8
Classe 1	7	8

Tabella 38 RFE 0.25 Bagging DT

True	Predicted	
	Classe 0	Classe 1
	Classe 0	11
Classe 1	7	8

Tabella 39 mrmr 11 Bagging DT

True	Predicted	
	Classe 0	Classe 1
	Classe 0	8
Classe 1	8	7

Tabella 40 mrmr 27 SVM linear

True	Predicted	
	Classe 0	Classe 1
	Classe 0	8
Classe 1	5	10

Tabella 41 AP KNN

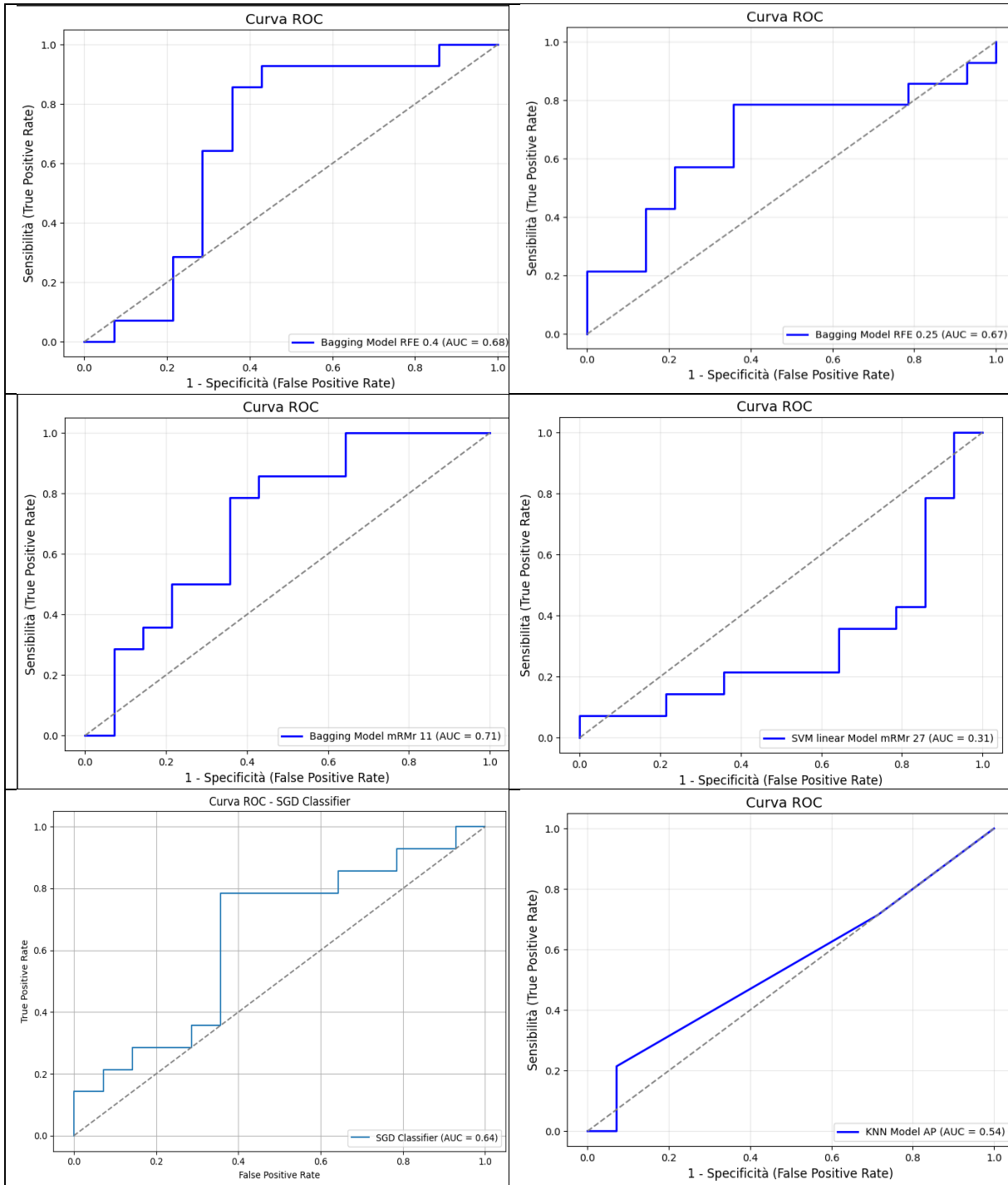
True	Predicted	
	Classe 0	Classe 1
	Classe 0	4
Classe 1	2	12

Tabella 42 Correlazione SGD

True	Predicted	
	Classe 0	Classe 1
	Classe 0	4
Classe 1	2	12

Tabella 43 PCA SVM Linear

Nella tabella successiva (tab. 44) si possono visualizzare le curve ROC dei classificatori selezionati:



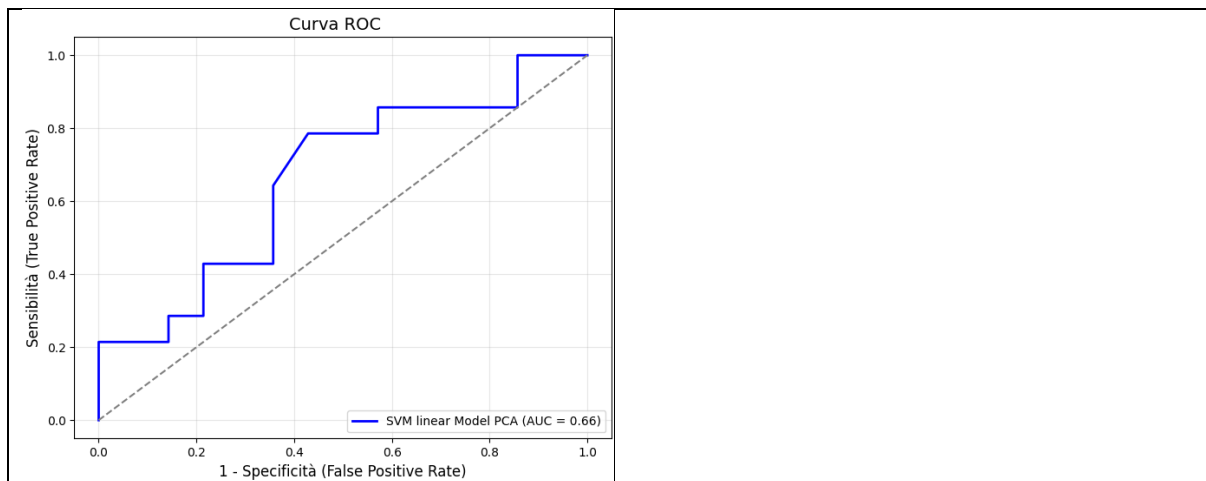


Tabella 44 Curve ROC

Nelle tabelle successive (tab. 45) invece è possibile visionare le suddette metriche per i classificatori migliori del dataset ottenuto dalle statistiche descrittive dell'istogramma.

Classificatore	Sensitivity %	Specificity %	PPV %	NPV %	AUC (ROC)
Gradient Boost	87	69	75	83	92
Gradient Boost	67	69	71	71	91
Bagging DT	53	69	67	60	84
Bagging DT	53	62	57	57	79
SVM RBF	47	62	54	53	48
SGD	86	50	63	78	67
SVM Lineare	100	0	50	0	53

Tabella 45 Metriche classificatori metodo di aggregazione istogramma

True	Predicted	
	Classe 0	Classe 1
	Classe 0	10
Classe 1	2	12

Tabella 46 RFE 0.19 Gradient Boost

True	Predicted	
	Classe 0	Classe 1
	Classe 0	10
Classe 1	4	10

Tabella 47 RFE 0.14 Gradient Boost

True	Predicted	
	Classe 0	Classe 1
	Classe 0	9
Classe 1	6	8

Tabella 48 mrmr 50 Bagging DT

True	Predicted	
	Classe 0	Classe 1
	Classe 0	8
Classe 1	6	8

Tabella 49 mrmr 181 SVM linear

True	Predicted	
	Classe 0	Classe 1
	Classe 0	8
Classe 1	7	7

Tabella 50 AP SVM RBF

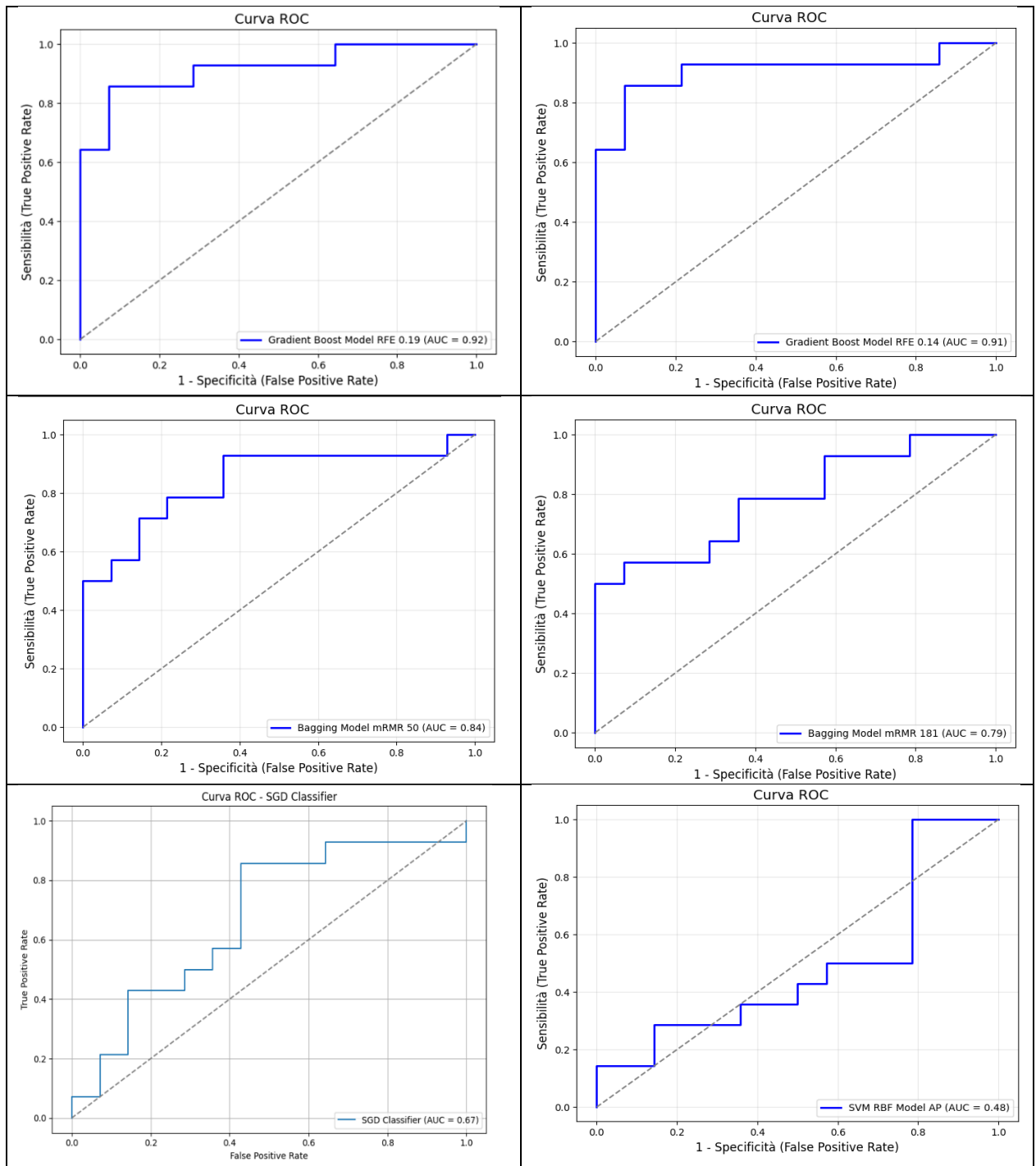
True	Predicted	
	Classe 0	Classe 1
	Classe 0	7
Classe 1	2	12

Tabella 51 Correlazione SGD

True	Predicted	
	Classe 0	Classe 1
	Classe 0	0
Classe 1	0	14

Tabella 52 PCA SVM Linear

Nella tabella successiva (tab. 53) si possono visualizzare le curve ROC dei classificatori selezionati:



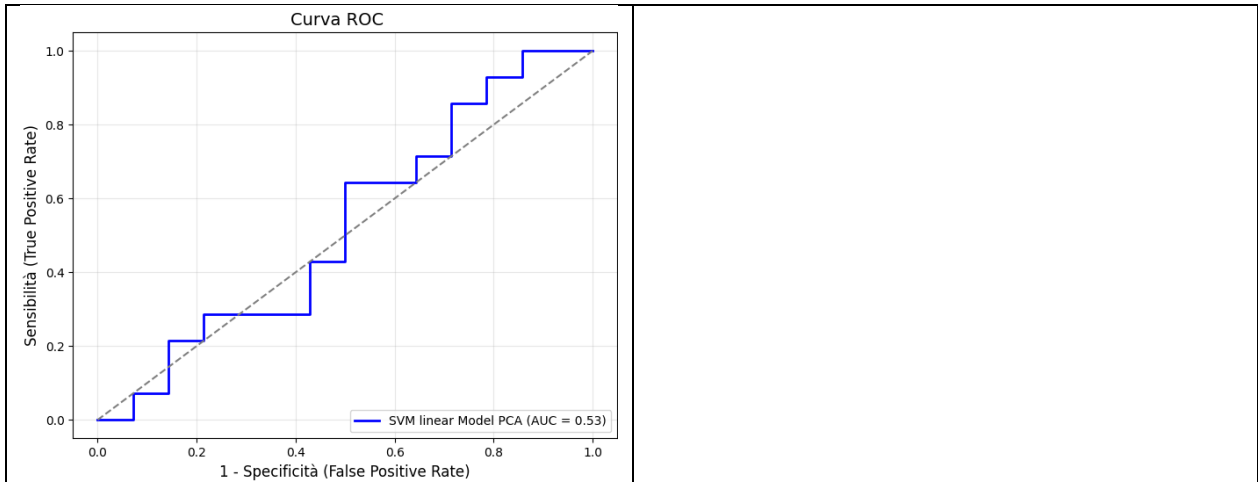


Tabella 53 Curve ROC

Il dataset che ottiene le prestazioni migliori risulta quello derivato dall' aggregazione dei pazienti mediante l'uso delle statistiche descrittive dell'istogramma. Il classificatore Gradient Boosting, utilizzato nel metodo di feature selection RFE, con il numero di feature identificate al di sopra della soglia 0.19, fornisce dei buoni risultati come evidenziato nelle metriche calcolate e riportate nelle tabelle (tab. 54)

True	Predicted	
	Classe 0	Classe 1
Classe 0	10	4
Classe 1	2	12

Sensitivity %	Specificity %	PPV %	NPV %
87	69	75	83

Tabella 54 Confusion matrix e metriche Gradient Boosting

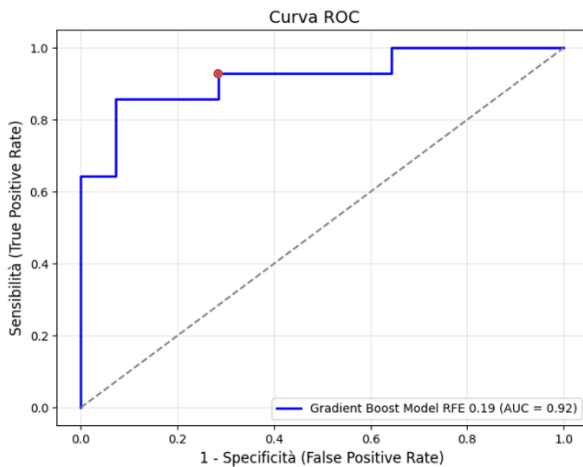


Figura 36 Curva Roc Gradient Boosting RFE 0.19

Nella figura adiacente (fig. 36) è possibile osservare la curva ROC che evidenzia le prestazioni del Gradient Boosting in termini di Sensibilità (True positive rate) e Specificità (1-False positive rate). La scelta della soglia è importante e dipende dal contesto clinico che si vuole attenzionare: un FPR elevato significa che molti pazienti non responder vengono classificati erroneamente come responder (Falsi positivi) e che quindi, non solo non hanno beneficiato del trattamento, ma non vengono neanche sottoposti all'intervento chirurgico, che generalmente viene eseguito in

caso di fallimento del trattamento, con potenziali effetti negativi sul loro esito clinico. Si individua, quindi, nel grafico, una soglia che rappresenti il massimo compromesso tra un'elevata Sensibilità e un'altrettanto alta Specificità (ovvero, un valore basso di FPR)



# Conclusioni

In questo studio è stata valutata la possibilità di sviluppare un biomarcatore per predire la risposta patologica nei pazienti con cancro al retto. A tal fine, sono state estratte le feature da immagini istologiche e, dopo aver eseguito diversi metodi di feature selection, sono stati allenati, con diversi classificatori, i modelli predittivi.

Il lavoro è stato suddiviso in due fasi, in quanto, la maggior parte delle immagini istologiche che si avevano a disposizione non presentavano la segmentazione manuale, utile per poter identificare i tessuti affetti da tumore da cui estrarre le feature.

Nella prima fase sono stati allenati tre modelli di rete neurale con i 20 pazienti di cui si avevano le segmentazioni manuali, sfruttando le tiles di dimensioni 224x224 pixel ad alta risoluzione (0.5  $\mu\text{m}/\text{pixel}$ ), a cui è stata applicata la color deconvolution: ResNet18, ResNet50, e VGG19, che era stata precedentemente allenata su un task simile. Dopo una fase di tuning dei parametri, le prestazioni migliori sono state ottenute dalla VGG19, con un'accuratezza nel Training set del 79% e nel Test set del 82%. Poiché il dataset risultava essere sbilanciato, è stata calcolata anche l'accuratezza bilanciata, pari al 74%.

Nella seconda fase dello studio sono state estratte le tiles dai nuovi pazienti e segmentati automaticamente dalla VGG19 precedentemente addestrata. Per ottenere feature più informative, derivanti da porzioni di tessuto più omogenee, si è deciso di estrarle direttamente dalle tiles e conseguire solo successivamente l'aggregazione, per lavorare con una rappresentazione globale del paziente. Sono stati proposti due metodi di aggregazione che si basano su statistiche robuste, per mantenere la variabilità e la distribuzione delle feature a livello di paziente:

- Calcolo della media per ciascuna feature
- Calcolo delle statistiche descrittive dell'istogramma per ciascuna feature: Media, Mediana, Skewness, Kurtosis, Percentili, range interquartili (IQR) ed Entropia

Sono stati valutati diversi metodi di feature selection, utilizzando diversi classificatori al fine di confrontare le loro prestazioni e identificare le combinazioni più efficaci. Dai risultati si può affermare che i classificatori allenati nel dataset ottenuto dal metodo di aggregazione basato sulle statistiche descrittive dell'istogramma, hanno prodotto dei risultati migliori rispetto a quelli allenati sul dataset ottenuto dalla media delle feature: ciò significa che l'insieme delle statistiche descrittive è stato in grado di includere più informazioni relative alle feature, permettendo ai modelli di distinguere meglio le classi. In particolare, il Gradient Boosting, combinato con la selezione delle feature tramite RFE con una soglia  $\geq 0.19$ , si è dimostrato il modello più performante non solo nel dataset dell'istogramma, ma anche rispetto agli altri classificatori combinati con le altre feature selection nel dataset della media. Invece, per quest'ultimo, la feature selection più performante è stata la mRMR, indicando che questo metodo sfrutta la semplicità del metodo di aggregazione della media ed ottenere risultati apprezzabili. Nella feature selection della correlazione assoluta tra feature e target, c'è stato un aumento delle prestazioni, indicando che le feature aggiunte nel metodo di aggregazione

dell'istogramma forniscono una maggiore variabilità, riducendo la ridondanza tra le feature. Nell'affinity propagation, le 142 feature individuate nel dataset dell'istogramma hanno portato maggiore informazione rispetto alle 8 individuate dal dataset della media. Nel metodo PCA invece, in entrambi i dataset, i modelli non sono riusciti ad imparare, mantenendo le prestazioni intorno al 50% di accuratezza: ciò potrebbe essere causato da una probabile sovrapposizione tra le classi nelle componenti principali, impedendo ai modelli di apprendere i pattern specifici; inoltre, la PCA si basa sulla varianza e potrebbe aver escluso feature altamente rilevanti ma a bassa varianza. Gli studi clinici sulla previsione della risposta alla terapia nei pazienti con cancro al retto [6] [7] [8] presenti in letteratura hanno portato dei buoni risultati con l'utilizzo delle feature radiomiche, estratte da diverse tipologie di immagini cliniche, implicandone l'impiego come potenti biomarcatori. Ciò in cui si differenzia questo studio è l'estrazione delle feature radiomiche a livello di tiles di immagini istologiche, permettendo di cogliere dettagli peculiari che potrebbero sfuggire a metodi che analizzano intere immagini (WSI), e l'utilizzo di metodi di aggregazione che offrono una rappresentazione robusta del paziente, mantenendo comunque variabilità senza perdere eccessivamente informazioni.

Un aspetto da attenzionare e migliore riguarda sicuramente il modello VGG19, allenato su un dataset limitato: questo influisce sulla robustezza e generalizzabilità del modello e di conseguenza anche sulla qualità delle feature estratte dai tessuti tumorali. Inoltre, sebbene il metodo di aggregazione dell'istogramma abbia portato dei buoni risultati, altri metodi di aggregazione potrebbero essere valutati e migliorare ulteriormente le prestazioni.

In conclusione, questo studio ha posto le basi per lo sviluppo di un biomarcatore predittivo, aprendo nuove prospettive nell'uso dell'intelligenza artificiale per migliorare la personalizzazione dei trattamenti nei pazienti con cancro al retto.

# Appendice

Metodo di aggregazione: Media

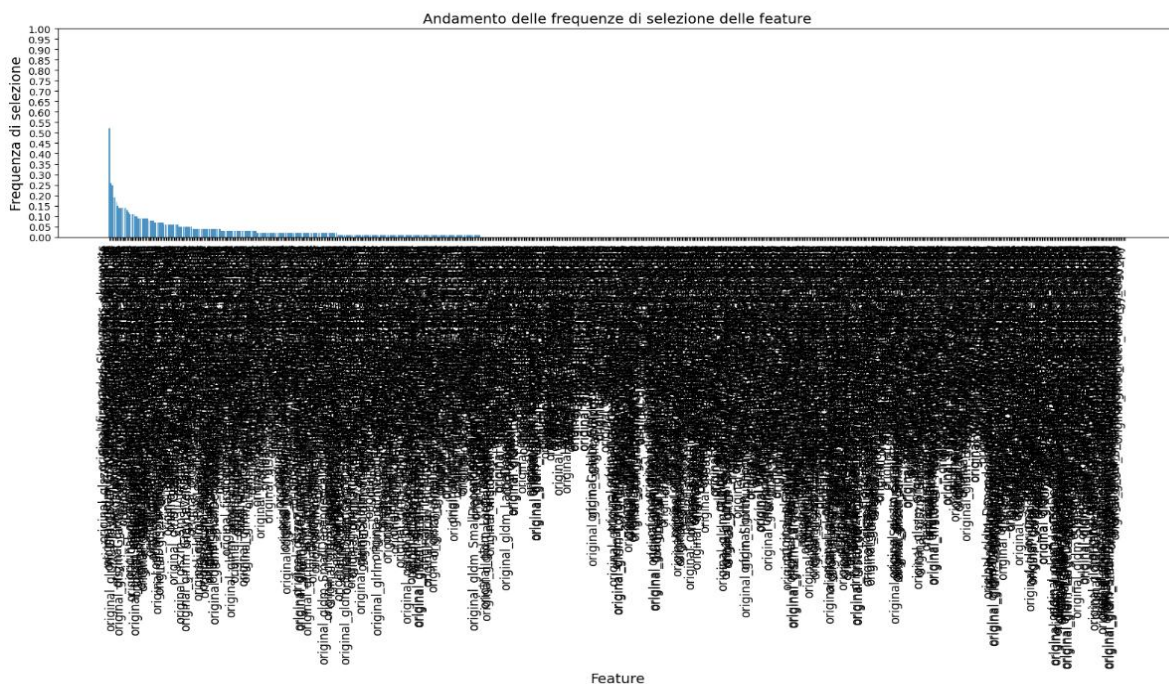
Feature selezionate con il metodo di correlazione assoluta tra feature e target, valutate dal classificatore SGD:

- original\_glrlm\_GrayLevelNonUniformityNormalized,
- original\_glszm\_GrayLevelNonUniformityNormalized,
- original\_glrlm\_LowGrayLevelRunEmphasis,
- original\_glszm\_LowGrayLevelZoneEmphasis,
- original\_glszm\_SmallAreaLowGrayLevelEmphasis,
- original\_glrlm\_ShortRunLowGrayLevelEmphasis,
- original\_firstorder\_Skewness,
- original\_firstorder\_Kurtosis,
- original\_gldm\_DependenceEntropy,
- original\_glrlm\_GrayLevelNonUniformity,
- original\_glszm\_GrayLevelNonUniformity,
- original\_firstorder\_10Percentile,
- original\_ngtdm\_Contrast,
- original\_glszm\_SizeZoneNonUniformityNormalized,
- original\_glrlm\_RunLengthNonUniformityNormalized,
- original\_glcm\_DifferenceAverage,
- original\_glcm\_Idmn,
- original\_glcm\_DifferenceVariance,
- original\_glrlm\_RunEntropy,
- original\_glszm\_ZoneEntropy,
- original\_glszm\_SmallAreaEmphasis,
- original\_glcm\_Contrast'
- original\_ngtdm\_Busyness,
- original\_firstorder\_Energy,
- original\_ngtdm\_Coarseness,
- original\_gldm\_DependenceNonUniformity,
- original\_glcm\_Imc2,
- original\_glszm\_ZoneVariance,
- original\_glrlm\_RunVariance,
- original\_glcm\_JointEntropy,
- original\_glcm\_Idn,
- original\_glcm\_ClusterProminence,
- original\_firstorder\_Mean,
- original\_firstorder\_TotalEnergy,

- original\_glszm\_LargeAreaEmphasis,
- original\_glrmlm\_LongRunEmphasis

## Metodo di aggregazione: Istogramma

Frequenza di selezione delle feature, metodo RFE:



- Feature: original\_firstorder\_Skewness\_skewness, Selection Frequency: 0.52
- Feature: original\_glszm\_SmallAreaEmphasis\_25th\_percentile, Selection Frequency: 0.26
- Feature: original\_firstorder\_Minimum\_kurtosis, Selection Frequency: 0.25
- Feature: original\_glszm\_GrayLevelVariance\_iqr, Selection Frequency: 0.19
- Feature: original\_glrmlm\_GrayLevelNonUniformity\_75th\_percentile, Selection Frequency: 0.19
- Feature: original\_glszm\_GrayLevelNonUniformity\_75th\_percentile, Selection Frequency: 0.17
- Feature: original\_glrmlm\_GrayLevelNonUniformityNormalized\_25th\_percentile, Selection Frequency: 0.15
- Feature: original\_firstorder\_Minimum\_iqr, Selection Frequency: 0.14
- Feature: original\_glszm\_SmallAreaEmphasis\_mean, Selection Frequency: 0.14
- Feature: original\_glrmlm\_GrayLevelVariance\_iqr, Selection Frequency: 0.14
- Feature: original\_glrmlm\_GrayLevelNonUniformity\_kurtosis, Selection Frequency: 0.14
- Feature: original\_glrmlm\_GrayLevelNonUniformityNormalized\_kurtosis, Selection Frequency: 0.14
- Feature: original\_glszm\_GrayLevelNonUniformityNormalized\_25th\_percentile, Selection Frequency: 0.14

- Feature: original\_glrmlm\_RunEntropy\_75th\_percentile, Selection Frequency: 0.13
- Feature: original\_firstorder\_Range\_iqr, Selection Frequency: 0.12
- Feature: original\_glrmlm\_LowGrayLevelRunEmphasis\_mean, Selection Frequency: 0.11
- Feature: original\_ngtdm\_Contrast\_25th\_percentile, Selection Frequency: 0.11
- Feature: original\_firstorder\_10Percentile\_25th\_percentile, Selection Frequency: 0.11
- Feature: original\_glrmlm\_ShortRunEmphasis\_mean, Selection Frequency: 0.11
- Feature: original\_glszm\_ZoneEntropy\_75th\_percentile, Selection Frequency: 0.1
- Feature: original\_glszm\_GrayLevelNonUniformityNormalized\_mean, Selection Frequency: 0.1
- Feature: original\_glszm\_GrayLevelNonUniformityNormalized\_skewness, Selection Frequency: 0.09
- Feature: original\_glrmlm\_ShortRunLowGrayLevelEmphasis\_iqr, Selection Frequency: 0.09
- Feature: original\_glszm\_LowGrayLevelZoneEmphasis\_kurtosis, Selection Frequency: 0.09
- Feature: original\_glszm\_GrayLevelNonUniformityNormalized\_75th\_percentile, Selection Frequency: 0.09
- Feature: original\_glrmlm\_GrayLevelNonUniformityNormalized\_skewness, Selection Frequency: 0.09
- Feature: original\_glrmlm\_LowGrayLevelRunEmphasis\_skewness, Selection Frequency: 0.09
- Feature: original\_glrmlm\_GrayLevelNonUniformityNormalized\_mean, Selection Frequency: 0.09
- Feature: original\_glszm\_LowGrayLevelZoneEmphasis\_mean, Selection Frequency: 0.09
- Feature: original\_glszm\_GrayLevelNonUniformityNormalized\_median, Selection Frequency: 0.08
- Feature: original\_glszm\_GrayLevelNonUniformityNormalized\_kurtosis, Selection Frequency: 0.08
- Feature: original\_glrmlm\_ShortRunEmphasis\_25th\_percentile, Selection Frequency: 0.08
- Feature: original\_glrmlm\_LowGrayLevelRunEmphasis\_kurtosis, Selection Frequency: 0.08
- Feature: original\_glrmlm\_GrayLevelVariance\_skewness, Selection Frequency: 0.07
- Feature: original\_firstorder\_Skewness\_median, Selection Frequency: 0.07
- Feature: original\_glrmlm\_ShortRunLowGrayLevelEmphasis\_mean, Selection Frequency: 0.07
- Feature: original\_glszm\_GrayLevelNonUniformity\_kurtosis, Selection Frequency: 0.07
- Feature: original\_glszm\_LowGrayLevelZoneEmphasis\_skewness, Selection Frequency: 0.07
- Feature: original\_firstorder\_Median\_kurtosis, Selection Frequency: 0.07
- Feature: original\_glrmlm\_ShortRunLowGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.07
- Feature: original\_glrmlm\_ShortRunLowGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.06
- Feature: original\_glszm\_LowGrayLevelZoneEmphasis\_iqr, Selection Frequency: 0.06
- Feature: original\_glszm\_LowGrayLevelZoneEmphasis\_25th\_percentile, Selection Frequency: 0.06
- Feature: original\_firstorder\_Skewness\_25th\_percentile, Selection Frequency: 0.06
- Feature: original\_gldm\_LowGrayLevelEmphasis\_iqr, Selection Frequency: 0.06
- Feature: original\_firstorder\_10Percentile\_median, Selection Frequency: 0.06
- Feature: original\_glrmlm\_RunPercentage\_skewness, Selection Frequency: 0.06
- Feature: original\_glcm\_JointEnergy\_25th\_percentile, Selection Frequency: 0.06

- Feature: original\_glszm\_GrayLevelNonUniformity\_skewness, Selection Frequency: 0.06
- Feature: original\_glszm\_ZonePercentage\_skewness, Selection Frequency: 0.06
- Feature: original\_glrlm\_GrayLevelNonUniformity\_skewness, Selection Frequency: 0.06
- Feature: original\_glrlm\_RunLengthNonUniformity\_iqr, Selection Frequency: 0.05
- Feature: original\_glrlm\_GrayLevelNonUniformityNormalized\_median, Selection Frequency: 0.05
- Feature: original\_glrlm\_ShortRunEmphasis\_median, Selection Frequency: 0.05
- Feature: original\_firstorder\_Variance\_median, Selection Frequency: 0.05
- Feature: original\_ngtdm\_Strength\_kurtosis, Selection Frequency: 0.05
- Feature: original\_glcm\_JointAverage\_25th\_percentile, Selection Frequency: 0.05
- Feature: original\_glszm\_SmallAreaLowGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.05
- Feature: original\_ngtdm\_Coarseness\_kurtosis, Selection Frequency: 0.05
- Feature: original\_ngtdm\_Busyness\_skewness, Selection Frequency: 0.05
- Feature: original\_gldm\_DependenceEntropy\_mean, Selection Frequency: 0.05
- Feature: original\_glrlm\_GrayLevelNonUniformityNormalized\_75th\_percentile, Selection Frequency: 0.04
- Feature: original\_glszm\_SmallAreaLowGrayLevelEmphasis\_iqr, Selection Frequency: 0.04
- Feature: original\_glrlm\_RunPercentage\_kurtosis, Selection Frequency: 0.04
- Feature: original\_firstorder\_10Percentile\_skewness, Selection Frequency: 0.04
- Feature: original\_gldm\_DependenceNonUniformity\_kurtosis, Selection Frequency: 0.04
- Feature: original\_glcm\_JointAverage\_iqr, Selection Frequency: 0.04
- Feature: original\_glcm\_SumEntropy\_skewness, Selection Frequency: 0.04
- Feature: original\_firstorder\_Kurtosis\_iqr, Selection Frequency: 0.04
- Feature: original\_glcm\_Correlation\_kurtosis, Selection Frequency: 0.04
- Feature: original\_glrlm\_RunLengthNonUniformityNormalized\_skewness, Selection Frequency: 0.04
- Feature: original\_glszm\_SmallAreaLowGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.04
- Feature: original\_glcm\_Idmn\_iqr, Selection Frequency: 0.04
- Feature: original\_glszm\_SmallAreaEmphasis\_skewness, Selection Frequency: 0.04
- Feature: original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_iqr, Selection Frequency: 0.04
- Feature: original\_glszm\_HighGrayLevelZoneEmphasis\_75th\_percentile, Selection Frequency: 0.04
- Feature: original\_glszm\_GrayLevelVariance\_skewness, Selection Frequency: 0.04
- Feature: original\_glrlm\_LongRunHighGrayLevelEmphasis\_skewness, Selection Frequency: 0.04
- Feature: original\_glrlm\_LowGrayLevelRunEmphasis\_25th\_percentile, Selection Frequency: 0.04
- Feature: original\_glszm\_SmallAreaLowGrayLevelEmphasis\_mean, Selection Frequency: 0.04
- Feature: original\_glrlm\_ShortRunEmphasis\_iqr, Selection Frequency: 0.04

- Feature: original\_glszm\_SmallAreaHighGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.04
- Feature: original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.03
- Feature: original\_firstorder\_Range\_25th\_percentile, Selection Frequency: 0.03
- Feature: original\_firstorder\_Entropy\_skewness, Selection Frequency: 0.03
- Feature: original\_ngtdm\_Busyness\_75th\_percentile, Selection Frequency: 0.03
- Feature: original\_firstorder\_Kurtosis\_75th\_percentile, Selection Frequency: 0.03
- Feature: original\_firstorder\_Kurtosis\_skewness, Selection Frequency: 0.03
- Feature: original\_firstorder\_Kurtosis\_kurtosis, Selection Frequency: 0.03
- Feature: original\_glcm\_Imc2\_skewness, Selection Frequency: 0.03
- Feature: original\_glszm\_GrayLevelVariance\_kurtosis, Selection Frequency: 0.03
- Feature: original\_firstorder\_InterquartileRange\_25th\_percentile, Selection Frequency: 0.03
- Feature: original\_firstorder\_10Percentile\_mean, Selection Frequency: 0.03
- Feature: original\_glcm\_Contrast\_iqr, Selection Frequency: 0.03
- Feature: original\_glrmlm\_ShortRunHighGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.03
- Feature: original\_glcm\_DifferenceVariance\_25th\_percentile, Selection Frequency: 0.03
- Feature: original\_glcm\_Autocorrelation\_kurtosis, Selection Frequency: 0.03
- Feature: original\_glszm\_LargeAreaHighGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.03
- Feature: original\_glszm\_SmallAreaEmphasis\_median, Selection Frequency: 0.03
- Feature: original\_glszm\_GrayLevelVariance\_75th\_percentile, Selection Frequency: 0.03
- Feature: original\_ngtdm\_Contrast\_kurtosis, Selection Frequency: 0.03
- Feature: original\_glrmlm\_GrayLevelVariance\_75th\_percentile, Selection Frequency: 0.03
- Feature: original\_firstorder\_Range\_mean, Selection Frequency: 0.03
- Feature: original\_glrmlm\_HighGrayLevelRunEmphasis\_75th\_percentile, Selection Frequency: 0.03
- Feature: original\_glszm\_SmallAreaEmphasis\_iqr, Selection Frequency: 0.03
- Feature: original\_glrmlm\_LowGrayLevelRunEmphasis\_75th\_percentile, Selection Frequency: 0.03
- Feature: original\_firstorder\_Skewness\_kurtosis, Selection Frequency: 0.03
- Feature: original\_glrmlm\_LowGrayLevelRunEmphasis\_iqr, Selection Frequency: 0.03
- Feature: original\_glszm\_ZonePercentage\_mean, Selection Frequency: 0.02
- Feature: original\_ngtdm\_Busyness\_mean, Selection Frequency: 0.02
- Feature: original\_gldm\_DependenceEntropy\_median, Selection Frequency: 0.02
- Feature: original\_ngtdm\_Coarseness\_skewness, Selection Frequency: 0.02
- Feature: original\_glcm\_MCC\_iqr, Selection Frequency: 0.02
- Feature: original\_glcm\_MaximumProbability\_skewness, Selection Frequency: 0.02
- Feature: original\_ngtdm\_Coarseness\_median, Selection Frequency: 0.02
- Feature: original\_glcm\_InverseVariance\_iqr, Selection Frequency: 0.02
- Feature: original\_firstorder\_MeanAbsoluteDeviation\_median, Selection Frequency: 0.02

- Feature: original\_firstorder\_Mean\_25th\_percentile, Selection Frequency: 0.02
- Feature: original\_ngtdm\_Busyness\_kurtosis, Selection Frequency: 0.02
- Feature: original\_firstorder\_Median\_25th\_percentile, Selection Frequency: 0.02
- Feature: original\_glcm\_Imc2\_iqr, Selection Frequency: 0.02
- Feature: original\_glcm\_Idmn\_kurtosis, Selection Frequency: 0.02
- Feature: original\_glszm\_ZonePercentage\_25th\_percentile, Selection Frequency: 0.02
- Feature: original\_glszm\_ZoneEntropy\_iqr, Selection Frequency: 0.02
- Feature: original\_glszm\_ZoneEntropy\_kurtosis, Selection Frequency: 0.02
- Feature: original\_glszm\_ZoneEntropy\_mean, Selection Frequency: 0.02
- Feature: original\_glszm\_ZoneEntropy\_median, Selection Frequency: 0.02
- Feature: original\_glcm\_Idmn\_skewness, Selection Frequency: 0.02
- Feature: original\_glcm\_Idm\_skewness, Selection Frequency: 0.02
- Feature: original\_grlm\_GrayLevelVariance\_kurtosis, Selection Frequency: 0.02
- Feature: original\_glszm\_SizeZoneNonUniformity\_mean, Selection Frequency: 0.02
- Feature: original\_firstorder\_Range\_75th\_percentile, Selection Frequency: 0.02
- Feature: original\_grlm\_LongRunEmphasis\_25th\_percentile, Selection Frequency: 0.02
- Feature: original\_firstorder\_RobustMeanAbsoluteDeviation\_iqr, Selection Frequency: 0.02
- Feature: original\_glszm\_SmallAreaHighGrayLevelEmphasis\_iqr, Selection Frequency: 0.02
- Feature: original\_glszm\_LowGrayLevelZoneEmphasis\_75th\_percentile, Selection Frequency: 0.02
- Feature: original\_glcm\_DifferenceVariance\_mean, Selection Frequency: 0.02
- Feature: original\_glcm\_Id\_mean, Selection Frequency: 0.02
- Feature: original\_glcm\_Id\_skewness, Selection Frequency: 0.02
- Feature: original\_glszm\_SizeZoneNonUniformity\_iqr, Selection Frequency: 0.02
- Feature: original\_firstorder\_Skewness\_75th\_percentile, Selection Frequency: 0.02
- Feature: original\_firstorder\_TotalEnergy\_iqr, Selection Frequency: 0.02
- Feature: original\_firstorder\_TotalEnergy\_median, Selection Frequency: 0.02
- Feature: original\_glszm\_SizeZoneNonUniformityNormalized\_skewness, Selection Frequency: 0.02
- Feature: original\_glszm\_SizeZoneNonUniformityNormalized\_75th\_percentile, Selection Frequency: 0.02
- Feature: original\_glszm\_SizeZoneNonUniformityNormalized\_25th\_percentile, Selection Frequency: 0.02
- Feature: original\_glcm\_DifferenceEntropy\_mean, Selection Frequency: 0.02
- Feature: original\_grlm\_RunLengthNonUniformity\_25th\_percentile, Selection Frequency: 0.02
- Feature: original\_grlm\_RunEntropy\_kurtosis, Selection Frequency: 0.02
- Feature: original\_grlm\_RunLengthNonUniformity\_mean, Selection Frequency: 0.02
- Feature: original\_grlm\_RunLengthNonUniformityNormalized\_25th\_percentile, Selection Frequency: 0.02
- Feature: original\_grlm\_ShortRunEmphasis\_kurtosis, Selection Frequency: 0.02
- Feature: original\_firstorder\_Variance\_iqr, Selection Frequency: 0.02
- Feature: original\_glszm\_HighGrayLevelZoneEmphasis\_median, Selection Frequency: 0.02



- Feature: original\_glrIm\_ShortRunHighGrayLevelEmphasis\_skewness, Selection Frequency: 0.02
- Feature: original\_glrIm\_ShortRunHighGrayLevelEmphasis\_iqr, Selection Frequency: 0.02
- Feature: original\_glrIm\_ShortRunHighGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.02
- Feature: original\_glcm\_Correlation\_iqr, Selection Frequency: 0.02
- Feature: original\_glrIm\_ShortRunEmphasis\_skewness, Selection Frequency: 0.02
- Feature: original\_glrIm\_RunPercentage\_iqr, Selection Frequency: 0.02
- Feature: original\_glrIm\_RunPercentage\_mean, Selection Frequency: 0.02
- Feature: original\_firstorder\_90Percentile\_median, Selection Frequency: 0.02
- Feature: original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.02
- Feature: original\_glcm\_JointEnergy\_median, Selection Frequency: 0.02
- Feature: original\_firstorder\_Energy\_skewness, Selection Frequency: 0.02
- Feature: original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.02
- Feature: original\_firstorder\_10Percentile\_75th\_percentile, Selection Frequency: 0.02
- Feature: original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_mean, Selection Frequency: 0.01
- Feature: original\_gldm\_LowGrayLevelEmphasis\_skewness, Selection Frequency: 0.01
- Feature: original\_gldm\_LargeDependenceEmphasis\_skewness, Selection Frequency: 0.01
- Feature: original\_gldm\_GrayLevelVariance\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_gldm\_GrayLevelNonUniformity\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_gldm\_DependenceVariance\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_gldm\_DependenceNonUniformityNormalized\_kurtosis, Selection Frequency: 0.01
- Feature: original\_gldm\_DependenceNonUniformity\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_glcm\_JointAverage\_skewness, Selection Frequency: 0.01
- Feature: original\_glcm\_JointAverage\_median, Selection Frequency: 0.01
- Feature: original\_glcm\_JointAverage\_mean, Selection Frequency: 0.01
- Feature: original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_iqr, Selection Frequency: 0.01
- Feature: original\_firstorder\_90Percentile\_iqr, Selection Frequency: 0.01
- Feature: original\_firstorder\_10Percentile\_kurtosis, Selection Frequency: 0.01
- Feature: original\_glrIm\_RunLengthNonUniformityNormalized\_median, Selection Frequency: 0.01
- Feature: original\_glrIm\_RunLengthNonUniformityNormalized\_mean, Selection Frequency: 0.01
- Feature: original\_glcm\_DifferenceAverage\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_glcm\_Correlation\_skewness, Selection Frequency: 0.01
- Feature: original\_glcm\_Correlation\_median, Selection Frequency: 0.01

- Feature: original\_glcm\_Correlation\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_glcm\_Correlation\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glrmlm\_ShortRunLowGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glrmlm\_ShortRunHighGrayLevelEmphasis\_median, Selection Frequency: 0.01
- Feature: original\_glrmlm\_ShortRunHighGrayLevelEmphasis\_mean, Selection Frequency: 0.01
- Feature: original\_glcm\_Contrast\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glcm\_ClusterTendency\_median, Selection Frequency: 0.01
- Feature: original\_glcm\_ClusterTendency\_iqr, Selection Frequency: 0.01
- Feature: original\_glcm\_ClusterTendency\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_glcm\_ClusterProminence\_skewness, Selection Frequency: 0.01
- Feature: original\_glcm\_Autocorrelation\_median, Selection Frequency: 0.01
- Feature: original\_glcm\_Contrast\_mean, Selection Frequency: 0.01
- Feature: original\_glcm\_Contrast\_kurtosis, Selection Frequency: 0.01
- Feature: original\_glrmlm\_RunLengthNonUniformityNormalized\_iqr, Selection Frequency: 0.01
- Feature: original\_glrmlm\_RunLengthNonUniformityNormalized\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_glrmlm\_RunEntropy\_skewness, Selection Frequency: 0.01
- Feature: original\_glrmlm\_RunEntropy\_median, Selection Frequency: 0.01
- Feature: original\_glrmlm\_RunEntropy\_mean, Selection Frequency: 0.01
- Feature: original\_glcm\_DifferenceVariance\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_glcm\_DifferenceEntropy\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_glszm\_LargeAreaEmphasis\_entropy, Selection Frequency: 0.01
- Feature: original\_glszm\_LargeAreaEmphasis\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glszm\_HighGrayLevelZoneEmphasis\_skewness, Selection Frequency: 0.01
- Feature: original\_glszm\_GrayLevelVariance\_mean, Selection Frequency: 0.01
- Feature: original\_glcm\_Autocorrelation\_iqr, Selection Frequency: 0.01
- Feature: original\_glcm\_Autocorrelation\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glszm\_GrayLevelVariance\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glrmlm\_LongRunLowGrayLevelEmphasis\_mean, Selection Frequency: 0.01
- Feature: original\_glcm\_Idm\_entropy, Selection Frequency: 0.01
- Feature: original\_glcm\_Id\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glcm\_DifferenceVariance\_kurtosis, Selection Frequency: 0.01
- Feature: original\_glrmlm\_LongRunHighGrayLevelEmphasis\_median, Selection Frequency: 0.01
- Feature: original\_firstorder\_Uniformity\_iqr, Selection Frequency: 0.01
- Feature: original\_firstorder\_TotalEnergy\_skewness, Selection Frequency: 0.01
- Feature: original\_firstorder\_Skewness\_mean, Selection Frequency: 0.01
- Feature: original\_glszm\_LargeAreaLowGrayLevelEmphasis\_iqr, Selection Frequency: 0.01
- Feature: original\_glszm\_LargeAreaLowGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glszm\_LargeAreaHighGrayLevelEmphasis\_mean, Selection Frequency: 0.01
- Feature: original\_firstorder\_Minimum\_mean, Selection Frequency: 0.01

- Feature: original\_glszm\_SizeZoneNonUniformityNormalized\_median, Selection Frequency: 0.01
- Feature: original\_glszm\_SizeZoneNonUniformityNormalized\_mean, Selection Frequency: 0.01
- Feature: original\_firstorder\_RootMeanSquared\_kurtosis, Selection Frequency: 0.01
- Feature: original\_firstorder\_RootMeanSquared\_iqr, Selection Frequency: 0.01
- Feature: original\_glszm\_SmallAreaHighGrayLevelEmphasis\_mean, Selection Frequency: 0.01
- Feature: original\_glszm\_SmallAreaHighGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_glszm\_SmallAreaHighGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glszm\_SmallAreaEmphasis\_kurtosis, Selection Frequency: 0.01
- Feature: original\_firstorder\_Range\_median, Selection Frequency: 0.01
- Feature: original\_firstorder\_Range\_kurtosis, Selection Frequency: 0.01
- Feature: original\_glrIm\_LongRunHighGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.01
- Feature: original\_glrIm\_LongRunEmphasis\_mean, Selection Frequency: 0.01
- Feature: original\_glrIm\_HighGrayLevelRunEmphasis\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glcM\_Idmn\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_glszm\_ZoneEntropy\_skewness, Selection Frequency: 0.01
- Feature: original\_firstorder\_Median\_mean, Selection Frequency: 0.01
- Feature: original\_glszm\_SmallAreaHighGrayLevelEmphasis\_skewness, Selection Frequency: 0.01
- Feature: original\_glrIm\_GrayLevelNonUniformityNormalized\_iqr, Selection Frequency: 0.01
- Feature: original\_glrIm\_GrayLevelNonUniformity\_median, Selection Frequency: 0.01
- Feature: original\_glcM\_Imc1\_median, Selection Frequency: 0.01
- Feature: original\_glrIm\_GrayLevelNonUniformity\_iqr, Selection Frequency: 0.01
- Feature: original\_glcM\_SumSquares\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glcM\_SumEntropy\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glcM\_SumAverage\_skewness, Selection Frequency: 0.01
- Feature: original\_ngtdm\_Coarseness\_iqr, Selection Frequency: 0.01
- Feature: original\_ngtdm\_Coarseness\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_ngtdm\_Coarseness\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_firstorder\_Mean\_iqr, Selection Frequency: 0.01
- Feature: original\_firstorder\_Maximum\_mean, Selection Frequency: 0.01
- Feature: original\_ngtdm\_Busyness\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glszm\_ZonePercentage\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_gldm\_DependenceEntropy\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_ngtdm\_Contrast\_mean, Selection Frequency: 0.01
- Feature: original\_ngtdm\_Contrast\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_ngtdm\_Complexity\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_ngtdm\_Coarseness\_mean, Selection Frequency: 0.01
- Feature: original\_glcM\_SumAverage\_iqr, Selection Frequency: 0.01

- Feature: original\_glcm\_SumAverage\_25th\_percentile, Selection Frequency: 0.01
- Feature: original\_glcm\_MaximumProbability\_median, Selection Frequency: 0.01
- Feature: original\_glcm\_MaximumProbability\_iqr, Selection Frequency: 0.01
- Feature: original\_glcm\_MCC\_skewness, Selection Frequency: 0.01
- Feature: original\_glcm\_MCC\_kurtosis, Selection Frequency: 0.01
- Feature: original\_glcm\_JointEntropy\_median, Selection Frequency: 0.01
- Feature: original\_glcm\_InverseVariance\_skewness, Selection Frequency: 0.01
- Feature: original\_glcm\_InverseVariance\_median, Selection Frequency: 0.01
- Feature: original\_glcm\_JointEntropy\_75th\_percentile, Selection Frequency: 0.01
- Feature: original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_90Percentile\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_90Percentile\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_90Percentile\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_90Percentile\_kurtosis, Selection Frequency: 0.0
- Feature: original\_firstorder\_90Percentile\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_90Percentile\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_Energy\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_firstorder\_InterquartileRange\_iqr, Selection Frequency: 0.0
- Feature: original\_firstorder\_InterquartileRange\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_InterquartileRange\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Entropy\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_Entropy\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_Entropy\_kurtosis, Selection Frequency: 0.0
- Feature: original\_firstorder\_Entropy\_iqr, Selection Frequency: 0.0
- Feature: original\_firstorder\_Entropy\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_Entropy\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Entropy\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Energy\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_Energy\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_Energy\_kurtosis, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_glcm\_JointAverage\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_InverseVariance\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_InverseVariance\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_InverseVariance\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_InverseVariance\_75th\_percentile, Selection Frequency: 0.0

- Feature: original\_glcm\_InverseVariance\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc2\_median, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc2\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc2\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_JointEntropy\_iqr, Selection Frequency: 0.0
- Feature: original\_glcm\_JointEntropy\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_JointEntropy\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_JointEnergy\_skewness, Selection Frequency: 0.0
- Feature: original\_glcm\_JointEnergy\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_JointEnergy\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_JointEnergy\_iqr, Selection Frequency: 0.0
- Feature: original\_glcm\_SumAverage\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_SumAverage\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_MaximumProbability\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_MaximumProbability\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_MaximumProbability\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_MaximumProbability\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_MaximumProbability\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_MCC\_median, Selection Frequency: 0.0
- Feature: original\_glcm\_MCC\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_MCC\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_MCC\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_MCC\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_JointEntropy\_skewness, Selection Frequency: 0.0
- Feature: original\_glcm\_JointEntropy\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_JointEntropy\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_JointAverage\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_SumSquares\_skewness, Selection Frequency: 0.0
- Feature: original\_glcm\_SumSquares\_median, Selection Frequency: 0.0
- Feature: original\_glcm\_SumSquares\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_SumSquares\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_SumSquares\_iqr, Selection Frequency: 0.0
- Feature: original\_glcm\_SumSquares\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_SumSquares\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_SumEntropy\_median, Selection Frequency: 0.0
- Feature: original\_glcm\_SumEntropy\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_SumEntropy\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_SumEntropy\_iqr, Selection Frequency: 0.0
- Feature: original\_glcm\_SumEntropy\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_SumEntropy\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_SumAverage\_median, Selection Frequency: 0.0

- Feature: original\_glcm\_SumAverage\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_SumAverage\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc2\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc2\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc2\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc1\_skewness, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc1\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc1\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc1\_iqr, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc1\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc1\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_Imc1\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_Idn\_skewness, Selection Frequency: 0.0
- Feature: original\_glcm\_Idn\_median, Selection Frequency: 0.0
- Feature: original\_glcm\_Idn\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_Idn\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glrlm\_GrayLevelNonUniformity\_entropy, Selection Frequency: 0.0
- Feature: original\_glrlm\_GrayLevelNonUniformity\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glrlm\_GrayLevelVariance\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_Idn\_iqr, Selection Frequency: 0.0
- Feature: original\_glcm\_Idn\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_Idn\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_Idn\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_Idmn\_median, Selection Frequency: 0.0
- Feature: original\_glcm\_Idmn\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_Idmn\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_Idmn\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_Idm\_median, Selection Frequency: 0.0
- Feature: original\_glcm\_Idm\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_Idm\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glrlm\_GrayLevelVariance\_entropy, Selection Frequency: 0.0
- Feature: original\_glrlm\_GrayLevelVariance\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glrlm\_GrayLevelNonUniformityNormalized\_entropy, Selection Frequency: 0.0
- Feature: original\_glrlm\_GrayLevelNonUniformity\_mean, Selection Frequency: 0.0
- Feature: original\_glrlm\_LongRunHighGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_glrlm\_LongRunHighGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glrlm\_LongRunHighGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glrlm\_LongRunEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_glrlm\_LongRunEmphasis\_median, Selection Frequency: 0.0

- Feature: original\_glrIm\_LongRunEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glrIm\_LongRunEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_glrIm\_LongRunEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_glrIm\_LongRunEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glrIm\_HighGrayLevelRunEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_glrIm\_HighGrayLevelRunEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_glrIm\_HighGrayLevelRunEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_glrIm\_HighGrayLevelRunEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glrIm\_HighGrayLevelRunEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_glrIm\_HighGrayLevelRunEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_glrIm\_GrayLevelVariance\_median, Selection Frequency: 0.0
- Feature: original\_glcM\_Idm\_iqr, Selection Frequency: 0.0
- Feature: original\_glcM\_Idm\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcM\_Idm\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcM\_Id\_median, Selection Frequency: 0.0
- Feature: original\_glcM\_Id\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcM\_Id\_iqr, Selection Frequency: 0.0
- Feature: original\_glcM\_Id\_entropy, Selection Frequency: 0.0
- Feature: original\_glcM\_Id\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcM\_DifferenceVariance\_skewness, Selection Frequency: 0.0
- Feature: original\_glcM\_DifferenceVariance\_median, Selection Frequency: 0.0
- Feature: original\_glrIm\_LongRunLowGrayLevelEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_glrIm\_LongRunLowGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_glrIm\_LongRunLowGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glrIm\_LongRunLowGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glrIm\_LongRunHighGrayLevelEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_glrIm\_LongRunHighGrayLevelEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_glcM\_DifferenceEntropy\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcM\_DifferenceEntropy\_iqr, Selection Frequency: 0.0
- Feature: original\_glcM\_DifferenceEntropy\_entropy, Selection Frequency: 0.0
- Feature: original\_glcM\_DifferenceEntropy\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcM\_DifferenceAverage\_skewness, Selection Frequency: 0.0
- Feature: original\_glcM\_DifferenceAverage\_median, Selection Frequency: 0.0
- Feature: original\_glcM\_DifferenceAverage\_mean, Selection Frequency: 0.0
- Feature: original\_glcM\_DifferenceAverage\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glrIm\_RunEntropy\_iqr, Selection Frequency: 0.0
- Feature: original\_glrIm\_RunEntropy\_entropy, Selection Frequency: 0.0
- Feature: original\_glrIm\_RunEntropy\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glrIm\_LowGrayLevelRunEmphasis\_median, Selection Frequency: 0.0

- Feature: original\_grlm\_LowGrayLevelRunEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_grlm\_LongRunLowGrayLevelEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_grlm\_LongRunLowGrayLevelEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_grlm\_LongRunLowGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_DifferenceAverage\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_DifferenceAverage\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_Correlation\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_Correlation\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_Contrast\_skewness, Selection Frequency: 0.0
- Feature: original\_glcm\_Contrast\_median, Selection Frequency: 0.0
- Feature: original\_grlm\_RunLengthNonUniformityNormalized\_entropy, Selection Frequency: 0.0
- Feature: original\_grlm\_RunLengthNonUniformity\_skewness, Selection Frequency: 0.0
- Feature: original\_grlm\_RunLengthNonUniformity\_median, Selection Frequency: 0.0
- Feature: original\_grlm\_RunLengthNonUniformity\_kurtosis, Selection Frequency: 0.0
- Feature: original\_grlm\_RunLengthNonUniformity\_entropy, Selection Frequency: 0.0
- Feature: original\_grlm\_RunLengthNonUniformity\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_DifferenceVariance\_iqr, Selection Frequency: 0.0
- Feature: original\_glcm\_DifferenceVariance\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_DifferenceEntropy\_skewness, Selection Frequency: 0.0
- Feature: original\_glcm\_DifferenceEntropy\_median, Selection Frequency: 0.0
- Feature: original\_grlm\_ShortRunEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_grlm\_ShortRunEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_grlm\_RunVariance\_skewness, Selection Frequency: 0.0
- Feature: original\_grlm\_RunVariance\_median, Selection Frequency: 0.0
- Feature: original\_grlm\_RunVariance\_mean, Selection Frequency: 0.0
- Feature: original\_grlm\_RunVariance\_kurtosis, Selection Frequency: 0.0
- Feature: original\_grlm\_RunVariance\_iqr, Selection Frequency: 0.0
- Feature: original\_grlm\_RunVariance\_entropy, Selection Frequency: 0.0
- Feature: original\_grlm\_RunVariance\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_grlm\_RunVariance\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_grlm\_RunPercentage\_median, Selection Frequency: 0.0
- Feature: original\_grlm\_RunPercentage\_entropy, Selection Frequency: 0.0
- Feature: original\_grlm\_RunPercentage\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_grlm\_RunPercentage\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_grlm\_RunLengthNonUniformityNormalized\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_DifferenceAverage\_iqr, Selection Frequency: 0.0
- Feature: original\_grlm\_ShortRunLowGrayLevelEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_grlm\_ShortRunLowGrayLevelEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_grlm\_ShortRunLowGrayLevelEmphasis\_entropy, Selection Frequency: 0.0



- Feature: original\_glcm\_Contrast\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_Contrast\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterTendency\_skewness, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterTendency\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterTendency\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterTendency\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterTendency\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterShade\_skewness, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterShade\_median, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterShade\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterShade\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glrIm\_ShortRunHighGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_glrIm\_ShortRunHighGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterShade\_iqr, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterShade\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterShade\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterShade\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterProminence\_median, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterProminence\_mean, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterProminence\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterProminence\_iqr, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterProminence\_entropy, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterProminence\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_ClusterProminence\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_Autocorrelation\_skewness, Selection Frequency: 0.0
- Feature: original\_glcm\_Autocorrelation\_mean, Selection Frequency: 0.0
- Feature: original\_glszm\_GrayLevelNonUniformity\_iqr, Selection Frequency: 0.0
- Feature: original\_glszm\_GrayLevelNonUniformity\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_GrayLevelNonUniformity\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glcm\_Autocorrelation\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Variance\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_Variance\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_Variance\_kurtosis, Selection Frequency: 0.0
- Feature: original\_firstorder\_Variance\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_Variance\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Variance\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Uniformity\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_Uniformity\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_Uniformity\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_Uniformity\_kurtosis, Selection Frequency: 0.0

- Feature: original\_glszm\_GrayLevelVariance\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_GrayLevelNonUniformityNormalized\_iqr, Selection Frequency: 0.0
- Feature: original\_glszm\_GrayLevelNonUniformityNormalized\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_GrayLevelNonUniformity\_median, Selection Frequency: 0.0
- Feature: original\_glszm\_GrayLevelNonUniformity\_mean, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaHighGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaHighGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaHighGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_HighGrayLevelZoneEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_glszm\_HighGrayLevelZoneEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glszm\_HighGrayLevelZoneEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_glszm\_HighGrayLevelZoneEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_HighGrayLevelZoneEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_GrayLevelVariance\_median, Selection Frequency: 0.0
- Feature: original\_glcm\_Autocorrelation\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaLowGrayLevelEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaLowGrayLevelEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaLowGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_firstorder\_Uniformity\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_Uniformity\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Uniformity\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_TotalEnergy\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_TotalEnergy\_kurtosis, Selection Frequency: 0.0
- Feature: original\_firstorder\_TotalEnergy\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_TotalEnergy\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_TotalEnergy\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaLowGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaLowGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaHighGrayLevelEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaHighGrayLevelEmphasis\_median, Selection Frequency: 0.0

- Feature: original\_glszm\_LargeAreaHighGrayLevelEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_firstorder\_RootMeanSquared\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_RootMeanSquared\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_RootMeanSquared\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_RootMeanSquared\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_RootMeanSquared\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_RootMeanSquared\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_RobustMeanAbsoluteDeviation\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_RobustMeanAbsoluteDeviation\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_RobustMeanAbsoluteDeviation\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_RobustMeanAbsoluteDeviation\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glszm\_SizeZoneNonUniformity\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_SizeZoneNonUniformity\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_SizeZoneNonUniformity\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_LowGrayLevelZoneEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_glszm\_LowGrayLevelZoneEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_LargeAreaLowGrayLevelEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_RobustMeanAbsoluteDeviation\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_RobustMeanAbsoluteDeviation\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Range\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_Range\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_Minimum\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_Minimum\_median, Selection Frequency: 0.0
- Feature: original\_glszm\_SmallAreaEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_SmallAreaEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_SizeZoneNonUniformityNormalized\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glszm\_SizeZoneNonUniformityNormalized\_iqr, Selection Frequency: 0.0
- Feature: original\_glszm\_SizeZoneNonUniformityNormalized\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_SizeZoneNonUniformity\_skewness, Selection Frequency: 0.0
- Feature: original\_glszm\_SizeZoneNonUniformity\_median, Selection Frequency: 0.0
- Feature: original\_glszm\_SizeZoneNonUniformity\_kurtosis, Selection Frequency: 0.0
- Feature: original\_firstorder\_Skewness\_iqr, Selection Frequency: 0.0
- Feature: original\_firstorder\_Skewness\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_Minimum\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Minimum\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Median\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_Median\_median, Selection Frequency: 0.0

- Feature: original\_firstorder\_Median\_iqr, Selection Frequency: 0.0
- Feature: original\_firstorder\_Median\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_Median\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Mean\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_Mean\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_Mean\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_Mean\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glszm\_SmallAreaLowGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_SmallAreaLowGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_SmallAreaHighGrayLevelEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_glszm\_SmallAreaHighGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_RobustMeanAbsoluteDeviation\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_ZoneVariance\_median, Selection Frequency: 0.0
- Feature: original\_glszm\_ZoneVariance\_mean, Selection Frequency: 0.0
- Feature: original\_glszm\_ZoneVariance\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glszm\_ZoneVariance\_iqr, Selection Frequency: 0.0
- Feature: original\_glszm\_ZoneVariance\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_ZoneVariance\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_ZoneVariance\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_ZonePercentage\_median, Selection Frequency: 0.0
- Feature: original\_glszm\_ZonePercentage\_kurtosis, Selection Frequency: 0.0
- Feature: original\_glszm\_ZonePercentage\_iqr, Selection Frequency: 0.0
- Feature: original\_glszm\_ZonePercentage\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_ZoneEntropy\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_ZoneEntropy\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_glszm\_SmallAreaLowGrayLevelEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_glszm\_SmallAreaLowGrayLevelEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_Minimum\_entropy, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Busyness\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_Mean\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_Mean\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_MeanAbsoluteDeviation\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_MeanAbsoluteDeviation\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_MeanAbsoluteDeviation\_kurtosis, Selection Frequency: 0.0
- Feature: original\_firstorder\_MeanAbsoluteDeviation\_iqr, Selection Frequency: 0.0
- Feature: original\_firstorder\_MeanAbsoluteDeviation\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_MeanAbsoluteDeviation\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_MeanAbsoluteDeviation\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Maximum\_skewness, Selection Frequency: 0.0

- Feature: original\_firstorder\_Maximum\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_Maximum\_kurtosis, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Busyness\_iqr, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Busyness\_entropy, Selection Frequency: 0.0
- Feature: original\_glszm\_ZoneVariance\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_Maximum\_iqr, Selection Frequency: 0.0
- Feature: original\_firstorder\_Maximum\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_Maximum\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Maximum\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_Kurtosis\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_Kurtosis\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_Kurtosis\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_Kurtosis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_firstorder\_InterquartileRange\_skewness, Selection Frequency: 0.0
- Feature: original\_firstorder\_InterquartileRange\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_InterquartileRange\_mean, Selection Frequency: 0.0
- Feature: original\_firstorder\_InterquartileRange\_kurtosis, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Complexity\_iqr, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Complexity\_entropy, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Complexity\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Coarseness\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceEntropy\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Strength\_skewness, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Strength\_median, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Strength\_mean, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Strength\_iqr, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Strength\_entropy, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Strength\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Strength\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Contrast\_skewness, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Contrast\_median, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Contrast\_iqr, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Contrast\_entropy, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Complexity\_skewness, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Complexity\_median, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Complexity\_mean, Selection Frequency: 0.0
- Feature: original\_ngtdm\_Complexity\_kurtosis, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformityNormalized\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformityNormalized\_mean, Selection Frequency: 0.0

- Feature: original\_gldm\_DependenceNonUniformityNormalized\_iqr, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformityNormalized\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformityNormalized\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformityNormalized\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformity\_skewness, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformity\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformity\_mean, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformity\_iqr, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformity\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformity\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceEntropy\_skewness, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceEntropy\_kurtosis, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceEntropy\_iqr, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceEntropy\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelVariance\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelNonUniformity\_skewness, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelNonUniformity\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelNonUniformity\_mean, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelNonUniformity\_kurtosis, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelNonUniformity\_iqr, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelNonUniformity\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelNonUniformity\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceVariance\_skewness, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceVariance\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceVariance\_mean, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceVariance\_kurtosis, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceVariance\_iqr, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceVariance\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceVariance\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_DependenceNonUniformityNormalized\_skewness, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_HighGrayLevelEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_gldm\_HighGrayLevelEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_HighGrayLevelEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_gldm\_HighGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_gldm\_HighGrayLevelEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_gldm\_HighGrayLevelEmphasis\_entropy, Selection Frequency: 0.0

- Feature: original\_gldm\_HighGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_HighGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelVariance\_skewness, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelVariance\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelVariance\_mean, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelVariance\_kurtosis, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelVariance\_iqr, Selection Frequency: 0.0
- Feature: original\_gldm\_GrayLevelVariance\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceHighGrayLevelEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceHighGrayLevelEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceHighGrayLevelEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceHighGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceHighGrayLevelEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceHighGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceHighGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceHighGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_Energy\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_firstorder\_10Percentile\_entropy, Selection Frequency: 0.0
- Feature: original\_firstorder\_10Percentile\_iqr, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceEmphasis\_iqr, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceEmphasis\_entropy, Selection Frequency: 0.0

- Feature: original\_gldm\_SmallDependenceEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_LowGrayLevelEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_LowGrayLevelEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_gldm\_LowGrayLevelEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_gldm\_LowGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_LowGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_LowGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_JointEnergy\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_JointEnergy\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_JointAverage\_kurtosis, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_entropy, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_75th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_25th\_percentile, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceEmphasis\_skewness, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceEmphasis\_median, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceEmphasis\_mean, Selection Frequency: 0.0
- Feature: original\_gldm\_SmallDependenceEmphasis\_kurtosis, Selection Frequency: 0.0
- Feature: original\_firstorder\_Energy\_iqr, Selection Frequency: 0.0
- Feature: original\_firstorder\_Energy\_entropy, Selection Frequency: 0.0

Feature selection: minimum Redundancy - maximum Relevance (mRmR)

Elenco 50 feature:

- original\_firstorder\_InterquartileRange\_kurtosis



- original\_firstorder\_Kurtosis\_25th\_percentile
- original\_firstorder\_Maximum\_25th\_percentile
- original\_firstorder\_MeanAbsoluteDeviation\_skewness
- original\_firstorder\_Median\_mean
- original\_firstorder\_Minimum\_median
- original\_firstorder\_Range\_kurtosis
- original\_firstorder\_Range\_skewness
- original\_firstorder\_RootMeanSquared\_skewness
- original\_firstorder\_Skewness\_skewness
- original\_firstorder\_TotalEnergy\_iqr
- original\_glcm\_ClusterProminence\_25th\_percentile
- original\_glcm\_DifferenceAverage\_25th\_percentile
- original\_glcm\_DifferenceEntropy\_kurtosis
- original\_glcm\_Idmn\_mean
- original\_glcm\_Imc2\_median
- original\_glcm\_Imc2\_skewness
- original\_glcm\_InverseVariance\_75th\_percentile
- original\_glcm\_JointEnergy\_median
- original\_glcm\_JointEntropy\_75th\_percentile
- original\_glcm\_SumSquares\_iqr
- original\_glrlm\_GrayLevelNonUniformity\_skewness
- original\_glrlm\_GrayLevelNonUniformityNormalized\_iqr
- original\_glrlm\_GrayLevelNonUniformityNormalized\_mean
- original\_glrlm\_GrayLevelVariance\_mean
- original\_glrlm\_LongRunLowGrayLevelEmphasis\_iqr
- original\_glrlm\_RunEntropy\_skewness
- original\_glrlm\_ShortRunEmphasis\_kurtosis
- original\_glrlm\_ShortRunHighGrayLevelEmphasis\_skewness
- original\_glszm\_GrayLevelNonUniformity\_75th\_percentile
- original\_glszm\_GrayLevelVariance\_iqr
- original\_glszm\_LowGrayLevelZoneEmphasis\_25th\_percentile
- original\_glszm\_SizeZoneNonUniformityNormalized\_75th\_percentile
- original\_glszm\_SmallAreaEmphasis\_25th\_percentile
- original\_glszm\_SmallAreaLowGrayLevelEmphasis\_median
- original\_glszm\_ZonePercentage\_kurtosis
- original\_ngtdm\_Busyness\_skewness
- original\_ngtdm\_Complexity\_kurtosis
- original\_ngtdm\_Contrast\_25th\_percentile
- original\_ngtdm\_Contrast\_skewness
- original\_gldm\_DependenceNonUniformity\_entropy
- original\_gldm\_DependenceNonUniformityNormalized\_kurtosis

- original\_gldm\_DependenceVariance\_kurtosis
- original\_gldm\_HighGrayLevelEmphasis\_25th\_percentile
- original\_gldm\_HighGrayLevelEmphasis\_mean
- original\_gldm\_LargeDependenceEmphasis\_mean
- original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_kurtosis
- original\_gldm\_LowGrayLevelEmphasis\_25th\_percentile
- original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_kurtosis
- original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_median

Elenco 181 feature:

- original\_firstorder\_10Percentile\_25th\_percentile,
- original\_firstorder\_10Percentile\_median,
- original\_firstorder\_10Percentile\_skewness,
- original\_firstorder\_90Percentile\_25th\_percentile,
- original\_firstorder\_90Percentile\_75th\_percentile,
- original\_firstorder\_90Percentile\_kurtosis,
- original\_firstorder\_90Percentile\_mean,
- original\_firstorder\_Energy\_iqr,
- original\_firstorder\_Entropy\_median,
- original\_firstorder\_Entropy\_skewness,
- original\_firstorder\_InterquartileRange\_25th\_percentile,
- original\_firstorder\_InterquartileRange\_kurtosis,
- original\_firstorder\_Kurtosis\_25th\_percentile,
- original\_firstorder\_Kurtosis\_kurtosis,
- original\_firstorder\_Kurtosis\_skewness,
- original\_firstorder\_Maximum\_25th\_percentile,
- original\_firstorder\_Maximum\_75th\_percentile,
- original\_firstorder\_Maximum\_iqr,
- original\_firstorder\_Maximum\_mean,
- original\_firstorder\_MeanAbsoluteDeviation\_median,
- original\_firstorder\_MeanAbsoluteDeviation\_skewness,
- original\_firstorder\_Mean\_25th\_percentile,
- original\_firstorder\_Median\_mean,
- original\_firstorder\_Minimum\_iqr,
- original\_firstorder\_Minimum\_mean,
- original\_firstorder\_Minimum\_median,
- original\_firstorder\_Range\_75th\_percentile,
- original\_firstorder\_Range\_iqr,
- original\_firstorder\_Range\_kurtosis,
- original\_firstorder\_Range\_median,
- original\_firstorder\_Range\_skewness,

- original\_firstorder\_RobustMeanAbsoluteDeviation\_kurtosis,
- original\_firstorder\_RootMeanSquared\_25th\_percentile,
- original\_firstorder\_RootMeanSquared\_75th\_percentile,
- original\_firstorder\_RootMeanSquared\_skewness,
- original\_firstorder\_Skewness\_25th\_percentile,
- original\_firstorder\_Skewness\_iqr,
- original\_firstorder\_Skewness\_median,
- original\_firstorder\_Skewness\_skewness,
- original\_firstorder\_TotalEnergy\_iqr,
- original\_firstorder\_Uniformity\_25th\_percentile,
- original\_firstorder\_Uniformity\_median,
- original\_firstorder\_Variance\_mean,
- original\_glcm\_Autocorrelation\_25th\_percentile,
- original\_glcm\_ClusterProminence\_25th\_percentile,
- original\_glcm\_ClusterProminence\_entropy,
- original\_glcm\_ClusterShade\_mean,
- original\_glcm\_ClusterShade\_median,
- original\_glcm\_ClusterTendency\_25th\_percentile,
- original\_glcm\_Contrast\_mean,
- original\_glcm\_Correlation\_median,
- original\_glcm\_DifferenceAverage\_25th\_percentile,
- original\_glcm\_DifferenceAverage\_skewness,
- original\_glcm\_DifferenceEntropy\_25th\_percentile,
- original\_glcm\_DifferenceEntropy\_kurtosis,
- original\_glcm\_DifferenceVariance\_25th\_percentile,
- original\_glcm\_Id\_entropy,
- original\_glcm\_Idm\_entropy,
- original\_glcm\_Idmn\_iqr,
- original\_glcm\_Idmn\_mean,
- original\_glcm\_Idmn\_skewness,
- original\_glcm\_Idn\_75th\_percentile,
- original\_glcm\_Idn\_median,
- original\_glcm\_Imc2\_iqr,
- original\_glcm\_Imc2\_median,
- original\_glcm\_Imc2\_skewness,
- original\_glcm\_InverseVariance\_75th\_percentile,
- original\_glcm\_InverseVariance\_kurtosis,
- original\_glcm\_JointAverage\_25th\_percentile,
- original\_glcm\_JointAverage\_mean,
- original\_glcm\_JointAverage\_median,
- original\_glcm\_JointAverage\_skewness,

- original\_glcm\_JointEnergy\_25th\_percentile,
- original\_glcm\_JointEnergy\_iqr,
- original\_glcm\_JointEnergy\_median,
- original\_glcm\_JointEntropy\_75th\_percentile,
- original\_glcm\_JointEntropy\_mean,
- original\_glcm\_MCC\_25th\_percentile,
- original\_glcm\_MCC\_75th\_percentile,
- original\_glcm\_MCC\_iqr,
- original\_glcm\_MaximumProbability\_iqr,
- original\_glcm\_MaximumProbability\_mean,
- original\_glcm\_SumSquares\_25th\_percentile,
- original\_glcm\_SumSquares\_iqr,
- original\_glcm\_SumSquares\_kurtosis,
- original\_glrlm\_GrayLevelNonUniformity\_25th\_percentile,
- original\_glrlm\_GrayLevelNonUniformity\_iqr,
- original\_glrlm\_GrayLevelNonUniformity\_skewness,
- original\_glrlm\_GrayLevelNonUniformityNormalized\_iqr,
- original\_glrlm\_GrayLevelNonUniformityNormalized\_mean,
- original\_glrlm\_GrayLevelNonUniformityNormalized\_median,
- original\_glrlm\_GrayLevelNonUniformityNormalized\_skewness,
- original\_glrlm\_GrayLevelVariance\_25th\_percentile,
- original\_glrlm\_GrayLevelVariance\_iqr,
- original\_glrlm\_GrayLevelVariance\_mean,
- original\_glrlm\_HighGrayLevelRunEmphasis\_75th\_percentile,
- original\_glrlm\_LongRunEmphasis\_kurtosis,
- original\_glrlm\_LongRunLowGrayLevelEmphasis\_25th\_percentile,
- original\_glrlm\_LongRunLowGrayLevelEmphasis\_iqr,
- original\_glrlm\_LongRunLowGrayLevelEmphasis\_median,
- original\_glrlm\_LongRunLowGrayLevelEmphasis\_skewness,
- original\_glrlm\_LowGrayLevelRunEmphasis\_25th\_percentile,
- original\_glrlm\_LowGrayLevelRunEmphasis\_kurtosis,
- original\_glrlm\_RunEntropy\_25th\_percentile,
- original\_glrlm\_RunEntropy\_skewness,
- original\_glrlm\_RunVariance\_mean,
- original\_glrlm\_ShortRunEmphasis\_75th\_percentile,
- original\_glrlm\_ShortRunEmphasis\_kurtosis,
- original\_glrlm\_ShortRunEmphasis\_skewness,
- original\_glrlm\_ShortRunHighGrayLevelEmphasis\_skewness,
- original\_glrlm\_ShortRunLowGrayLevelEmphasis\_75th\_percentile,
- original\_glrlm\_ShortRunLowGrayLevelEmphasis\_median,
- original\_glszm\_GrayLevelNonUniformity\_75th\_percentile,

- original\_glszm\_GrayLevelNonUniformity\_skewness,
- original\_glszm\_GrayLevelNonUniformityNormalized\_iqr,
- original\_glszm\_GrayLevelNonUniformityNormalized\_mean,
- original\_glszm\_GrayLevelVariance\_iqr,
- original\_glszm\_GrayLevelVariance\_skewness,
- original\_glszm\_HighGrayLevelZoneEmphasis\_kurtosis,
- original\_glszm\_HighGrayLevelZoneEmphasis\_mean,
- original\_glszm\_LargeAreaEmphasis\_kurtosis,
- original\_glszm\_LargeAreaLowGrayLevelEmphasis\_75th\_percentile,
- original\_glszm\_LargeAreaLowGrayLevelEmphasis\_entropy,
- original\_glszm\_LowGrayLevelZoneEmphasis\_25th\_percentile,
- original\_glszm\_LowGrayLevelZoneEmphasis\_iqr,
- original\_glszm\_LowGrayLevelZoneEmphasis\_kurtosis,
- original\_glszm\_SizeZoneNonUniformity\_mean,
- original\_glszm\_SizeZoneNonUniformityNormalized\_75th\_percentile,
- original\_glszm\_SizeZoneNonUniformityNormalized\_kurtosis,
- original\_glszm\_SizeZoneNonUniformityNormalized\_median,
- original\_glszm\_SmallAreaEmphasis\_25th\_percentile,
- original\_glszm\_SmallAreaEmphasis\_iqr,
- original\_glszm\_SmallAreaHighGrayLevelEmphasis\_iqr,
- original\_glszm\_SmallAreaLowGrayLevelEmphasis\_iqr,
- original\_glszm\_SmallAreaLowGrayLevelEmphasis\_mean,
- original\_glszm\_SmallAreaLowGrayLevelEmphasis\_median,
- original\_glszm\_ZoneEntropy\_25th\_percentile,
- original\_glszm\_ZoneEntropy\_75th\_percentile,
- original\_glszm\_ZoneEntropy\_entropy,
- original\_glszm\_ZonePercentage\_kurtosis,
- original\_glszm\_ZoneVariance\_25th\_percentile,
- original\_glszm\_ZoneVariance\_skewness,
- original\_ngtdm\_Busyness\_iqr,
- original\_ngtdm\_Busyness\_mean,
- original\_ngtdm\_Busyness\_median,
- original\_ngtdm\_Busyness\_skewness,
- original\_ngtdm\_Complexity\_iqr,
- original\_ngtdm\_Complexity\_kurtosis,
- original\_ngtdm\_Complexity\_mean,
- original\_ngtdm\_Contrast\_25th\_percentile,
- original\_ngtdm\_Contrast\_kurtosis,
- original\_ngtdm\_Contrast\_skewness,
- original\_gldm\_DependenceEntropy\_75th\_percentile,
- original\_gldm\_DependenceEntropy\_median,

- original\_gldm\_DependenceEntropy\_skewness,
- original\_gldm\_DependenceNonUniformity\_entropy,
- original\_gldm\_DependenceNonUniformity\_kurtosis,
- original\_gldm\_DependenceNonUniformityNormalized\_kurtosis,
- original\_gldm\_DependenceNonUniformityNormalized\_mean,
- original\_gldm\_DependenceVariance\_75th\_percentile,
- original\_gldm\_DependenceVariance\_kurtosis,
- original\_gldm\_DependenceVariance\_median,
- original\_gldm\_DependenceVariance\_skewness,
- original\_gldm\_GrayLevelVariance\_skewness,
- original\_gldm\_HighGrayLevelEmphasis\_25th\_percentile,
- original\_gldm\_HighGrayLevelEmphasis\_mean,
- original\_gldm\_LargeDependenceEmphasis\_25th\_percentile,
- original\_gldm\_LargeDependenceEmphasis\_mean,
- original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_kurtosis,
- original\_gldm\_LowGrayLevelEmphasis\_25th\_percentile,
- original\_gldm\_LowGrayLevelEmphasis\_75th\_percentile,
- original\_gldm\_LowGrayLevelEmphasis\_iqr,
- original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_25th\_percentile,
- original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_kurtosis,
- original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_mean,
- original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_skewness,
- original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_25th\_percentile,
- original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_iqr,
- original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_kurtosis,
- original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_median,
- original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_skewness

## Feature selection: Affinity Propagation

- Cluster 0: original\_firstorder\_10Percentile\_entropy
- Cluster 1: original\_firstorder\_90Percentile\_entropy
- Cluster 2: original\_firstorder\_Energy\_entropy
- Cluster 3: original\_firstorder\_Energy\_kurtosis
- Cluster 4: original\_firstorder\_Entropy\_entropy
- Cluster 5: original\_firstorder\_InterquartileRange\_entropy
- Cluster 6: original\_firstorder\_Kurtosis\_entropy
- Cluster 7: original\_firstorder\_Maximum\_entropy
- Cluster 8: original\_firstorder\_Mean\_entropy
- Cluster 9: original\_firstorder\_Median\_entropy

- Cluster 10: original\_firstorder\_Minimum\_entropy
- Cluster 11: original\_firstorder\_Minimum\_mean
- Cluster 12: original\_firstorder\_Range\_entropy
- Cluster 13: original\_firstorder\_Range\_mean
- Cluster 14: original\_firstorder\_RobustMeanAbsoluteDeviation\_entropy
- Cluster 15: original\_firstorder\_RobustMeanAbsoluteDeviation\_mean
- Cluster 16: original\_firstorder\_RootMeanSquared\_entropy
- Cluster 17: original\_firstorder\_Skewness\_25th\_percentile
- Cluster 18: original\_firstorder\_Skewness\_entropy
- Cluster 19: original\_firstorder\_TotalEnergy\_entropy
- Cluster 20: original\_firstorder\_TotalEnergy\_iqr
- Cluster 21: original\_firstorder\_TotalEnergy\_median
- Cluster 22: original\_firstorder\_TotalEnergy\_skewness
- Cluster 23: original\_firstorder\_Uniformity\_entropy
- Cluster 24: original\_firstorder\_Variance\_entropy
- Cluster 25: original\_firstorder\_Variance\_kurtosis
- Cluster 26: original\_glcm\_Autocorrelation\_entropy
- Cluster 27: original\_glcm\_Autocorrelation\_mean
- Cluster 28: original\_glcm\_ClusterProminence\_entropy
- Cluster 29: original\_glcm\_ClusterShade\_entropy
- Cluster 30: original\_glcm\_ClusterShade\_median
- Cluster 31: original\_glcm\_ClusterTendency\_entropy
- Cluster 32: original\_glcm\_Contrast\_entropy
- Cluster 33: original\_glcm\_Correlation\_entropy
- Cluster 34: original\_glcm\_Correlation\_mean
- Cluster 35: original\_glcm\_DifferenceAverage\_entropy
- Cluster 36: original\_glcm\_DifferenceAverage\_mean
- Cluster 37: original\_glcm\_DifferenceEntropy\_entropy
- Cluster 38: original\_glcm\_DifferenceVariance\_entropy
- Cluster 39: original\_glcm\_Id\_entropy
- Cluster 40: original\_glcm\_Id\_kurtosis
- Cluster 41: original\_glcm\_Idm\_75th\_percentile
- Cluster 42: original\_glcm\_Idm\_entropy
- Cluster 43: original\_glcm\_Idmn\_entropy
- Cluster 44: original\_glcm\_Idmn\_iqr
- Cluster 45: original\_glcm\_Idmn\_kurtosis
- Cluster 46: original\_glcm\_Idn\_entropy
- Cluster 47: original\_glcm\_Idn\_mean
- Cluster 48: original\_glcm\_Imc1\_entropy
- Cluster 49: original\_glcm\_Imc1\_iqr
- Cluster 50: original\_glcm\_Imc2\_entropy

- Cluster 51: original\_glcM\_Imc2\_skewness
- Cluster 52: original\_glcM\_InverseVariance\_entropy
- Cluster 53: original\_glcM\_JointAverage\_entropy
- Cluster 54: original\_glcM\_JointEnergy\_entropy
- Cluster 55: original\_glcM\_JointEnergy\_skewness
- Cluster 56: original\_glcM\_JointEntropy\_entropy
- Cluster 57: original\_glcM\_JointEntropy\_median
- Cluster 58: original\_glcM\_MCC\_entropy
- Cluster 59: original\_glcM\_MaximumProbability\_entropy
- Cluster 60: original\_glcM\_SumAverage\_entropy
- Cluster 61: original\_glcM\_SumEntropy\_entropy
- Cluster 62: original\_glcM\_SumEntropy\_kurtosis
- Cluster 63: original\_glcM\_SumEntropy\_skewness
- Cluster 64: original\_glcM\_SumSquares\_entropy
- Cluster 65: original\_glcM\_SumSquares\_iqr
- Cluster 66: original\_glrM\_GrayLevelNonUniformity\_entropy
- Cluster 67: original\_glrM\_GrayLevelNonUniformity\_mean
- Cluster 68: original\_glrM\_GrayLevelNonUniformityNormalized\_75th\_percentile
- Cluster 69: original\_glrM\_GrayLevelNonUniformityNormalized\_entropy
- Cluster 70: original\_glrM\_GrayLevelVariance\_entropy
- Cluster 71: original\_glrM\_HighGrayLevelRunEmphasis\_entropy
- Cluster 72: original\_glrM\_HighGrayLevelRunEmphasis\_median
- Cluster 73: original\_glrM\_LongRunEmphasis\_entropy
- Cluster 74: original\_glrM\_LongRunEmphasis\_median
- Cluster 75: original\_glrM\_LongRunHighGrayLevelEmphasis\_entropy
- Cluster 76: original\_glrM\_LongRunLowGrayLevelEmphasis\_75th\_percentile
- Cluster 77: original\_glrM\_LongRunLowGrayLevelEmphasis\_entropy
- Cluster 78: original\_glrM\_LowGrayLevelRunEmphasis\_entropy
- Cluster 79: original\_glrM\_RunEntropy\_entropy
- Cluster 80: original\_glrM\_RunLengthNonUniformity\_entropy
- Cluster 81: original\_glrM\_RunLengthNonUniformityNormalized\_entropy
- Cluster 82: original\_glrM\_RunLengthNonUniformityNormalized\_kurtosis
- Cluster 83: original\_glrM\_RunLengthNonUniformityNormalized\_skewness
- Cluster 84: original\_glrM\_RunPercentage\_entropy
- Cluster 85: original\_glrM\_RunVariance\_entropy
- Cluster 86: original\_glrM\_ShortRunEmphasis\_entropy
- Cluster 87: original\_glrM\_ShortRunHighGrayLevelEmphasis\_entropy
- Cluster 88: original\_glrM\_ShortRunHighGrayLevelEmphasis\_iqr
- Cluster 89: original\_glrM\_ShortRunHighGrayLevelEmphasis\_median
- Cluster 90: original\_glrM\_ShortRunLowGrayLevelEmphasis\_entropy
- Cluster 91: original\_glrM\_ShortRunLowGrayLevelEmphasis\_mean



- Cluster 92: original\_glrIm\_ShortRunLowGrayLevelEmphasis\_skewness
- Cluster 93: original\_glszm\_GrayLevelNonUniformity\_entropy
- Cluster 94: original\_glszm\_GrayLevelNonUniformityNormalized\_entropy
- Cluster 95: original\_glszm\_GrayLevelVariance\_entropy
- Cluster 96: original\_glszm\_GrayLevelVariance\_mean
- Cluster 97: original\_glszm\_HighGrayLevelZoneEmphasis\_entropy
- Cluster 98: original\_glszm\_LargeAreaEmphasis\_entropy
- Cluster 99: original\_glszm\_LargeAreaEmphasis\_iqr
- Cluster 100: original\_glszm\_LargeAreaHighGrayLevelEmphasis\_entropy
- Cluster 101: original\_glszm\_LargeAreaLowGrayLevelEmphasis\_entropy
- Cluster 102: original\_glszm\_LowGrayLevelZoneEmphasis\_entropy
- Cluster 103: original\_glszm\_SizeZoneNonUniformity\_entropy
- Cluster 104: original\_glszm\_SizeZoneNonUniformityNormalized\_entropy
- Cluster 105: original\_glszm\_SizeZoneNonUniformityNormalized\_mean
- Cluster 106: original\_glszm\_SmallAreaEmphasis\_entropy
- Cluster 107: original\_glszm\_SmallAreaHighGrayLevelEmphasis\_entropy
- Cluster 108: original\_glszm\_SmallAreaHighGrayLevelEmphasis\_kurtosis
- Cluster 109: original\_glszm\_SmallAreaHighGrayLevelEmphasis\_skewness
- Cluster 110: original\_glszm\_SmallAreaLowGrayLevelEmphasis\_entropy
- Cluster 111: original\_glszm\_ZoneEntropy\_entropy
- Cluster 112: original\_glszm\_ZoneEntropy\_median
- Cluster 113: original\_glszm\_ZonePercentage\_entropy
- Cluster 114: original\_glszm\_ZoneVariance\_entropy
- Cluster 115: original\_ngtdm\_Busyness\_entropy
- Cluster 116: original\_ngtdm\_Busyness\_median
- Cluster 117: original\_ngtdm\_Coarseness\_entropy
- Cluster 118: original\_ngtdm\_Complexity\_entropy
- Cluster 119: original\_ngtdm\_Contrast\_entropy
- Cluster 120: original\_ngtdm\_Strength\_entropy
- Cluster 121: original\_gldm\_DependenceEntropy\_entropy
- Cluster 122: original\_gldm\_DependenceEntropy\_iqr
- Cluster 123: original\_gldm\_DependenceNonUniformity\_entropy
- Cluster 124: original\_gldm\_DependenceNonUniformityNormalized\_entropy
- Cluster 125: original\_gldm\_DependenceVariance\_75th\_percentile
- Cluster 126: original\_gldm\_DependenceVariance\_entropy
- Cluster 127: original\_gldm\_DependenceVariance\_skewness
- Cluster 128: original\_gldm\_GrayLevelNonUniformity\_entropy
- Cluster 129: original\_gldm\_GrayLevelVariance\_entropy
- Cluster 130: original\_gldm\_GrayLevelVariance\_median
- Cluster 131: original\_gldm\_HighGrayLevelEmphasis\_entropy
- Cluster 132: original\_gldm\_LargeDependenceEmphasis\_entropy

- Cluster 133: original\_gldm\_LargeDependenceHighGrayLevelEmphasis\_entropy
- Cluster 134: original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_entropy
- Cluster 135: original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_kurtosis
- Cluster 136: original\_gldm\_LargeDependenceLowGrayLevelEmphasis\_mean
- Cluster 137: original\_gldm\_LowGrayLevelEmphasis\_entropy
- Cluster 138: original\_gldm\_SmallDependenceEmphasis\_entropy
- Cluster 139: original\_gldm\_SmallDependenceEmphasis\_mean
- Cluster 140: original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_entropy
- Cluster 141: original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_entropy

Feature selection: Correlazione assoluta tra features e target

- original\_glrIm\_GrayLevelNonUniformityNormalized\_median
- original\_glszm\_GrayLevelNonUniformityNormalized\_median
- original\_glrIm\_GrayLevelNonUniformity\_skewness
- original\_glszm\_GrayLevelNonUniformity\_skewness
- original\_glrIm\_GrayLevelNonUniformityNormalized\_mean
- original\_glszm\_GrayLevelNonUniformityNormalized\_mean
- original\_glszm\_GrayLevelNonUniformityNormalized\_75th\_percentile
- original\_glrIm\_GrayLevelNonUniformityNormalized\_75th\_percentile
- original\_glszm\_GrayLevelNonUniformityNormalized\_25th\_percentile
- original\_glrIm\_GrayLevelNonUniformityNormalized\_25th\_percentile
- original\_glrIm\_GrayLevelNonUniformity\_75th\_percentile
- original\_firstorder\_Minimum\_iqr
- original\_gldm\_DependenceNonUniformity\_kurtosis
- original\_glszm\_GrayLevelNonUniformity\_75th\_percentile
- original\_firstorder\_Skewness\_median
- original\_glszm\_SmallAreaLowGrayLevelEmphasis\_75th\_percentile
- original\_glrIm\_ShortRunLowGrayLevelEmphasis\_75th\_percentile
- original\_glrIm\_LowGrayLevelRunEmphasis\_mean
- original\_glszm\_LowGrayLevelZoneEmphasis\_mean
- original\_glrIm\_ShortRunLowGrayLevelEmphasis\_mean
- original\_glszm\_SmallAreaLowGrayLevelEmphasis\_mean
- original\_firstorder\_Skewness\_mean
- original\_gldm\_DependenceEntropy\_median
- original\_gldm\_DependenceNonUniformityNormalized\_kurtosis
- original\_firstorder\_Skewness\_25th\_percentile
- original\_firstorder\_Kurtosis\_mean
- original\_gldm\_DependenceEntropy\_mean
- original\_firstorder\_Kurtosis\_75th\_percentile
- original\_glrIm\_GrayLevelNonUniformityNormalized\_iqr
- original\_glszm\_GrayLevelNonUniformityNormalized\_iqr

- original\_firstorder\_Kurtosis\_median
- original\_glszm\_LowGrayLevelZoneEmphasis\_75th\_percentile
- original\_glrlm\_LowGrayLevelRunEmphasis\_75th\_percentile
- original\_firstorder\_10Percentile\_median
- original\_glrlm\_LowGrayLevelRunEmphasis\_25th\_percentile
- original\_glszm\_LowGrayLevelZoneEmphasis\_25th\_percentile
- original\_glrlm\_GrayLevelNonUniformity\_mean
- original\_firstorder\_Kurtosis\_25th\_percentile
- original\_gldm\_DependenceEntropy\_75th\_percentile
- original\_firstorder\_Skewness\_75th\_percentile
- original\_gldm\_SmallDependenceHighGrayLevelEmphasis\_kurtosis
- original\_firstorder\_10Percentile\_25th\_percentile
- original\_gldm\_DependenceEntropy\_25th\_percentile
- original\_glszm\_GrayLevelNonUniformity\_mean
- original\_firstorder\_10Percentile\_mean
- original\_glcm\_Idm\_kurtosis
- original\_glcm\_Idmn\_75th\_percentile
- original\_glcm\_Imc2\_75th\_percentile
- original\_glszm\_SizeZoneNonUniformity\_kurtosis
- original\_glcm\_Id\_kurtosis
- original\_glrlm\_GrayLevelNonUniformity\_median
- original\_glrlm\_RunLengthNonUniformity\_kurtosis
- original\_gldm\_SmallDependenceEmphasis\_kurtosis
- original\_glrlm\_GrayLevelNonUniformity\_iqr
- original\_glrlm\_GrayLevelVariance\_kurtosis
- original\_glszm\_GrayLevelVariance\_kurtosis
- original\_glszm\_GrayLevelNonUniformity\_median
- original\_glszm\_SizeZoneNonUniformityNormalized\_25th\_percentile
- original\_glrlm\_RunLengthNonUniformityNormalized\_25th\_percentile
- original\_ngtdm\_Contrast\_25th\_percentile
- original\_glszm\_ZonePercentage\_kurtosis
- original\_glrlm\_RunEntropy\_25th\_percentile
- original\_glszm\_SmallAreaLowGrayLevelEmphasis\_25th\_percentile
- original\_glrlm\_ShortRunLowGrayLevelEmphasis\_25th\_percentile
- original\_glrlm\_ShortRunHighGrayLevelEmphasis\_kurtosis
- original\_glszm\_SmallAreaHighGrayLevelEmphasis\_kurtosis
- original\_glcm\_DifferenceAverage\_median
- original\_glrlm\_RunPercentage\_kurtosis
- original\_glszm\_ZoneEntropy\_25th\_percentile
- original\_glcm\_DifferenceVariance\_25th\_percentile
- original\_glrlm\_ShortRunEmphasis\_iqr

- original\_ngtdm\_Contrast\_mean
- original\_glcm\_Contrast\_25th\_percentile
- original\_glrlm\_RunLengthNonUniformityNormalized\_mean
- original\_glszm\_SizeZoneNonUniformityNormalized\_mean
- original\_glcm\_DifferenceAverage\_25th\_percentile
- original\_glszm\_LargeAreaLowGrayLevelEmphasis\_kurtosis
- original\_glrlm\_LongRunLowGrayLevelEmphasis\_kurtosis
- original\_glcm\_DifferenceAverage\_mean
- original\_glszm\_GrayLevelNonUniformity\_iqr
- original\_glszm\_SizeZoneNonUniformityNormalized\_75th\_percentile
- original\_glrlm\_RunLengthNonUniformityNormalized\_75th\_percentile
- original\_gldm\_DependenceVariance\_kurtosis
- original\_glcm\_Idmn\_median
- original\_gldm\_LargeDependenceEmphasis\_kurtosis
- original\_glrlm\_ShortRunLowGrayLevelEmphasis\_iqr
- original\_glszm\_SmallAreaLowGrayLevelEmphasis\_iqr
- original\_glcm\_Idn\_kurtosis
- original\_glcm\_Idmn\_mean
- original\_glszm\_SmallAreaEmphasis\_25th\_percentile
- original\_glszm\_LargeAreaEmphasis\_25th\_percentile
- original\_glrlm\_LongRunEmphasis\_25th\_percentile
- original\_glrlm\_LongRunEmphasis\_median
- original\_glszm\_LargeAreaEmphasis\_median
- original\_glcm\_DifferenceVariance\_mean
- original\_glcm\_JointEnergy\_25th\_percentile
- original\_firstorder\_10Percentile\_skewness
- original\_glszm\_SizeZoneNonUniformityNormalized\_median
- original\_glrlm\_RunLengthNonUniformityNormalized\_median
- original\_firstorder\_10Percentile\_75th\_percentile
- original\_glrlm\_RunEntropy\_75th\_percentile
- original\_glrlm\_RunEntropy\_mean
- original\_glrlm\_RunEntropy\_median
- original\_glszm\_SmallAreaLowGrayLevelEmphasis\_median
- original\_glrlm\_ShortRunLowGrayLevelEmphasis\_median
- original\_glrlm\_LowGrayLevelRunEmphasis\_median
- original\_glszm\_LowGrayLevelZoneEmphasis\_median
- original\_glcm\_JointEnergy\_median
- original\_glszm\_ZoneEntropy\_75th\_percentile
- original\_glszm\_ZoneEntropy\_median
- original\_glszm\_ZoneEntropy\_mean
- original\_glcm\_JointEntropy\_75th\_percentile

- original\_gldm\_SmallDependenceLowGrayLevelEmphasis\_kurtosis
- original\_glszm\_SmallAreaEmphasis\_mean
- original\_firstorder\_Skewness\_skewness
- original\_glszm\_SmallAreaEmphasis\_iqr
- original\_glszm\_ZoneVariance\_median
- original\_glrlm\_RunVariance\_median
- original\_glcm\_ClusterProminence\_median
- original\_firstorder\_90Percentile\_iqr
- original\_glszm\_SmallAreaEmphasis\_median
- original\_glrlm\_RunLengthNonUniformityNormalized\_iqr
- original\_glszm\_SizeZoneNonUniformityNormalized\_iqr
- original\_glcm\_Contrast\_mean
- original\_glcm\_Imc2\_median
- original\_glcm\_DifferenceAverage\_75th\_percentile

# Bibliografia

1. <https://www.airc.it/cancro/informazioni-tumori/guida-ai-tumori/tumore-colon-retto>
2. <https://www.aiom.it/tumori-nel-2021-in-italia-13-di-morti-negli-uomini-rispetto-alleuropa-post-covid-piu-neoplasie-al-seno-e-colon-operate-in-stadio-gia-avanzato/>
3. <https://www.siccr.org/informazioni-general/cancro-del-retto/>
4. <https://www.ospedaleniguarda.it/news/leggi/tumore-del-retto-localmente-avanzato-remissione-completa-per-1-persona-su-4-anche-senza-chirurgia-niguarda-promotore-dello-studio-no-cut>
5. G. Chiloiro et al., “Delta Radiomics Can Predict Distant Metastasis in Locally Advanced Rectal Cancer: The Challenge to Personalize the Cure,” *Front. Oncol.*, 2020, doi: 10.3389/fonc.2020.595012.
6. Rosati, Giannini et al. “Development and validation of an AI-based pathomics biomarker to predict response to first-line treatment in metastatic colorectal cancers” Paper No. ICBB 122 DOI: 10.11159/icbb24.122
7. Alrahawy M, Aker M, Issa M, et al. (December 06, 2022) Textural Analysis as a Predictive Biomarker in Rectal Cancer. *Cureus* 14(12): e32241. DOI 10.7759/cureus.32241
8. Chee CG, Kim YH, Lee KH, Lee YJ, Park, JH, Lee HS, et al. (2017) CT texture analysis in patients with locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy: A potential imaging biomarker for treatment response and prognosis. *PLoS ONE* 12(8): e0182883. <https://doi.org/10.1371/journal.pone.0182883>
9. Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis C-A, et al. (2019) Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med* 16(1): e1002730. <https://doi.org/10.1371/journal.pmed.1002730>
10. <https://www.ospedaleniguarda.it/news/leggi/tumore-del-retto-localmente-avanzato-remissione-completa-per-1-persona-su-4-anche-senza-chirurgia-niguarda-promotore-dello-studio-no-cut>
11. Bankhead, P. et al. QuPath: Open source software for digital pathology image analysis. *Scientific Reports* (2017). <https://doi.org/10.1038/s41598-017-17204-5>
12. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Xiaojun G, et al., editors. A method for normalizing histology slides for quantitative analysis. 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2009 June 28 2009-July 1 2009.
13. [https://en.wikipedia.org/wiki/Data\\_augmentation](https://en.wikipedia.org/wiki/Data_augmentation).
14. Wang, SC. (2003). Artificial Neural Network. In: *Interdisciplinary Computing in Java Programming*. The Springer International Series in Engineering and Computer Science, vol 743. Springer, Boston, MA. [https://doi.org/10.1007/978-1-4615-0377-4\\_5](https://doi.org/10.1007/978-1-4615-0377-4_5)
15. F. Q. Lauzon, "An introduction to deep learning," 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA), Montreal, QC, Canada, 2012, pp. 1438-1439, doi: 10.1109/ISSPA.2012.6310529.
16. [Kaiming He et al., “Deep Residual Learning for Image Recognition” https://doi.org/10.48550/arXiv.1512.03385](https://doi.org/10.48550/arXiv.1512.03385)
17. <https://en.wikipedia.org/wiki/Backpropagation>

18. [Karen Simonyan](https://doi.org/10.48550/arXiv.1409.1556) et al., “Very Deep Convolutional Networks for Large-Scale Image Recognition”, <https://doi.org/10.48550/arXiv.1409.1556>
19. <https://keras.io/about/>
20. Kingma, Diederik P., Jimmy Ba., "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980, 2014
21. Vishal Yathishm, “Loss Functions and Their Use In Neural Networks”, <https://towardsdatascience.com/loss-functions-and-their-use-in-neural-networks-a470e703f1e9>
22. SAGAR SHARMA, “Activation Functions in Neural Networks”, <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>
23. [https://www.tensorflow.org/api\\_docs/python/tf/keras/applications/ResNet50](https://www.tensorflow.org/api_docs/python/tf/keras/applications/ResNet50)
24. Weiss, K., Khoshgoftaar, T.M. & Wang, D. A survey of transfer learning. *J Big Data* 3, 9 (2016). <https://doi.org/10.1186/s40537-016-0043-6>
25. <https://www.image-net.org/>
26. Yingjie Tian, Yuqi Zhang, A comprehensive survey on regularization strategies in machine learning, *Information Fusion*, Volume 80, 2022, Pages 146-166, ISSN 1566-2535, <https://doi.org/10.1016/j.inffus.2021.11.005>.
27. Damaskinos, G., El Mhamdi, E.M., Guerraoui, R., Patra, R. & Taziki, M.. (2018). Asynchronous Byzantine Machine Learning (the case of SGD). Available from <https://proceedings.mlr.press/v80/damaskinos18a.html>.
28. Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis C-A, et al. (2019) Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med* 16(1): e1002730. <https://doi.org/10.1371/journal.pmed.1002730>
29. <https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=it>
30. <https://www.educative.io/answers/what-is-balanced-accuracy>
31. Montesinos López, O.A., Montesinos López, A., Crossa, J. (2022). Overfitting, Model Tuning, and Evaluation of Prediction Performance. In: *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer, Cham. [https://doi.org/10.1007/978-3-030-89010-0\\_4](https://doi.org/10.1007/978-3-030-89010-0_4)
32. <https://pyradiomics.readthedocs.io/en/latest/>
33. <https://scikit-learn.org/1.5/modules/generated/sklearn.preprocessing.MinMaxScaler.html>
34. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2001
35. Dash, Manoranjan, and Huan Liu. "Feature selection for classification." *Intelligent data analysis* 1.1-4 (1997): 131-156.
36. “A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction”, *JASTT*, vol. 1, no. 1, pp. 56–70, May 2020, doi: 10.38094/jastt1224.
37. Darst, B.F., Malecki, K.C. & Engelman, C.D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet* 19 (Suppl 1), 65 (2018). <https://doi.org/10.1186/s12863-018-0633-8>
38. Tangirala, Suryakanthi. "Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm." *International Journal of Advanced Computer Science and Applications* 11.2 (2020): 612-619.

39. Kirti Jain, Rajni Jindal, Sampling and noise filtering methods for recommender systems: A literature review, *Engineering Applications of Artificial Intelligence*, Volume 122, 2023, 106129, ISSN 0952-1976, <https://doi.org/10.1016/j.engappai.2023.106129>.
40. Berrar, Daniel. "Cross-validation." (2019): 542-545.
41. Milos Radovic et al., "Minimum redundancy maximum relevance feature selection approach for temporal gene expression data", *BMC Bioinformatics* (2017), DOI 10.1186/s12859-016-1423-9
42. Hui Liu, Chapter 7 - Description methods of spatial wind along railways, 2021, Pages 251-282, ISBN 9780128237069, <https://doi.org/10.1016/B978-0-12-823706-9.00007-7>.
43. S. S. R. Moustafa, M. S. Abdalzaher, F. Khan, M. Metwaly, E. A. Elawadi and N. S. Al-Arifi, "A Quantitative Site-Specific Classification Approach Based on Affinity Propagation Clustering," in *IEEE Access*, vol. 9, pp. 155297-155313, 2021, doi: 10.1109/ACCESS.2021.3128284.
44. Ma, Yuan Zhe. (2014). A Tutorial on Principal Component Analysis. 10.13140/2.1.1593.1684.
45. Bayes, Thomas & Price, Richard (1763). "[An Essay towards solving a Problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, A.M.F.R.S.](#)" *Philosophical Transactions of the Royal Society of London*. 53: 370-418. doi:10.1098/rstl.1763.0053.
46. Sun, Jingwen & Du, Weixing & Shi, Niancai. (2018). A Survey of kNN Algorithm. *Information Engineering and Applied Computing*. 1. 10.18063/ieac.v1i1.770.
47. Patil, Dharmaraj & Patil, Jayantrao. (2018). Malicious URLs Detection Using Decision Tree Classifiers and Majority Voting Technique. *Cybernetics and Information Technologies*. 18. 11-29. 10.2478/cait-2018-0002.
48. M. A. Cano Lengua and E. A. Papa Quiroz, "A Systematic Literature Review on Support Vector Machines Applied to Classification," *2020 IEEE Engineering International Research Conference (EIRCON)*, Lima, Peru, 2020, pp. 1-4, doi: 10.1109/EIRCON51178.2020.9254028.
49. Rokach, Lior & Maimon, Oded. (2005). Decision Trees. 10.1007/0-387-25465-X\_9.
50. Bartosz Krawczyk, Leandro L. Minku, João Gama, Jerzy Stefanowski, Michał Woźniak, Ensemble learning for data stream analysis: A survey, *Information Fusion*, Volume 37, 2017, Pages 132-156, ISSN 1566-2535, <https://doi.org/10.1016/j.inffus.2017.02.004>.
51. Leo Breiman, Random Forests, Statistics Department, University of California, Berkeley, CA 94720