



**Politecnico
di Torino**

Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Biomedica

A.a. 2020/2021

Sessione di Laurea Dicembre 2021

**Predizione della risposta alla
terapia di singole metastasi
epatiche da tumore del colon-retto
per mezzo di metodi di machine
learning**

Relatore:

Prof.ssa Gabriella Balestra

Candidato:

Sara Burdisso

Correlatori:

Prof.ssa Samanta Rosati

Ing. Valentina Giannini

Sommario

Lo scopo del lavoro di tesi è stato quello di sviluppare un algoritmo di machine learning per predire la risposta al primo ciclo di chemioterapia di singole metastasi epatiche da cancro coloretale (CRC) trattate con FOLFOX o FOLFIRI.

Sono state estratte 107 features radiomiche dalle segmentazioni manuali di 396 metastasi epatiche su scansioni TC, acquisite al momento della diagnosi, appartenenti a 87 pazienti provenienti dall'ASST Grande Ospedale Metropolitano Niguarda e dall'Istituto di Candiolo - Fondazione del Piemonte per l'Oncologia – IRCCS. Per il lavoro si avevano a disposizione anche le scansioni TC successive al primo ciclo di chemioterapia con cui è stata calcolata la variazione percentuale del diametro maggiore di ciascuna lesione utile a definire la risposta alla terapia delle metastasi.

Nella prima fase del lavoro sono state effettuate diverse prove sfruttando diversi classificatori e metodi di features selection al fine di classificare le lesioni in “non responder” (R-), se il loro diametro maggiore aumentava più del 20% in una scansione TC eseguita dopo un ciclo di trattamento, “responder” (R+) se il loro diametro maggiore diminuiva almeno del 30% e “stabili” (S) altrimenti. Sono state valutate le prestazioni, in termini di corretti classificati, fornite dalle seguenti coppie classificatore-metodo di feature selection: classificatori bayesiano ed SVM allenati con subset di variabili individuati effettuando feature selection basata su algoritmi genetici, classificatore random forest e feature selection basata sull'importance, classificatori naive Bayes, SVM, random forest e decision tree e feature selection basata sulla correlazione e classificatori SVM e reti neurali artificiali e algoritmo QuikReduct.

In una seconda fase sono state clusterizzate le lesioni, tramite reti SOM, in ognuna delle 3 classi e sono state ripetute le prove sfruttando feature selection basata sull'importance e classificatore random forest e feature selection basata sulla correlazione e classificatori naive Bayes, SVM, random forest e decision tree al fine di classificare le lesioni nei diversi cluster e poi ricondurre la classificazione alle classi R+, S ed R-.

Con tutti i modelli di RF è stato osservato il fenomeno di overfitting poiché le prestazioni erano pari al 100% di corretti classificati sul dataset di Training ma diminuivano attestandosi in media attorno a 52% (45/86), 49% (32/65), 75% (9/12) rispettivamente sulle classi R+, S ed R- sul Test Set e 44% (26/59), 38% (19/50) e 41% (14/34) rispettivamente sulle classi R+, S ed R- sul Validation Set.

Le prestazioni raggiunte con il miglior modello di DT (allenato con un subset di features individuato effettuando feature selection basata sulla correlazione) si attestavano intorno all'80% di corretti classificati sulle 3 classi sul Training Set ma si riducevano al 47% (40/85), 41% (27/65) e 64% (7/11) sul Test Set e a 39% (23/59), 34% (17/50), 50% (17/34) sul Validation Set rispettivamente sulle classi R+, S ed R-.

Il modello di classificatore bayesiano più performante, allenato anch'esso con un subset di features individuato effettuando feature selection basata sulla correlazione, forniva prestazioni che si attestavano intorno all'80% (24-25/30) di corretti classificati sulle classi R+ ed R- ed erano pari al 60% (18/30) sulla classe S sul Training Set ma si riducevano al

56% (48/85), 37% (24/65) e 73% (8/11) sul Test Set e a 42% (25/59), 26% (13/50), 47% (16/34) sul Validation Set rispettivamente sulle classi R+, S ed R-.

Con il modello di SVM migliore (SVM gaussiano con C pari a 50 e allenato con un subset di variabili individuato effettuando feature selection basata sulla correlazione) infine sono state ottenute prestazioni pari a 58% (18/31), 77% (23/30) e 77% (24/31) rispettivamente sulle classi R+, S ed R- sul Training Set, pari a 54% (46/85), 41% (27/65) e 64% (7/11) rispettivamente sulle classi R+, S ed R- sul Test Set e pari a 37% (22/59), 52% (26/50), 41% (14/34) rispettivamente sulle classi R+, S ed R- sul Validation Set.

Le prestazioni ottenute non risultano essere sufficientemente buone per poter utilizzare tali modelli nella pratica clinica e dimostrano la necessità di ampliare il lavoro svolto. Vista, tuttavia, la complessità del problema data dall'ampia numerosità delle features a disposizione, ridotta dimensione del dataset in termini di numero di lesioni e la decisione di classificare le lesioni in 3 classi, i risultati ottenuti sono comunque un buon punto di partenza per successive prove e approfondimenti.

Indice

Introduzione	1
Materiali e metodi.....	4
Pazienti.....	4
Estrazione delle features	5
Creazione dei dataset.....	6
Feature selection e costruzione dei classificatori	8
Problema a 3 classi	9
Algoritmi genetici.....	9
Risultati.....	11
Random Forest [24]	14
Risultati.....	16
Feature selection basata su correlazione.....	22
Risultati.....	22
QuikReduct Algorithm	34
Risultati.....	35
Problema a più classi: clusterizzazione delle lesioni	45
Risultati relativi ai modelli di RF	51
Risultati relativi ai modelli allenati effettuando feature selection basata sulla correlazione	57
Confronto tra le prestazioni ottenute	77
Riconcettualizzazione del problema: data mining	85
Risultati	86
Discussioni e sviluppi futuri.....	93
Appendice.....	96
Bibliografia.....	151

Introduzione

Il tumore del colon-retto (CRC) è dovuto alla proliferazione incontrollata delle cellule della mucosa che rivestono la parte interna dell'intestino. Si distingue generalmente in tumore del colon vero e proprio e tumore del retto, ovvero dell'ultimo tratto dell'intestino [2].

Nel 2020 sono stati stimati più di 1.9 milioni di nuovi casi di cancro del colon-retto e 935 000 decessi. Nel complesso, il tumore del colon-retto è al 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 %), ma è al secondo posto in termini di mortalità (9.4 % rispetto a tutti i decessi per cancro) preceduto solo dal tumore ai polmoni (18.0 %).

Il CRC può essere considerato un marcatore dello sviluppo socioeconomico; i tassi di incidenza del cancro del colon e del retto tendono infatti ad essere bassi nella maggior parte delle regioni dell'Africa e nell'Asia centro-meridionale [Figura 1] [1].

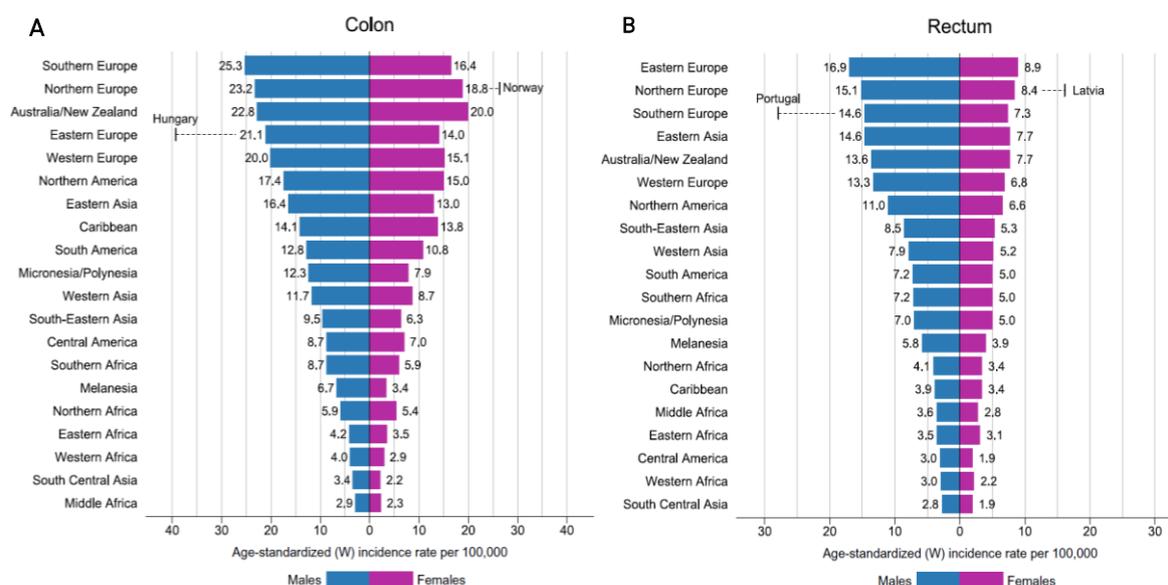


Figura 1: Tassi di incidenza specifici per regione, standardizzati in base all'età, per i tumori del colon (A) e retto (B) nel 2020 [1].

I fattori di rischio del tumore del colon-retto sono legati a:

- dieta: obesità, vita sedentaria e alto consumo di grassi e proteine animali;
- fattori genetici: è possibile avere una predisposizione ad ammalarsi di CRC se nella famiglia d'origine si sono manifestate malattie come le poliposi adenomatose ereditarie e il carcinoma ereditario del colon-retto su base non poliposica;

- altri fattori non ereditari: età (la malattia è maggiormente diffusa in persone fra i 60 e i 75 anni), il fumo e le malattie infiammatorie croniche intestinali.

Il tumore del colon-retto viene diagnosticato più precocemente grazie alle sempre più diffuse campagne di screening sulla popolazione più a rischio. La diagnosi viene effettuata con palpazione dell'addome alla ricerca di eventuali masse nell'intestino, nel fegato e nei linfonodi, con eventuale esplorazione rettale manuale e indagini strumentali quali la colonscopia che permette di fare l'analisi istologica del tessuto e l'ecografia addominale, la RM (risonanza magnetica) e la TC (Tomografia Computerizzata) di addome e torace con mezzo di contrasto che consentono di valutare l'estensione del tumore e la presenza di eventuali metastasi [2]. Circa il 20-25% dei pazienti presenta infatti metastasi al fegato alla diagnosi iniziale [4].

La terapia più comune per le metastasi epatiche consiste nella loro resezione chirurgica ma la maggior parte dei pazienti ha una malattia metastatica che inizialmente non è adatta a resezione (85%). Le metastasi epatiche inizialmente non resecabili possono diventare tali dopo un trattamento chemioterapico che ne riduce le dimensioni. In questo caso, l'obiettivo è quello di raggiungere un'elevata risposta alla terapia per convertire le metastasi non resecabili in resecabili. I 2 trattamenti chemioterapici più utilizzati sono FOLFOX (5-fluorouracile/leucovorin/oxaliplatino) e FOLFIRI (5-fluorouracile/leucovorin/irinotecano) che hanno un diverso profilo di tossicità ma svolgono un'attività chemioterapica simile [3 - 4].

Circa il 50% dei pazienti non risponde al trattamento di prima linea [4] e altri, dopo un'iniziale risposta, vanno incontro a progressione. Sarebbe importante quindi conoscere quali pazienti potrebbero beneficiare della terapia prima ancora della somministrazione in modo da permettere ai clinici di valutare eventuali trattamenti alternativi evitando ai pazienti terapie inutili.

La risposta ad una terapia è spesso valutata usando il RECIST (Response Evaluation Criteria in Solid Tumors), un criterio che valuta la risposta di una lesione misurando i cambiamenti nei diametri assiali più lunghi delle lesioni dopo il trattamento [14]. Le immagini mediche possono tuttavia fornire ulteriori informazioni aggiuntive sul tumore. Nell'ultimo decennio, infatti, il campo dell'imaging medico si è notevolmente sviluppato portando, fra le altre cose, allo sviluppo della radiomica, la pratica che analizza le immagini e ne estrae informazioni quantitative che non potrebbero essere colte dalla semplice analisi visiva di un operatore. Tali informazioni e dati estratti possono poi essere utilizzati come input per l'allenamento di sistemi di machine learning [6].

Molti studi hanno applicato la radiomica per predire la risposta alla terapia di metastasi epatiche da CRC. In particolare Dohan et al. hanno costruito una "radiomic signature" per predire l'overall survival (OS) e l'assenza di risposta alla terapia utilizzando le TC pre e post trattamento. Lo score, che combina le features radiomiche estratte dalla lesione di dimensioni maggiori di ciascun paziente, criteri di densità e dimensioni, permette la previsione di uno scarso OS prima del RECIST. Questo risultato dimostra la necessità di

usare features quantitative che riflettono cambiamenti tumorali e non solo parametri legati alle dimensioni [11].

Anche lo studio di Rao et al. ha osservato, per mezzo di un'analisi di regressione logistica univariata, che le analisi di features radiomiche possono essere migliori discriminatori di risposta di lesioni epatiche da CRC rispetto alle misure di dimensione e volume [9].

Ahn et al. hanno estratto le features radiomiche dalla lesione epatica da CRC di dimensioni maggiori al fine di predire la risposta del paziente alla chemioterapia. Per mezzo di un'analisi di regressione logistica multivariata è stato osservato che le metastasi epatiche con una maggiore eterogeneità (nella TC al baseline) sono associate ad una scarsa risposta [4].

Al contrario lo studio di Giannini et al., su pazienti con metastasi epatiche da CRC trattati con terapia mirata HER2, ha osservato che le lesioni più uniformi tendevano ad avere una risposta peggiore ed erano correlate con una sopravvivenza più breve [7].

Klassen et al. hanno estratto 370 features dalle TC pre-trattamento di 196 metastasi epatiche da tumore gastroesofageo e, attraverso un modello di Random Forest, hanno valutato la possibilità di prevedere la risposta alla chemioterapia delle singole metastasi [8].

A differenza degli studi precedenti, gli ultimi 2 citati, insieme a [12-13], hanno analizzato il comportamento di singole metastasi epatiche e non della lesione di maggiori dimensioni come rappresentativa del comportamento globale del tumore, mostrando che la risposta individuale delle metastasi epatiche varia notevolmente all'interno di uno stesso paziente. In tali pazienti, nota la risposta alla terapia di ciascuna metastasi, si potrebbe attuare una strategia specifica per le diverse metastasi, sottoponendo ad esempio a biopsia o asportando solo le lesioni che non rispondono alla terapia.

Scopo del lavoro di tesi è quello di valutare la possibilità di predire la risposta al primo ciclo di terapia con FOLFOX/FOLFIRI di singole metastasi epatiche da CRC per mezzo di sistemi di machine learning, che sfruttano come input le features radiomiche estratte dalle TC pre-trattamento di tali lesioni epatiche.

Materiali e metodi

Pazienti

I pazienti a disposizione per il lavoro di tesi erano 87 provenienti da due centri differenti: dall'Istituto di Candiolo - Fondazione del Piemonte per l'Oncologia – IRCCS (centro 1) e dall'ASST Grande Ospedale Metropolitano Niguarda (centro 2). I pazienti arruolati, a cui è stato diagnosticato un CRC con almeno una metastasi epatica con diametro ≥ 10 mm, sono stati sottoposti ad almeno un ciclo di chemioterapia con FOLFOX o FOLFIRI e possedevano una scansione TC con mezzo di contrasto al momento della diagnosi (baseline) e una scansione TC successiva al primo ciclo di chemioterapia (12 settimane = time-point 1= TP1). Sono state considerate fino ad un massimo di 10 metastasi epatiche per paziente [Tabella 1]. Le scansioni TC sono state acquisite con diversi scanner (Somatome Definition Flash – Siemens, Sensation 64 – Siemens, Brilliance 64 – Philips) e protocolli. Per le acquisizioni delle scansioni dell'ospedale Niguarda lo spessore mediano della slice era pari a 3 mm (range 1-5 mm) mentre per quelle dell'istituto di Candiolo era pari a 2.5 mm (range 1-5 mm); la dimensione della matrice era 512×512 in entrambi i centri; la tensione del tubo a raggi X era pari a 100-140 kV; l'intensità di corrente era pari a 180- 230 mA; il field of view era pari a 300-350 mm e infine la dimensione mediana dei pixel era pari a 0.762×0.762 mm (range 0.598×0.598 - 0.977×0.977 mm) per le scansioni dei pazienti in cura presso l'ospedale Niguarda e 0.683×0.683 mm (range 0.625×0.625 - 1.060×1.060 mm) per le scansioni dei pazienti in cura presso l'istituto di Candiolo.

I dati sono stati completamente resi anonimi in ogni centro, assegnando un identificativo unico casuale ad ogni paziente in modo da garantirne la protezione della privacy.

Una radiologa specializzanda ha segmentato manualmente le lesioni presenti nelle immagini TC di ognuno degli 87 pazienti selezionati, escludendo le metastasi confluenti, tramite ITK-SNAP, un'applicazione software utilizzata per segmentare strutture in immagini mediche 3D. È stato poi misurato il diametro maggiore di ogni metastasi epatica presente nelle TC al baseline e al TP1 ed è stata calcolata la variazione percentuale del diametro ($V(\%)$) [equazione (1)]:

$$V(\%) = \frac{\text{diametro}_{baseline} - \text{diametro}_{TP1}}{\text{diametro}_{baseline}} \quad (1)$$

Con questa informazione le lesioni sono state classificate come:

- responder (R+) se $V(\%) \geq 30.0\%$
- stabili (S+) se $0\% \leq V(\%) < 30.0\%$
- stabili (S-) se $-20.0\% \leq V(\%) < 0\%$
- non responder (R-) se $V(\%) < -20.0\%$

		Candiolo	Niguarda	Totale
Pazienti		26	61	87
Lesioni		120	276	396
Lesioni R+		48	127	175
Lesioni S+		38	87	125
Lesioni S-		5	15	20
Lesioni R-		29	47	76
N° lesioni epatiche per paziente	1 lesione	6	10	16
	2-5 lesioni	10	30	40
	6-10 lesioni	9	20	29
	>10 lesioni	1	1	2
Età pazienti	Mediana	68.5	64	65
	Min	40	24	24
	Max	82	81	82

Tabella 1: caratteristiche dei dati a disposizione.

Estrazione delle features

Le segmentazioni delle lesioni in ogni slice delle scansioni TC hanno prodotto le maschere binarie di segmentazione che sono state utilizzate per l'estrazione delle features dalle lesioni, passaggio chiave della radiomica [6]. L'estrazione delle features è avvenuta per mezzo di Pyradiomics, un pacchetto open-source scritto in linguaggio python che permette l'estrazione di features radiomiche da immagini mediche [5].

Sono state estratte dal volume di ciascuna lesione le seguenti 107 features:

- 18 features di **primo ordine**: energy, total energy, entropy, minimum, 10th percentile, 90th percentile, maximum, mean, median, interquartile range, range, mean absolute deviation, robust mean absolute deviation, root mean squared, skewness, kurtosis, variance e uniformity;
- 14 features **shape-based (3D)**: mesh volume, voxel volume, surface area, surface volume ratio, sphericity, maximum 3D diameter, elongation, flatness, least axis length, major axis length, maximum 2D diameter column, maximum 2D diameter row, maximum 2D diameter slice e minor axis length;
- 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;

- 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;
- 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

Per l'estrazione delle features le immagini sono state discretizzate utilizzando un numero fisso di bins ($n=32$), le maschere delle lesioni sono state risegmentate tra il 1° e il 99° percentile delle ROI per rimuovere gli outlier ed è stato forzato il calcolo delle features per slice (force2D).

Creazione dei dataset

Tramite il procedimento di estrazione delle features è stato ottenuto un dataset di dimensione 396 x 111, con la struttura riportata in [Tabella 2](#), di cui:

- 175 lesioni R+ (classe 1)
- 145 lesioni S+/S- (classe 2)
- 76 lesioni R- (classe 3)

Feature selection e costruzione dei classificatori

Le fasi di features selection, costruzione e verifica delle performance dei classificatori sono state effettuate per mezzo del Toolbox MATLAB *Statistics and Machine Learning Toolbox*TM [15] e per mezzo della libreria LIBSVM per la costruzione dei modelli di support vector machine [16].

La selezione delle features (FS) è un importante step di pre-processing per i problemi di classificazione che mira a ridurre la dimensione dei dati eliminando le variabili irrilevanti/ridondanti dai dataset originali. Se N è il numero totale di features a disposizione nel dataset, i metodi di FS tentano di cercare un sottoinsieme di features "ottimale" tra i $2^N - 1$ sottoinsiemi candidati [18].

I metodi di features selection si dividono in [19]:

- Filter: valutano il sottoinsieme di features indipendentemente dall'algoritmo di machine learning basandosi sulle caratteristiche intrinseche dei dati. Il principale svantaggio di questi metodi è che trascurano l'integrazione tra il sottoinsieme selezionato e le prestazioni dell'algoritmo di apprendimento.
- Wrapper: utilizzano un algoritmo di apprendimento esterno per valutare le variabili selezionate. Lo svantaggio principale dei metodi wrapper è la complessità di calcolo rispetto agli approcci filter.
- Embedded: incorporano la selezione delle features nell'algoritmo di apprendimento e usano le sue proprietà per guidare la valutazione delle features.
- Hybrid: integrano 2 metodi diversi (generalmente wrapper e filter), due metodi con gli stessi criteri, o due approcci di selezione delle features.

Problema a 3 classi

Nella prima fase del lavoro sono stati allenati dei classificatori con le 107 features o un subset di features estratte dalle lesioni utilizzando come label le 3 classi: R+, S (S+ e S-) ed R-. Spesso si affronta un problema binario accorpendo le lesioni di tipo stabile alle lesioni R+ o R-; in questo lavoro è stato deciso di non assegnare in modo arbitrario le lesioni S ad una delle 2 classi e di affrontare un problema con 3 classi.

Algoritmi genetici

Per la prima prova è stata effettuata features selection tramite algoritmi genetici (GAs) [7]. I GAs sono algoritmi di tipo euristico utilizzati per risolvere problemi di ottimizzazione, in questo caso relativi al trovare il sottoinsieme di features più informative e rilevanti per la caratterizzazione delle lesioni epatiche, modellizzando il concetto di evoluzione genetica.

Un algoritmo genetico parte da un iniziale set di possibili sottoinsiemi di features (soluzioni) detto popolazione, generato randomicamente.

Ogni soluzione è codificata da una stringa binaria di n bits in cui ogni bit rappresenta una feature: “0” nella posizione i -esima della stringa indica che la feature i -esima non è selezionata mentre “1” indica che la feature i -esima fa parte del sottoinsieme di features selezionate. La bontà di ogni soluzione è valutata con una funzione di fitness [equazione (3)] che, in questo lavoro, misura la capacità del corrispondente sottoinsieme di features di ottenere un modello in grado di classificare le metastasi del Construction Set. Nel lavoro sono stati implementati parallelamente 2 GA basati rispettivamente sui classificatori Naive Bayes e support vector machine (SVM) [17]. Gli algoritmi genetici hanno anche il potenziale di selezionare i parametri del modello da allenare (SVM in questo caso) simultaneamente al sottoinsieme di features ottimale.

Le soluzioni relative al GA basato sul classificatore bayesiano erano costituite da un vettore di 107 bits mentre quelle relative al GA basato sul classificatore SVM erano costituite dallo stesso vettore di 107 bits a cui sono stati aggiunti 4 bits per testare diversi parametri del metodo (i primi 2 restituivano un valore che in decimale codifica il valore di C (1, 10, 50 e 100) e i secondi 2 restituivano un valore che in decimale codifica il tipo di kernel (lineare, gaussiano e polinomiale di grado 2 e 3)).

Partendo da una popolazione iniziale di 1000 soluzioni è stato applicato l'operatore di selezione con cui è stato selezionato un sottoinsieme di soluzioni in modo random pari al 50% del totale di soluzioni iniziali (500 genitori). Alle 500 soluzioni sono stati applicati gli operatori di crossover (crea 2 nuovi individui attraverso la combinazione di 2 soluzioni “genitore”) e mutazione (processo di cambiamento casuale dei valori dei bits in una soluzione) per generare nuove soluzioni.

È stata calcolata la funzione di fitness associata alle nuove soluzioni: utilizzando solo le features selezionate dalla soluzione corrente è stato allenato un classificatore utilizzando come approccio il leave-one-out (LOO). In ogni iterazione del LOO tutte le metastasi di uno stesso paziente sono state escluse dal training set, un classificatore è stato allenato utilizzando le metastasi rimanenti e le metastasi escluse sono state utilizzate come test set. Questa procedura è stata ripetuta per tutti i pazienti nel construction set. Ottenuta una classificazione per ogni lesione di ogni paziente, è stata ricavata la Confusion matrix (CM) [Figura 3] su cui è stata calcolata la funzione di fitness.

		Classe vera		
		R+	S	R-
Classe predetta	R+	CC_{R+}	err_1	err_2
	S	err_3	CC_S	err_4
	R-	err_5	err_6	CC_{R-}

Figura 3: confusion matrix

La funzione di fitness è stata calcolata con l'equazione (3):

$$fitness = 1 - \frac{\frac{cc_{R+}}{tot_{R+}} + \frac{cc_{R-}}{tot_{R-}} + \frac{cc_S}{tot_S}}{3} + 0.3 \left(1 - \frac{cc_{R-}}{cc_{R-} + err_2 + err_4} \cdot \frac{cc_{R-}}{cc_{R-} + err_5 + err_6} \right) \quad (3)$$

dove il primo termine è legato ai corretti classificati (cc) ottenuti rispetto al totale delle lesioni presenti nelle classi (tot) e il secondo termine rappresenta una penalità aggiuntiva in caso di imprecisioni nel classificare le lesioni R-.

È stata poi valutata la fitness minore, ovvero la fitness che corrispondeva al miglior set di features (e ai migliori parametri del metodo nel caso del GA basato su SVM) e se questa risultava essere minore della fitness migliore trovata nell'iterazione precedente la corrispondente soluzione veniva salvata.

L'iterazione terminava con la costruzione della popolazione iniziale di 1000 individui mettendo insieme le nuove soluzioni e la popolazione iniziale di 1000 individui e selezionando 1000 "genitori" in modo random per procedere con la successiva iterazione. Queste operazioni sono state ripetute fino a quando venivano raggiunte 500 iterazioni o se non si verificava nessun cambiamento della migliore soluzione attuale per 50 iterazioni consecutive.

L'intero algoritmo è stato ripetuto 10 volte partendo dalla stessa popolazione iniziale di soluzioni in modo da ottenere 10 sottoinsiemi di features. Il miglior sottoinsieme di features è stato identificato come quello con il più basso valore di fitness e il minor numero di features selezionate.

In sintesi sono stati impostati i seguenti parametri per l'esecuzione dell'algoritmo genetico:

- numero di ripetizioni: 10
- numero di individui della popolazione iniziale: 1000
- numero di genitori: 500 selezionati in modo random
- numero di iterazioni (prima condizione di stop): 500
- seconda condizione di stop: se per 50 iterazioni consecutive la fitness non migliora
- probabilità di crossover: 1 (4 punti di taglio per il GA basato su classificatore bayesiano e 5 punti di taglio per il GA basato su classificatore SVM)
- probabilità di mutazione: 0.3 (decresce a passi di 0.002 fino a 0.1)

Una volta selezionato il miglior sottoinsieme di features individuato dal GA basato sul classificatore bayesiano, un modello dello stesso tipo è stato allenato utilizzando il Construction Set e validato sul Validation Set con le features individuate dalla soluzione; nel

caso della soluzione individuata dal GA basato sul SVM, un modello di SVM con i parametri (kernel e C) individuati dalla soluzione è stato allenato utilizzando il Construction Set e validato sul Validation Set con le features individuate dalla soluzione.

Risultati

Nelle figure 4 A e 4 B sono riportati gli andamenti della funzione di fitness durante la prima ripetizione degli algoritmi genetici basati rispettivamente sui classificatori bayesiano ed SVM. Le ripetizioni successive alla prime presentavano andamenti simili ai grafici riportati per entrambi i GA.

Nel caso del GA basato su classificatore bayesiano, durante le prime 4 ripetizioni, il valore di fitness non è cambiato ed è rimasto pari a quello trovato nella popolazione iniziale.

Nel caso del GA basato su classificatore SVM il valore della fitness migliore ha subito una lieve diminuzione.

Nelle Tabelle 3 e 4 sono riportati i valori di fitness e il numero di features selezionate dalle soluzioni migliori individuate in ogni ripetizione del GA. Nel caso del GA basato su classificatore bayesiano la soluzione migliore individuata dalle prime 4 ripetizioni dell'algoritmo è rimasta pari a quella individuata nella popolazione iniziale e quindi non si è proceduto con le successive ripetizioni.

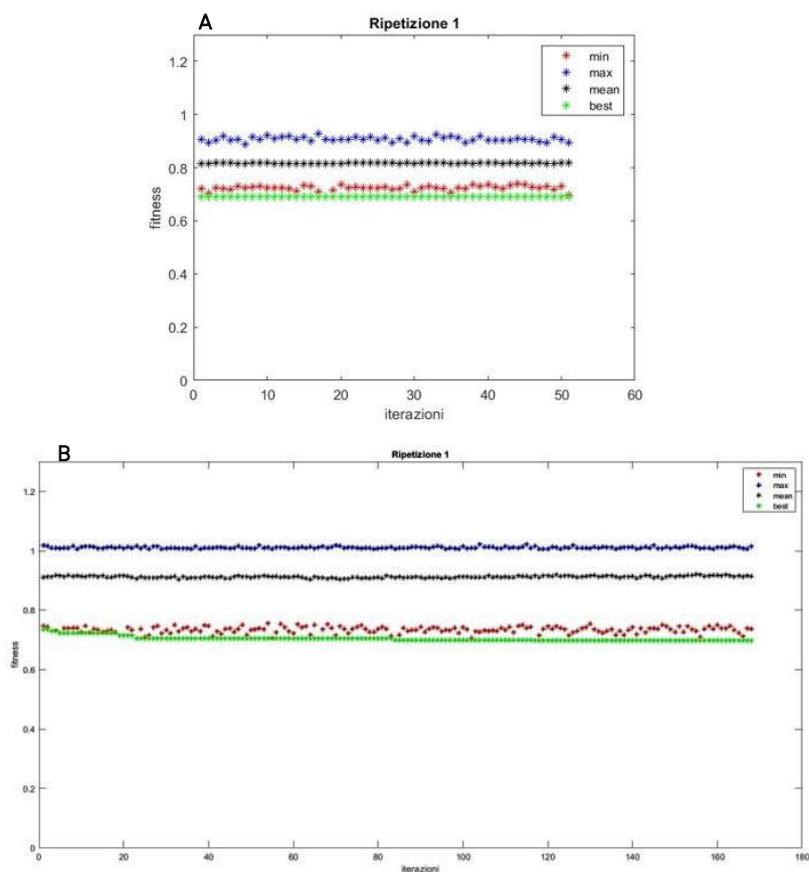


Figura 4: Andamento della fitness durante le iterazioni della prima ripetizione del GA basato sul classificatore Bayesiano (Figura 4 A) e del GA basato sul classificatore SVM (Figura 4 B). Sono riportati i valori massimo, minimo, medio e migliore assunto della fitness in ogni iterazione.

Ripetizione	Fitness - soluzione migliore	N° features selezionate - soluzione migliore
1	0.690	24
2	0.690	24
3	0.690	24
4	0.690	24

Tabella 3: valore della fitness e numero di features selezionate corrispondenti alla miglior soluzione individuata in ciascuna ripetizione del GA basato sul classificatore bayesiano.

Ripetizione	Fitness - soluzione migliore	N° features selezionate - soluzione migliore	Parametri modello – soluzione migliore
1	0.698	49	Lineare, $C = 100$
2	0.695	61	Lineare, $C = 50$
3	0.696	57	Lineare, $C = 50$
4	0.692	57	Lineare, $C = 50$
5	0.680	46	Lineare, $C = 50$
6	0.689	43	Lineare, $C = 100$
7	0.695	47	Lineare, $C = 50$
8	0.703	53	Lineare, $C = 50$
9	0.697	53	Lineare, $C = 50$
10	0.695	53	Lineare, $C = 100$

Tabella 4: valore della fitness, numero di features selezionate e parametri del modello corrispondenti alla miglior soluzione individuata in ciascuna ripetizione del GA basato sul classificatore SVM.

Nel caso del GA basato sul classificatore bayesiano è stata presa in considerazione l'unica soluzione individuata nelle 4 ripetizioni; nel caso del GA basato su SVM sono state selezionate le soluzioni 5 e la 6 che avevano rispettivamente il valore di fitness minore e il minor numero di features selezionate tra i 10 sottoinsiemi individuati dall' algoritmo.

Nelle tabelle [5](#) e [27](#) in appendice sono riportate le caratteristiche dei 3 sottoinsiemi di features migliori individuati.

In [figura 5](#) sono riportate le confusion matrix relative alle prestazioni su Construction (CS) e Validation Set (VS) dei 3 modelli allenati sul Construction Set con le features selezionate dalle soluzioni individuate.

GA	Valore della fitness	N° di features selezionate	Parametri del modello selezionati
Basato su classificatore bayesiano	0.690	24	/
Basato su SVM	0.680	46	$C = 50$ Kernel: lineare
Basato su SVM	0.689	43	$C = 100$ Kernel: lineare

Tabella 5: Caratteristiche delle soluzioni migliori individuate.

A					B					C				
CS		Classe vera			CS		Classe vera			CS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	45	11	2	Classe predetta	R+	87	22	8	Classe predetta	R+	85	23	8
	S	45	65	15		S	26	71	6		S	29	70	5
	R-	27	19	24		R-	4	2	27		R-	3	2	28
VS					VS					VS				
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	19	21	10	Classe predetta	R+	24	17	18	Classe predetta	R+	37	28	25
	S	20	21	9		S	32	29	14		S	17	17	7
	R-	20	8	15		R-	3	4	2		R-	5	5	2

Figura 5: confusion matrix in termini assoluti relative alle prestazioni su Construction (CS) e Validation Set (VS) del modello di bayesiano (Confusion matrix A), del modello di SVM lineare ($C = 50$) (Confusion matrix B) e del modello di SVM lineare ($C = 100$) (Confusion matrix C) allenati sul Construction Set con le features selezionate dalle soluzioni individuate dal GA.

Le prestazioni ottenute con i 2 modelli di SVM risultavano essere buone sul Construction Set ma peggioravano sul Validation Set in particolare sulla classe R-. Le prestazioni ottenute con il classificatore bayesiano risultavano invece peggiori sul Construction Set rispetto ai 2 modelli di SVM ma questo modello dimostrava maggior capacità di generalizzare rispetto agli stessi modelli poiché le percentuali di corretti classificati sulle 3 classi su Construction Set e Validation Set, seppur con un peggioramento sul secondo dataset, risultavano simili.

Le prestazioni in generale però non risultano buone per nessuno dei 3 modelli e il GA non risulta essere il metodo di feature selection ottimale per il problema vista l'elevata ampiezza dello spazio delle soluzioni da esplorare (soluzioni del problema non vincolato = $2^{107}-1 = \sim 10^{31}$).

Random Forest [24]

Per la seconda prova è stato utilizzato come classificatore il random forest (RF) effettuando feature selection basata sull'importance delle features stimata dal modello stesso.

Il random forest è un classificatore che combina diversi modelli di decision tree al fine di ottenere prestazioni migliori rispetto all'utilizzo di un singolo modello.

Il decision tree (DT) [25] è un classificatore basato su una struttura ad albero [Figura 6] costituito da nodi in cui il primo rappresenta la radice da cui si dipartono diversi rami. I nodi-foglia terminali costituiscono le classi e i nodi intermedi rappresentano i test da effettuare sui valori assunti dalle features di un elemento da classificare. La classificazione di un nuovo elemento avviene tracciando un percorso dalla radice fino al nodo-foglia che definisce la classe dell'elemento stesso.

Le tecniche di costruzione di un DT sono di tipo top-down e iterative. Si parte dall'intero dataset di training che viene poi diviso in subset di dimensioni minori (D) mentre la costruzione procede.

La partizione si ferma quando una delle seguenti condizioni è verificata:

- tutti gli elementi nella partizione D appartengono alla stessa classe;
- non ci sono attributi rimanenti con cui poter partizionare ulteriormente gli elementi: in questo caso si converte il nodo in un nodo-foglia corrispondente alla classe più presente nella partizione D;
- una partizione (D_j) di D è vuota (nessun elemento è presente): si crea un nodo-foglia corrispondente alla classe più presente nella partizione D.

La partizione del dataset D avviene sulla base di *splitting rules/attribute selection measures* che individuano il miglior criterio per separare il dataset D. Il caso migliore si ottiene quando ogni partizione è pura (ovvero tutti gli elementi di una partizione appartengono alla stessa classe). Il criterio di separazione migliore è quindi quello che seleziona l'attributo che permette di ottenere subset più puri possibile.

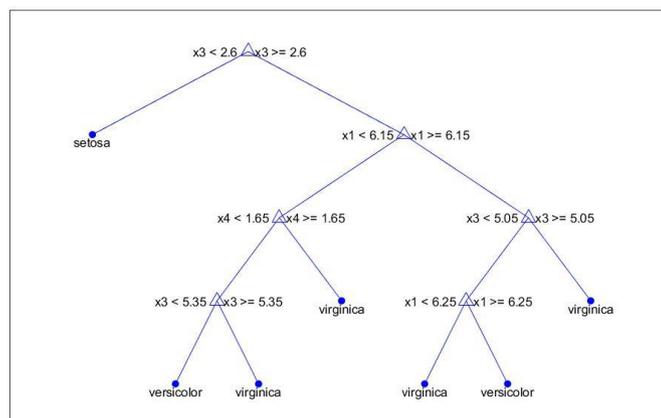


Figura 6: esempio di un DT allenato con il dataset Fisher's iris costituito da 150 elementi appartenenti a 3 classi (setosa, versicolor e virginica) e 4 features.

Ogni albero nel random forest è costruito estraendo, per campionamento con sostituzione, una parte dei dati di input (bootstrap replica). Circa 1/3 degli elementi viene escluso dal dataset di training per ottenere una stima dell'errore di classificazione man mano che gli

alberi vengono aggiunti alla “foresta” e anche per ottenere stime dell'importanza/importance delle variabili. Gli elementi non inclusi in una replica sono definiti "out of bag" per quell'albero.

Per classificare un nuovo elemento tramite RF, occorre dare in input a ciascuno degli alberi della “foresta” tale elemento e ogni albero fornirà una classificazione. La classe finale coincide con la classificazione che ha ottenuto più voti su tutti gli alberi della “foresta” (majority voting).

Nel lavoro di tesi sono state valutate le prestazioni di diversi modelli di RF costruiti con un diverso numero di features selezionate in base alla loro importanza.

Visto l'instabilità ottenuta con l'approccio del LOO, per valutare l'insieme ottimale di features da utilizzare, il Construction Set è stato ulteriormente suddiviso in 2 dataset: un Training Set utilizzato per costruire il modello con determinati parametri e features e un Test Set utilizzato per fare il tuning di tali parametri e capire quanto questi permettano la costruzione di modelli con buona capacità di generalizzare se utilizzati su elementi diversi da quelli presenti nel Training Set.

La suddivisione è stata effettuata tramite clustering per mezzo delle reti neurali SOM (Self-organizing maps). Le reti SOM sono reti neurali non supervisionate costituite da un unico layer in cui ciascun neurone ha, durante l'addestramento, un comportamento competitivo. Ogni volta che un elemento del Training Set viene dato in input alla rete ciascun neurone calcola, attraverso misure di similarità, la distanza tra i valori dei pesi (che sono inizializzati in modo random) associati al neurone stesso e i valori delle features dell'elemento di input; il neurone con la distanza minore risulta essere il vincitore e i valori dei suoi pesi (e degli eventuali neuroni vicini) vengono modificati in modo da diventare più simili ai valori caratteristici degli elementi di input. Al termine dell'allenamento ogni neurone risulterà essere specializzato nel riconoscere specifici elementi del Training Set.

Le reti SOM consentono di individuare dei pattern significativi nell'insieme dei dati di input [26].

Sono state implementate 3 reti SOM di dimensioni 3x3 (9 neuroni) per clusterizzare i soggetti appartenenti ad ognuna delle 3 classi (R+, S ed R-). La dimensione delle reti è stata scelta in modo che il numero di neuroni fosse circa pari ad un ordine di grandezza inferiore rispetto al numero di elementi nei dati di input.

Per ogni rete allenata con gli elementi di una delle 3 classi è stato costruito il dendrogramma dei pesi dei neuroni utilizzando come distanza tra i cluster le distanze tra gli elementi più lontani e come metrica la distanza euclidea. Il dendrogramma è stato tagliato usando un taglio naturale al fine di ottenere la clusterizzazione dei neuroni. Sono stati individuati gli elementi contenuti all'interno di ogni cluster di neuroni osservando, per ogni neurone, gli elementi di input per cui il neurone stesso è risultato il vincitore.

È stato deciso di costruire un Training Set bilanciato (con circa 30 lesioni per ogni classe) in modo da evitare che il classificatore allenato con questo dataset potesse apprendere maggiormente la classe più rappresentata nel dataset stesso.

Per la costruzione del Training Set sono state quindi estratte in modo proporzionale dai cluster ottenuti circa 30 lesioni per ogni classe.

Nel lavoro di tesi sono stati allenati diversi modelli di RF con il Training Set bilanciato (per mezzo della funzione matlab *treebagger*) con i seguenti parametri:

- numero di alberi pari a 100;
- frazione di osservazioni che sono selezionate randomicamente per ogni bootstrap replica pari a 1 (tutti gli elementi);
- numero di variabili selezionate (*NumPredictorsToSample*) in modo random per ogni nodo degli alberi pari a $\sqrt{n^\circ features}$, $\frac{n^\circ features}{2}$ e $\frac{3}{4} \cdot n^\circ features$, dove $n^\circ features$ indica il numero di features con cui è stato costruito il modello di RF;
- algoritmo usato per selezionare la miglior feature in ogni nodo: *curvature test* che seleziona la feature che minimizza il p-value del Chi-square test of independence tra ogni features e la classe.
- $n^\circ features$ ($n^\circ features \leq 107$): sono stati costruiti diversi modelli di RF con un diverso numero di features selezionate in base alla loro importanza.

È stato costruito, come primo passo, un modello di RF sul Training Set per ogni *NumPredictorsToSample* da testare utilizzando tutte le 107 features estratte dalle lesioni.

Per ognuno dei 3 modelli è stata valutata l'importanza delle 107 features.

Sono stati costruiti, per ogni *NumPredictorsToSample*, diversi modelli di RF effettuando feature selection basata sull'importance: ogni modello è stato costruito sul Training Set con un $n^\circ features$ decrescente imponendo delle soglie con valore crescente sulla loro importance.

Sono state valutate le prestazioni dei RF sul Training Set e sul Set Test e sono stati individuati, per ogni *NumPredictorsToSample* da testare, quei modelli e quindi la soglia che permettevano di avere buone e non sbilanciate numerosità di corretti classificati sulle 3 classi sul Test Set.

Sono state valutate le prestazioni dei modelli individuati anche sul Validation Set e sono stati identificati, per ogni *NumPredictorsToSample* da testare, i RF che permettevano di avere buone prestazioni anche sul dataset di validazione.

È stata effettuata un'analisi finale per confrontare i modelli migliori individuati per ogni *NumPredictorsToSample*.

Risultati

In [figura 7](#) sono riportati i dendrogrammi dei pesi delle reti SOM 3x3 per ogni classe con i relativi tagli effettuati per la suddivisione in cluster; in [figura 8](#) sono riportate le mappe in cui è presente l'organizzazione della rete SOM e il numero di vittorie di ciascun neurone e le relative bolle che individuano i cluster.

In [tabella 6](#) sono riportati il numero del cluster (prima colonna), i corrispondenti neuroni appartenenti al cluster stesso (seconda colonna) e la numerosità dei cluster (terza colonna) individuati per ogni classe.

In [tabella 7](#) sono riportate le numerosità di Training Set e Test Set, ottenuti estraendo da ogni cluster in modo proporzionale alla dimensione dei cluster stessi circa 30 lesioni per ogni classe, che sono stati utilizzati per tutte le prove descritte anche nei capitoli [Feature selection basata su Correlazione](#) e [QuickReduct Algorithm](#).

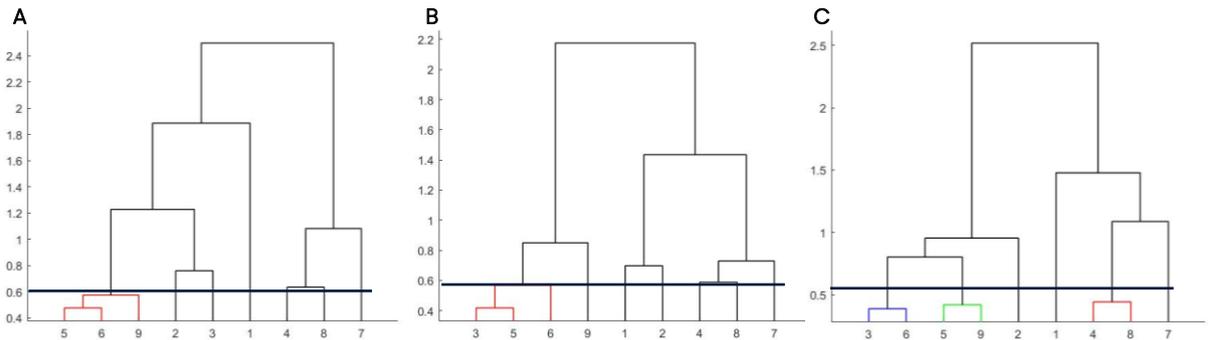


Figura 7: dendrogrammi dei pesi delle reti SOM 3x3 per le classi R+ (Figura A), S (Figura B) ed R- (Figura C) con i relativi tagli effettuati per la suddivisione in cluster.

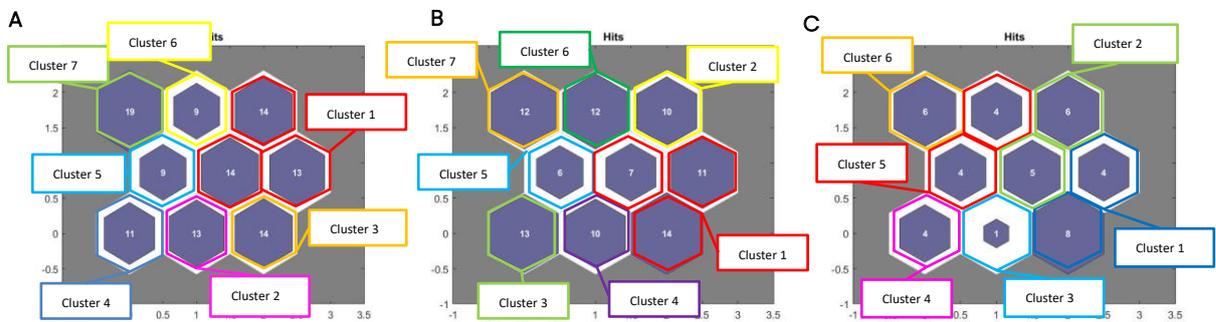


Figura 8: mappe con l'organizzazione delle reti SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) e le relative bolle che individuano i cluster per le classi R+ (Figura A), S (Figura B) ed R- (Figura C).

A			B			C		
Cluster	Neuroni	N° lesioni classe R+	Cluster	Neuroni	N° lesioni classe S	Cluster	Neuroni	N° lesioni classe R-
1	5, 6, 9	41	1	3, 5, 6	32	1	3, 6	12
2	2	13	2	9	10	2	5, 9	11
3	3	14	3	1	13	3	2	1
4	1	11	4	2	10	4	1	4
5	4	9	5	4	6	5	4, 8	8
6	8	9	6	8	12	6	7	6
7	7	19	7	7	12			

Tabella 6: numero del cluster (prima colonna), corrispondenti neuroni appartenenti al cluster stesso (seconda colonna) e la numerosità di ogni cluster in termini di lesioni (terza colonna) per le classi R+ (Tabella A), S (Tabella B) ed R- (Tabella C).

Classe	N° lesioni	
	TRS SOM 1	TS SOM 1
R+	30	86
S	30	65
R-	31	11

Tabella 7: numero delle lesioni contenute nel Training Set bilanciato e nel Test Set per ognuna delle 3 classi.

Nelle figure [60](#), [62](#) e [64](#) in Appendice sono riportate le importance delle 107 features ottenute con i modelli di RF allenati con *NumPredictorsToSample* rispettivamente pari a $\sqrt{n^\circ features}$, $\frac{n^\circ features}{2}$ e $\frac{3}{4} \cdot n^\circ features$ (con $n^\circ features = 107$).

Tutti i modelli costruiti presentavano performance sul Training Set pari al 100% di corretti classificati. Per questo motivo non sono stati riportati grafici di confronto delle prestazioni sul Training Set.

In appendice (Figure [61](#), [63](#) e [65](#)) sono riportati i grafici relativi al numero di corretti classificati ottenuti sul Test Set con i modelli di RF costruiti con i 3 *NumPredictorsToSample* e con diverso numero di features.

Tutti i grafici di confronto presenti in questo elaborato riportano il numero di corretti classificati separatamente per ogni classe, essendo il numero di lesioni diverso in ognuna di esse. Solo per il Training Set, il numero di corretti classificati delle 3 classi è stato riportato in uno stesso grafico essendo il dataset bilanciato. Il valore massimo dell'asse delle ordinate di ciascun grafico indica il numero totale di lesioni della specifica classe presente nel dataset.

Confrontando le prestazioni in termini di corretti classificati sul Test Set, sono stati individuati, per ogni *NumPredictorsToSample* testato, i modelli che fornivano buone e non sbilanciate numerosità di corretti classificati sulle 3 classi nel Test Set (in verde nelle figure [61](#), [63](#) e [65](#) in appendice).

Dal confronto delle prestazioni sul Test Set di questi modelli (Figura [66](#) in Appendice), queste risultavano simili utilizzando le diverse soglie sull'importanza delle features. Sono state quindi valutate le prestazioni sul Validation Set dei modelli trovati e sono stati individuati i modelli che consentivano di ottenere le migliori prestazioni sul Validation Set su tutte e 3 le classi (in verde in Figura [66](#) in Appendice).

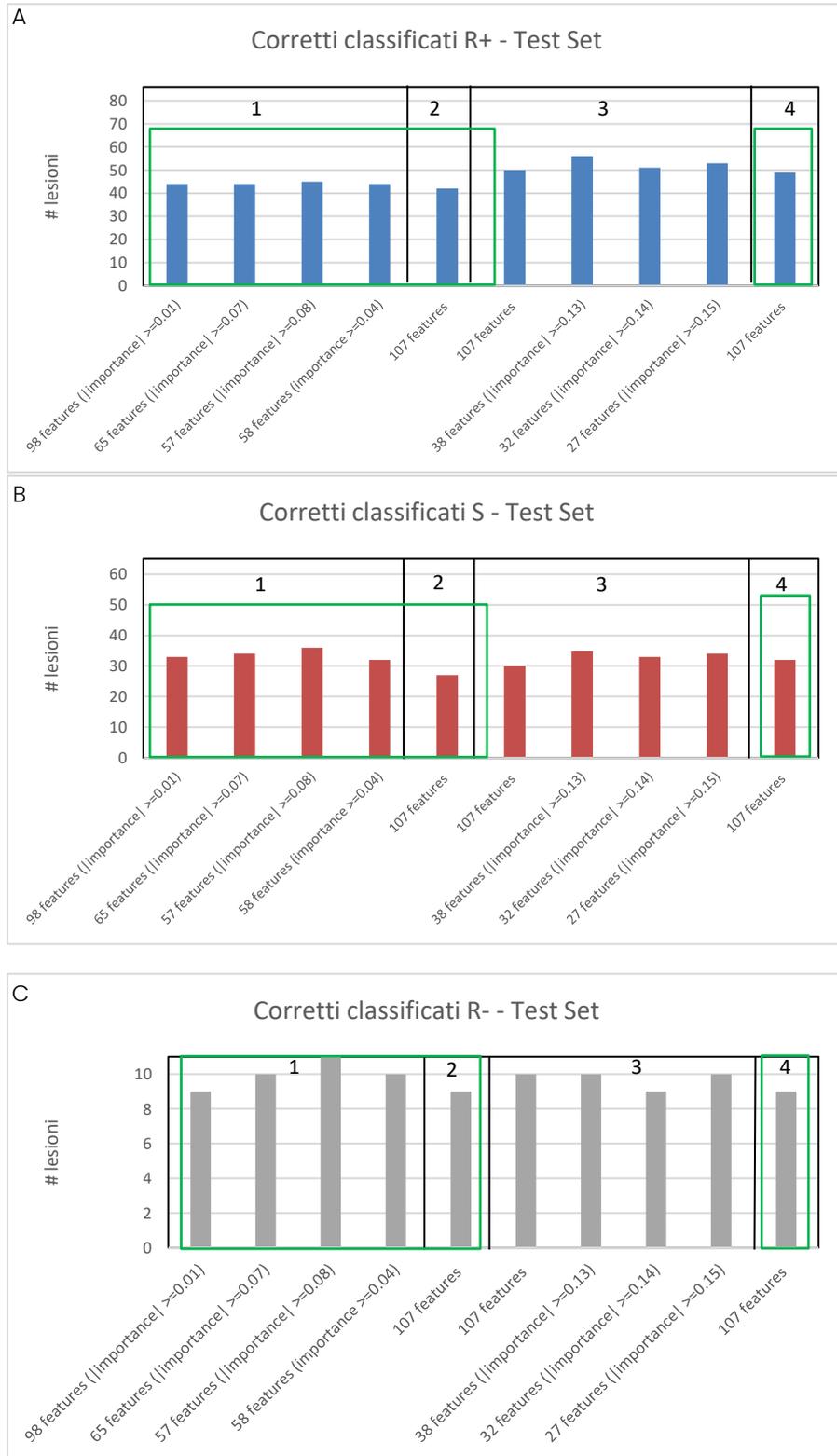
È stata effettuata un'analisi finale per confrontare i modelli migliori individuati per ogni *NumPredictorsToSample* testato.

In [figura 9](#) sono riportati i grafici di confronto tra i numeri di corretti classificati ottenuti sul Test Set e sul Validation Set con i modelli di RF individuati e in verde sono stati indicati i modelli che consentivano di ottenere le migliori performance sui 2 dataset.

In [figura 10](#) sono riportate le confusion matrix relative alle prestazioni su Training Set, Test Set e Validation Set dei modelli individuati. Questi presentavano prestazioni confrontabili su tutti i 3 dataset quindi il modello migliore risultava essere quello che permetteva di ottenere tali prestazioni con un numero inferiore di features (modello con 57 features ($|importance| \geq 0.08$) – *NumPredictorsToSample* = 10).

Le prestazioni non risultano complessivamente buone poiché con tutti i modelli individuati (in verde in [figura 9](#)) si verifica il fenomeno di *overfitting*: le prestazioni sul Training Set risultano essere sempre pari al 100% di corretti classificati mentre si riducono sulle classi R+ ed S sul Test Set attestandosi in media intorno al 50% (45/86 e 32/65) di corretti

classificati e su tutte e 3 le classi sul Validation Set attestandosi in media intorno al 39% (23/59), 44% (22/50) e 35% (12/34) di corretti classificati rispettivamente sulle classi R+, S ed R-.



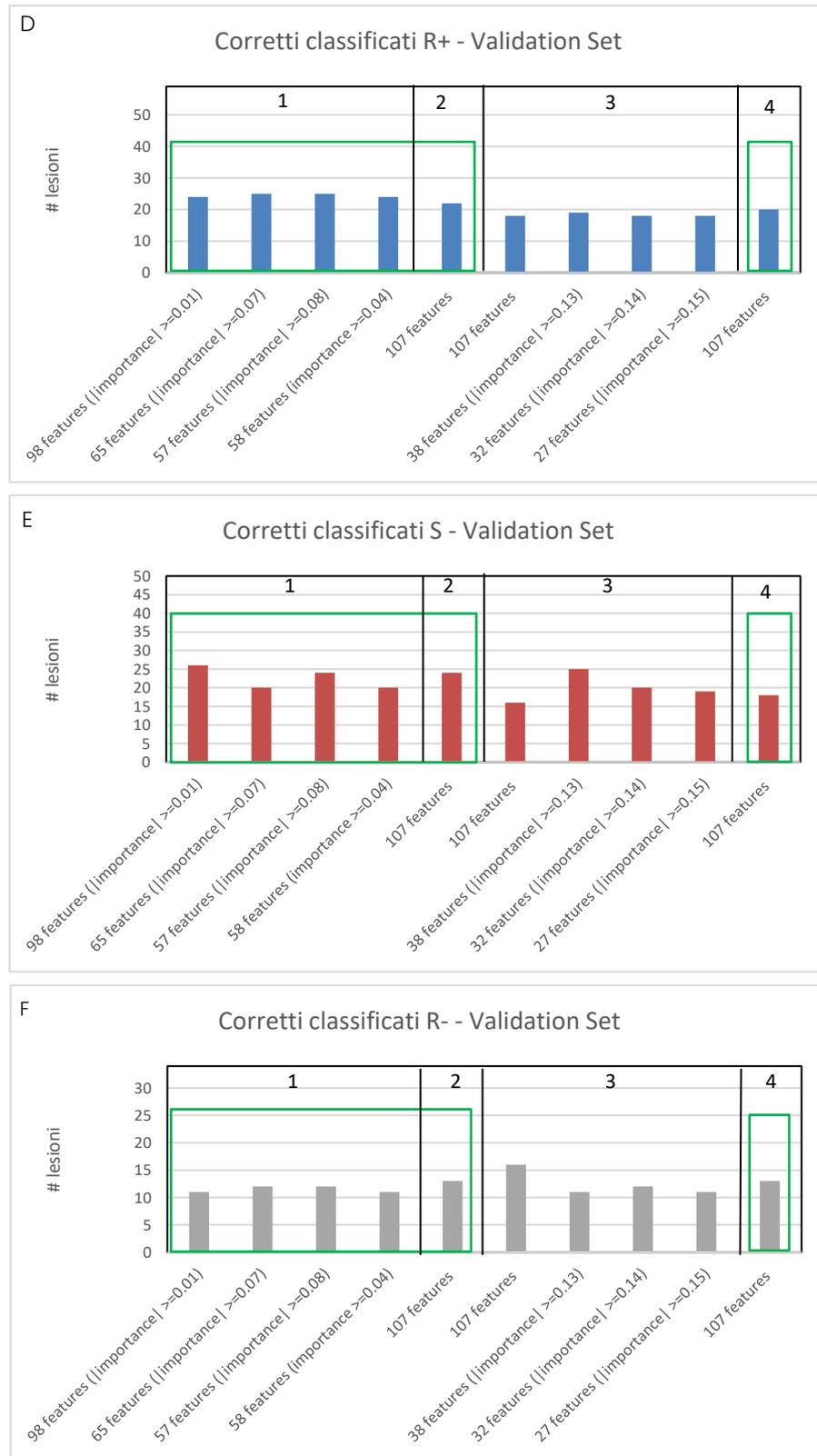


Figura 9: confronto del n° di corretti classificati ottenuti con i modelli migliori individuati per ogni *NumPredictorsToSample* (1: $\sqrt{n^\circ \text{ features}}$; 2: $n^\circ \text{ features}/2$; 3: $\frac{3}{4} \cdot n^\circ \text{ features}$; 4: $n^\circ \text{ features}$) sul Test Set (Figure A-B-C) e sul Validation Set (Figure D-E-F)

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	30	0	0	Classe predetta	R+	44	14	2	Classe predetta	R+	24	15	10
	S	0	30	0		S	22	33	0		S	12	26	13
	R-	0	0	31		R-	20	18	9		R-	23	9	11

Confusion matrix con le prestazioni fornite dal RF allenato con 98 Features ($|importance| \geq 0.01$) e $NumPredictorsToSample$ pari a 10

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	30	0	0	Classe predetta	R+	44	14	1	Classe predetta	R+	25	18	15
	S	0	30	0		S	22	34	0		S	14	20	7
	R-	0	0	31		R-	20	17	10		R-	20	12	12

Confusion matrix con le prestazioni fornite dal RF allenato con 65 Features ($|importance| \geq 0.07$) e $NumPredictorsToSample$ pari a 9

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	30	0	0	Classe predetta	R+	45	12	0	Classe predetta	R+	25	16	11
	S	0	30	0		S	24	36	0		S	14	24	11
	R-	0	0	31		R-	17	17	11		R-	20	10	12

Confusion matrix con le prestazioni fornite dal RF allenato con 57 Features ($|importance| \geq 0.08$) e $NumPredictorsToSample$ pari a 8

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	30	0	0	Classe predetta	R+	44	15	1	Classe predetta	R+	24	19	15
	S	0	30	0		S	22	32	0		S	16	20	8
	R-	0	0	31		R-	20	18	10		R-	19	11	11

Confusion matrix con le prestazioni fornite dal RF allenato con 58 Features ($importance \geq 0.04$) e $NumPredictorsToSample$ pari a 8

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	30	0	0	Classe predetta	R+	42	16	2	Classe predetta	R+	22	13	12
	S	0	30	0		S	24	27	0		S	13	24	9
	R-	0	0	31		R-	20	22	9		R-	24	13	13

Confusion matrix con le prestazioni fornite dal RF allenato con 107 e $NumPredictorsToSample$ pari a 54

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	30	0	0	Classe predetta	R+	49	18	2	Classe predetta	R+	20	19	11
	S	0	30	0		S	18	32	0		S	14	18	10
	R-	0	0	31		R-	19	15	9		R-	25	13	13

Confusion matrix con le prestazioni fornite dal RF allenato con 107 e $NumPredictorsToSample$ pari a 107

Figura 10: Confusion matrix relative alle prestazioni su Training Set, Test Set e Validation Set dei modelli di RF migliori individuati.

Feature selection basata su correlazione

Per la terza prova sono stati utilizzati come classificatori RF, DT, SVM e classificatore bayesiano eseguendo la fase di features selection sfruttando la correlazione di Spearman.

Sono stati valutati 2 approcci di features selection basati sulla correlazione. Nel primo è stato calcolato il coefficiente di correlazione di Spearman (funzione *corr* di matlab) tra ogni colonna delle features nel Training Set (TRS_SOM_1) e la colonna della classe (R+, S ed R-). Essendo un problema con 3 classi ed essendo l'ordine delle classi arbitrario (1=R+, 2=S e 3=R-), è stata calcolata la mediana della feature *i*-esima per ogni classe al fine di individuare la classe "centrale" (classe 2) e nel caso questa non fosse uguale alla classe S allora il coefficiente di correlazione veniva calcolato cambiando l'ordine delle classi. Sono stati poi allenati diversi classificatori sul Training Set (TRS_SOM_1) aggiungendo una alla volta le features, da quella più correlata a quella meno correlata con la classe. Per ogni classificatore costruito sono state valutate le sue prestazioni sul Test Set (TS_SOM_1).

Il secondo approccio prevedeva invece di calcolare il modulo del coefficiente di correlazione di Spearman tra le features ottenendo la matrice di correlazione al fine di individuare le coppie di features altamente correlate. Sono stati impostati dei valori - soglia che variavano da 1 a 0.955 a passi di -0.001 e da 0.955 a 0.85 a passi di -0.01 e sono state trovate in maniera iterativa le coppie di features correlate più di ogni soglia. Per ogni coppia individuata, è stata esclusa dal Training Set la variabile con il coefficiente di correlazione con la classe minore (calcolato con lo stesso metodo descritto per il primo approccio).

Per entrambi gli approcci, le prestazioni dei modelli sono state graficate utilizzando come metrica l'equazione 4:

$$prestazioni = \frac{1}{3} \cdot \left(\frac{cc_1}{tot_1} + \frac{cc_2}{tot_2} + \frac{cc_3}{tot_3} \right) \quad (4)$$

dove cc_i indica il numero di corretti classificati ottenuti per la classe *i*-esima e tot_i indica il numero totale di lesioni presenti nella medesima classe.

È stato infine selezionato il sottoinsieme di features a seguito del quale l'aggiunta, nel caso del primo approccio, o la rimozione, per il secondo approccio, di ulteriori features provocava un peggioramento nelle prestazioni sul Test Set.

Per il primo approccio è stata valutata una seconda modalità di selezione delle features che consisteva nel selezionare il sottoinsieme di features significativamente correlate con la classe ($p\text{-value} < 0.05$) sfruttando il valore di $p\text{-value}$, restituito dalla funzione *corr*, che testa l'ipotesi di nessuna correlazione contro l'ipotesi alternativa di una correlazione non nulla.

Risultati

Per il classificatore SVM, prima di effettuare feature selection basata sulla correlazione, sono state valutate le performance di differenti strutture allenate con tutte le 107 features e si è proceduto con la feature selection solo con le strutture che fornivano le performance migliori. In [figura 67](#) in appendice sono riportate le confusion matrix con le prestazioni su Training Set e Test Set ottenute allenando diverse strutture di SVM con diversi kernel (lineare, polinomiale e gaussiano) e valori di *C* (1, 10, 50 e 100) sul Training Set (TRS_SOM_1).

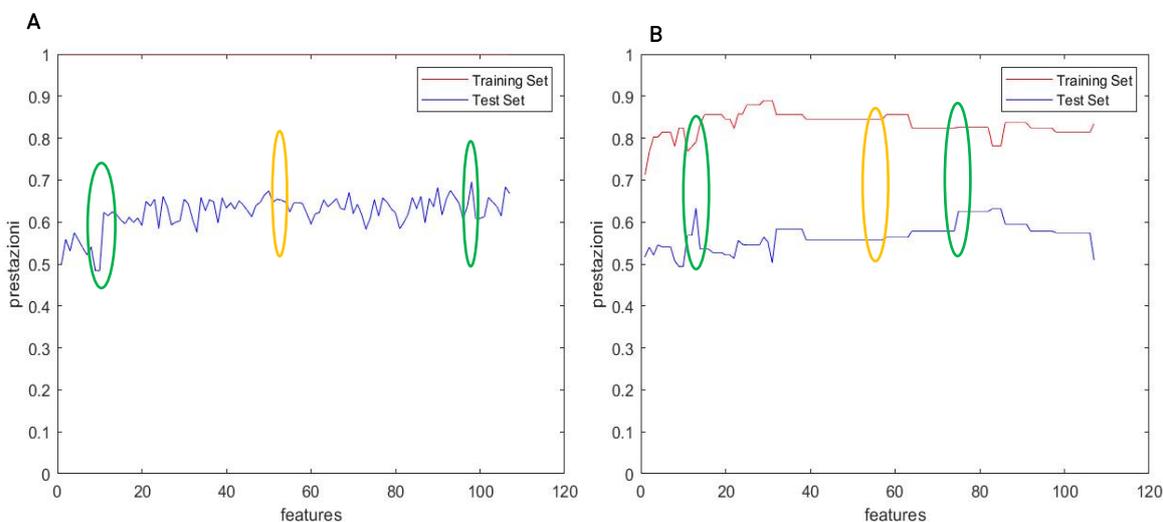
È stata effettuata feature selection basata sulla correlazione utilizzando le strutture di SVM lineare e gaussiano rispettivamente con C pari a 10 e 100.

Nelle figure [68](#) e [69](#) in Appendice sono riportati rispettivamente i coefficienti di correlazione con la classe delle 107 features in ordine decrescente e i valori di p -value della correlazione delle 107 features in ordine crescente. Si nota come i valori di correlazione delle features con la classe siano bassi e siano presenti 53 features con p -value < 0.05 .

In [figura 11](#) sono riportati gli andamenti delle prestazioni dei diversi modelli allenati aggiungendo una feature alla volta in ordine decrescente di correlazione con la classe. Sono inoltre indicate le prestazioni in corrispondenza dell'aggiunta della variabile che genera un picco nelle prestazioni sul Test Set (in verde) e sono evidenziate le prestazioni in corrispondenza dell'aggiunta dell'ultima feature significativamente correlata con la classe (in giallo).

In [figura 12](#) sono riportati gli andamenti delle prestazioni dei diversi modelli allenati scartando le features meno correlate con la classe appartenenti alle coppie di features correlate tra loro più di una certa soglia. Sono inoltre indicate le soglie in corrispondenza delle quali si verifica un picco nelle prestazioni sul Test Set (in verde).

Nei casi in cui non era evidente un aumento seguito da un peggioramento o da un andamento costante delle prestazioni sul Test Set, sono stati individuati il picco in corrispondenza del maggiore aumento delle prestazioni e un picco in corrispondenza di un successivo incremento.



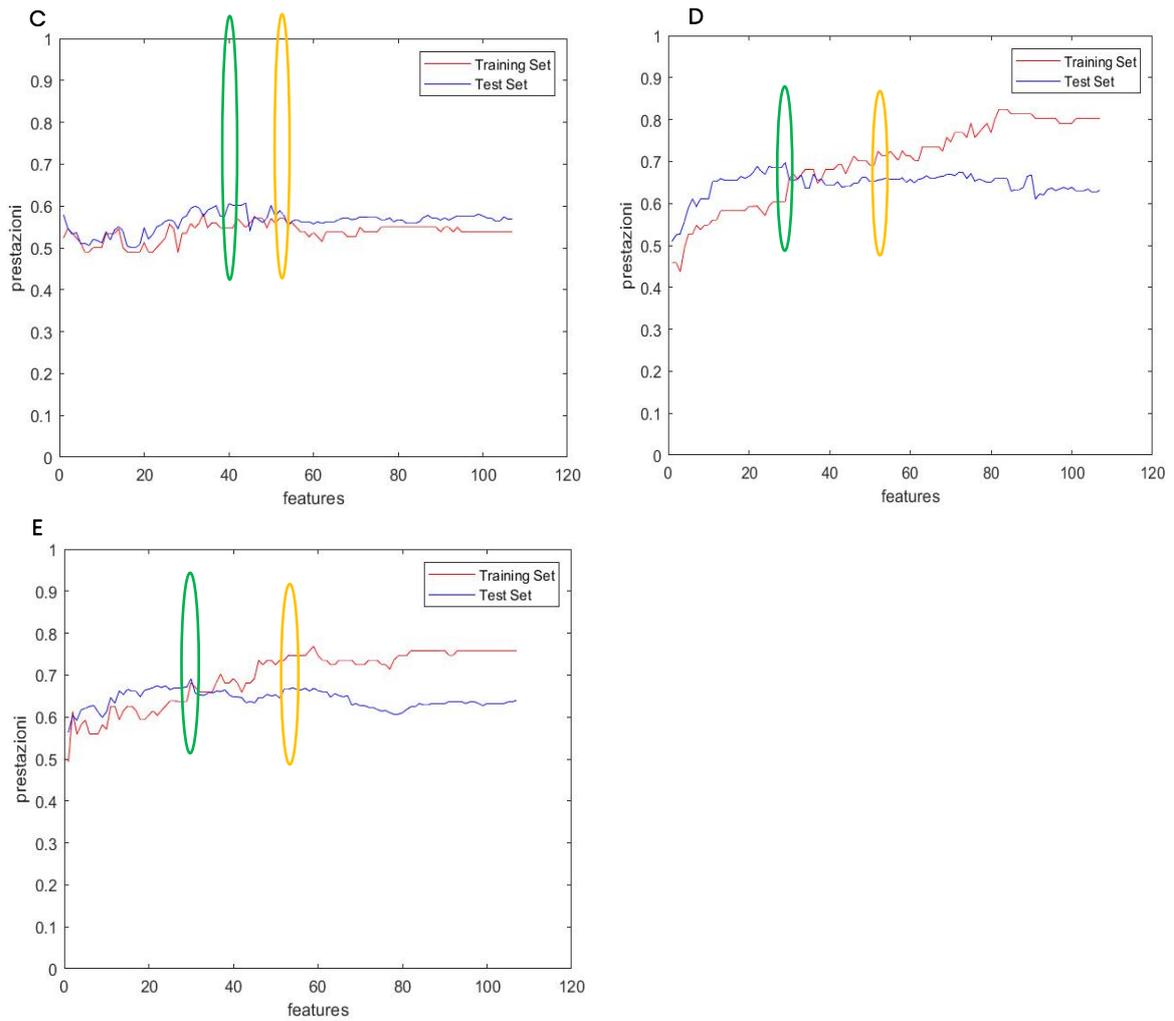
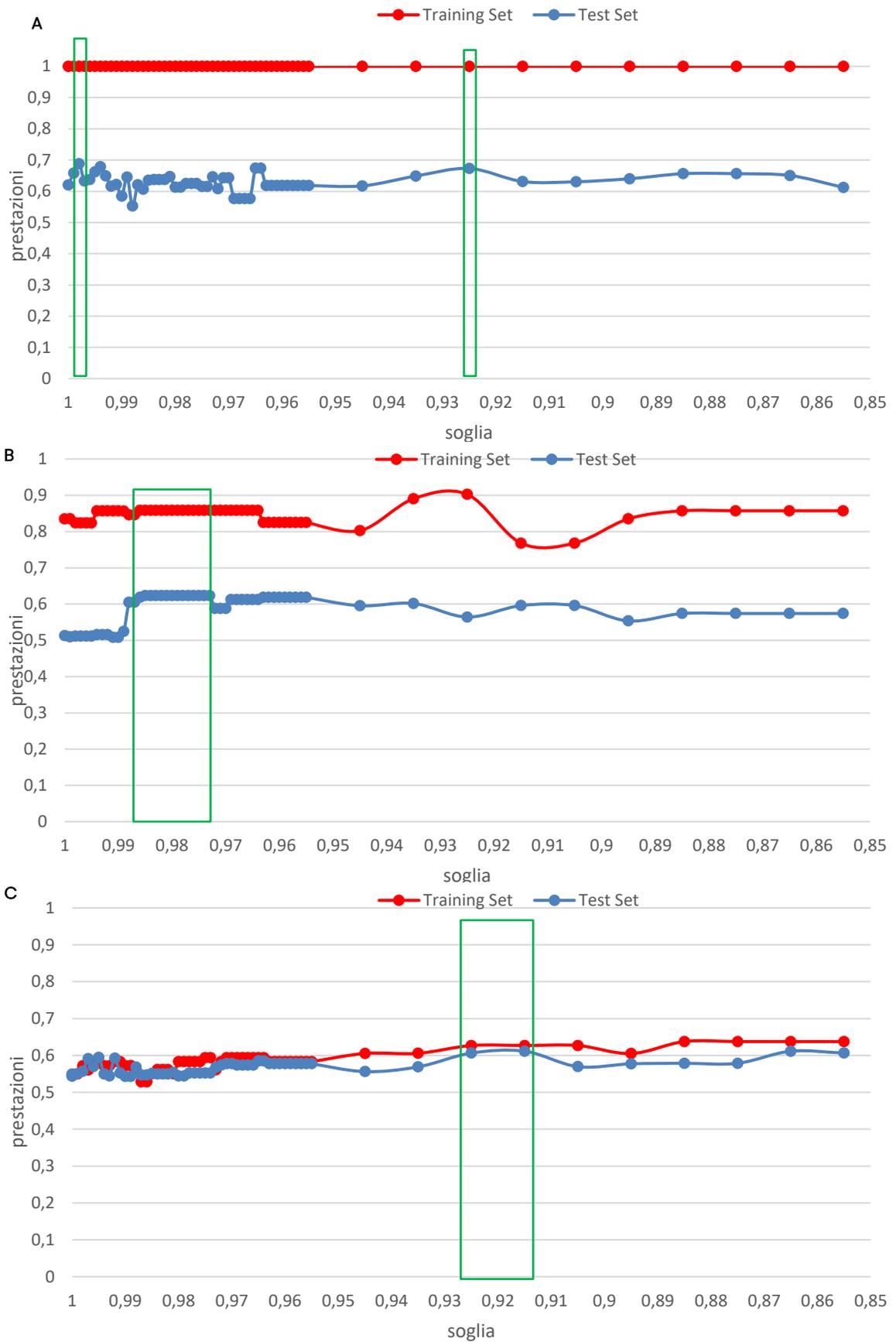


Figura 11: curve relative all'andamento delle prestazioni dei modelli di RF (Figura A), DT (Figura B), bayesiano (Figura C) ed SVM lineare (Figura D) e gaussiano (Figura E) allenati con un numero di features crescente (primo approccio). In verde sono indicati i subset di features in corrispondenza dei quali si verifica un picco nelle prestazioni sul Test Set e in giallo le prestazioni in corrispondenza del subset di features con p -value < 0.05



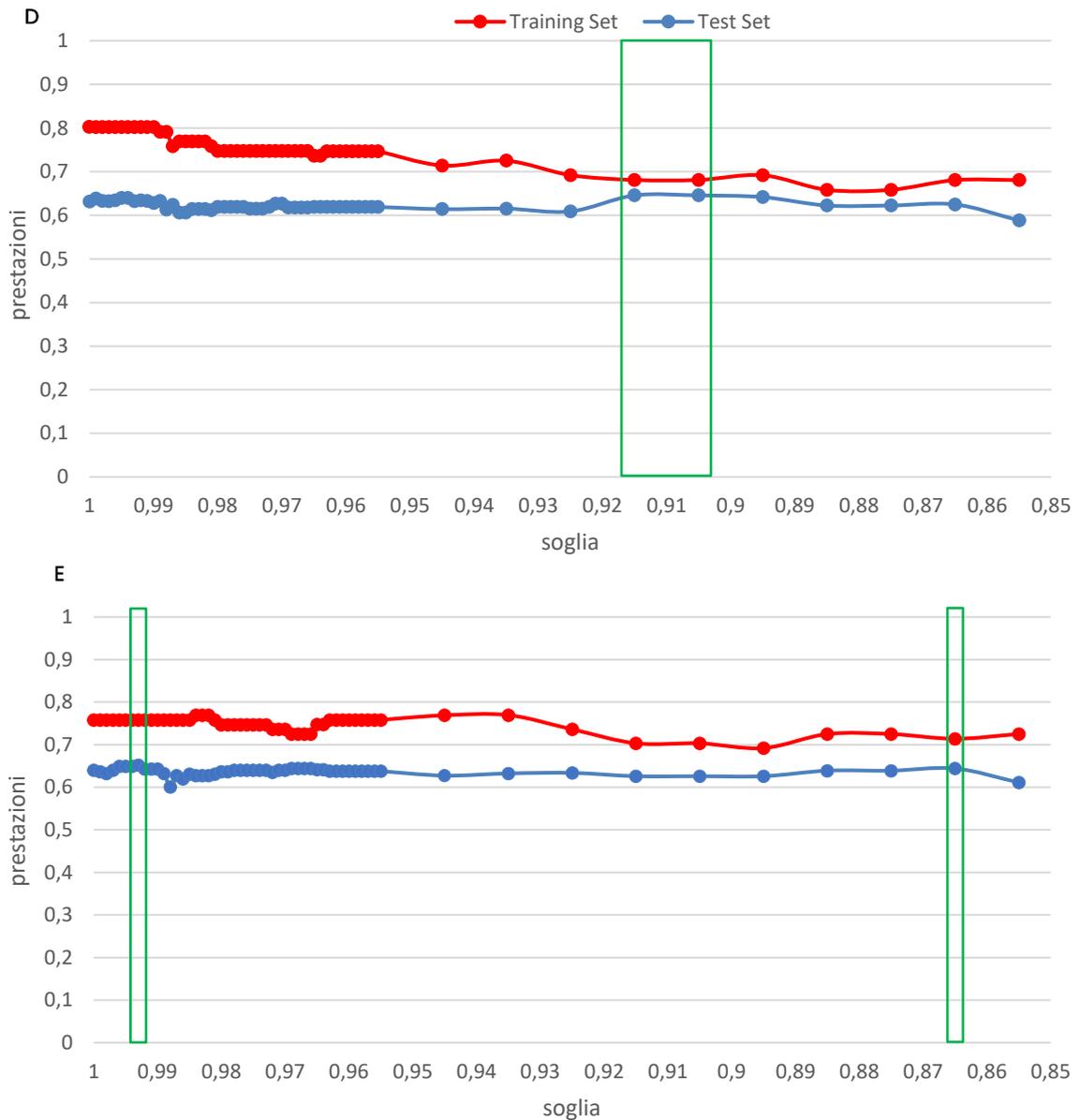


Figura 12: curve relative all'andamento delle prestazioni dei modelli di RF (Figura A), DT (Figura B), bayesiano (Figura C) ed SVM lineare (Figura D) e gaussiano (Figura E) allenati scartando le features meno correlate con la classe appartenenti alle coppie di features correlate fra loro più di una certa soglia (secondo approccio). In verde sono indicati i subset di features in corrispondenza dei quali si verifica un picco nelle prestazioni sul Test.

Sono state valutate le prestazioni dei modelli individuati anche sul Validation Set. Per ogni tipologia di classificatore testato sono stati costruiti i grafici, riportati nelle figure da [13](#) a [16](#), al fine di confrontare le prestazioni dei modelli allenati con un diverso numero di features e individuare i modelli che fornivano le migliori prestazioni sui 3 dataset (in verde nelle figure [13](#) a [16](#)). Nel caso dei modelli di RF sono stati riportati i grafici di confronto delle prestazioni solo per il Test Set e il Validation Set poiché le prestazioni sul Training Set risultavano essere pari al 100% di corretti classificati.

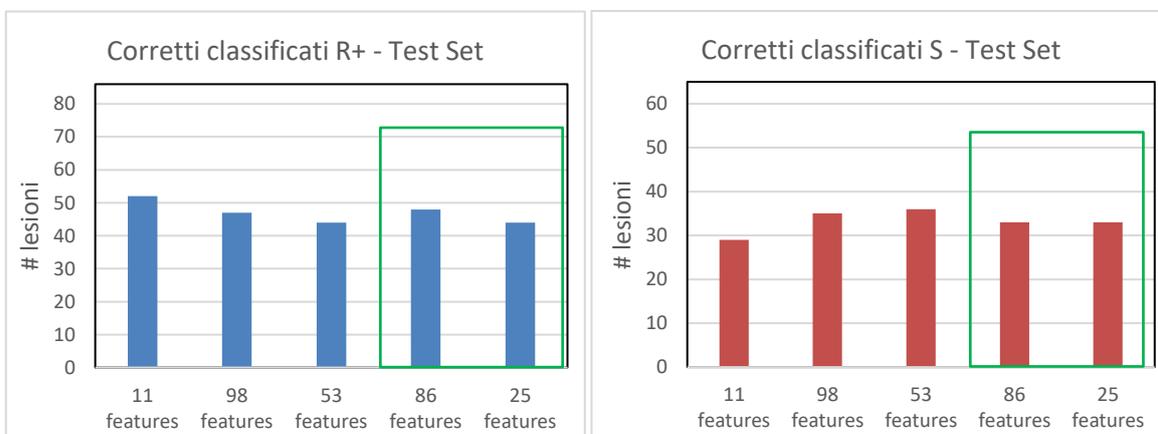
Per i modelli di RF sono stati individuati come migliori (figura 13 in verde) i modelli che fornivano un maggiore numero di corretti classificati relativo alle classi R+ ed R- sul Validation Set. Le performance di tali modelli non risultano complessivamente buone visto il notevole calo delle prestazioni su Test Set e Validation Set rispetto a quelle ottenute sul Training Set: le prestazioni sulle classi R+ ed S sul Test Set sono in media appena superiori al 50% di corretti classificati e quelle sul Validation Set sulla classe R+ si attestano in media intorno al 30% di corretti classificati, sulla classe R- intorno al 41% e sulla classe S intorno al 43%.

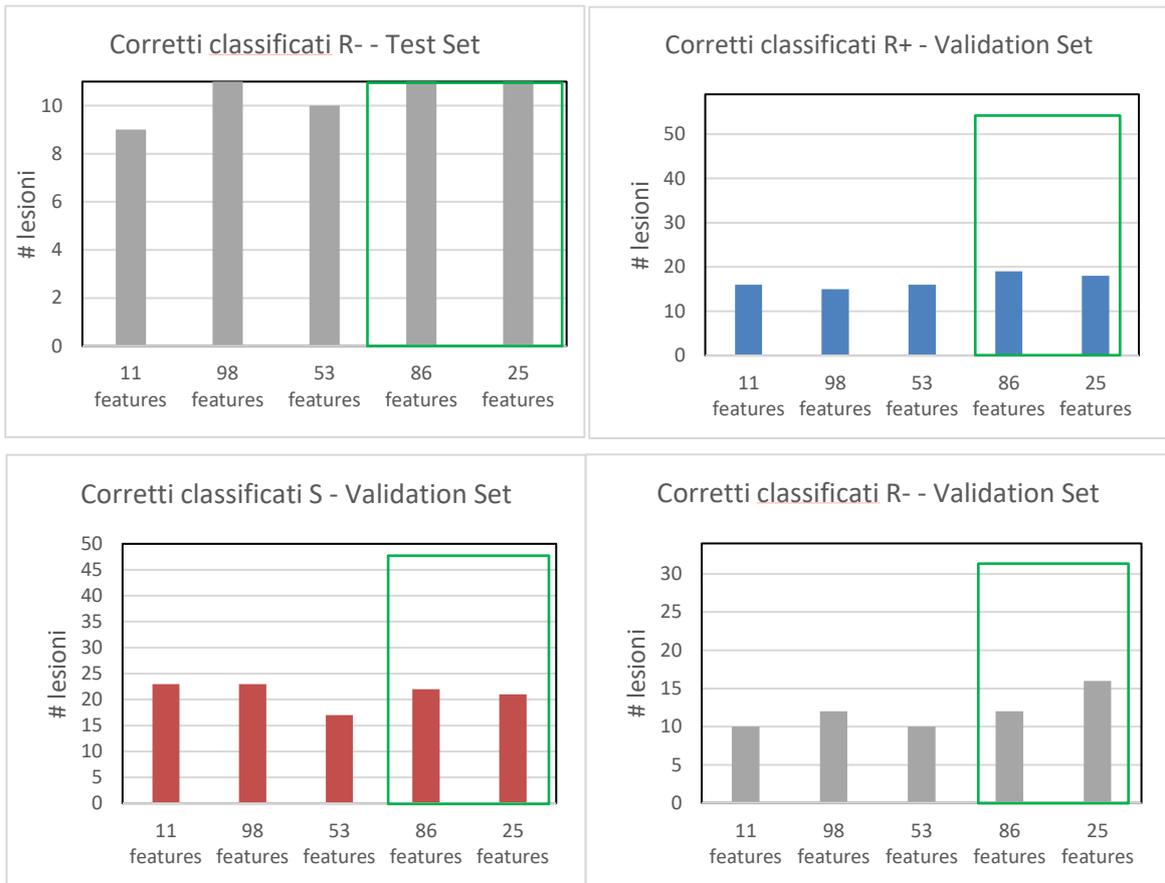
Per i modelli di DT è stato individuato come migliore (figura 14 in verde) il modello che forniva un buon numero di corretti classificati su tutte e 3 le classi sul Validation Set. Anche in questo caso si osserva un calo delle prestazioni relative alle classi R+ ed S sul Test Set che risultano essere pari al 45% e 51% di corretti classificati e relative a tutte e 3 le classi sul Validation Set che risultano essere pari rispettivamente al 20%, 62% e 56%.

Relativamente ai modelli di classificatore bayesiano sono stati individuati come migliori (figura 15 in verde) i 2 che fornivano un maggiore numero di corretti classificati relativi alla classe R+ sul Validation Set. Le prestazioni sulla classe R+ non sono buone essendo inferiori al 50 % di corretti classificati su tutti e 3 i dataset con entrambi i modelli. Le prestazioni relative alla classe S, rispetto a quelle ottenute sul Training Set, peggiorano su Test Set e Validation Set e su quest'ultimo dataset peggiorano anche quelle relative alla classe R- con entrambi i modelli.

Relativamente ai modelli di SVM, infine, è stato individuato come migliore (figura 16 in verde) il modello che forniva il più alto numero di corretti classificati relativi alla classe R+ nel Validation Set. Le prestazioni relative alle 3 classi tendono a peggiorare su Test Set e Validation Set rispetto a quelle ottenute sul Training Set ma in maniera minore rispetto ai modelli precedenti.

Ad eccezione dei modelli di SVM, i modelli più performanti sono stati individuati tra quelli ottenuti sfruttando il secondo approccio di feature selection basato sulla correlazione. Questo potrebbe essere dovuto al fatto che il primo approccio non considera l'eventuale capacità discriminante di una feature se combinata con un'altra.





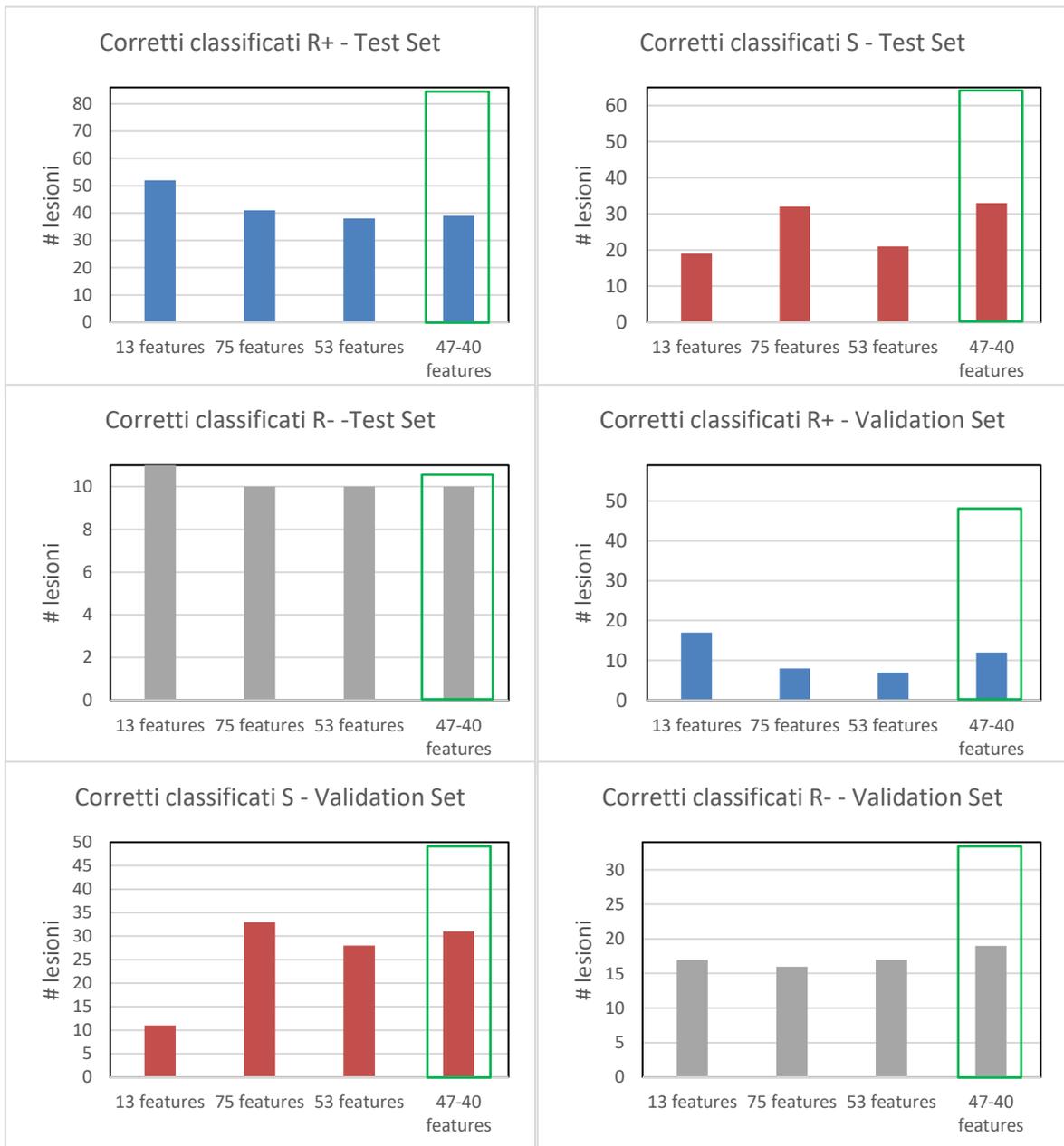
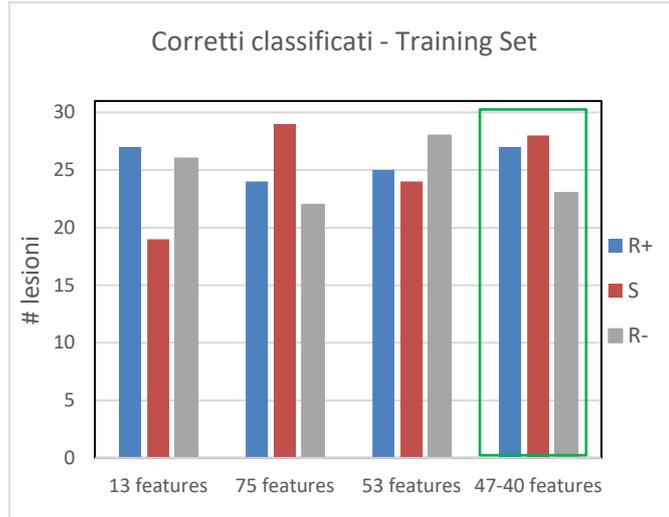
TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	30	0	0	Classe predetta	R+	48	14	0	Classe predetta	R+	19	16	15
	S	0	30	0		S	20	33	0		S	16	22	7
	R-	0	0	31		R-	18	18	11		R-	24	12	12

Confusion matrix con le prestazioni fornite dal RF allenato con 86 Features (*correlazione* < 0.998)

TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	30	0	0	Classe predetta	R+	44	15	0	Classe predetta	R+	18	13	12
	S	0	30	0		S	28	33	0		S	21	21	6
	R-	0	0	31		R-	14	17	11		R-	20	16	16

Confusion matrix con le prestazioni fornite dal RF allenato con 25 Features (*correlazione* < 0.925)

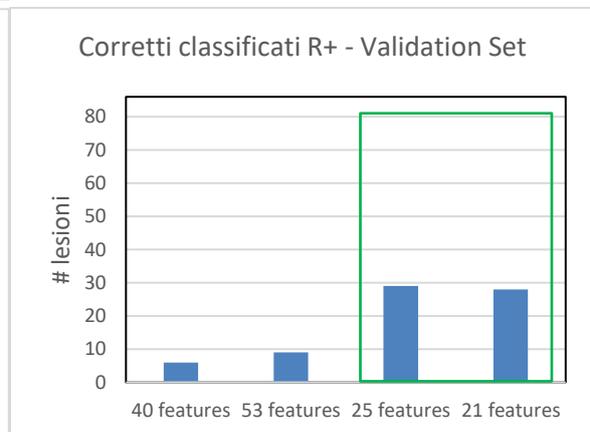
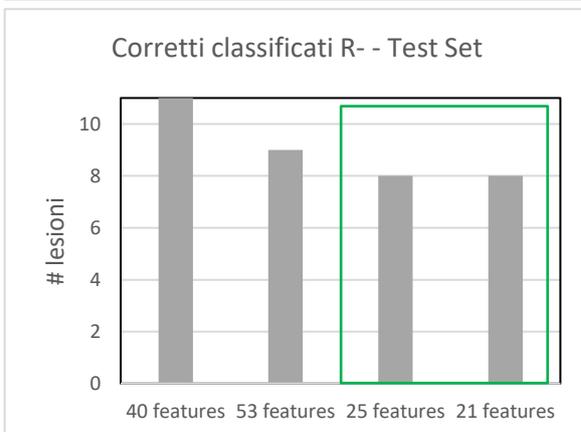
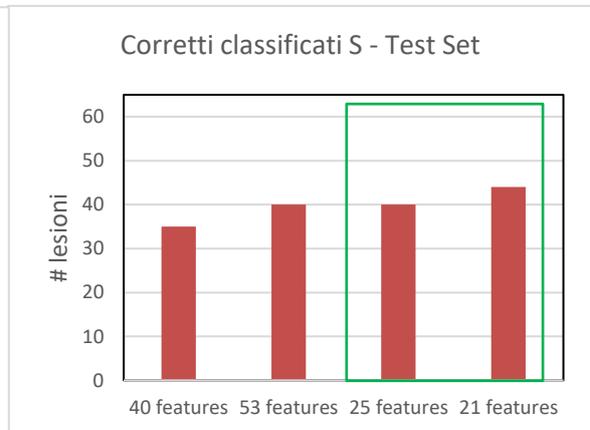
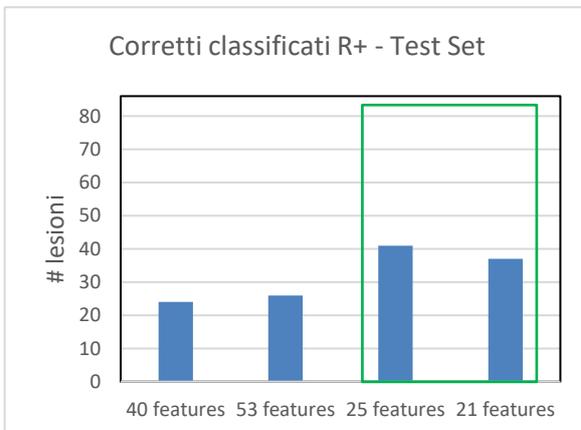
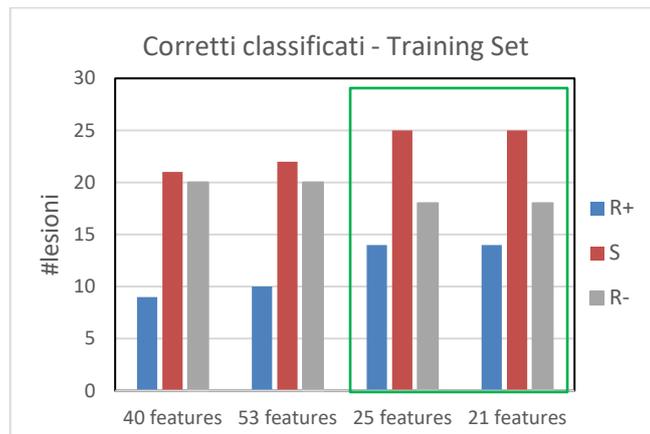
Figura 13: confronto del n° di corretti classificati ottenuti con i modelli di RF migliori individuati (approccio 1 e 2). In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.

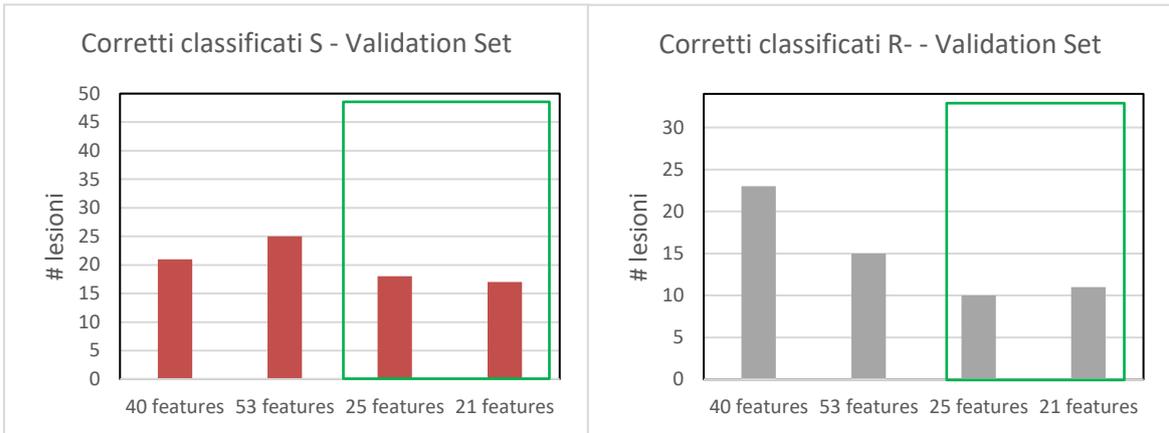


TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	27	2	5	Classe predetta	R+	39	23	0	Classe predetta	R+	12	7	3
	S	0	28	3		S	30	33	1		S	31	31	12
	R-	3	0	23		R-	17	9	10		R-	16	12	19

Confusion matrix con le prestazioni fornite dal DT allenato con 47-40 Features (*correlazione* < 0.985-0.973)

Figura 14: confronto del n° di corretti classificati ottenuti con i modelli di DT migliori individuati (approccio 1 e 2). In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.





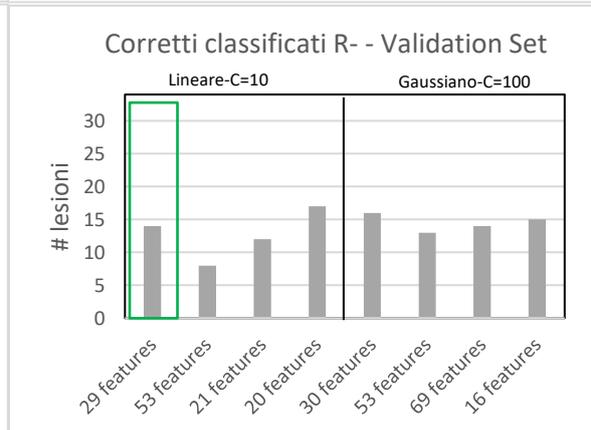
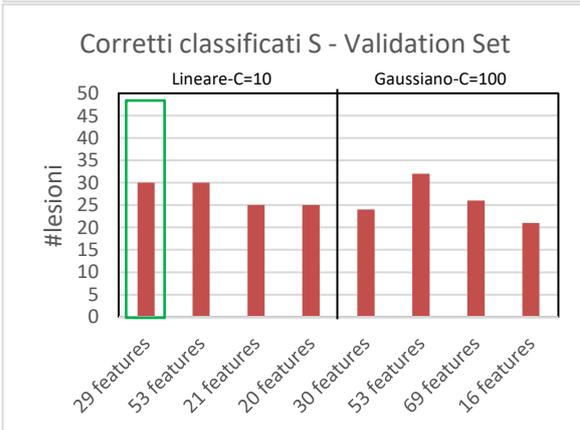
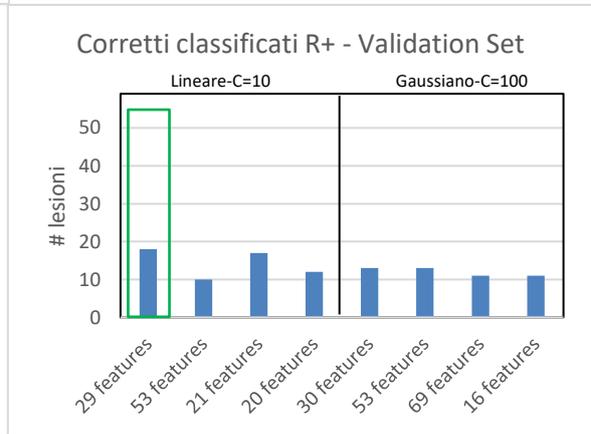
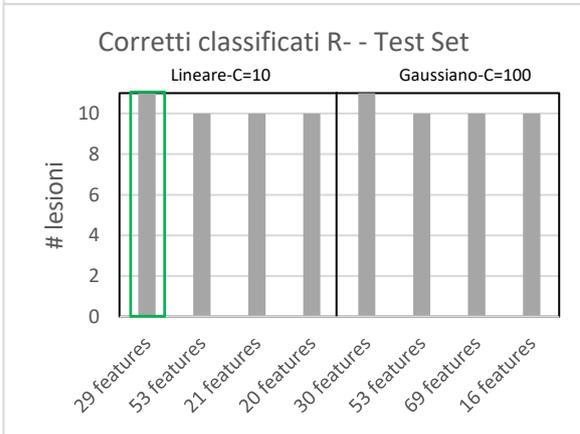
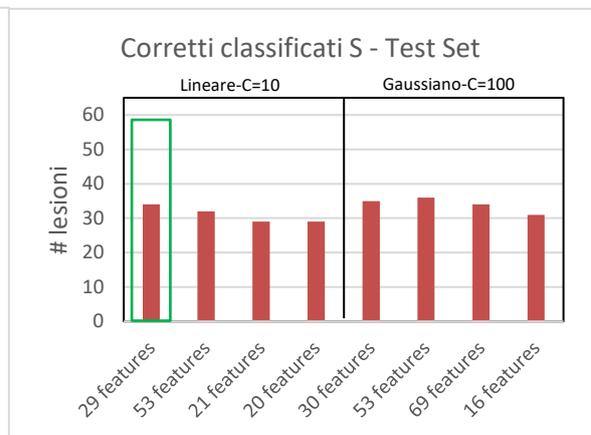
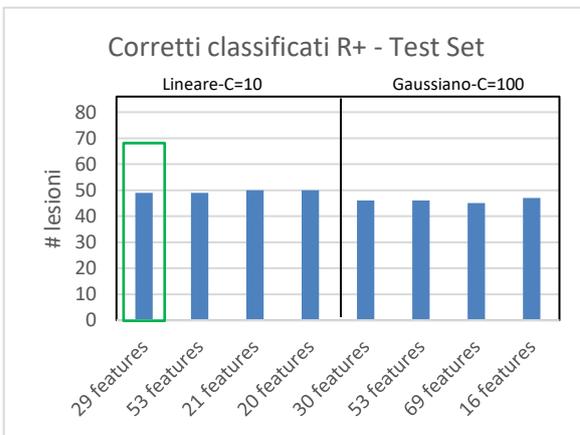
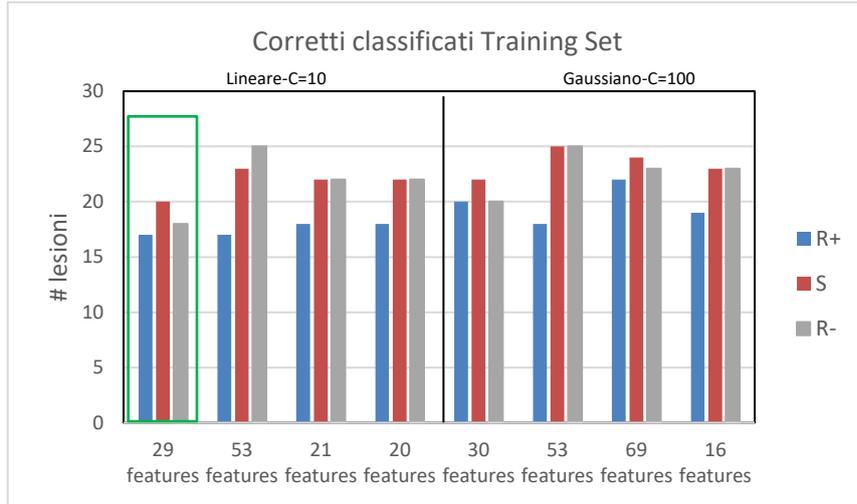
TRS	Classe vera			TS	Classe vera			VS	Classe vera			
	R+	S	R-		R+	S	R-		R+	S	R-	
Classe predetta	R+	14	1	2	R+	41	13	0	R+	29	23	15
	S	9	25	11	S	36	40	3	S	18	18	9
	R-	7	4	18	R-	9	12	8	R-	12	9	10

Confusion matrix con le prestazioni fornite dal bayesiano allenato con 25 Features (*correlazione* < 0.925)

TRS	Classe vera			TS	Classe vera			VS	Classe vera			
	R+	S	R-		R+	S	R-		R+	S	R-	
Classe predetta	R+	14	0	2	R+	37	8	0	R+	28	26	15
	S	9	25	11	S	39	44	3	S	20	17	8
	R-	7	5	18	R-	10	13	8	R-	11	7	11

Confusion matrix con le prestazioni fornite dal bayesiano allenato con 21 Features (*correlazione* < 0.915)

Figura 15: confronto del n° di corretti classificati ottenuti con i modelli di bayesiano migliori individuati (approccio 1 e 2). In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.



TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	17	6	4	Classe predetta	R+	49	16	0	Classe predetta	R+	18	13	7
	S	7	20	9		S	27	34	0		S	29	30	13
	R-	6	4	18		R-	10	15	11		R-	12	7	14

Confusion matrix con le prestazioni fornite dal SVM allenato con 29 Features

Figura 16: confronto del n° di corretti classificati ottenuti con i modelli di SVM migliori individuati (approccio 1 e 2). In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.

QuikReduct Algorithm

È stata effettuata infine features selection tramite l'algoritmo QuickReduct (QRA).

Il QRA è un algoritmo basato sulla Rough Set Theory (RST) che genera un Reduct, ovvero un sottoinsieme minimo di features che sono le più informative, basandosi sulla funzione di dipendenza.

Dato un insieme di attributi Q , questo dipende totalmente da un insieme di attributi P , se tutti i valori degli attributi di Q sono determinati unicamente dai valori degli attributi di P . Nella RST, la dipendenza è definita dall'equazione (5):

$$\gamma_P(Q) = \frac{|POS_P(Q)|}{|\mathbb{U}|} \quad (5)$$

Dove:

- $P, Q \subset \mathbb{A}$, con \mathbb{A} un insieme finito non vuoto di n attributi
- \mathbb{U} è un insieme non vuoto di N oggetti (l'universo)
- $|S|$ indica la cardinalità dell'insieme S
- $POS_P(Q) = \text{positive region} = \bigcup_{X \in \mathbb{U}/Q} \underline{P}X$ dove X è il set di oggetti ($X \subseteq \mathbb{U}$). L'approssimazione P-lower $\underline{P}X = P - \text{lower} = \{x | [x]_P \subseteq X\}$ è l'insieme degli oggetti che appartengono inevitabilmente ad una classe.
- $|POS_P(Q)|$ è uguale al numero di elementi del dataset che sono distinguibili rispetto al sottoinsieme di attributi P . Dato un subset di features A e 2 elementi d e $d' \in \mathbb{U}$, con classi c e c' , i vettori di dati \mathbf{v} e \mathbf{v}' sono generati in modo che se $A = a_1, a_2, \dots, a_n$, $\mathbf{v} = \langle d_{a_1}, d_{a_2}, \dots, d_{a_n} \rangle$ e $\mathbf{v}' = \langle d'_{a_1}, d'_{a_2}, \dots, d'_{a_n} \rangle$ allora d e d' sono distinguibili se $\mathbf{v} \neq \mathbf{v}' \vee c = c'$.

L'algoritmo QuickReduct, riportato in [Figura 17](#), inizia con un insieme di features vuoto e calcola il grado di dipendenza di ogni attributo. Scelta la feature con il grado di dipendenza maggiore, la aggiunge al subset vuoto. Vengono poi aggiunti a turno, uno alla volta, quegli attributi che provocano il maggior aumento del grado di dipendenza dell'insieme, fino a che questo produce il suo massimo valore possibile per il set di dati [\[22-23\]](#).

```

QUICKREDUCT(C, D)
Input: C, the set of all conditional features; D, the set of
decision features
Output: R, the feature subset
(1) R ← {}
(2) while  $\gamma_R(D) \neq \gamma_C(D)$ 
(3)   T ← R
(4)   foreach  $x \in (C - R)$ 
(5)     if  $\gamma_{R \cup \{x\}}(D) > \gamma_T(D)$ 
(6)       T ← R ∪ {x}
(7)   R ← T
(8) return R
    
```

Figura 17: Algoritmo QuickReduct [\[22\]](#)

Poiché in ciascuna iterazione dell'algoritmo potrebbero esserci 2 o più features che portano allo stesso incremento di γ e la scelta di quale feature aggiungere al subset sarebbe arbitraria, il QRA è stato modificato in modo che in questi casi siano portati avanti tutti i subset contemporaneamente.

Per utilizzare l'algoritmo occorre che i valori delle features siano interi. In caso di valori continui, come nel problema in esame, questi devono essere discretizzati.

Per la discretizzazione dei dati è stato utilizzato l'algoritmo ChiMerge che consente, a differenza di metodi di discretizzazione come equal-width-intervals (divide la linea dei numeri tra il valore minimo e il valore massimo in N intervalli di uguale dimensione) e equal-frequency-intervals (i confini degli intervalli sono scelti in modo che ogni intervallo contenga approssimativamente lo stesso numero di elementi), di individuare i confini degli intervalli considerando la classe degli elementi nel dataset.

L'algoritmo procede, considerando una variabile alla volta, valutando il valore di χ^2 (equazione 6) per ogni coppia di intervalli adiacenti per verificare se la classe di tutti gli elementi nei due intervalli sia indipendente dai valori che la variabile assume per quegli elementi e fondendo quegli intervalli con il valore minimo di χ^2 .

La formula per calcolare il valore χ^2 è:

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(A_{ij} - E_{ij})^2}{E_{ij}} \quad (6)$$

dove:

- $m = 2$ (i 2 intervalli che vengono confrontati)
- $k = 3$ (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 $A_{ij} = \frac{R_i \cdot C_j}{N}$

L'algoritmo si ferma quando il valore di χ^2 per tutti gli intervalli adiacenti è maggiore di una determinata soglia. Il valore di χ^2 -soglia è determinato selezionando un livello di significatività desiderato e un numero di gradi di libertà pari al numero totale di classi - 1 [20-21]. In questo lavoro, è stato fissato un livello di significatività pari a 0.95 e i gradi di libertà pari a 2 poiché si considerano 3 classi (R+ S ed R-).

I limiti dei valori discretizzati per ogni feature sono stati calcolati applicando il ChiMerge al Construction Set.

L'operazione di FS tramite QRA è stata effettuata sul Construction Set discretizzato.

Risultati

La [tabella 28](#) in Appendice mostra i risultati della discretizzazione in termini di numero di valori distinti iniziali per ognuna delle 107 features (terza colonna) e numero di classi discrete identificate dall'algoritmo ChiMerge (quarta colonna).

La [tabella 29](#) in Appendice mostra i 46 subset di features individuati dal QRA in termini di numero (seconda colonna) e nome delle features selezionate (terza colonna). Tutti i subset erano costituiti da 4 features.

Il grafico combinato in figura [70](#) in Appendice mostra in azzurro i valori dei coefficienti di correlazione di Spearman di ciascuna feature nel Construction Set con la classe (calcolati come riportato nel capitolo [Feature selection basata su correlazione](#)) e in arancione il numero di volte in cui la corrispondente feature è presente nei 46 subset di variabili individuati dal QRA. La maggior parte delle variabili è presente in 0-2 subset, la feature *Long Run Emphasis* (glrlm) è presente in 12 subset, *Mean Absolute Deviation* si ripete in 35 soluzioni e infine le features *Low Gray Level Zone Emphasis* (glszm) e *Maximum Probability* (glcm) sono presenti in tutti i 46 subset.

È stato osservato che il QRA ha selezionato anche variabili poco correlate con la classe e visto che anche i coefficienti di correlazione di Spearman tra le features e la classe risultavano essere bassi è stata osservata un'assenza di relazione lineare tra classe e features. È stato quindi deciso di utilizzare, a seguito del QRA, metodi di classificazione non lineari: SVM gaussiano e polinomiale e reti neurali artificiali (NN).

Come sottoinsieme di features derivante dal QRA è stata scelta la soluzione 14 poiché conteneva le 4 features (Long Run Emphasis (glrlm), Mean Absolute Deviation, Low Gray Level Zone Emphasis (glszm) e Maximum Probability (glcm)) maggiormente selezionate nei 46 subsets trovati dal QRA.

Relativamente al SVM, sono stati allenati diversi modelli di SVM gaussiano e polinomiale sul Training Set (TRS_SOM_1), con la soluzione individuata dal QRA e con tutte le combinazioni dei parametri riportate in [tabella 8](#) e valutati sul Test Set (TS_SOM_1) con i valori continui normalizzati con il min-max scaling al fine di individuare la struttura più performante. Sono state inoltre valutate le prestazioni degli stessi modelli di SVM allenati sul Training Set (TRS_SOM_1) con i valori discretizzati con il ChiMerge.

	SVM polinomiale	SVM gaussiano
<i>C</i>	0.01, 0.1, 1, 5, 10, 50, 100	0.01, 0.1, 1, 5, 10, 50, 100
<i>Gamma</i>	1	0.01, 0.1, 1, 5, 10, 50, 100
<i>Coefficiente</i>	1	/
<i>Grado</i>	2, 3, 4, 5, 6, 7, 8, 9, 10	/

Tabella 8: parametri relativi ai modelli di SVM polinomiale e gaussiano testati.

Nelle figure [18](#) e [20](#) sono state graficate le prestazioni su Training Set e Test Set dei diversi modelli di SVM allenati sul Training Set utilizzando come metrica l'equazione (4) e sono state evidenziate le strutture più performanti. Nelle figure [19](#) e [21](#) sono riportate le confusion matrix relative alle prestazioni su Training Set, Test Set e Validation Set ottenute con questi modelli.

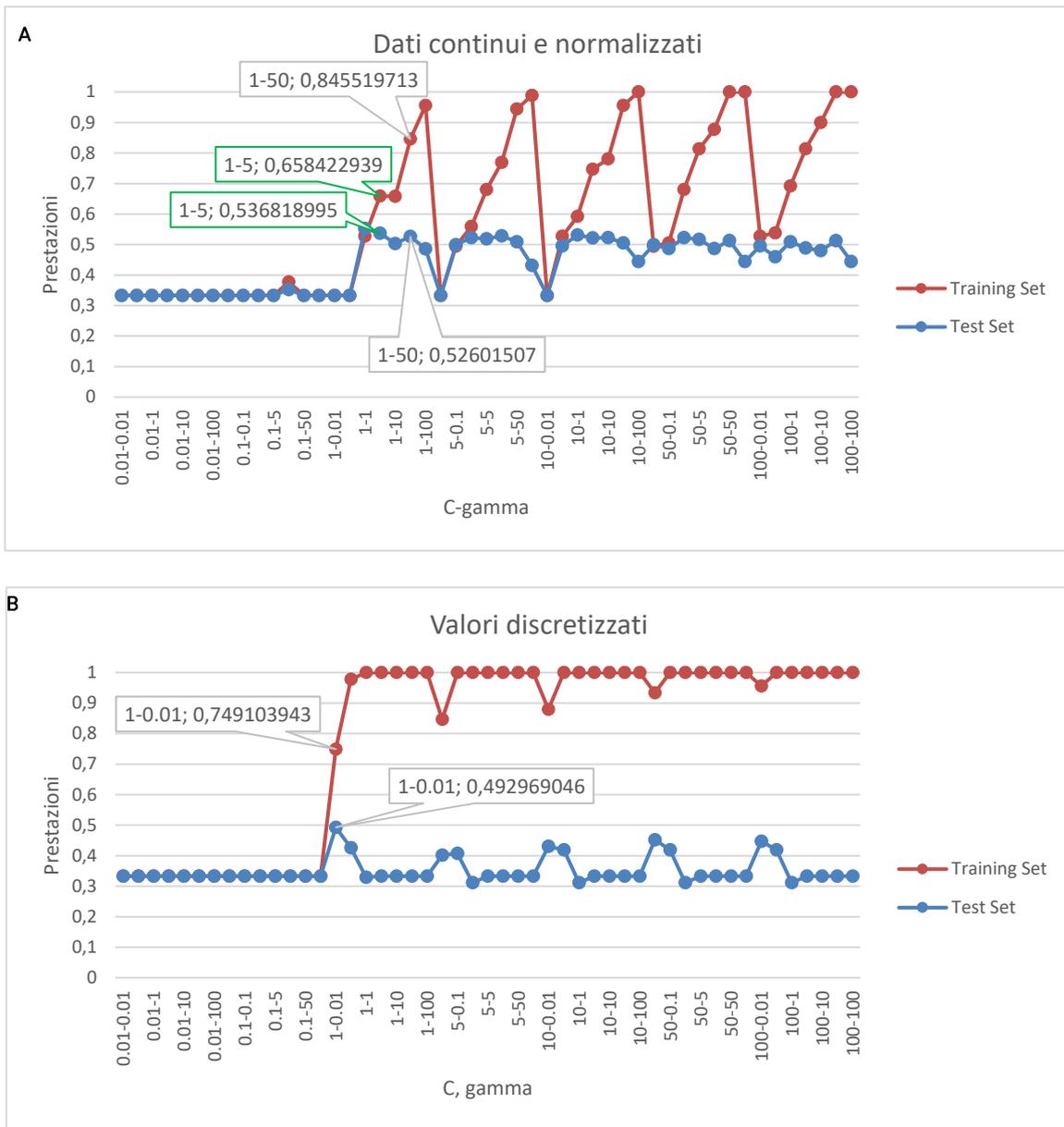


Figura 18: prestazioni su Training Set e Test Set dei diversi modelli di SVM gaussiano allenati con i dati continui (Figure A) e con i dati discreti (Figure B). Sono evidenziate le strutture più performanti.

TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	17	6	1	Classe predetta	R+	27	7	1	Classe predetta	R+	43	34	23
	S	8	20	7		S	45	37	2		S	12	10	9
	R-	5	4	23		R-	14	21	8		R-	4	6	2

Confusion matrix con le prestazioni fornite dal SVM gaussiano ($C = 1$ e $\gamma = 5$) allenato sul TRS_SOM_1 con valori continui normalizzati.

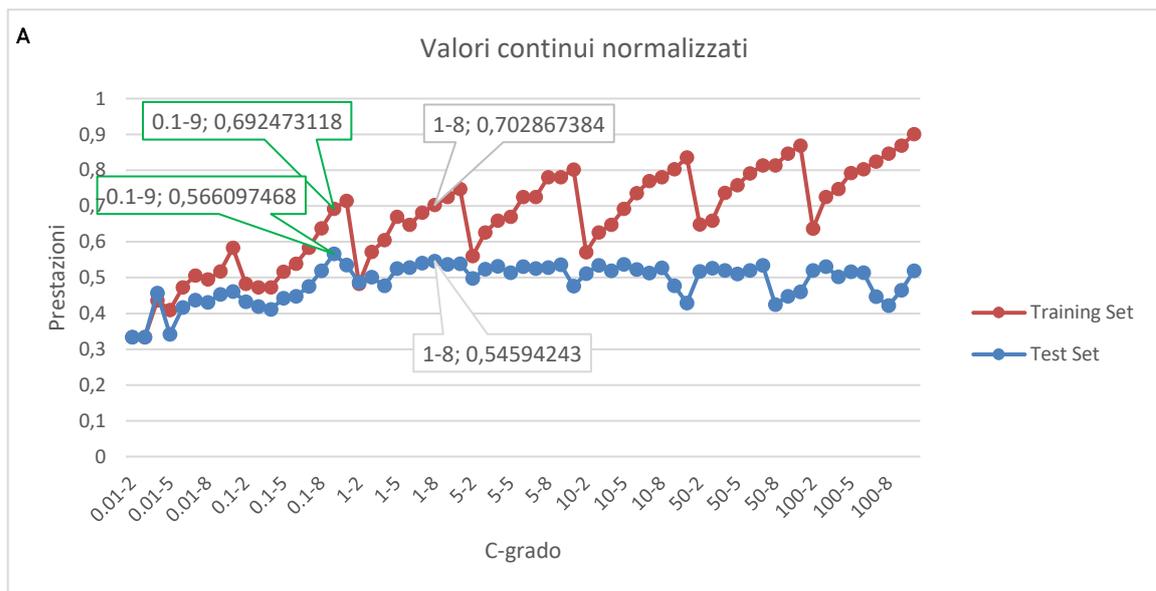
TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	25	3	1	Classe predetta	R+	36	13	1	Classe predetta	R+	48	37	29
	S	4	24	2		S	34	34	3		S	7	9	3
	R-	1	3	28		R-	16	18	7		R-	4	4	2

Confusion matrix con le prestazioni fornite dal SVM gaussiano ($C = 1$ e $\gamma = 50$) allenato sul TRS_SOM_1 con valori continui normalizzati.

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	23	2	5	Classe predetta	R+	30	13	3	Classe predetta	R+	32	28	18
	S	6	27	8		S	41	38	2		S	25	21	12
	R-	1	1	18		R-	15	14	6		R-	2	1	4

Confusion matrix con le prestazioni fornite dal SVM gaussiano ($C = 1$ e $\gamma = 0.01$) allenato sul TRS_SOM_1 con valori discretizzati.

Figura 19: confusion matrix con le prestazioni fornite dai modelli di SVM gaussiani più performanti.



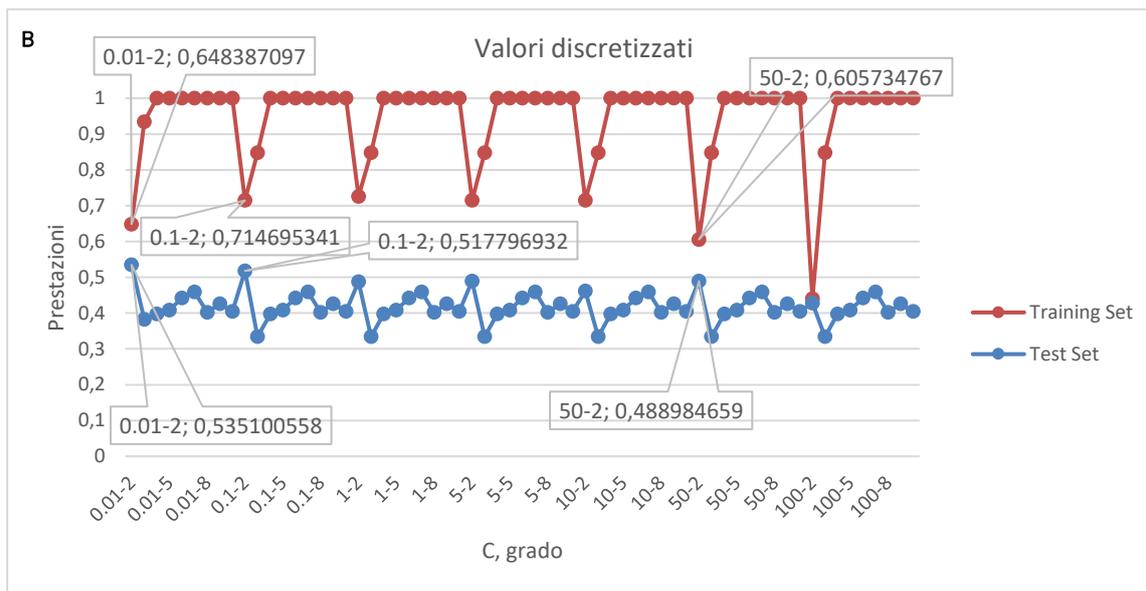


Figura 20: prestazioni su Training Set e Test Set dei diversi modelli di SVM polinomiale allenati con i dati continui (Figure A) e con i dati discreti (Figure B). Sono evidenziate le strutture più performanti.

TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	12	0	0	Classe predetta	R+	20	3	1	Classe predetta	R+	30	23	20
	S	15	30	10		S	56	48	2		S	27	21	12
	R-	3	0	21		R-	10	14	8		R-	2	6	2

Confusion matrix con le prestazioni fornite dal SVM polinomiale ($C = 0.1$ e $grado = 9$) allenato sul TRS_SOM_1 con valori continui normalizzati.

TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	20	8	1	Classe predetta	R+	32	10	1	Classe predetta	R+	43	31	22
	S	7	21	7		S	38	35	2		S	12	12	10
	R-	3	1	23		R-	16	20	8		R-	4	7	2

Confusion matrix con le prestazioni fornite dal SVM polinomiale ($C = 1$ e $grado = 8$) allenato sul TRS_SOM_1 con valori continui normalizzati.

TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	16	3	2	Classe predetta	R+	28	14	1	Classe predetta	R+	36	32	23
	S	11	23	9		S	42	30	1		S	20	16	10
	R-	3	4	20		R-	16	21	9		R-	3	2	1

Confusion matrix con le prestazioni fornite dal SVM polinomiale ($C = 0.01$ e $grado = 2$) allenato sul TRS_SOM_1 con valori discretizzati.

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	20	1	3	Classe predetta	R+	34	22	2	Classe predetta	R+	44	42	26
	S	8	24	7		S	36	28	1		S	11	5	7
	R-	2	5	21		R-	16	15	8		R-	4	3	1

Confusion matrix con le prestazioni fornite dal SVM polinomiale ($C = 0.1$ e $grado = 2$) allenato sul TRS_SOM_1 con valori discretizzati.

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	24	7	8	Classe predetta	R+	41	19	3	Classe predetta	R+	44	41	30
	S	3	16	8		S	29	23	1		S	11	8	4
	R-	3	7	15		R-	16	23	7		R-	4	1	0

Confusion matrix con le prestazioni fornite dal SVM polinomiale ($C = 50$ e $grado = 2$) allenato sul TRS_SOM_1 con valori discretizzati.

Figura 21: confusion matrix con le prestazioni fornite dai modelli di SVM polinomiale più performanti.

Come ci si aspettava, all'aumentare dei parametri $\gamma/grado$ i modelli tendono ad essere soggetti al fenomeno di overfitting; le prestazioni sul Training Set risultano essere prossime o pari al 100% di corretti classificati mentre le prestazioni sul Test Set risultano essere peggiori. Sono stati quindi individuati quei modelli con performance confrontabili sui 2 dataset.

Confrontando le prestazioni ottenute con i modelli allenati con il subset di features individuato dal QRA con valori continui e con valori discreti si osserva che nel secondo caso si ottengono prestazioni molto simili al primo caso ma utilizzando valori di $\gamma/grado$ inferiori.

Le prestazioni ottenute non risultano essere buone, in particolare sul Validation Set in cui i diversi modelli tendono a classificare più della metà delle lesioni in classe R+ (i modelli classificano in classe R+ un numero di lesioni compreso tra 73 e 114).

Relativamente alle reti neurali sono state allenate 9 strutture di NN sul Training Set (TRS_SOM_1) con i valori continui, con la soluzione 14 individuata dal QRA e con soglia fissa per assegnare i soggetti ad una classe (R+/S/R-) (tabella 9):

- 1 layer nascosto costituito da 20 neuroni; 10 neuroni e 5 neuroni
- 2 layer nascosti costituiti rispettivamente da 20 e 10 neuroni; 10 e 10 neuroni; 10 e 5 neuroni
- 3 layer nascosti costituiti rispettivamente da 20, 10 e 5 neuroni; 10, 10 e 10 neuroni; 10, 5 e 3 neuroni

Le prestazioni delle reti sono state valutate sul Test Set (TS_SOM_1) al fine di individuare la struttura più performante.

Sono state inoltre valutate le prestazioni delle stesse reti allenate sul Training Set (TRS_SOM_1) con i valori discretizzati con il ChiMerge.

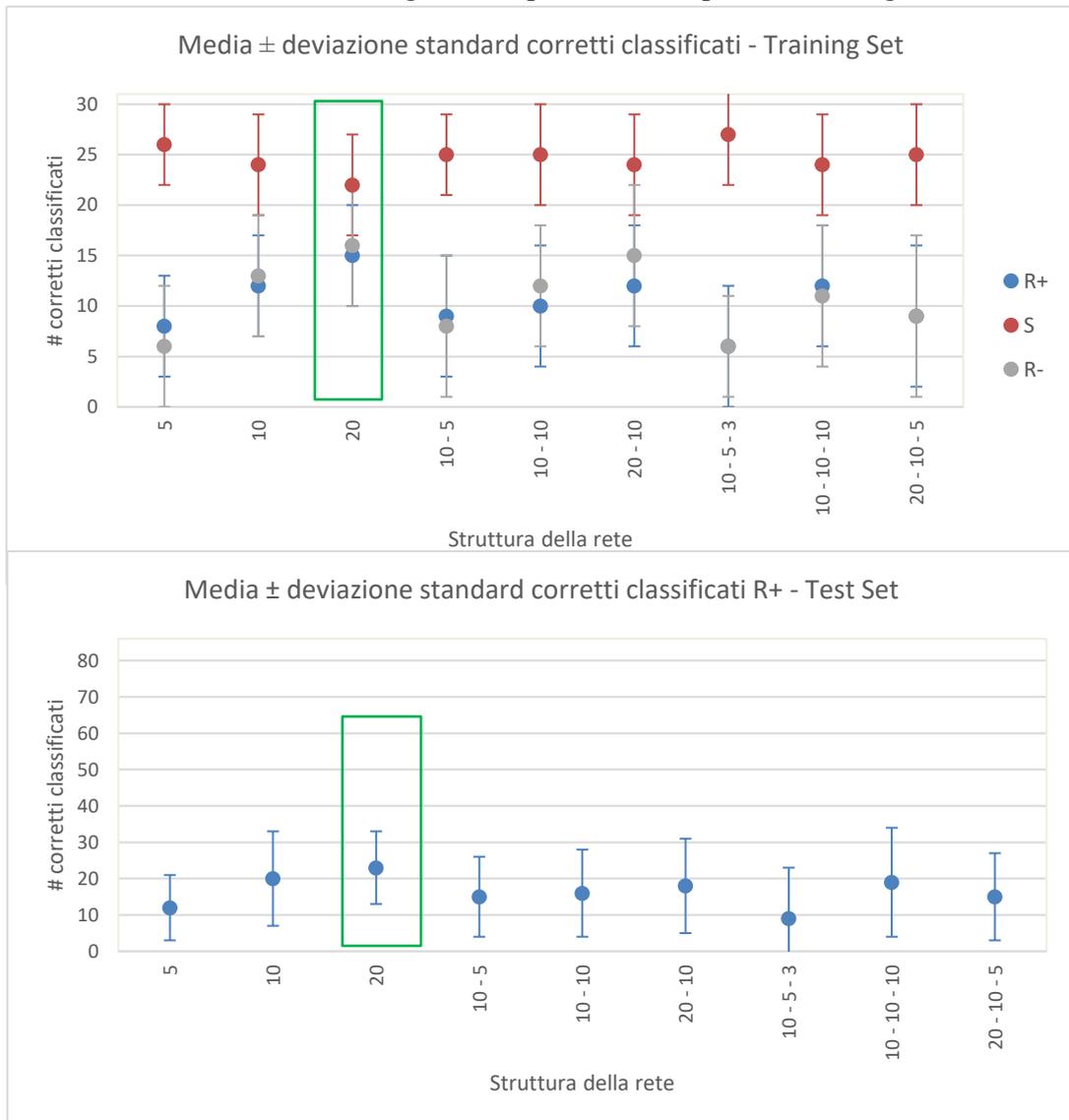
Output della rete y	Classe
$Y \leq 1.5$	R+ (1)
$1.5 < Y \leq 2.5$	S (2)
$Y > 2.5$	R- (3)

Tabella 9: Soglia fissa utilizzata per assegnare i soggetti ad una classe in bassa all'output y fornito dalla rete.

Dato che l'inizializzazione dei pesi delle reti è random, per ogni struttura testata sono state ripetute le fasi di inizializzazione, training e verifica su Training Set e Test Set 100 volte. Per ogni ripetizione, sono state salvate le confusion matrix su Training Set e Test Set e, al termine delle 100 ripetizioni, sono stati calcolati i valori medi e le deviazioni standard delle 9 celle di ciascuna confusion matrix.

In appendice nelle figure 71 e 72 sono riportate le tabelle con la media \pm la deviazione standard di ogni cella per ogni struttura testata.

Nelle figure 22 e 24 sono riportati i grafici di confronto dei valori di media \pm deviazione standard dei corretti classificati ottenuti. Nelle figure 23 e 25 sono riportate le confusion matrix relative alle prestazioni su Training Set, Test Set e Validation Set ottenute con le strutture di NN allenate sul Training Set che presentavano prestazioni migliori.



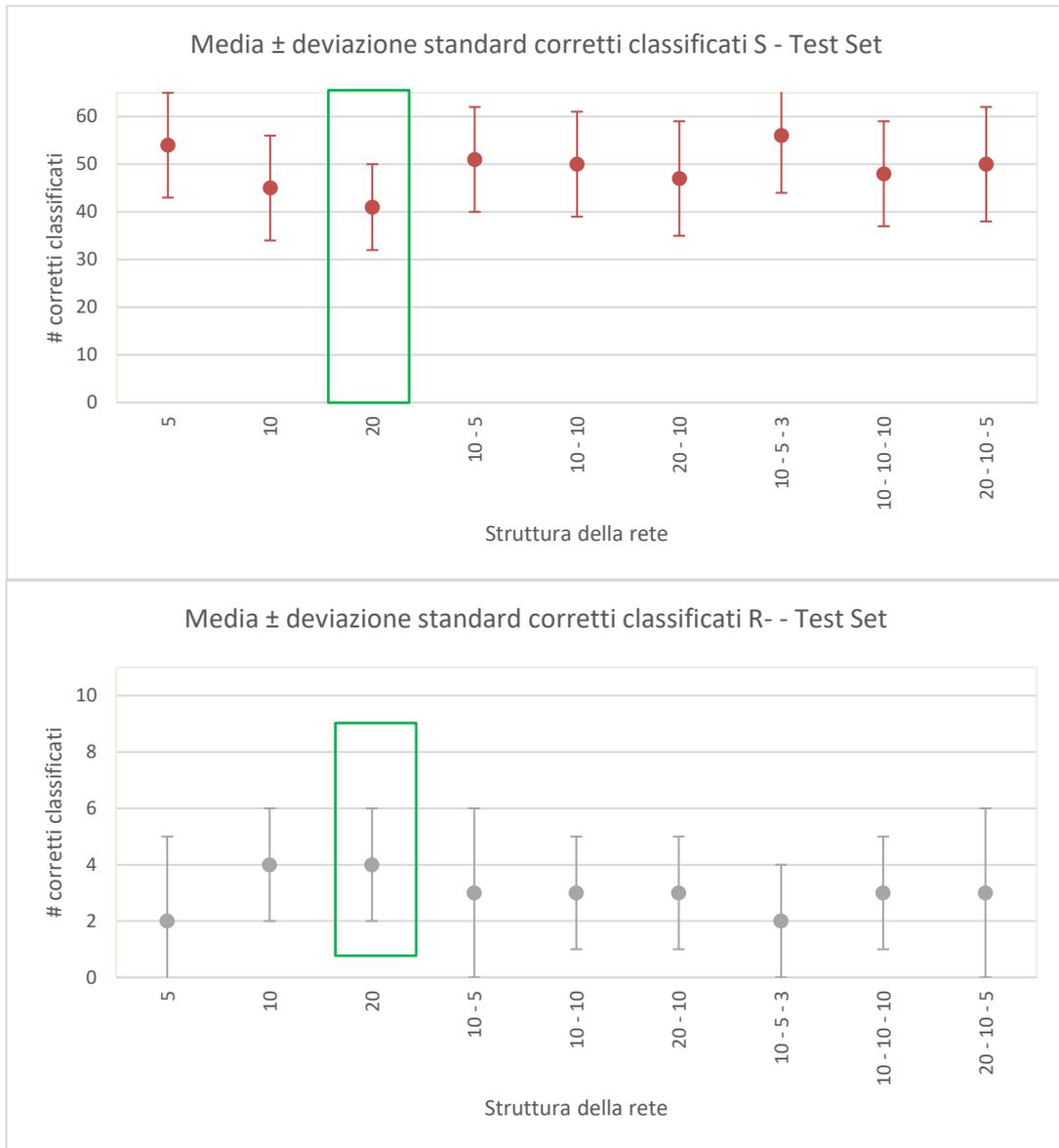
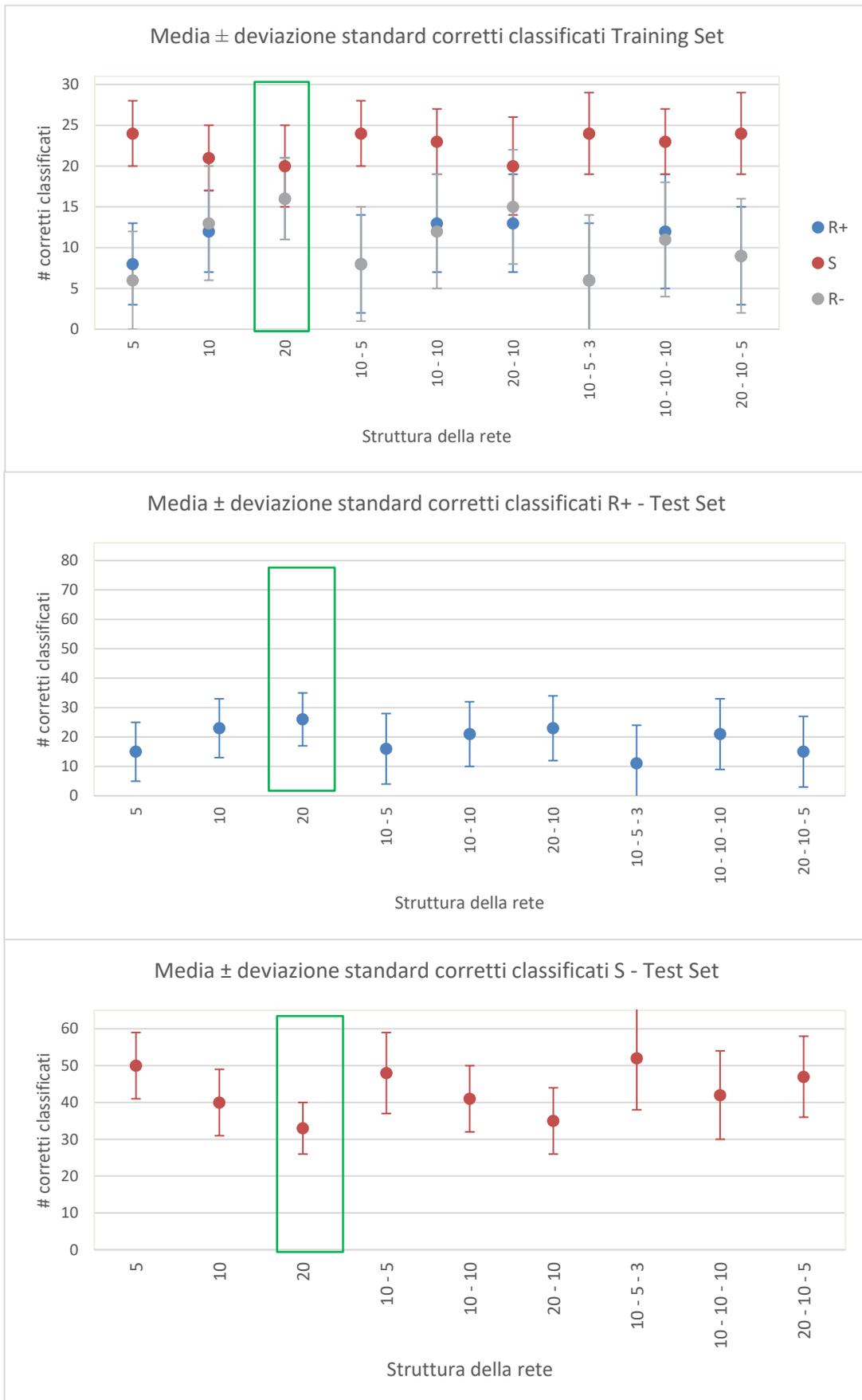


Figura 22: grafici di confronto relativi ai valori medi \pm deviazione standard di corretti classificati ottenuti con le strutture allenate sul Training Set con le features individuate dalla soluzione 14 (dataset con valori continui) e testate su Training Set (Figura A) e Test Set (Figure B, C e D). Sono evidenziate le strutture più performanti.

TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	20	4	2	Classe predetta	R+	32	12	1	Classe predetta	R+	36	27	23
	S	10	21	6		S	39	38	4		S	19	18	9
	R-	0	5	23		R-	15	15	6		R-	4	5	2

Figura 23: confusion matrix con le prestazioni fornite dalla NN con 1 layer nascosto costituito da 20 neuroni allenata su TRS_SOM_1 con valori continui.



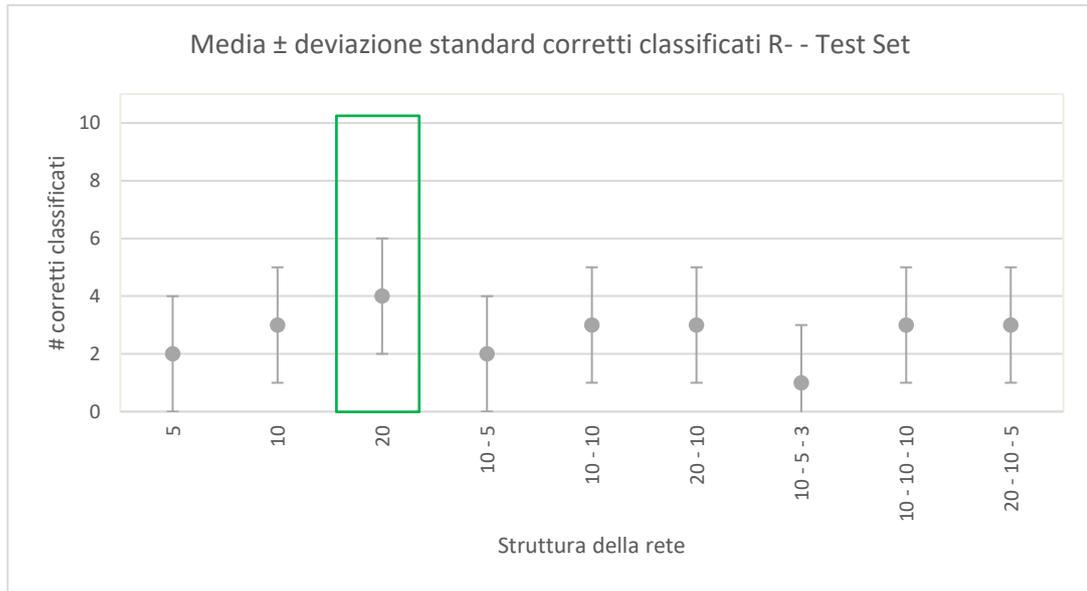


Figura 24: grafici di confronto relativi ai valori medi \pm deviazione standard di corretti classificati ottenuti con le strutture allenate sul Training Set con le features individuate dalla soluzione 14 (dataset con valori discretizzati) e testate su Training Set (Figura A) e Test Set (Figure B, C e D). Sono evidenziate le strutture più performanti.

TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	22	0	1	Classe predetta	R+	23	11	2	Classe predetta	R+	29	25	18
	S	7	25	6		S	30	33	4		S	19	15	9
	R-	1	5	24		R-	33	21	5		R-	11	10	7

Figura 25: confusion matrix con le prestazioni fornite dalla NN con 1 layer nascosto costituito da 20 neuroni allenata su TRS_SOM_1 con valori discretizzati

Le reti tendono a classificare la maggior parte delle lesioni in classe 2 (S) ma, all’aumentare della complessità della struttura della rete, i numeri di corretti classificati sulle 3 classi tendono a bilanciarsi sia su Training Set che Test Set.

Non ci sono differenze significative tra le prestazioni ottenute con i modelli allenati con valori continui e con valori discreti con il subset di features individuato dalla soluzione 14. Con entrambi i dataset, la rete costituita da 1 layer nascosto con 20 neuroni forniva un numero di corretti classificati relativo alle 3 classi più bilanciato. Le prestazioni delle reti risultano buone sul Training Set ma peggiorano su Test Set e Validation Set.

Problema a più classi: clusterizzazione delle lesioni

Vista la presenza del fenomeno di overfitting ottenuta considerando il problema a 3 classi, è stato effettuato un confronto degli errori commessi su Training, Test e Validation Set da alcuni dei modelli migliori allenati con un diverso numero di features precedentemente illustrati.

Nella [tabella 10](#) sono riportate le intersezioni degli errori commessi dai seguenti modelli (le cui confusion matrix sono riportate nei paragrafi [Random Forest](#) e [Feature selection basata su correlazione](#)):

1. RF – 57 features - importance
2. DT – 40 features - correlazione
3. Classificatore bayesiano – 21 features - correlazione
4. SVM lineare – 29 features - correlazione

dove l'intersezione è data da $\frac{n}{m}$, con n pari al numero di errori comuni ai 2 modelli e m pari al numero minore di errori commessi dai 2 modelli.

Il simbolo “/” in una cella indica che uno dei 2 modelli confrontati non ha commesso errori e quindi l'intersezione è vuota.

Seppur il confronto sia stato effettuato tra modelli di tipo diverso e allenati con un numero diverso di features, su tutti i 3 dataset i modelli condividono più della metà degli errori.

Intersezione degli errori dei 4 modelli (modello i – modello j)

	1-2	1-3	1-4	2-3	2-4	3-4
Errori Training Set	/	/	/	8/13	9/13	23/34
Errori Test Set	45/70	50/70	54/68	46/73	48/68	49/68
Errori Validation Set	52/81	60/82	61/81	47/81	60/81	53/81

Tabella 10: tabella relativa alle intersezioni degli errori commessi da 4 modelli di tipo diverso e allenati con un numero di features differente su Training set, Test set e Validation set.

È stata inoltre effettuata una clusterizzazione di tutte le lesioni del Training Set tramite una rete SOM 3x3 allenata sul Training Set da cui è stato osservato che uno stesso neurone risultava essere il «vincitore» per lesioni appartenenti a classi diverse ([Figura 26](#) e [tabella 11](#)).

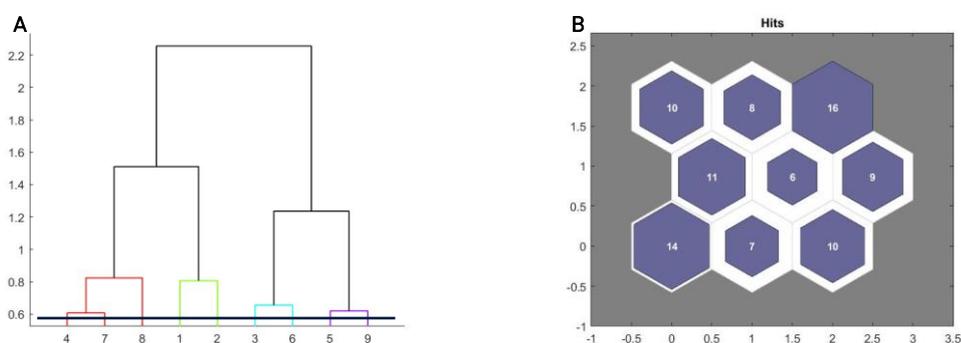


Figura 26: dendrogramma dei pesi di una rete SOM 3x3 allenata con tutte le lesioni del Training Set con il relativo taglio in modo da creare un cluster per ognuno dei 9 neuroni (Figura A) e mappe con l'organizzazione della rete SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) (Figura B).

Neuroni	Numerosità classe R+	Numerosità classe S	Numerosità classe R-
1	6	3	5
2	1	3	3
3	2	2	6
4	1	4	6
5	2	3	1
6	3	3	3
7	1	4	5
8	5	3	0
9	9	5	2

Tabella 11: numero di lesioni per ogni classe per cui ciascun neurone è risultato essere il vincitore

È stato quindi deciso di operare una differente concettualizzazione del problema: le lesioni, separatamente per ogni classe, sono state clusterizzate e sono state valutate le prestazioni di diversi modelli allenati per classificare le lesioni nei diversi cluster. La classificazione finale è stata poi ricondotta a R+, R- ed S a seconda del cluster di appartenenza delle lesioni.

Partendo dal Construction Set e da tutte le 107 features, sono state allenate 3 reti SOM 3x3 con vicinato 1, una per ogni classe, e le lesioni sono state clusterizzate costruendo il dendrogramma dei pesi utilizzando come distanza tra i cluster le distanze tra gli elementi più lontani e come metrica la distanza euclidea.

Nelle figure [27](#) e [28](#) sono riportati rispettivamente i dendrogrammi dei pesi delle reti SOM 3x3 per ogni classe con i relativi tagli effettuati per la suddivisione in cluster e le mappe in cui è presente l'organizzazione delle reti SOM, il numero di vittorie di ciascun neurone e le relative bolle che individuano i cluster.

In [tabella 12](#) sono riportati il numero del cluster (prima colonna), i corrispondenti neuroni appartenenti al cluster stesso (seconda colonna) e la numerosità dei cluster (terza colonna) individuati.

Come già svolto in precedenza, da ciascun cluster individuato per ogni classe sono state estratte, in modo proporzionale alla dimensione dei cluster stessi, circa 30 lesioni per classe (R+, S, R-) ottenendo un nuovo Training Set bilanciato (TRS_SOM_2) e un nuovo Test Set (TS_SOM_2).

Il numero del cluster di appartenenza di ciascuna lesione è stato memorizzato per effettuare la successiva classificazione.

Nelle tabelle [13](#) e [14](#) sono riportati il numero di lesioni appartenenti a ciascuna classe nel Training Set e Test Set e il numero di lesioni presenti nei 2 dataset appartenenti a ciascun cluster.

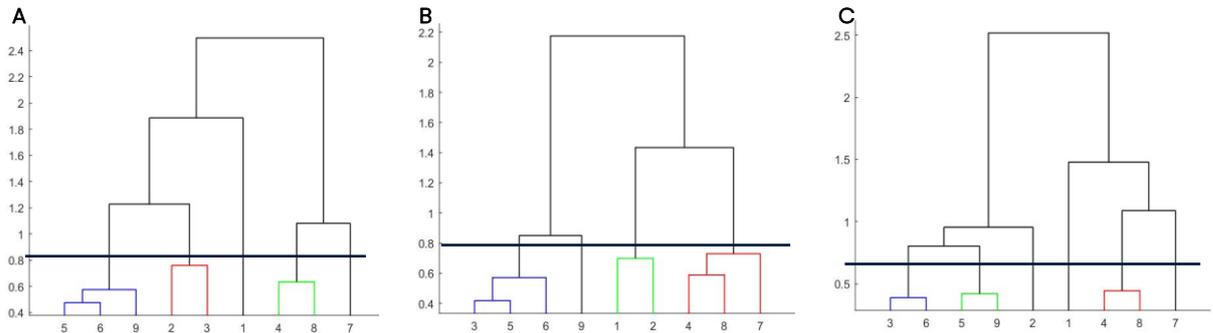


Figura 27: dendrogrammi dei pesi delle reti SOM 3x3 per le classi R+ (Figura A), S (Figura B) ed R- (Figura C) con i relativi tagli effettuati per clusterizzare le lesioni.

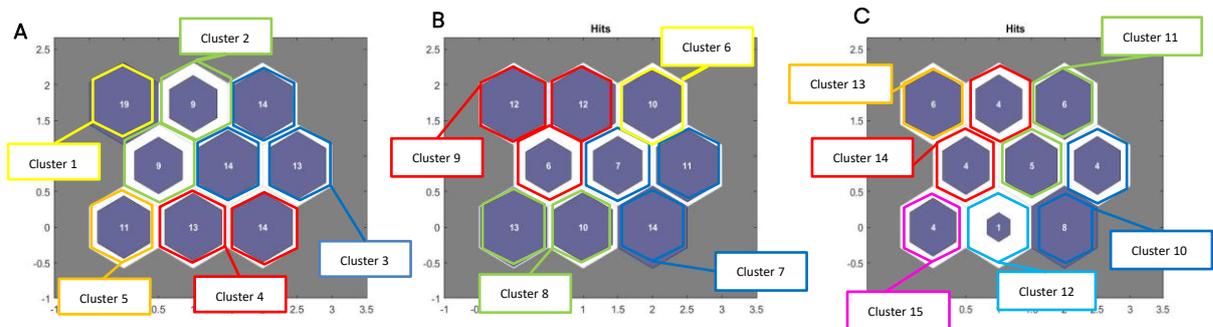


Figura 28: mappe con l'organizzazione delle reti SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) e le relative bolle che individuano i cluster per le classi R+ (Figura A), S (Figura B) ed R- (Figura C).

A			B			C		
Cluster	Neuroni	N° lesioni classe R+	Cluster	Neuroni	N° lesioni classe S	Cluster	Neuroni	N° lesioni classe R-
3	5, 6, 9	41	7	3, 5, 6	32	10	3, 6	12
4	2, 3	27	6	9	10	11	5, 9	11
5	1	11	8	1, 2	23	12	2	1
2	4, 8	18	9	4, 7, 8	30	15	1	4
1	7	19				14	4, 8	8
						13	7	6

Tabella 12: numero del cluster (prima colonna; il numero è relativo all'ordine di creazione dei cluster nel codice matlab), corrispondenti neuroni appartenenti al cluster stesso (seconda colonna) e numero di lesioni appartenenti ad ogni cluster (terza colonna) per le classi R+ (Tabella A), S (Tabella B) ed R- (Tabella C).

N° Cluster	Lesioni nel TRS_SOM_2	Lesioni nel TS_SOM_2
1	5	14
2	5	13
3	11	30
4	7	20
5	3	8
6	3	7
7	10	22
8	7	16
9	10	20
10	9	3
11	8	3
12	1	/
13	4	2
14	6	2
15	3	1

Tabella 14: numero delle lesioni appartenenti a ciascun cluster contenute in Training Set bilanciato e Test Set (in azzurro sono indicati i cluster relativi alle lesioni R+, in rosso quelli relativi alle lesioni S e in grigio i relativi alle lesioni R-)

Classe	N° lesioni TRS_SOM_2	N° lesioni TS_SOM_2
R+	31	85
S	30	65
R-	31	11

Tabella 13: numero delle lesioni contenute in Training Set bilanciato e Test Set per ognuna delle 3 classi

Ciascun classificatore è stato quindi allenato sul TRS_SOM_2 per riconoscere 15 cluster e la classe finale di ogni lesione è stata ricondotta alla classificazione R+, R- ed S a seconda del cluster di appartenenza al fine di valutare le prestazioni su Training Set, Test Set e Validation Set.

Visto l'elevato numero di features a disposizione, è stata effettuata una seconda clusterizzazione delle lesioni seguendo lo stesso procedimento riportato in precedenza con la differenza che le 3 reti SOM 3x3 sono state allenate con il Construction Set costituito dalle sole features significativamente correlate ($p\text{-value} < 0.05$) con la classe (R+, S, R-) in modo da evitare di fornire troppa informazione alle reti SOM durante il loro allenamento.

È stato calcolato quindi il coefficiente di correlazione di Spearman tra ogni feature nel Construction Set e la classe, con lo stesso procedimento riportato nel paragrafo [Feature selection basata su correlazione](#). In [figura 70](#) in Appendice sono riportati in azzurro i valori dei coefficienti di correlazione di Spearman di ciascuna feature con la classe e in [figura 73](#) sono riportati i relativi valori di p-value. Sono presenti 75 features con $p\text{-value} < 0.05$.

Nelle figure [29](#) e [30](#) sono riportati rispettivamente i dendrogrammi dei pesi delle reti SOM 3x3 per ogni classe con i relativi tagli effettuati per la suddivisione in cluster, le mappe con l'organizzazione delle reti SOM e il numero di vittorie di ciascun neurone.

In [tabella 15](#) sono riportati il numero del cluster (prima colonna), i corrispondenti neuroni appartenenti al cluster stesso (seconda colonna) e la numerosità dei cluster (terza colonna) individuati per ogni classe.

Da ciascun cluster individuato per ogni classe sono state estratte, in modo proporzionale alla dimensione dei cluster stessi, circa 30 lesioni per classe (R+, S, R-) ottenendo un nuovo Training Set (TRS_SOM_3) e un nuovo Test (TS_SOM_3).

Il numero del cluster di appartenenza di ciascuna lesione è stato memorizzato per effettuare la successiva classificazione.

Nelle tabelle 16 e 17 sono riportati rispettivamente il numero di lesioni appartenenti a ciascuna classe presenti nel Training Set e Test Set e il numero di lesioni appartenenti a ciascun cluster nei 2 dataset.

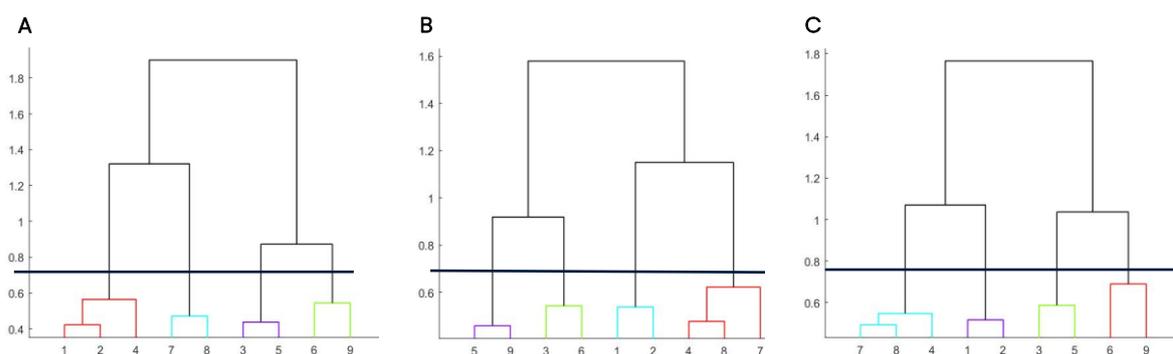


Figura 30: dendrogrammi dei pesi delle reti SOM 3x3 per le classi R+ (Figura A), S (Figura B) e R- (Figura C) con i relativi tagli effettuati per clusterizzare le lesioni.

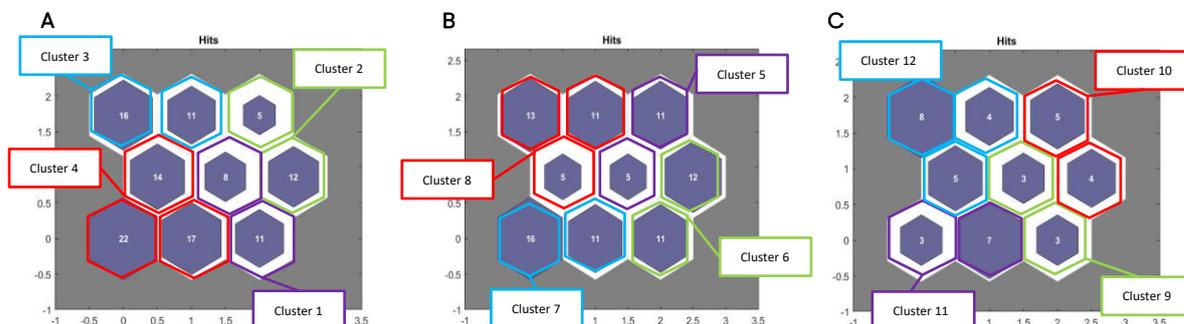


Figura 30: mappe con l'organizzazione delle reti SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) e le relative bolle che individuano i cluster per le classi R+ (Figura A), S (Figura B) e R- (Figura C).

Cluster	Neuroni	N° lesioni classe R+	Cluster	Neuroni	N° lesioni classe S	Cluster	Neuroni	N° lesioni classe R-
4	1, 2, 4	53	5	5, 9	16	12	7, 8, 4	17
3	7, 8	27	6	3, 6	23	11	1, 2	10
1	3, 5	19	7	1, 2	27	9	3, 5	6
2	6, 9	17	8	4, 7, 8	29	10	6, 9	9

Tabella 15: numero del cluster (prima colonna; il numero è relativo all'ordine di creazione dei cluster nel codice matlab), corrispondenti neuroni appartenenti al cluster stesso (seconda colonna) e numero di lesioni appartenenti ad ogni cluster (terza colonna) per le classi R+ (Tabella A), S (Tabella B) e R- (Tabella C).

N° Cluster	Lesioni nel TRS SOM_3	Lesioni nel TS SOM_3
1	5	14
2	4	13
3	7	20
4	14	39
5	5	11
6	7	16
7	9	18
8	9	20
9	4	2
10	7	2
11	7	3
12	12	5

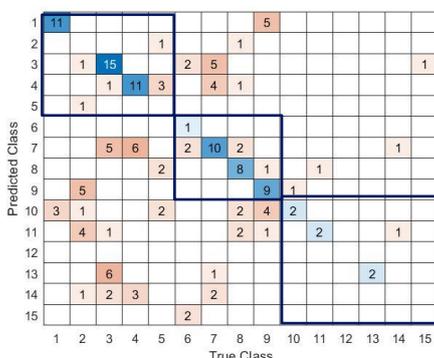
Tabella 17: numero delle lesioni appartenenti a ciascun cluster contenute in Training Set bilanciato e Test Set (in azzurro sono indicati i cluster relativi alle lesioni R+, in rosso quelli relativi alle lesioni S e in grigio i relativi alle lesioni R-)

Classe	N° lesioni TRS SOM_3	N° lesioni TS SOM_3
R+	31	85
S	30	65
R-	31	11

Tabella 16: numero delle lesioni contenute in Training Set bilanciato e Test Set per ognuna delle 3 classi.

Ciascun classificatore è stato quindi allenato sul TRS_SOM_3 per riconoscere 12 cluster e la classe finale di ogni lesione è stata ricondotta alla classificazione a R+, R-, S a seconda del cluster di appartenenza al fine di valutare le prestazioni su Training Set, Test Set e Validation Set.

Con i dataset creati sono state ripetute le stesse prove descritte nei capitoli [Random Forest](#) e [Feature selection basata su correlazione](#) con la differenza che i modelli sono stati allenati per classificare i cluster. Sono state quindi ottenute delle confusion matrix 15x15 nel caso dei modelli allenati sul TRS_SOM_2 e 12x12 nel caso di modelli allenati sul TRS_SOM_3. Queste confusion matrix sono state poi riportate a confusion matrix 3x3 riconducendo la classificazione delle lesioni alle classi R+, S ed R- come riportato nell'esempio in [figura 31](#).



TS	Classe vera			
	R+	S	R-	
Classe predetta	R+	44	18	1
	S	18	33	3
	R-	23	14	7

Figura 31: esempio di confusion matrix 3x3 e relativa ad una confusion matrix 15x15 con la classificazione fornita da un classificatore allenato sul TRS_SOM_2 e con classe target i 15 cluster individuati per ogni lesione. Nella confusion matrix 15x15 sono contornate le predizioni, in termini di cluster, fatte dal classificatore che non portano ad un errore in termini di classe della lesione.

Risultati relativi ai modelli di RF

Nelle figure [74](#), [76](#), [78](#) e [81](#), [83](#) e [85](#) in Appendice sono riportate le importance delle 107 features ottenute con i modelli di RF allenati sul Training Set (TRS_SOM_2 e TRS_SOM_3) con *NumPredictorsToSample* rispettivamente pari a $\sqrt{n^{\circ}features}$; $\frac{n^{\circ}features}{2}$; $\frac{3}{4} \cdot n^{\circ}features$ (con $n^{\circ}features = 107$), 100 alberi e frazione di osservazioni che sono selezionate casualmente per ogni bootstrap replica pari a 1.

Tutti i modelli costruiti avevano performance sul Training Set pari al 100% di corretti classificati per questo non sono stati riportati grafici di confronto delle prestazioni.

In appendice (figure [75](#), [77](#), [79](#) e [82](#), [84](#), [86](#)) sono riportati i grafici relativi al numero di corretti classificati ottenuti sul Test Set (TS_SOM_2 e TS_SOM_3) con i modelli di RF costruiti con i 3 *NumPredictorsToSample* e con diverso numero di features.

Confrontando le prestazioni in termini di corretti classificati sul Test Set, sono stati individuati, per ogni *NumPredictorsToSample* testato, i modelli che fornivano buone e non sbilanciate numerosità di corretti classificati sulle 3 classi (in verde nelle figure [75](#), [77](#), [79](#) e [82](#), [84](#), [86](#) in appendice).

Dal confronto delle prestazioni sul Test Set dei modelli individuati (figure [80](#) e [87](#) in Appendice), queste risultavano confrontabili utilizzando le diverse soglie sull'importanza delle features. Sono state quindi valutate le prestazioni sul Validation Set dei modelli e sono stati individuati i modelli che consentivano di ottenere le migliori prestazioni sul Validation Set su tutte e 3 le classi (in verde nelle figure [80](#) e [87](#) in Appendice).

È stata effettuata un'analisi finale per confrontare i modelli migliori individuati per ogni *NumPredictorsToSample* testato.

Nelle figure [32](#) e [34](#) sono riportati i grafici di confronto tra i numeri di corretti classificati ottenuti sul Test Set e sul Validation Set con i modelli di RF individuati e in verde sono stati indicati i modelli che consentivano di ottenere le migliori performance sui 2 dataset.

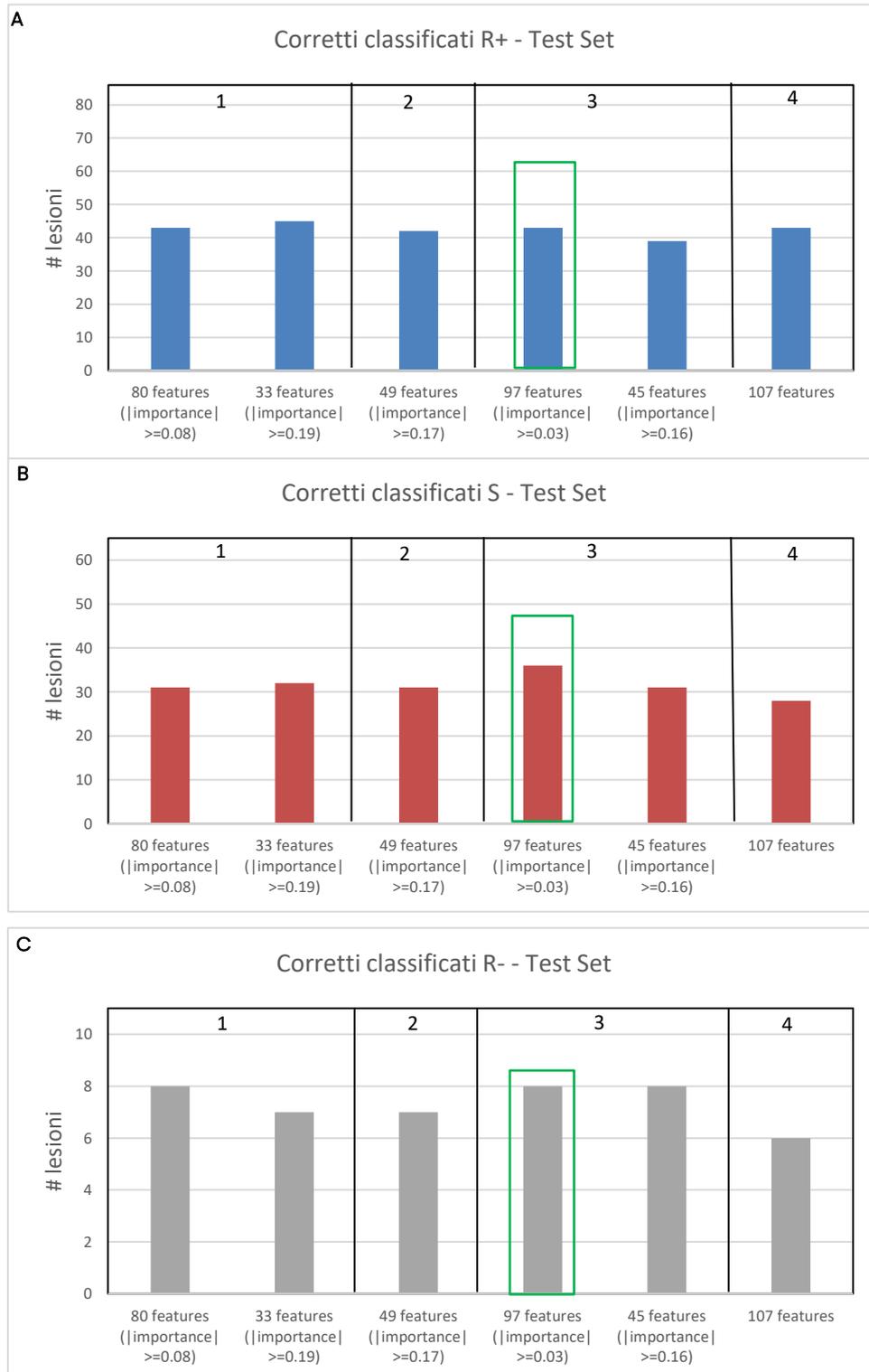
Nelle figure [33](#) e [35](#) sono riportate le confusion matrix relative alle prestazioni su Training Set, Test Set e Validation Set dei modelli individuati.

I modelli allenati sul TRS_SOM_2 individuati presentavano prestazioni non buone sulla classe S nel Validation Set quindi è stato scelto come modello migliore quello che permetteva di ottenere le migliori prestazioni sulla classe S sul Validation Set (modello con 97 features ($|importance| \geq 0.03$) - *NumPredictorsToSample* = 73).

Le prestazioni non risultano complessivamente buone poiché con tutti i modelli allenati sul TRS_SOM_2 si verifica il fenomeno di overfitting. Con il modello migliore individuato (in verde in figura [32](#)) le prestazioni sul Training Set risultano essere pari al 100% di corretti classificati mentre si riducono quelle relative alle classi R+ ed S sul Test Set attestandosi intorno al 50% di corretti classificati e quelle relative alle classi R+ ed S sul Validation Set attestandosi intorno al 30% di corretti classificati. Il modello individuato inoltre tende a classificare la maggior parte delle lesioni in classe R- sul Validation Set.

I modelli allenati sul TRS_SOM_3 individuati presentavano prestazioni confrontabili su tutti i 3 dataset quindi il modello migliore risultava essere quello che permetteva di ottenere tali prestazioni con un numero inferiore di features (modello con 45 features ($|importance| \geq 0.13$) - *NumPredictorsToSample* = 23). Le prestazioni non risultano complessivamente buone poiché con tutti i modelli allenati sul TRS_SOM_3 individuati come migliori (in verde

in figura 34) si verifica il fenomeno di overfitting: le prestazioni sul Training Set risultano essere sempre pari al 100% di corretti classificati mentre si riducono quelle relative alle classi R+ ed S sul Test Set attestandosi in media intorno al 50% di corretti classificati e quelle relative alle 3 classi sul Validation Set attestandosi in media intorno al 51% di corretti classificati per la classe R+, intorno al 42% per la classe S ed intorno al 41% per la classe R-.



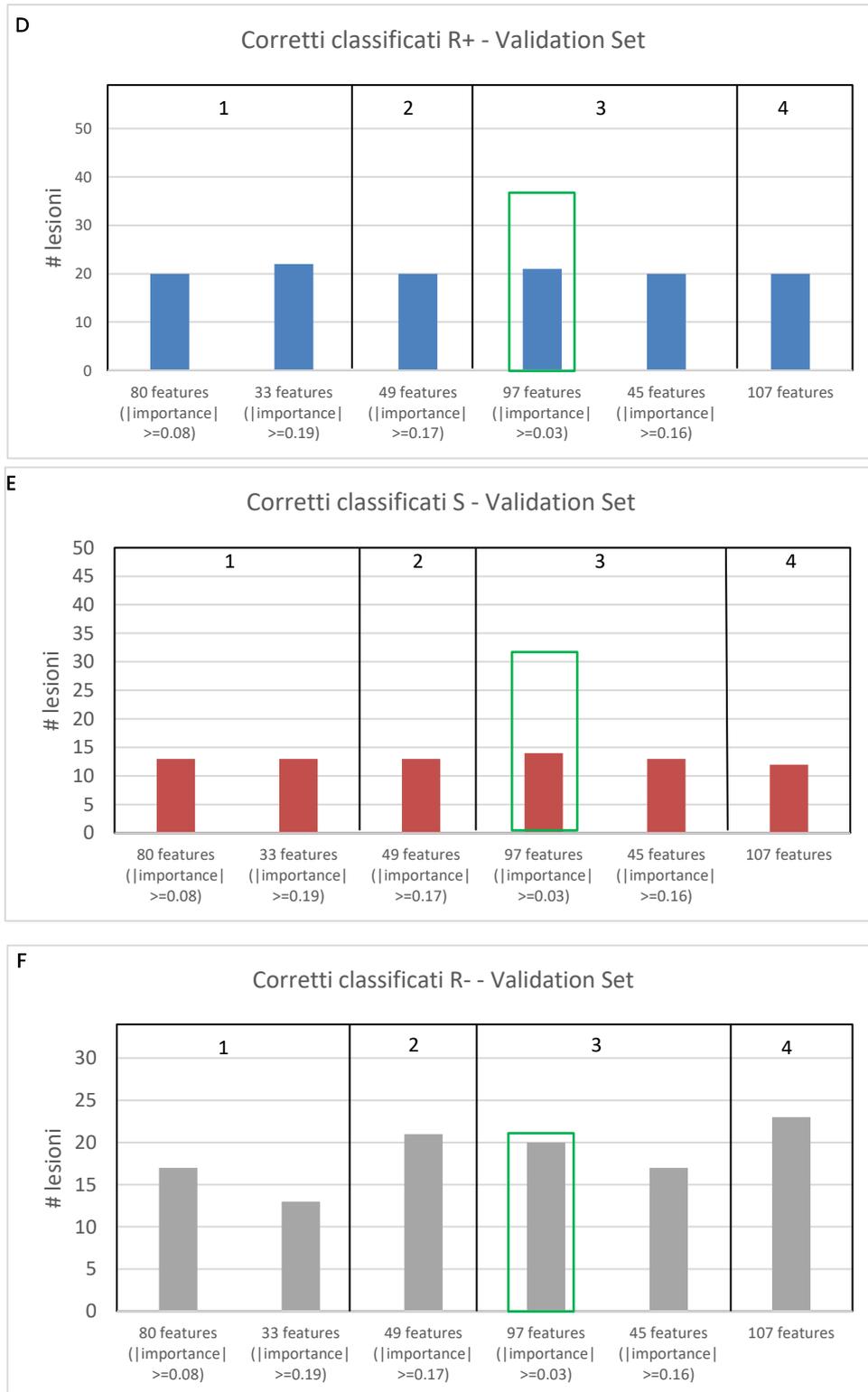
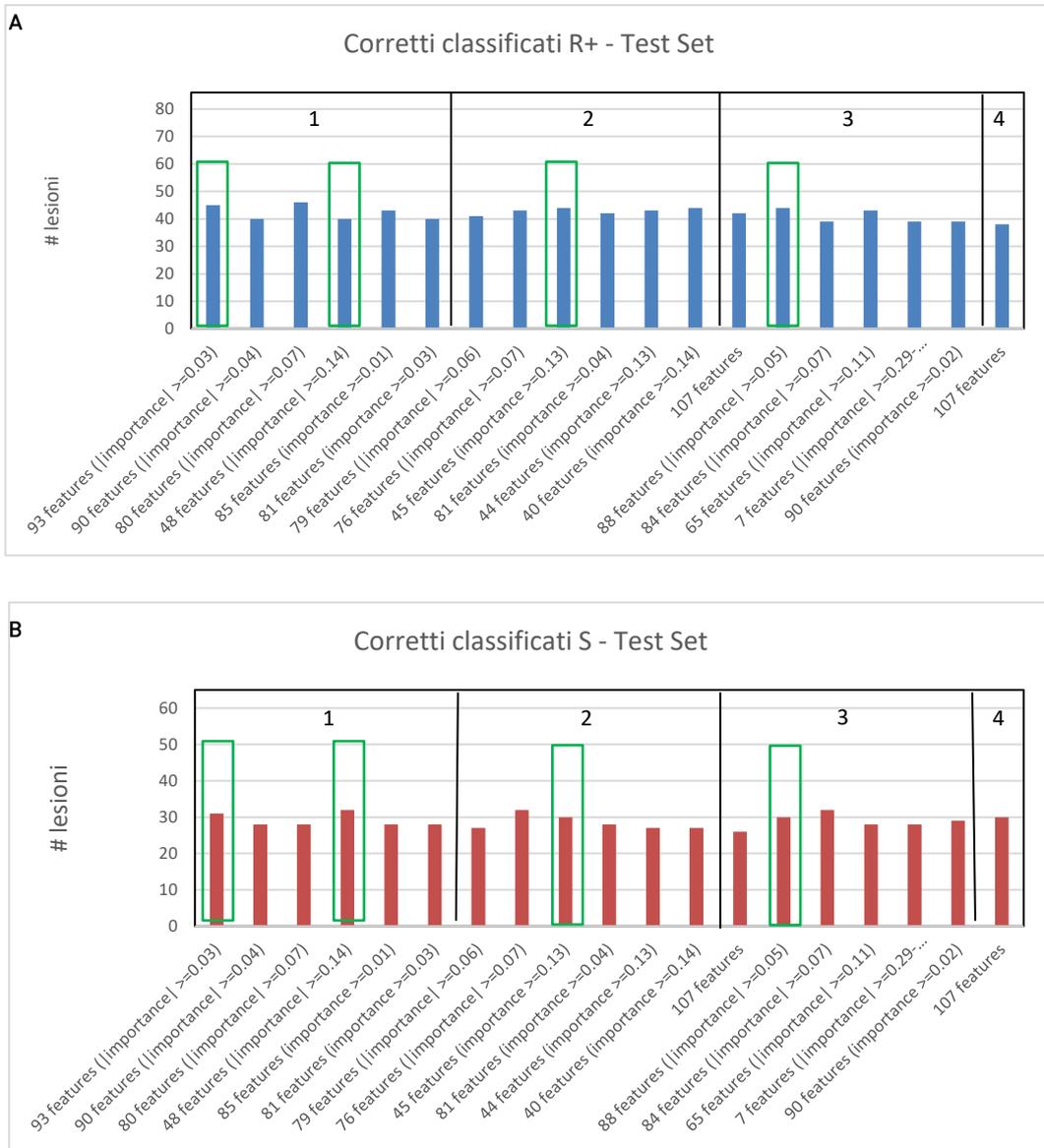
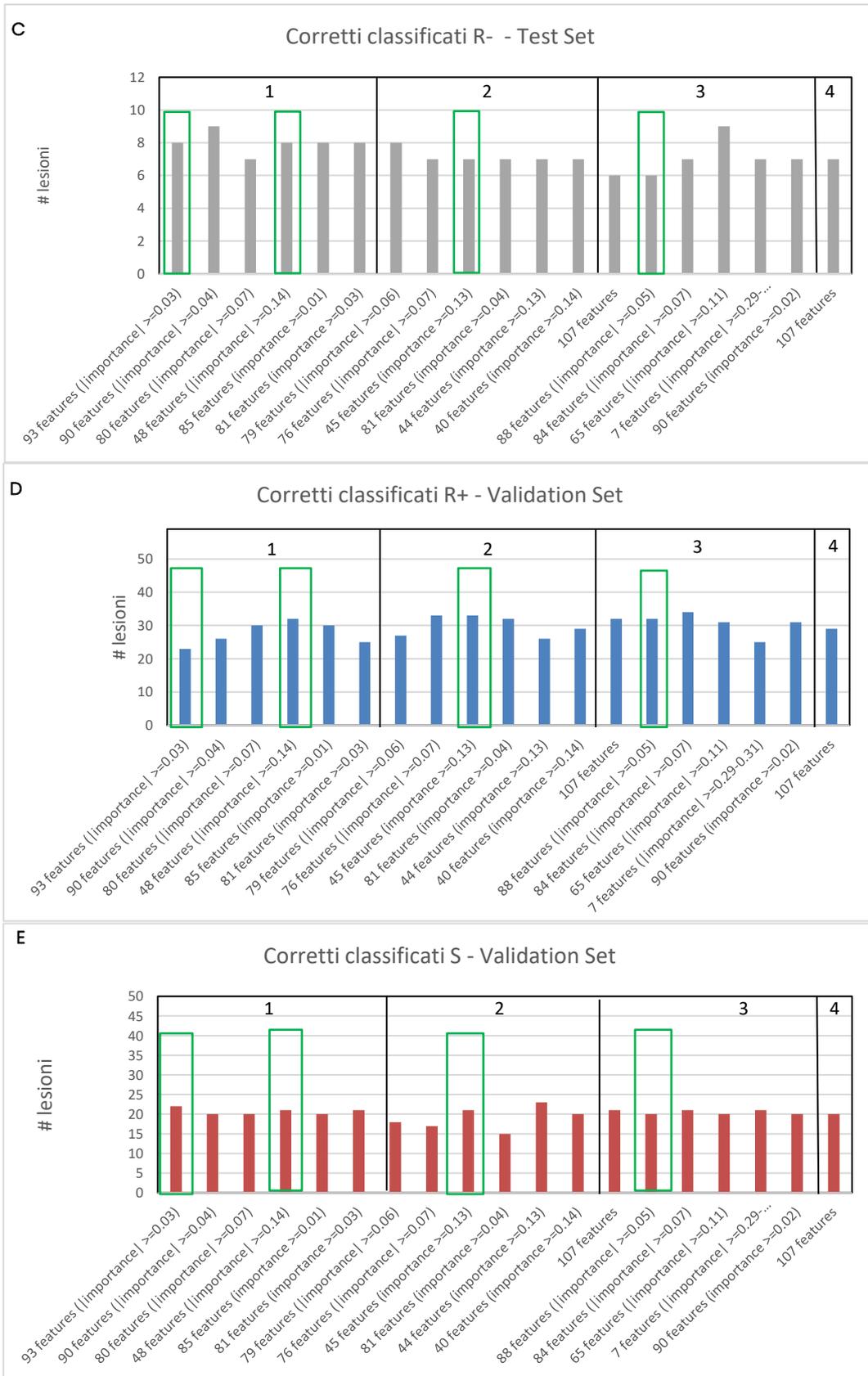


Figura 32: confronto del n° di corretti classificati ottenuti con i modelli (allenati sul TRS_SOM_2) migliori individuati per ogni *NumPredictorsToSample* (1: $\sqrt{n^\circ features}$; 2: $n^\circ features/2$; 3: $3/4 * n^\circ features$; 4: $n^\circ features$) sul Test Set (Figure A-B-C) e sul Validation Set (Figura D-E-F)

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	31	0	0	Classe predetta	R+	43	18	1	Classe predetta	R+	21	14	9
	S	0	30	0		S	25	36	2		S	8	14	5
	R-	0	0	31		R-	17	11	8		R-	30	22	20

Figura 33: confusion matrix con le prestazioni fornite dal RF allenato con 97 Features ($|importance| \geq 0.03$) e $NumPredictorsToSample = 73$





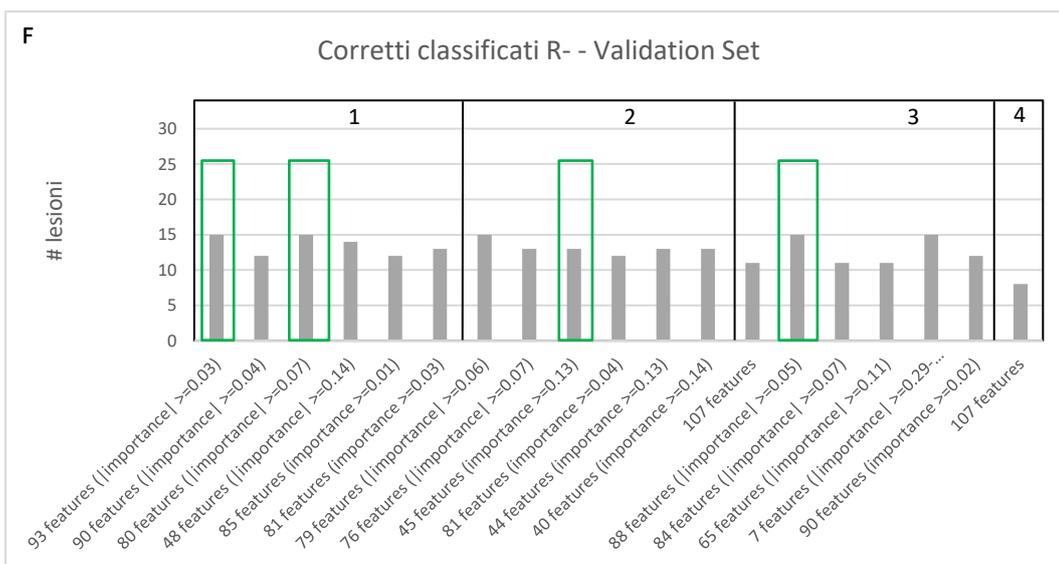


Figura 34: confronto del n° di corretti classificati ottenuti con i modelli (allenati sul TRS_SOM_3) migliori individuati per ogni NumPredictorsToSample (1: $\sqrt{n^\circ \text{ features}}$; 2: $n^\circ \text{ features} / 2$; 3: $3/4 * n^\circ \text{ features}$; 4: $n^\circ \text{ features}$) sul Test Set (Figure A-B-C) e sul Validation Set (Figura D-E-F).

TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	30	0	0	Classe predetta	R+	45	27	2	Classe predetta	R+	23	20	12
	S	0	30	0		S	25	31	2		S	18	22	7
	R-	0	0	30		R-	16	7	8		R-	18	8	15

Confusion matrix con le prestazioni fornite dal RF allenato con 93 Features ($|importance| \geq 0.03$) e $NumPredictorsToSample = 10$

TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	30	0	0	Classe predetta	R+	40	27	3	Classe predetta	R+	32	19	10
	S	0	30	0		S	23	32	1		S	10	21	10
	R-	0	0	30		R-	23	6	8		R-	17	10	14

Confusion matrix con le prestazioni fornite dal RF allenato con 48 Features ($|importance| \geq 0.14$) e $NumPredictorsToSample = 7$

TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	30	0	0	Classe predetta	R+	44	27	2	Classe predetta	R+	33	18	12
	S	0	30	0		S	26	30	3		S	9	21	9
	R-	0	0	30		R-	16	8	7		R-	17	11	13

Confusion matrix con le prestazioni fornite dal RF allenato con 45 Features ($|importance| \geq 0.13$) e $NumPredictorsToSample = 23$

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	30	0	0	Classe predetta	R+	44	27	3	Classe predetta	R+	32	20	10
	S	0	30	0		S	25	30	3		S	12	20	9
	R-	0	0	30		R-	17	8	6		R-	15	10	15

Confusion matrix con le prestazioni fornite dal RF allenato con 88 Features ($|importance| \geq 0.05$) e $NumPredictorsToSample = 66$

Figura 35: confusion matrix relative alle prestazioni su Training Set, Test Set e Validation Set dei modelli migliori individuati allenati su TRS_SOM_3.

Risultati relativi ai modelli allenati effettuando feature selection basata sulla correlazione

Per il classificatore SVM, prima di effettuare feature selection basata sulla correlazione, sono state valutate le performance di differenti strutture con tutte le 107 features e si è proceduto con la feature selection solo con le strutture che fornivano le performance migliori. Nelle figure [88](#) e [91](#) in appendice sono riportate le confusion matrix con le prestazioni su Training Set (TRS_SOM_2 e TRS_SOM_3) e Test Set (TS_SOM_2 e TS_SOM_3) ottenute allenando diverse strutture di SVM con diversi kernel (lineare, polinomiale e gaussiano) e valori di C (1, 10, 50 e 100) sul Training Set (TRS_SOM_2 e TRS_SOM_3).

È stata effettuata feature selection con la correlazione utilizzando le strutture di SVM lineare e gaussiano rispettivamente con $C = 1$ e $C = 50$ nel caso dei modelli allenati su TRS_SOM_2 e strutture di SVM lineare e gaussiano rispettivamente con $C = 1$ e $C = 100$ nel caso dei modelli allenati su TRS_SOM_3.

Nelle figure [89](#), [90](#) e [92](#), [93](#) in Appendice sono riportati i coefficienti di correlazione delle 107 features in ordine decrescente e i valori di p -value della correlazione delle 107 features in ordine crescente individuati rispettivamente con il TRS_SOM_2 e TRS_SOM_3. Si nota come i valori di correlazione delle features con la classe siano bassi. Sono presenti 53 features con p -value < 0.05 nel TRS_SOM_2 e 62 features nel TRS_SOM_3.

Nelle figure [36](#) e [42](#) sono riportati gli andamenti delle prestazioni dei diversi modelli allenati aggiungendo una feature alla volta in ordine decrescente di correlazione con la classe. Sono inoltre indicate le prestazioni in corrispondenza dell'aggiunta della variabile che genera un picco nelle prestazioni Test Set (in verde) e sono evidenziate le prestazioni in corrispondenza dell'aggiunta dell'ultima feature significativamente correlata con la classe (in giallo). Nei casi in cui non era evidente un aumento seguito da un peggioramento o da un andamento costante delle prestazioni sul Test Set, sono stati individuati il picco in corrispondenza del maggiore aumento delle prestazioni e un picco in corrispondenza di un successivo incremento.

Nelle figure [37](#) e [43](#) sono riportati gli andamenti delle prestazioni dei diversi modelli allenati scartando le features meno correlate con la classe appartenenti alle coppie di features correlate fra loro più di una certa soglia. Sono inoltre indicate le soglie in corrispondenza delle quali si verifica un picco nelle prestazioni sul Test Set (in verde).

Sono state valutate le prestazioni dei modelli individuati anche sul Validation Set.

Per ogni tipologia di classificatore testato sono stati costruiti i grafici, riportati nelle figure [38](#), [39](#), [40](#), [41](#), [44](#), [45](#), [46](#) e [47](#), al fine di confrontare le prestazioni dei modelli allenati con un diverso numero di features e individuare i modelli che fornivano le migliori prestazioni sui 3 dataset (in verde nelle figure [38](#), [39](#), [40](#), [41](#), [44](#), [45](#), [46](#) e [47](#)).

Nel caso dei modelli di RF sono stati riportati i grafici di confronto delle prestazioni solo per il Test e il Validation Set poiché le prestazioni sul Training risultavano pari al 100% di corretti classificati.

Ad eccezione dei modelli di SVM lineare, in cui si osservano dei sottoinsiemi di features per cui le prestazioni su Training Set e Test Set sono confrontabili, con i diversi modelli allenati sul TRS_SOM_2, si osserva un calo delle prestazioni sul Test Set rispetto a quelle ottenute sul Training Set.

Per i modelli di RF allenati sul TRS_SOM_2 sono stati individuati come migliori (figura [38](#) in verde) i modelli che fornivano un maggiore numero di corretti classificati sulle 3 classi sul Validation Set. Le prestazioni medie ottenute con i due modelli relativamente alle classi R+ ed S sul Test Set sono rispettivamente pari al 50% e 48% di corretti classificati e le prestazioni medie sul Validation Set relative alle classi R+, S ed R- Set sono rispettivamente pari a 53%, 35% e 41% di corretti classificati.

Per i modelli di DT allenati sul TRS_SOM_2 sono stati individuati come migliori (figura [39](#) in verde) i modelli che fornivano un maggiore numero di corretti classificati sulle 3 classi su Test Set e Validation Set. I due modelli individuati forniscono prestazioni relative alle classi R+ ed S sul Test Set pari in media al 49% e 40% di corretti classificati e prestazioni medie sul Validation Set relative alle classi R+, S ed R- pari a 43%, 30% e 50% di corretti classificati.

Relativamente ai modelli di classificatore bayesiano allenati sul TRS_SOM_2 è stato individuato come più performante (figura [40](#) in verde) il modello che forniva un maggiore numero di corretti classificati relativo alle 3 classi sui 3 dataset. Il modello individuato fornisce prestazioni relative alle classi R+ ed S sul Test Set pari al 42% e 54% di corretti classificati e prestazioni sul Validation Set relative alle classi R+, S ed R- pari a 36%, 54% e 29% di corretti classificati.

Relativamente ai modelli di SVM allenati sul TRS_SOM_2, infine, è stato individuato come migliore (figura [41](#) in verde) il modello che forniva il più alto numero di corretti classificati relativo alla classe R- del Validation Set. Le prestazioni relative alle 3 classi tendono a peggiorare su Test Set e Validation Set rispetto a quelle ottenute sul Training Set ma in maniera minore rispetto ai modelli precedenti.

Anche con queste prove, ad eccezione dei modelli di SVM, i modelli più performanti sono stati individuati tra quelli ottenuti sfruttando il secondo approccio di feature selection basato sulla correlazione.

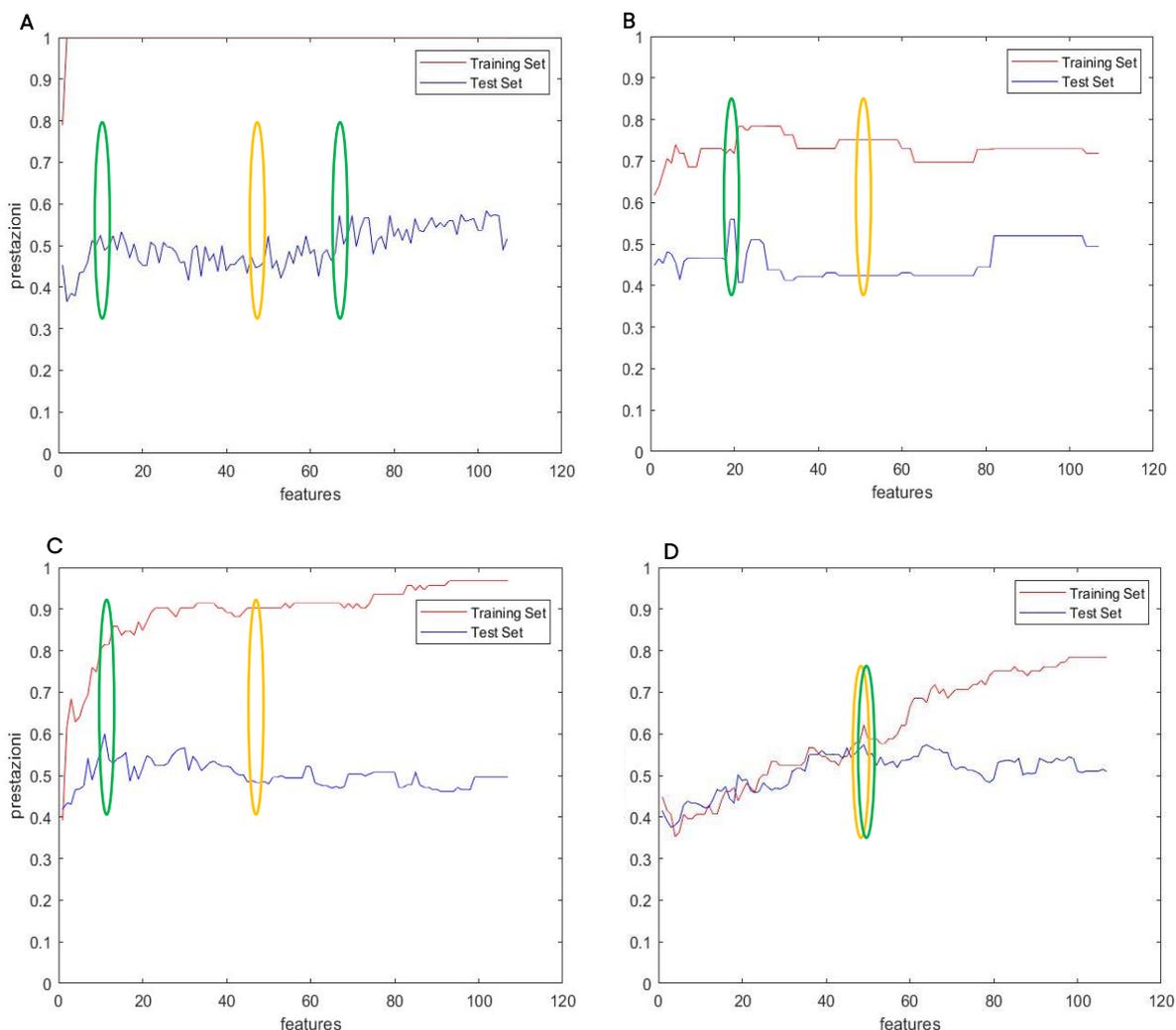
Con i diversi modelli allenati sul TRS_SOM_3 con un diverso numero di features si osserva un calo delle prestazioni su Test Set rispetto a quelle ottenute sul Training Set.

Per i modelli di RF allenati sul TRS_SOM_3 sono stati individuati come migliori (figura [44](#) in verde) i modelli che fornivano un maggiore numero di corretti classificati sulle 3 classi su Test Set e Validation Set. Le prestazioni medie ottenute con i due modelli relativamente alle classi R+ ed S sul Test Set sono rispettivamente pari al 55% e 47% di corretti classificati e le prestazioni medie sul Validation Set relative alle classi R+, S ed R- sono rispettivamente pari a 47%, 37% e 38% di corretti classificati.

Per i modelli di DT allenati sul TRS_SOM_3 è stato individuato come migliore (figura 45 in verde) il modello che forniva un maggiore numero di corretti classificati relativo alle classi S ed R- sul Validation Set. Il modello individuato fornisce prestazioni relative alle classi R+ ed S sul Test Set pari al 38% e 49% di corretti classificati e prestazioni medie sul Validation Set relative alle classi R+, S ed R- Set pari a 52%, 42% e 32% di corretti classificati.

Relativamente ai modelli di classificatore bayesiano allenati sul TRS_SOM_3 è stato individuato come più performante (figura 46 in verde) il modello che forniva un maggiore numero di corretti classificati relativo alla classe S sul Validation Set. Le prestazioni fornite dal modello sul Test relative alle classi R+ ed S sono rispettivamente pari a 56% e 37% di corretti classificati; quelle relative alle classi R+, S ed R- sul Validation Set sono rispettivamente pari a 42%, 26% e 47% di corretti classificati.

Relativamente ai modelli di SVM allenati sul TRS_SOM_3, infine, sono stati individuati come migliori (figura 47 in verde) i modelli che fornivano il più alto numero di corretti classificati relativo alle classi S ed R- sul Validation Set. Le prestazioni medie sul Test relative alle classi R+ ed S sono rispettivamente pari a 59% e 37% di corretti classificati; quelle relative alle classi R+, S ed R- sul Validation Set sono rispettivamente pari a 52%, 38% e 32% di corretti classificati.



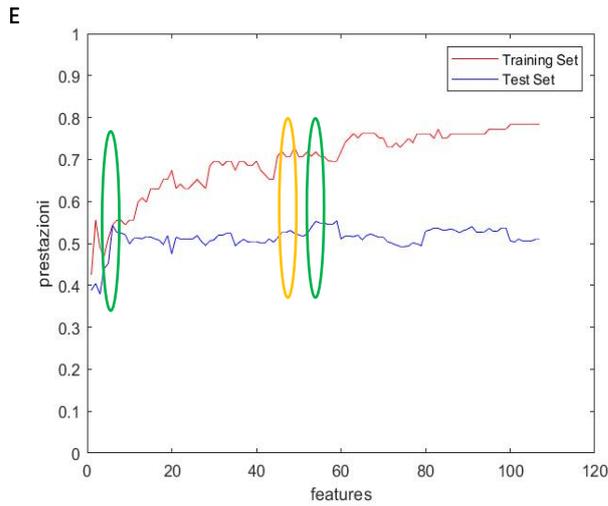
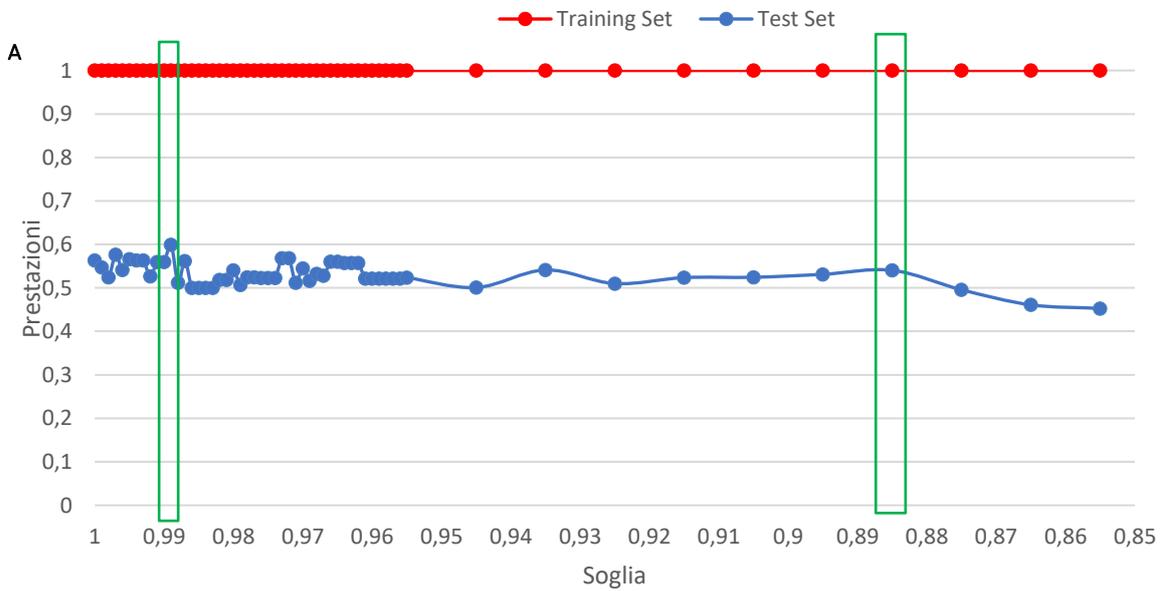
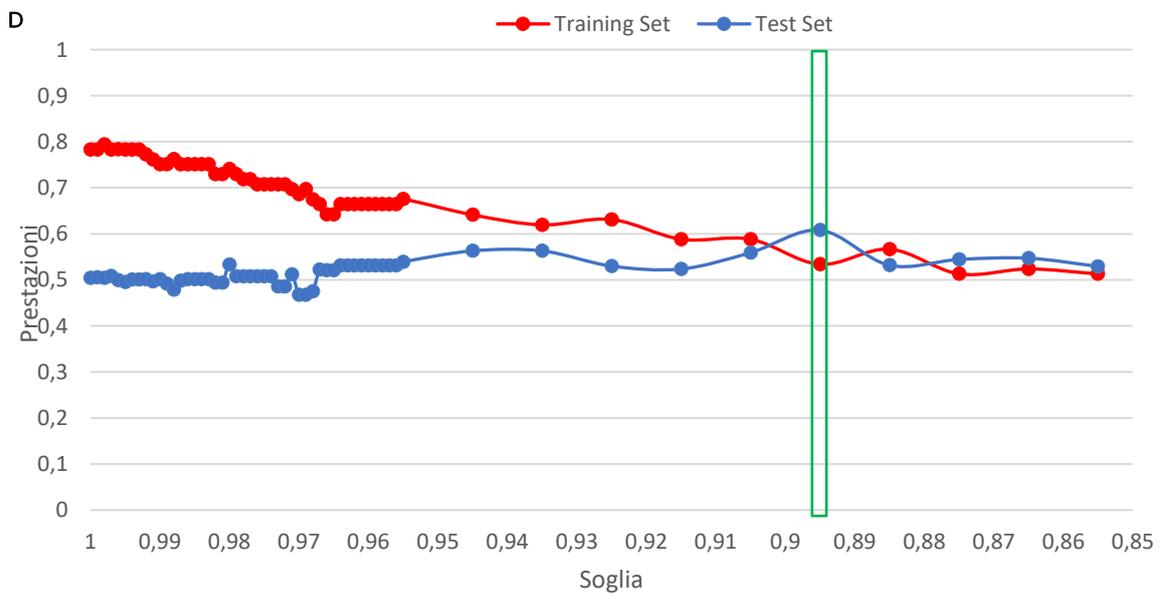
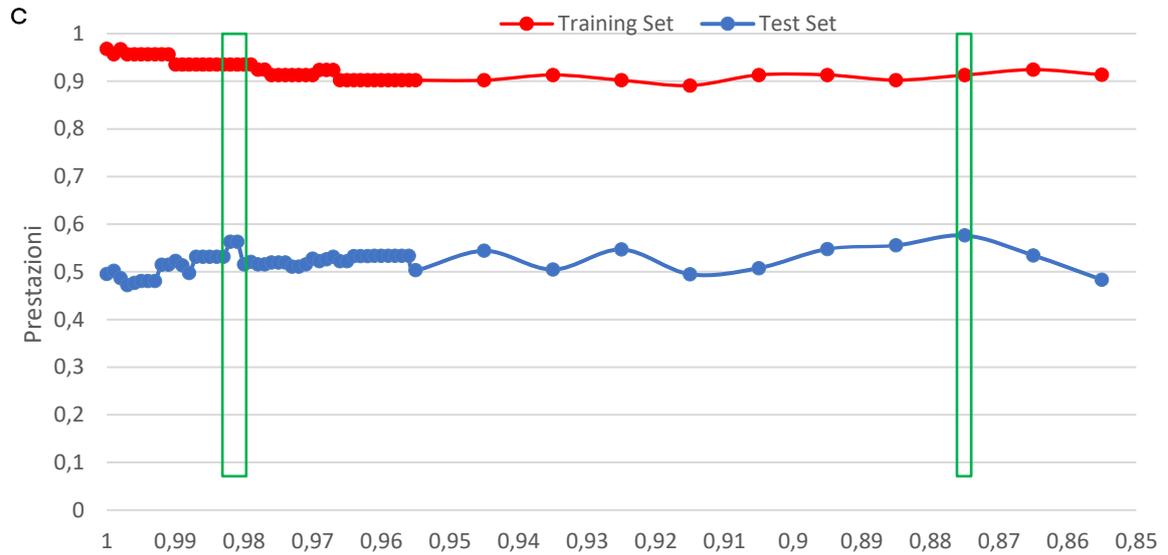
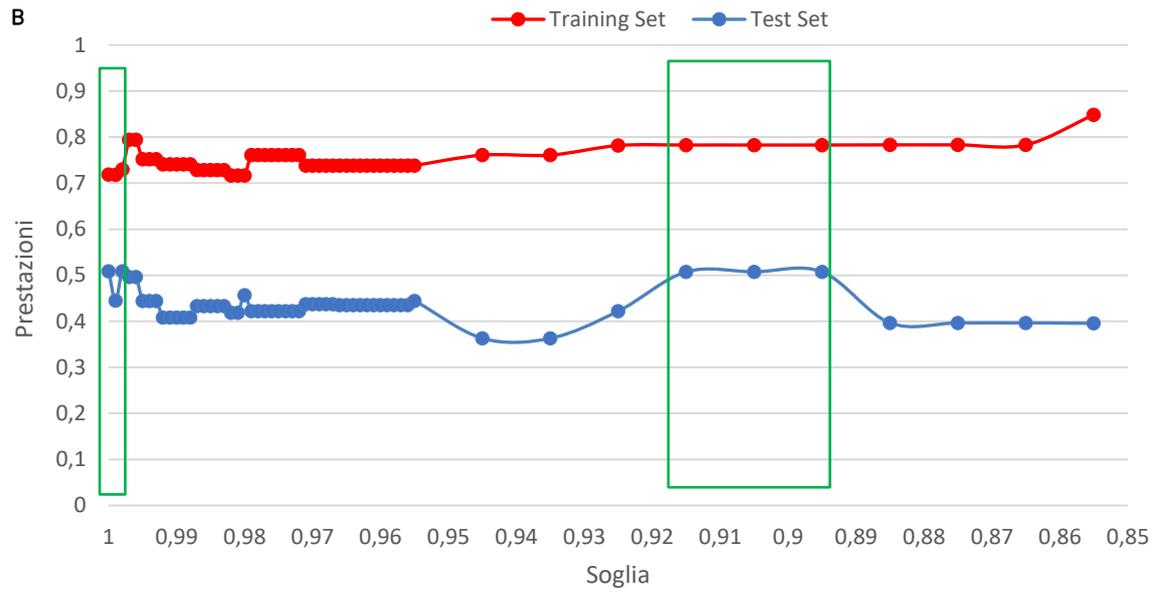


Figura 36: curve relative all'andamento delle prestazioni dei modelli di RF (Figura A), DT (Figura B), bayesiano (Figura C) ed SVM lineare (Figura D) e gaussiano (Figura E) allenati su TRS_SOM_2 con un numero di features crescente (primo approccio). In verde sono indicati i subset di features in corrispondenza dei quali si verifica un picco nelle prestazioni sul Test Set e in giallo le prestazioni in corrispondenza del subset di features con $p\text{-value} \leq 0.05$





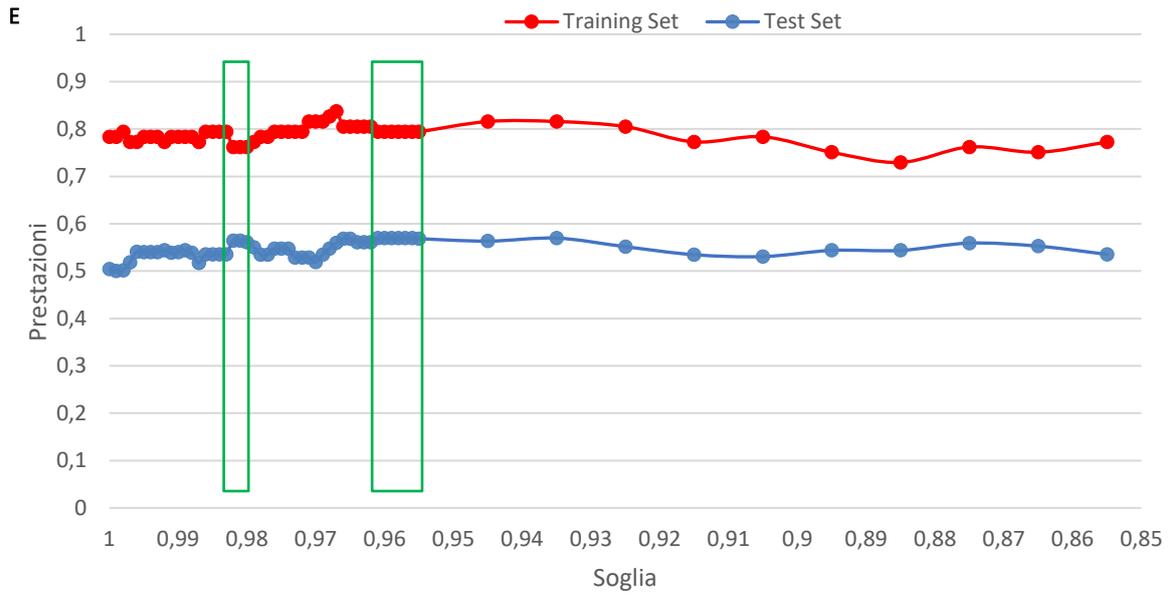
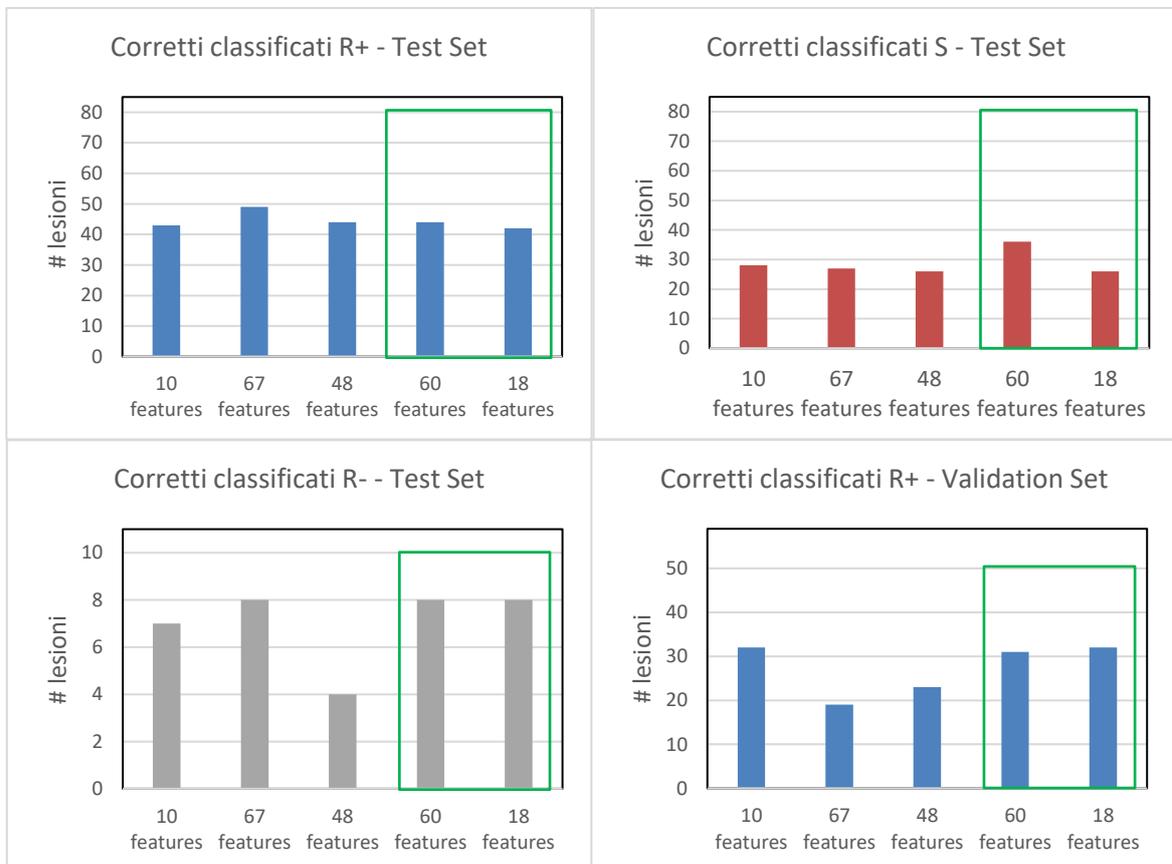
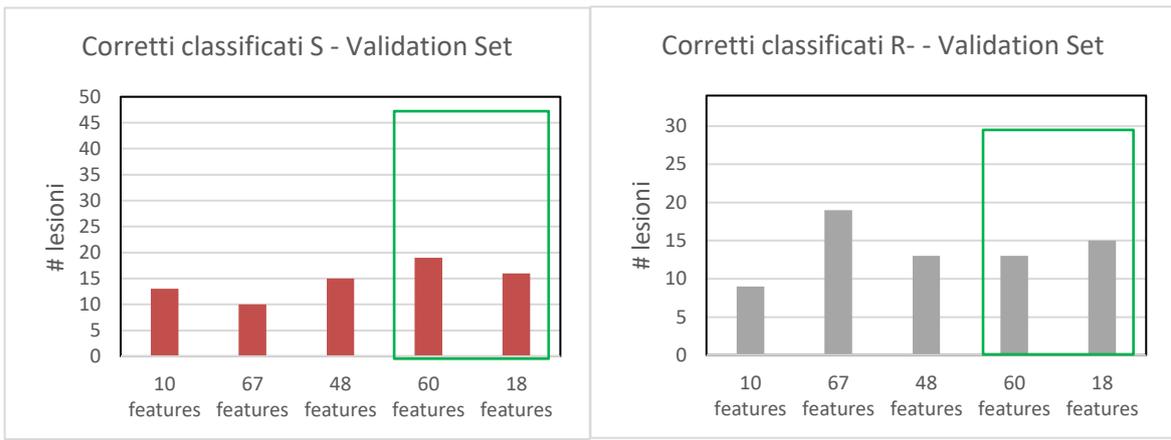


Figura 37: curve relative all’andamento delle prestazioni dei modelli di RF (Figura A), DT (Figura B), bayesiano (Figura C) ed SVM lineare (Figura D) e gaussiano (Figura E) allenati su TRS_SOM_2 scartando le features meno correlate con la classe appartenenti alle coppie di features correlate fra loro più di una certa soglia (secondo approccio). In verde sono indicati i subset di features in corrispondenza dei quali si verifica un picco nelle prestazioni sul Test.





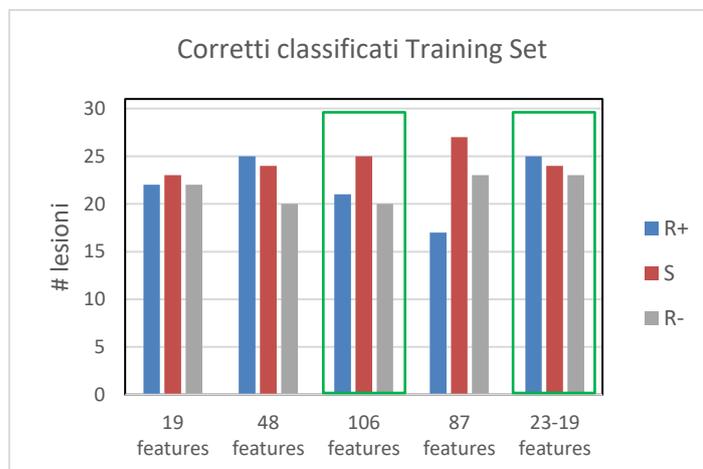
TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	31	0	0	Classe predetta	R+	44	16	1	Classe predetta	R+	31	16	14
	S	0	30	0		S	25	36	2		S	12	19	7
	R-	0	0	31		R-	16	13	8		R-	16	15	13

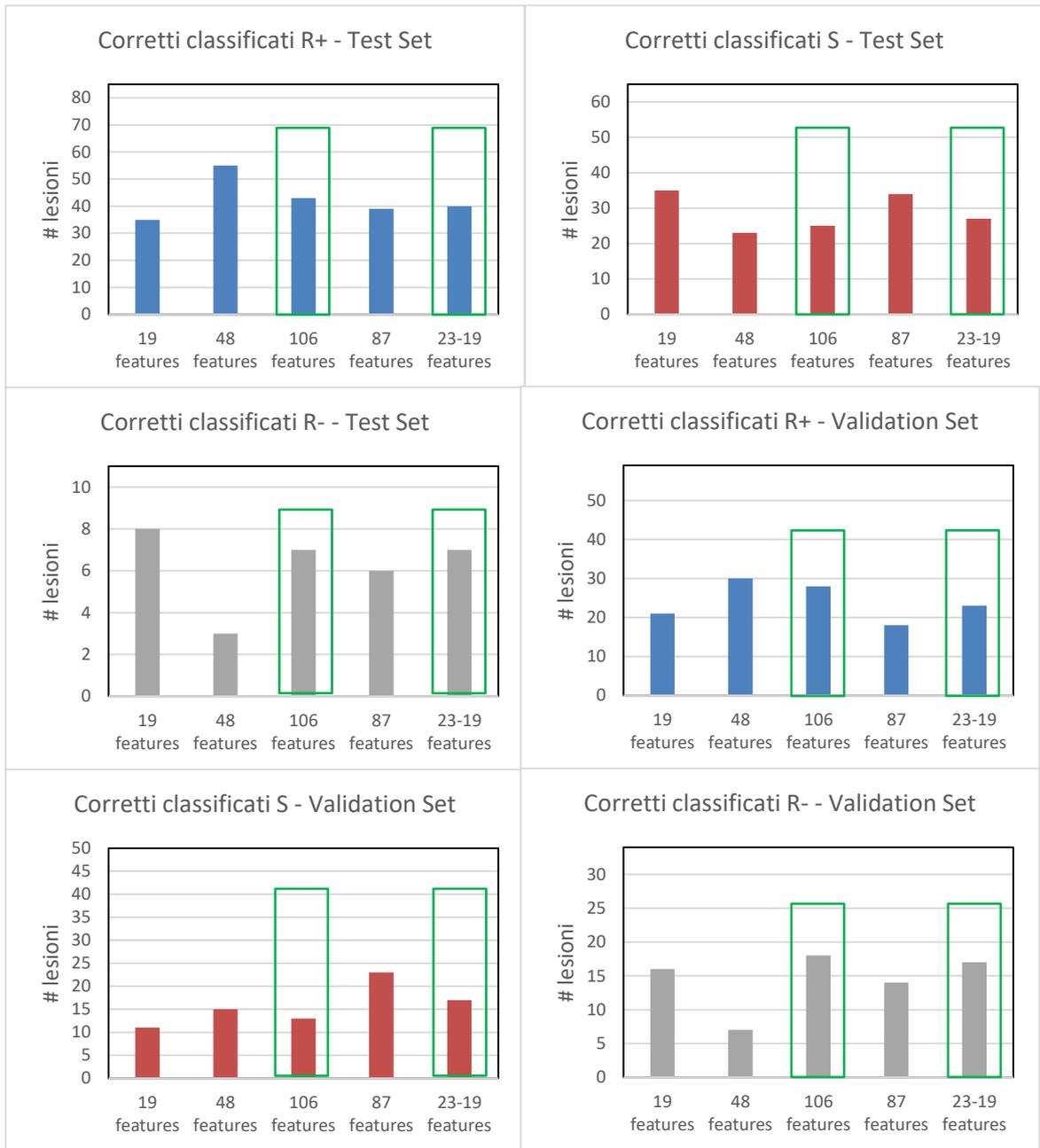
Confusion matrix con le prestazioni fornite dal RF allenato con 60 Features (*correlazione* < 0.989)

TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	31	0	0	Classe predetta	R+	42	23	1	Classe predetta	R+	32	26	14
	S	0	30	0		S	26	26	2		S	12	16	5
	R-	0	0	31		R-	17	16	8		R-	15	8	15

Confusion matrix con le prestazioni fornite dal RF allenato con 18 Features (*correlazione* < 0.885)

Figura 38: confronto del n° di corretti classificati ottenuti con i modelli di RF migliori individuati (approccio 1 e 2) allenati su TRS_SOM_2. In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.





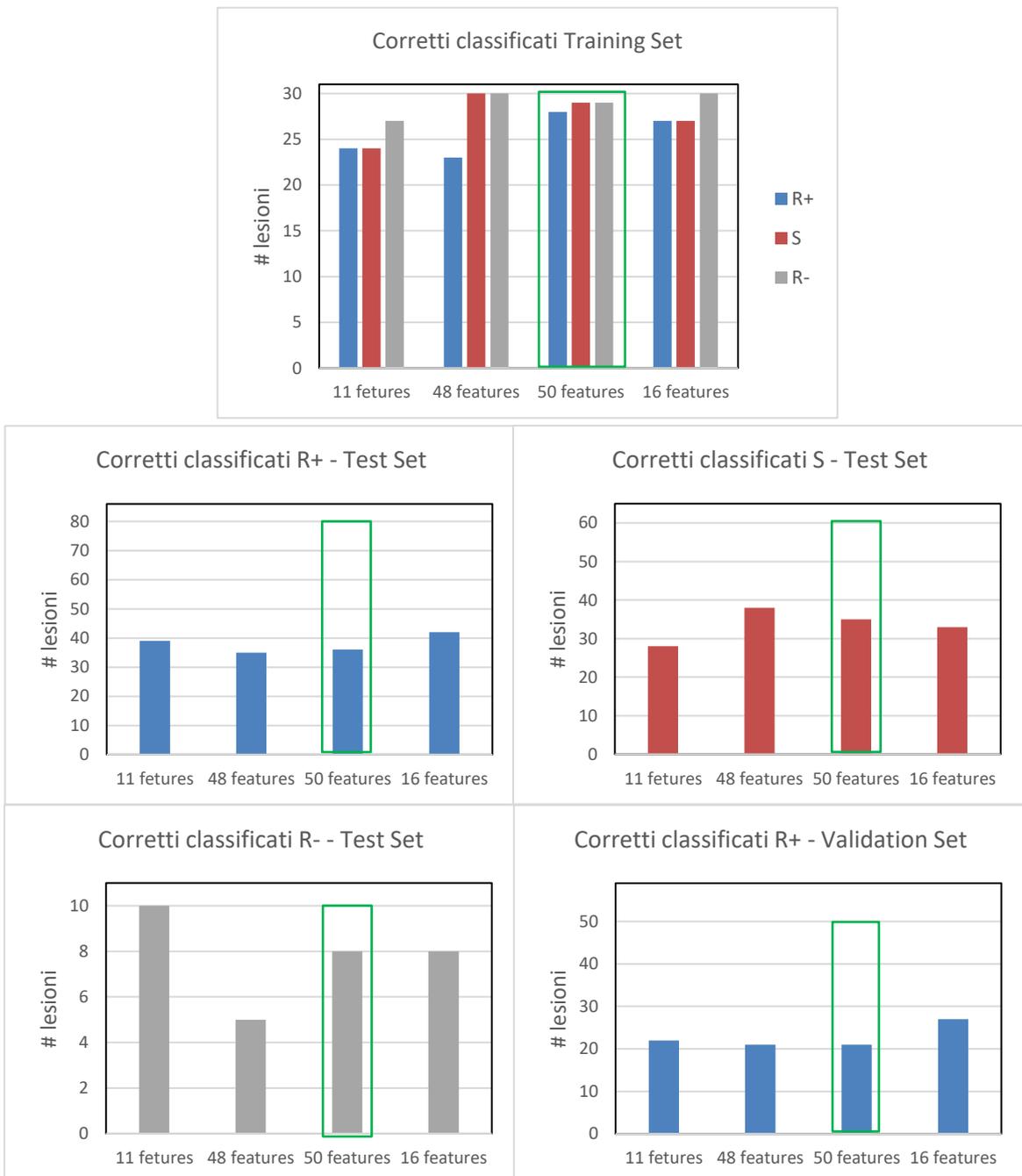
TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	21	4	7	Classe predetta	R+	43	23	1	Classe predetta	R+	28	16	8
	S	2	25	4		S	23	25	3		S	10	13	8
	R-	8	1	20		R-	19	17	7		R-	21	21	18

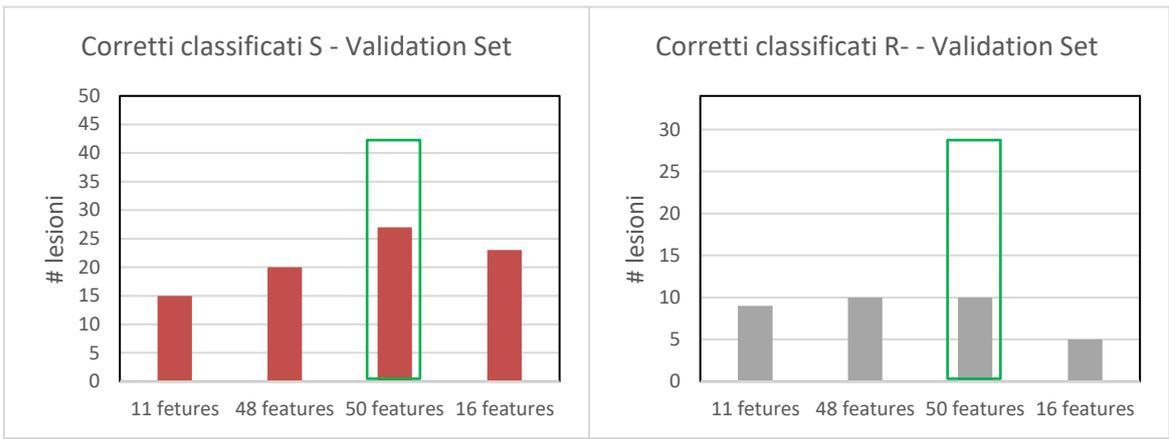
Confusion matrix con le prestazioni fornite dal DT allenato con 106 Features (*correlazione* < 1)

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	25	4	5	Classe predetta	R+	40	21	2	Classe predetta	R+	23	24	9
	S	6	24	3		S	30	27	2		S	17	17	8
	R-	0	2	23		R-	15	17	7		R-	19	9	17

Confusion matrix con le prestazioni fornite dal DT allenato con 23-19 Features (*correlazione* < 0.915-0.895)

Figura 39: confronto del n° di corretti classificati ottenuti con i modelli di DT migliori individuati (approccio 1 e 2) allenati su TRS_SOM_2. In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.

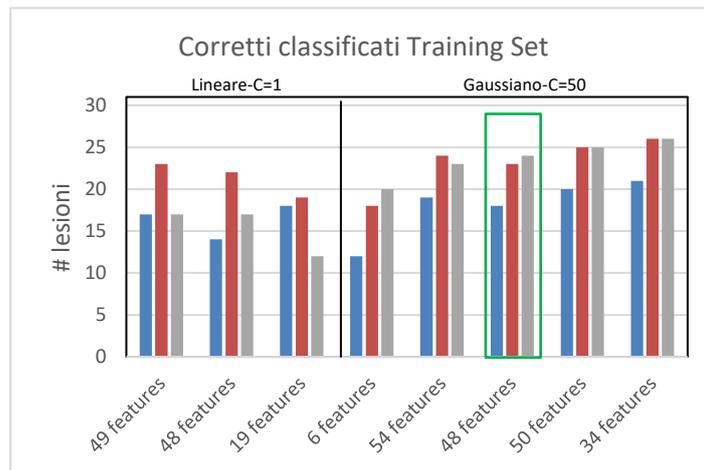


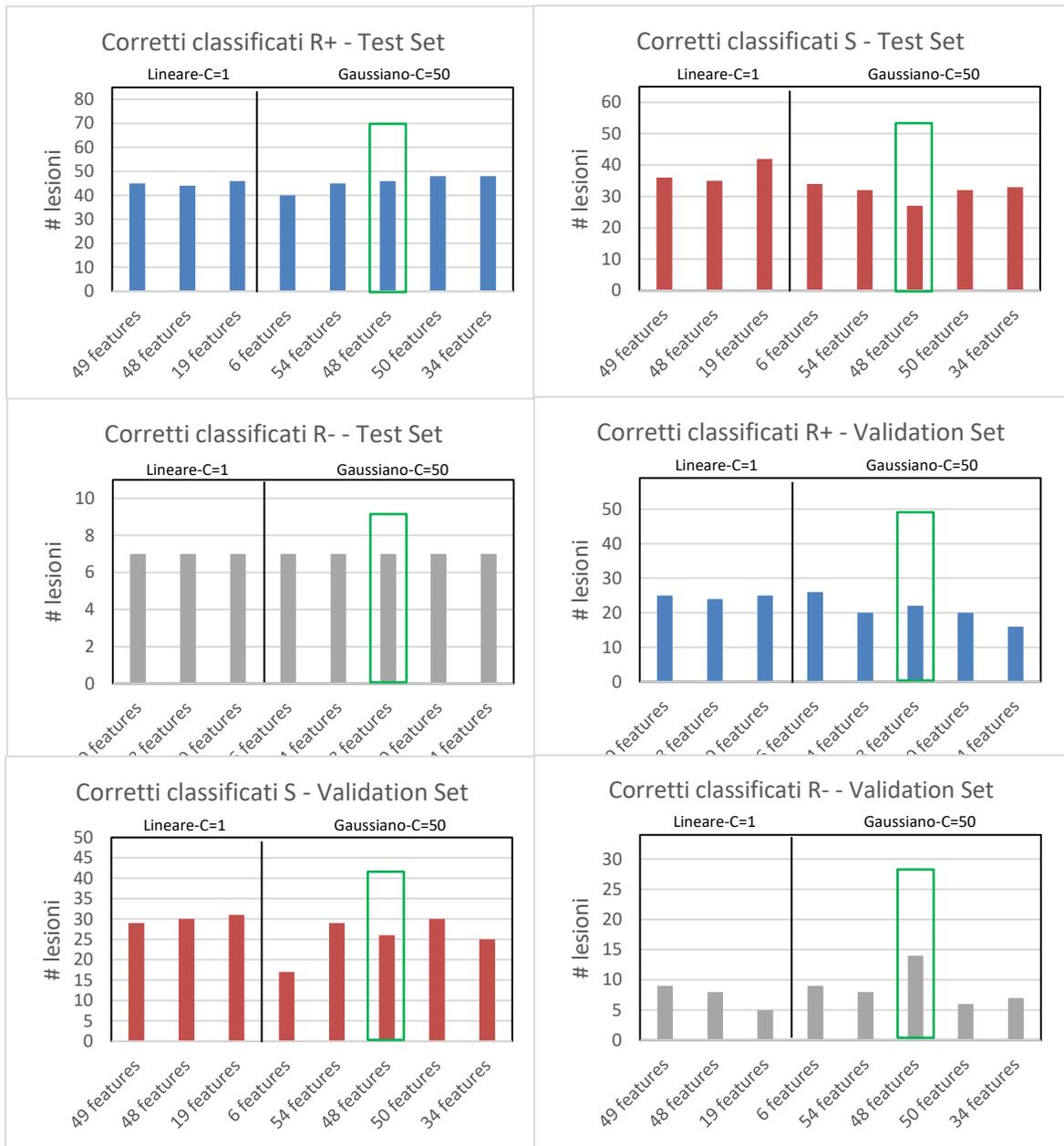


TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	28	1	2	Classe predetta	R+	36	17	2	Classe predetta	R+	21	14	13
	S	3	29	0		S	34	35	1		S	27	27	11
	R-	0	0	29		R-	15	13	8		R-	11	9	10

Confusion matrix con le prestazioni fornite dal bayesiano allenato con 50 Features (*correlazione* < 0.982-0.981)

Figura 40: confronto del n° di corretti classificati ottenuti con i modelli di bayesiano migliori individuati (approccio 1 e 2) allenati su TRS_SOM_2. In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.





TRS				TS				VS			
Classe predetta	Classe vera			Classe predetta	Classe vera			Classe predetta	Classe vera		
	R+	S	R-		R+	S	R-		R+	S	R-
	R+	18	3		4	R+	46		23	2	R+
S	10	23	3	S	22	27	2	S	26	26	13
R-	3	4	24	R-	17	15	7	R-	11	8	14

Confusion matrix con le prestazioni fornite dal SVM gaussiano allenato con 48 Features ($p\text{-value} < 0.05$)

Figura 41: confronto del n° di corretti classificati ottenuti con i modelli di SVM migliori individuati (approccio 1 e 2) allenati su STR_SOM_2. In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.

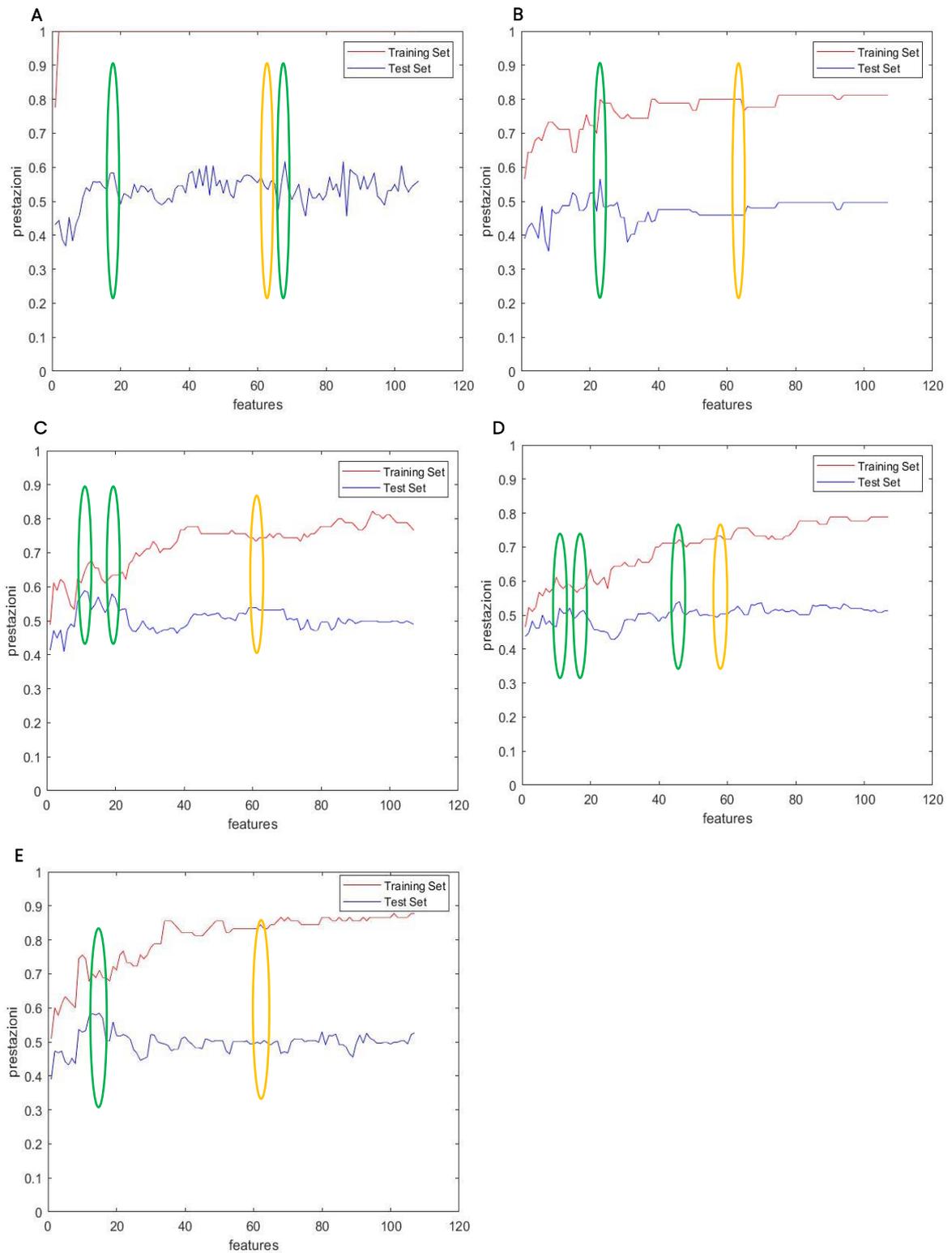
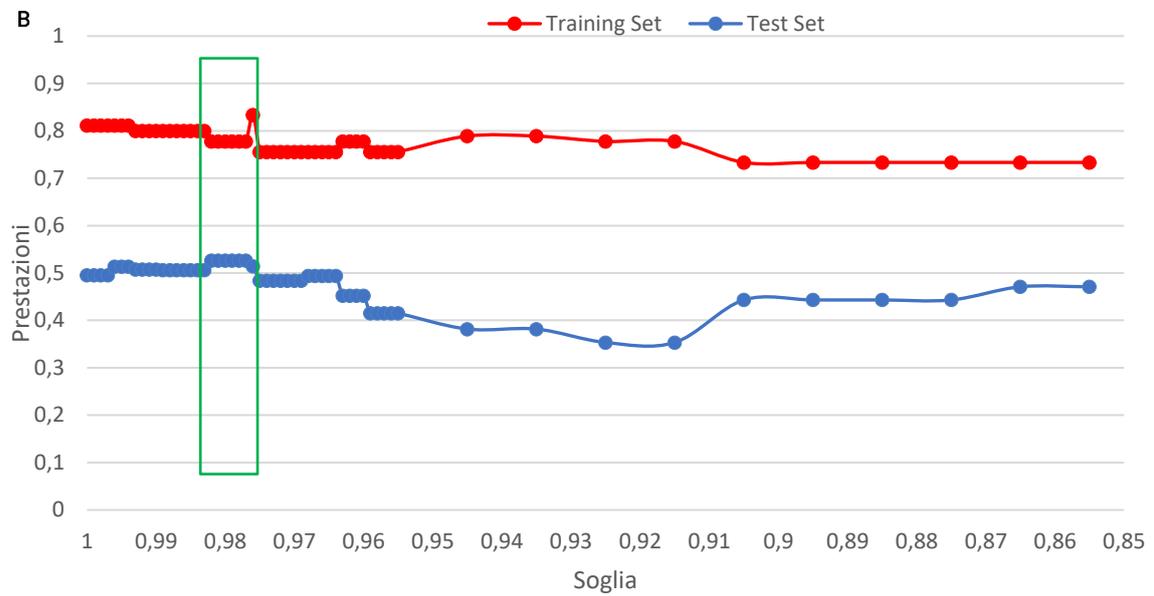
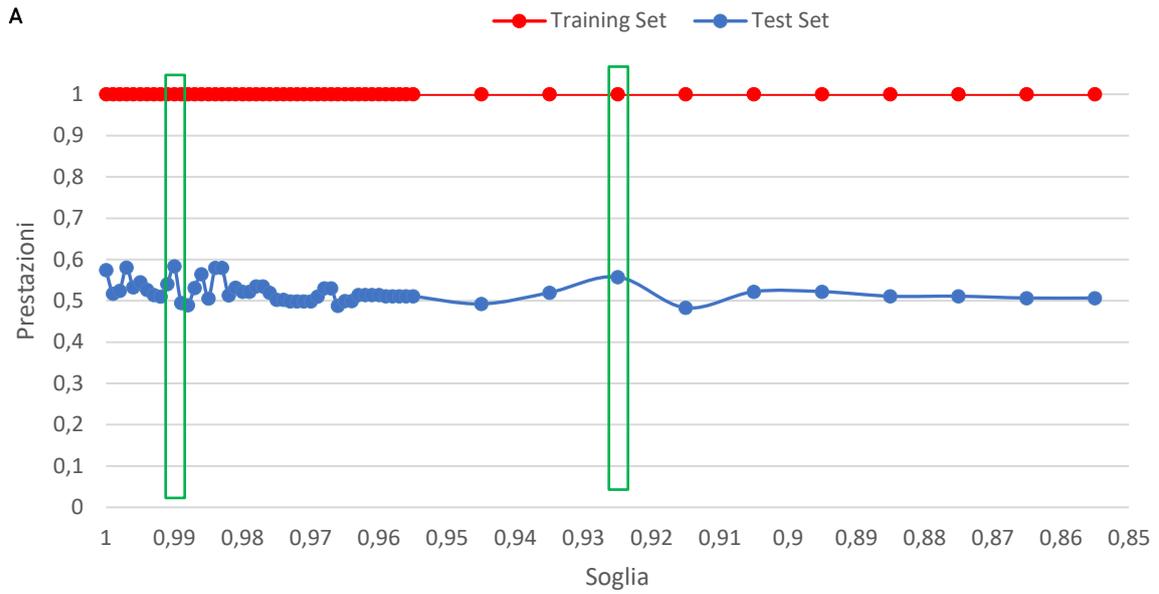


Figura 42: curve relative all'andamento delle prestazioni dei modelli di RF (Figura A), DT (Figura B), bayesiano (Figura C) ed SVM lineare (Figura D) e gaussiano (Figura E) allenati su TRS_SOM_3 con un numero di features crescente (primo approccio). In verde sono indicati i subset di features in corrispondenza dei quali si verifica un picco nelle prestazioni sul Test Set e in giallo le prestazioni in corrispondenza del subset di features con $p\text{-value} \leq 0.05$



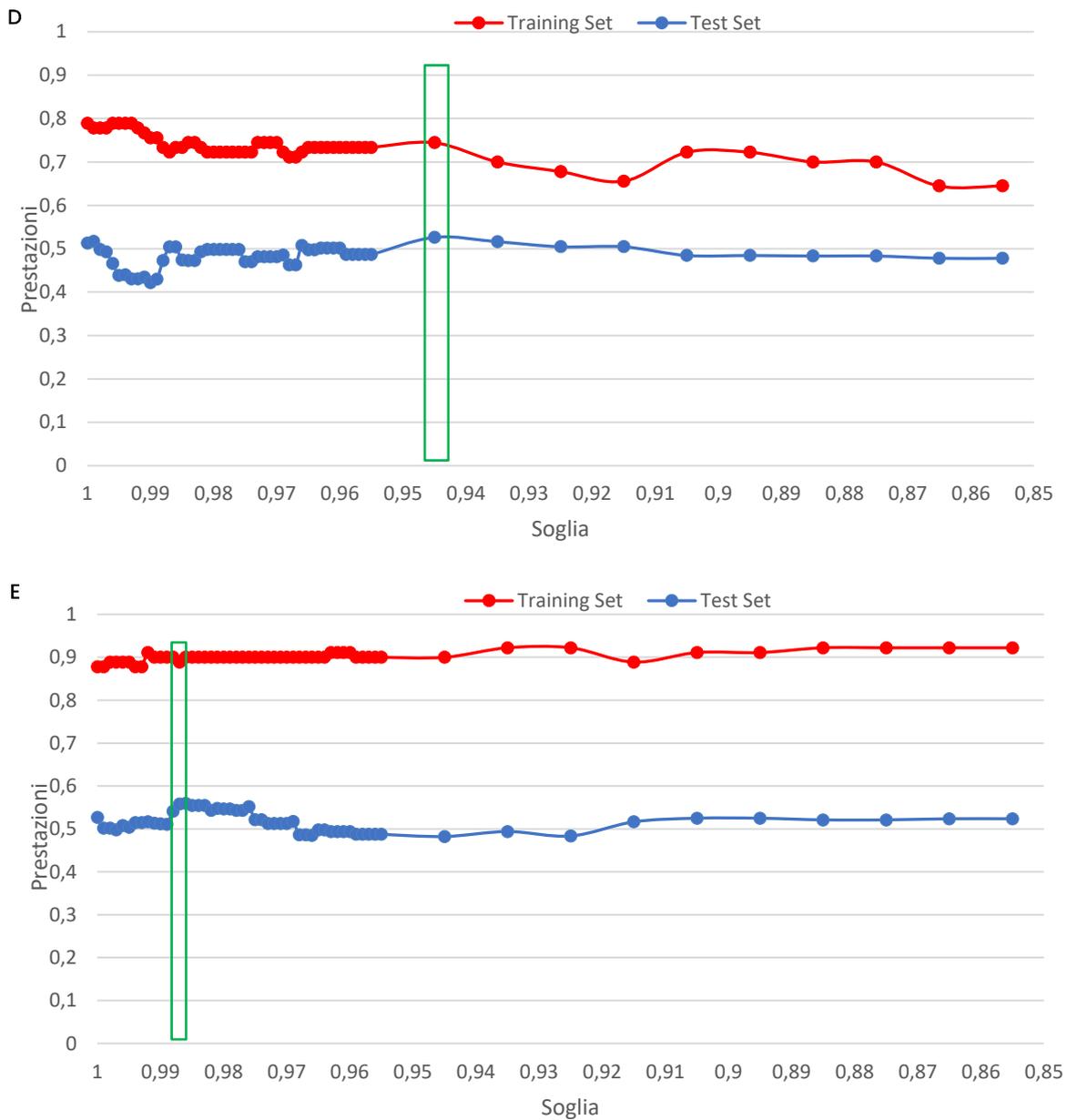
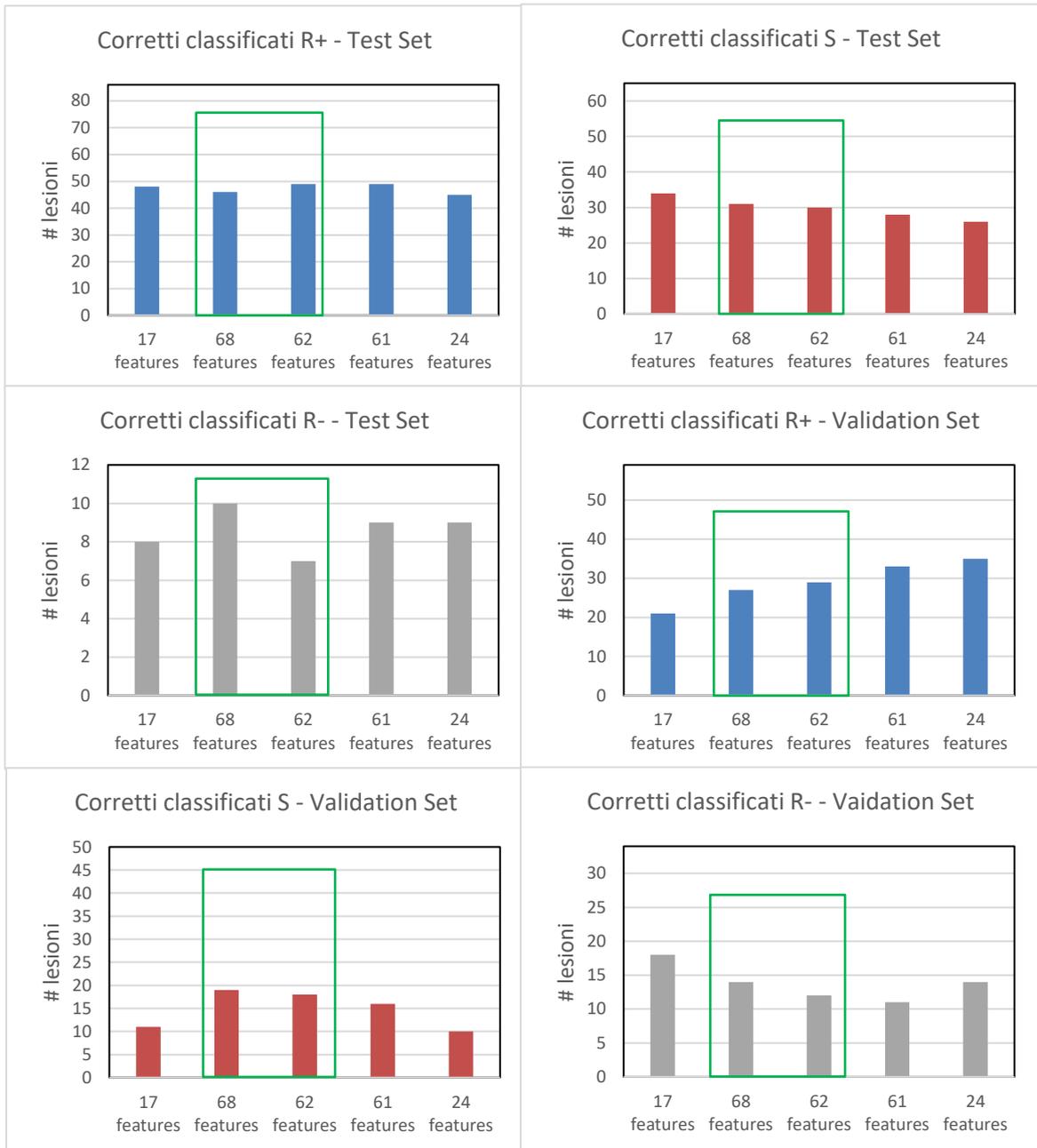


Figura 43: curve relative all'andamento delle prestazioni dei modelli di RF (Figura A), DT (Figura B), bayesiano (Figura C) ed SVM lineare (Figura D) e gaussiano (Figura E) allenati su TRS_SOM_3 scartando le features meno correlate con la classe appartenenti alle coppie di features correlate fra loro più di una certa soglia (secondo approccio). In verde sono indicati i subset di features in corrispondenza dei quali si verifica un picco nelle prestazioni sul Test.



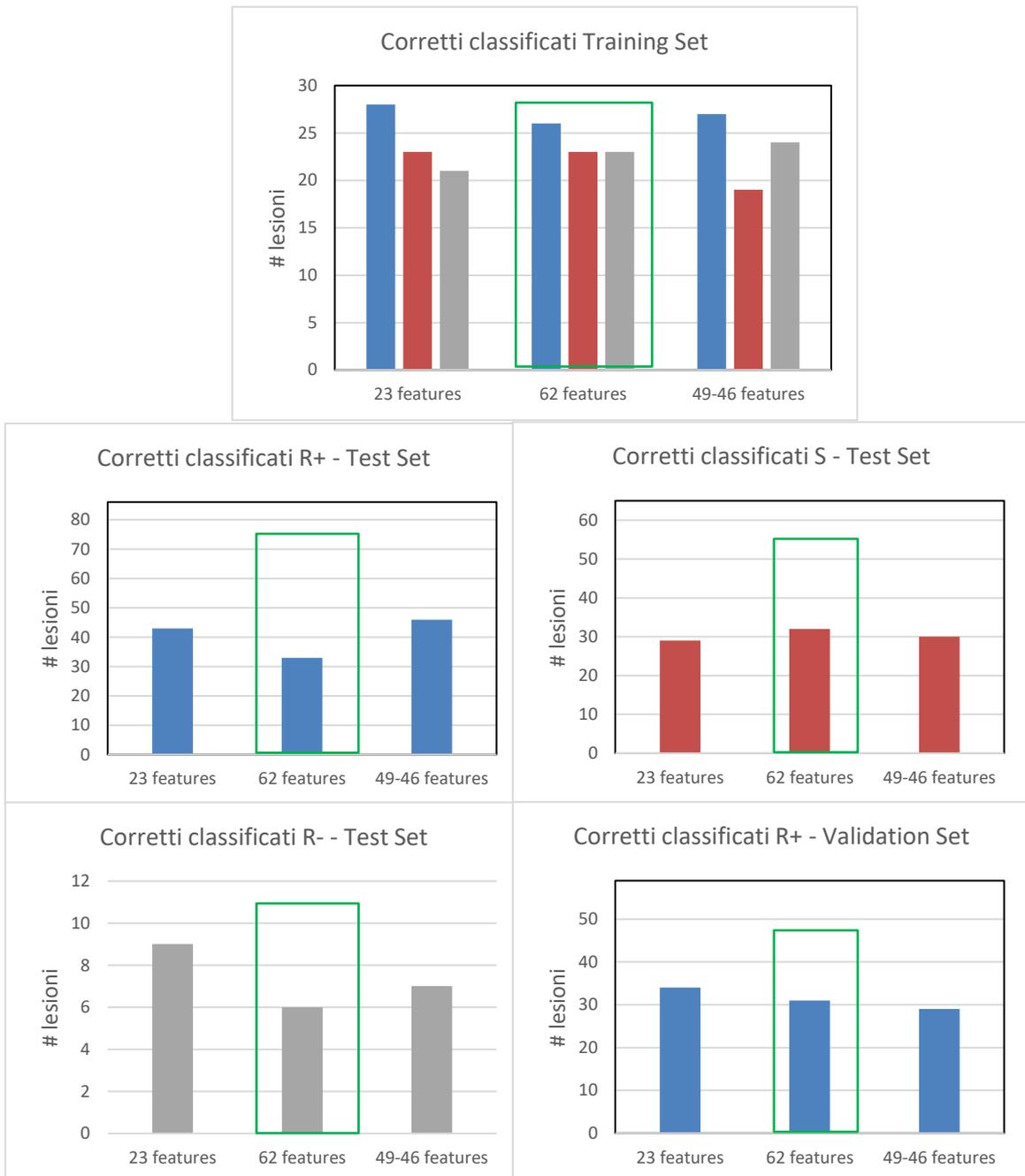
TRS	Classe vera			TS	Classe vera			VS	Classe vera					
	R+	S	R-		R+	S	R-		R+	S	R-			
Classe predetta	R+	30	0	0	Classe predetta	R+	46	25	2	Classe predetta	R+	27	19	8
	S	0	30	0		S	22	31	0		S	17	19	12
	R-	0	0	30		R-	18	9	10		R-	15	12	14

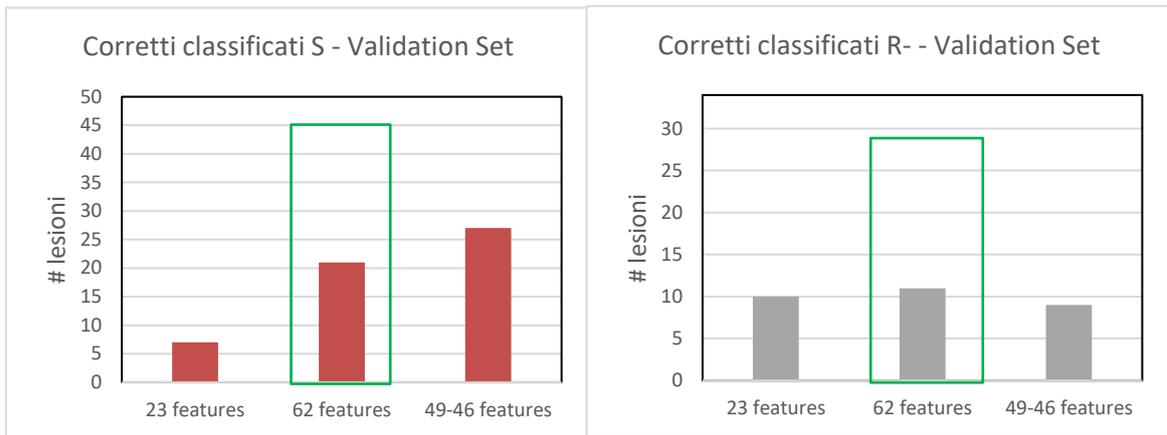
Confusion matrix con le prestazioni fornite dal RF allenato con 68 Features

TRS				TS				VS						
		Classe vera					Classe vera					Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	30	0	0	Classe predetta	R+	49	26	2	Classe predetta	R+	29	25	10
	S	0	30	0		S	17	30	3		S	14	18	12
	R-	0	0	30		R-	20	9	7		R-	16	7	12

Confusion matrix con le prestazioni fornite dal RF allenato con 62 Features ($p\text{-value} < 0.05$)

Figura 44: confronto del n° di corretti classificati ottenuti con i modelli di RF migliori individuati (approccio 1 e 2) allenati su TRS_SOM_3. In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.

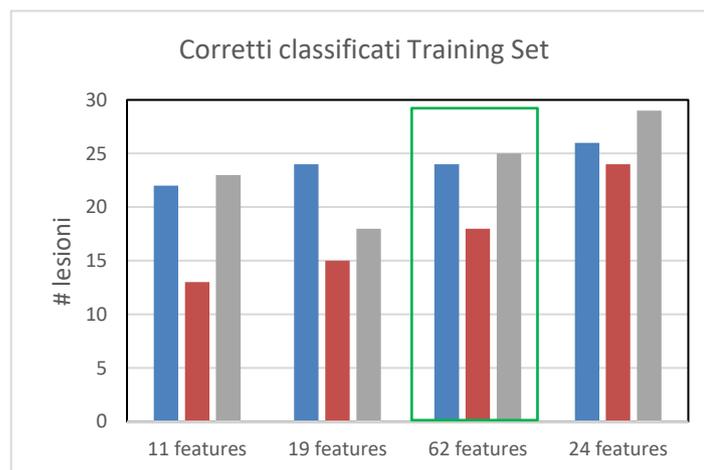


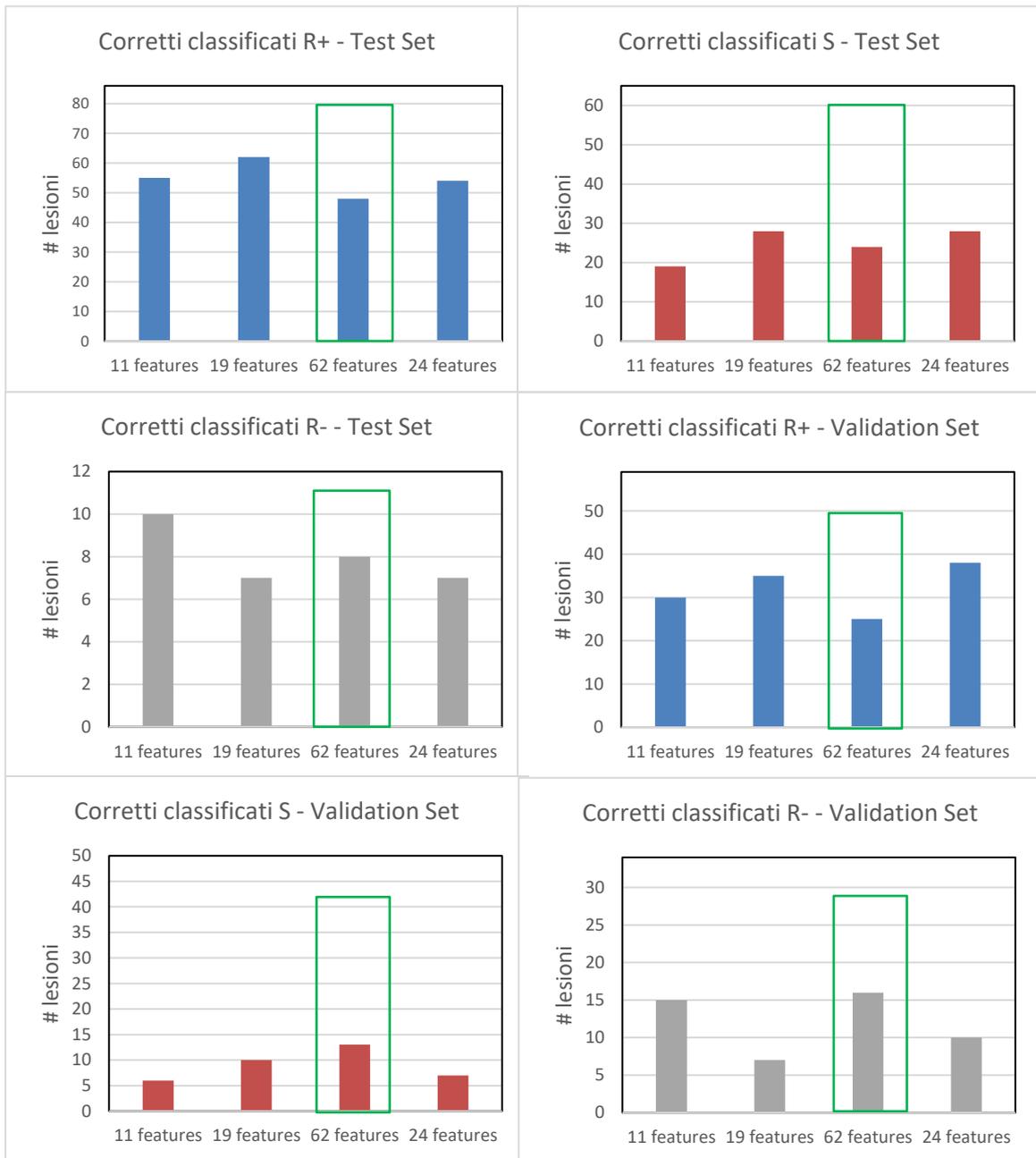


TRS				TS				VS			
Classe predetta	Classe vera			Classe predetta	Classe vera			Classe predetta	Classe vera		
	R+	S	R-		R+	S	R-		R+	S	R-
R+	26	4	0	R+	33	25	3	R+	31	23	11
S	1	23	7	S	32	32	3	S	20	21	12
R-	3	3	23	R-	21	8	6	R-	8	6	11

Confusion matrix con le prestazioni fornite dal DT allenato con 62 Features ($p\text{-value} < 0.05$)

Figura 45: confronto del n° di corretti classificati ottenuti con i modelli di DT migliori individuati (approccio 1 e 2) allenati su TRS_SOM_3. In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.

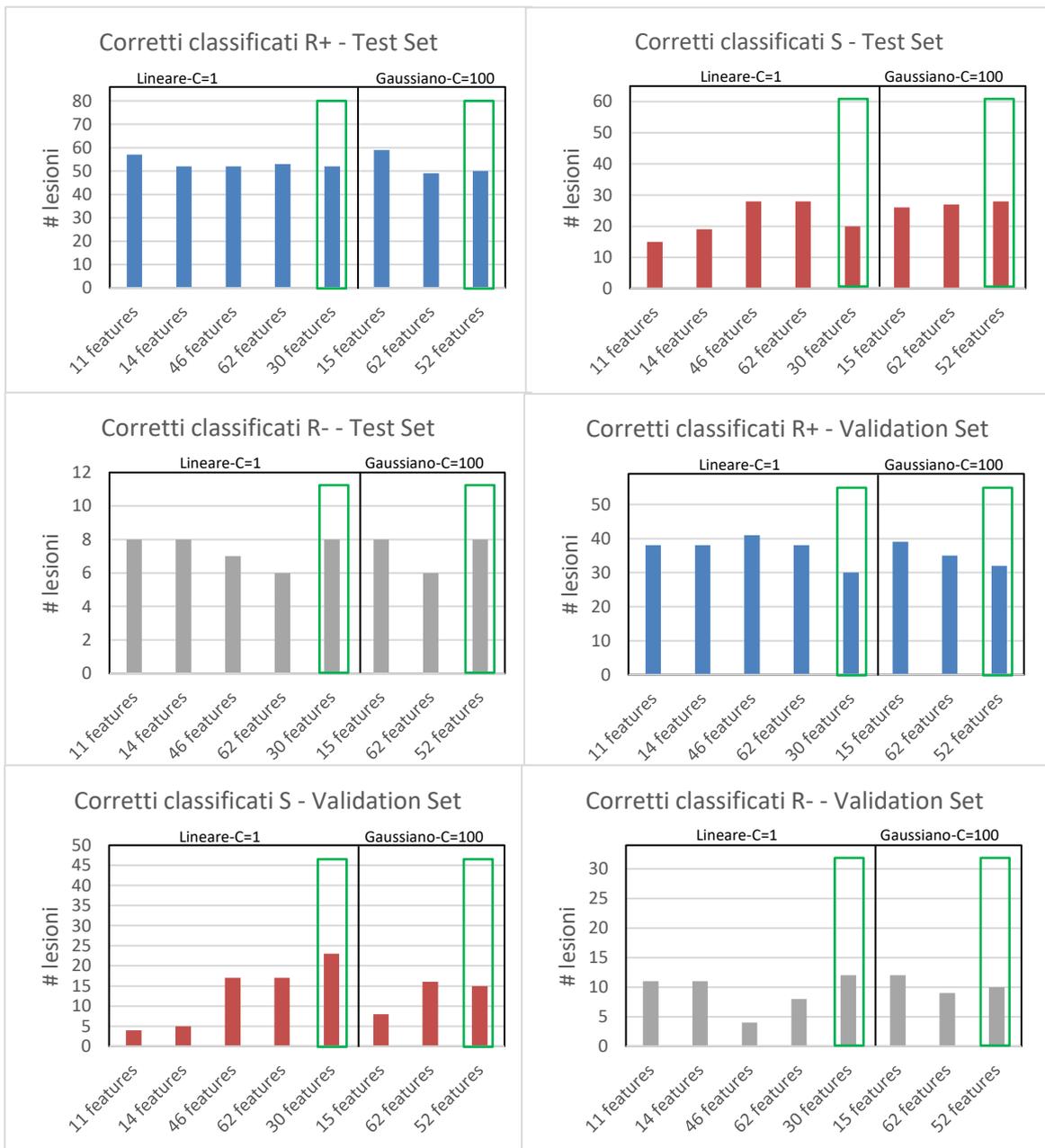
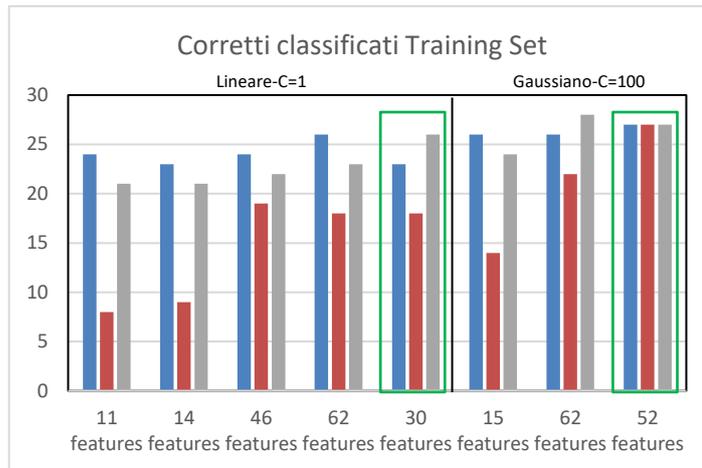




TRS	Classe vera			TS	Classe vera			VS	Classe vera			
	R+	S	R-		R+	S	R-		R+	S	R-	
Classe predetta	R+	24	8	2	R+	48	26	3	R+	25	23	9
	S	5	18	3	S	20	24	1	S	16	13	9
	R-	1	4	25	R-	18	15	8	R-	18	14	16

Confusion matrix con le prestazioni fornite dal bayesiano allenato con 62 Features (p -value < 0.05)

Figura 46: confronto del n° di corretti classificati ottenuti con i modelli di bayesiano migliori individuati (approccio 1 e 2) allenati su TRS_SOM_3. In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.



TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	23	9	2	Classe predetta	R+	52	29	3	Classe predetta	R+	30	16	11
	S	4	18	2		S	14	20	1		S	17	23	11
	R-	3	3	26		R-	20	16	8		R-	12	11	12

Confusion matrix con le prestazioni fornite dal SVM lineare allenato con 30 Features (*correlazione* < 0.945)

TRS		Classe vera			TS		Classe vera			VS		Classe vera		
		R+	S	R-			R+	S	R-			R+	S	R-
Classe predetta	R+	27	3	1	Classe predetta	R+	50	28	2	Classe predetta	R+	32	23	10
	S	3	27	2		S	21	28	2		S	16	15	14
	R-	0	0	27		R-	15	9	8		R-	11	12	10

Confusion matrix con le prestazioni fornite dal SVM gaussiano allenato con 52 Features (*correlazione* < 0.986)

Figura 47: confronto del n° di corretti classificati ottenuti con i modelli di SVM migliori individuati (approccio 1 e 2) allenati su STR_SOM_3. In verde sono indicate le migliori prestazioni sui 3 dataset. Nelle confusion matrix sono riportate le prestazioni su Training Set, Test Set e Validation Set dei modelli allenati con i subset di features individuati.

Confronto tra le prestazioni ottenute

Nel seguente capitolo verranno effettuati dei confronti tra le prestazioni ottenute con i modelli costruiti e verranno individuati i modelli con prestazioni migliori.

In figura [48](#) sono riportati le prestazioni dei modelli migliori individuati nei capitoli [Algoritmi genetici](#), [Random Forest](#), [Feature selection basata su Correlazione](#) e [QuickReduct Algorith](#). Le prestazioni dei modelli di bayesiano ed SVM allenati con i subset di features ottenuti con gli algoritmi genetici sono state riportate solo nei grafici relativi al Validation Set essendo i modelli allenati su tutto il construction Set.

Si osserva un forte overfitting visto il calo delle prestazioni su Test Set e Validation Set in particolare con tutti i modelli di Random forest, che forniscono prestazioni sul Training Set, pari al 100 % di corretti classificati su tutte e 3 le classi ma prestazioni medie pari a 53 % (46/86) e 52 % (34/65) di corretti classificati rispettivamente sulle classi R+ ed S sul Test Set e a 36 % (21/59), 44 % (22/50) e 38 % (13/34) di corretti classificati rispettivamente sulle classi R+, S ed R- sul Validation Set.

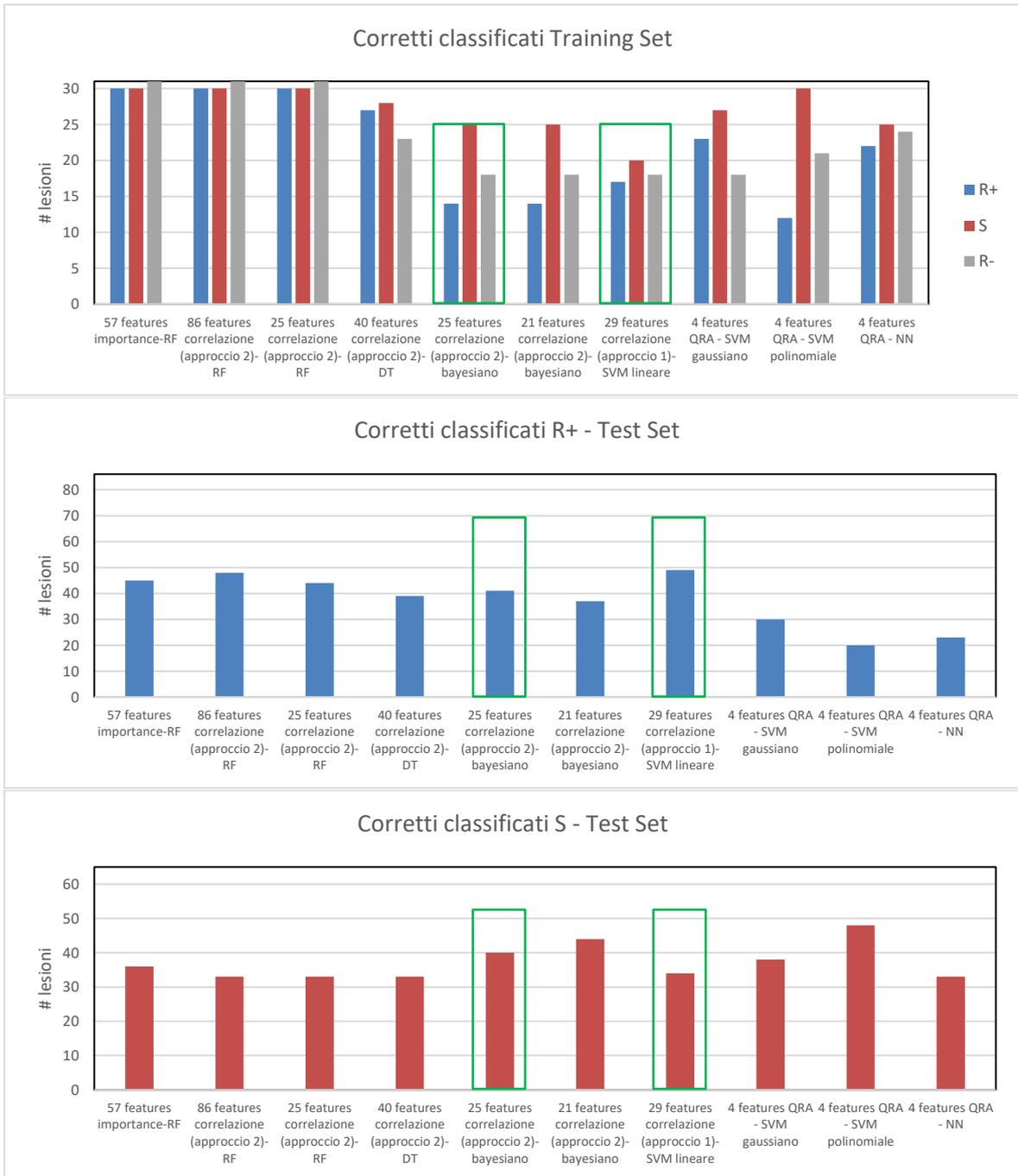
Le prestazioni ottenute con i modelli di classificatore bayesiano sul Validation Set allenati effettuando features selection basata sulla correlazione sono confrontabili rispetto a quelle ottenute con lo stesso modello allenato con il sottoinsieme di features individuato dal GA con cui si ottiene un numero di corretti classificati inferiore sulla classe R+ (19/59 contro i 28-29/59 ottenuti con features selection basata sulla correlazione) e un numero di corretti classificati maggiore sulle classi S (21/50 contro 17-18/50) ed R- (15/34 contro 10-11/34). Le prestazioni sul Validation Set ottenute con il classificatore SVM allenato effettuando features selection basata sulla correlazione risultano essere migliori rispetto a quelle ottenute con i modelli di SVM allenati con i sottoinsiemi di features individuati dal GA e dal QRA con cui si ottiene un numero di corretti classificati relativi alla classe R- sul Validation Set molto basso (2/34 e 2-4/34).

I modelli migliori risultano essere il modello di SVM lineare allenato con le 29 features significativamente correlate con la classe poiché le prestazioni su Training Set e Test Set risultano essere simili e il modello di bayesiano allenato con il sottoinsieme di features effettuando FS basata sulla correlazione (25 features). Con entrambi i modelli si osserva però un calo delle prestazioni sul Validation Set.

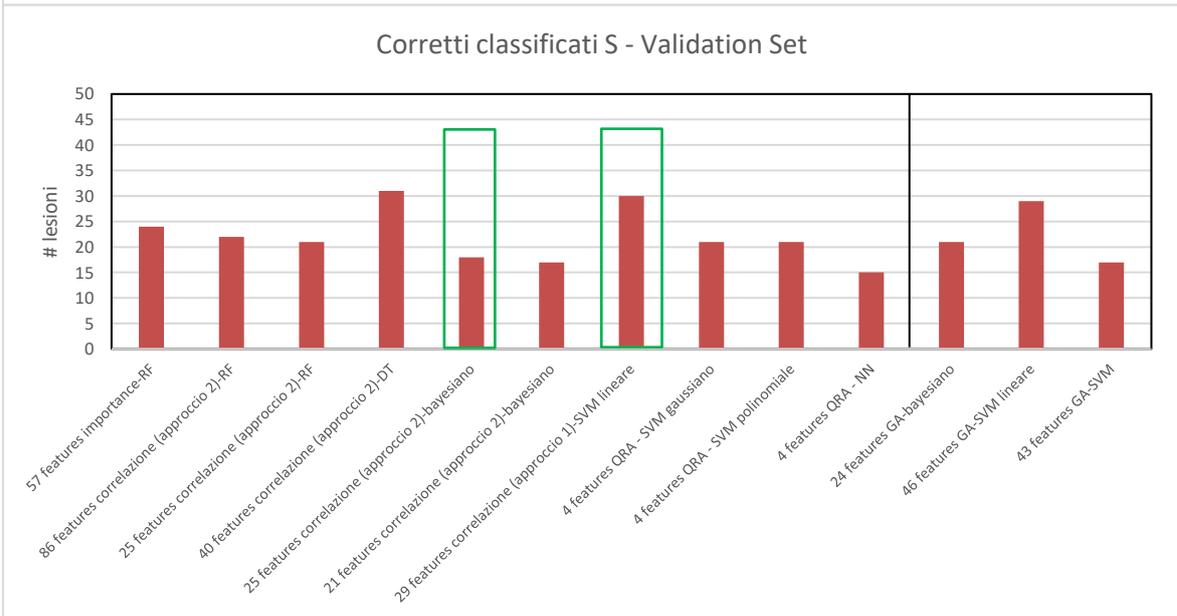
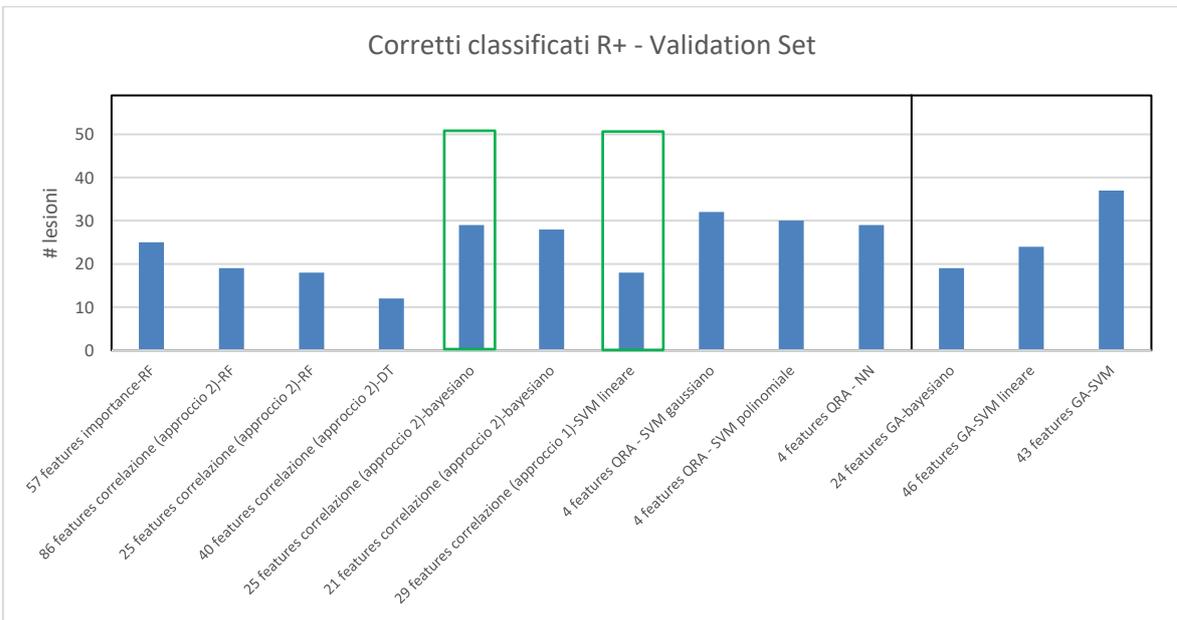
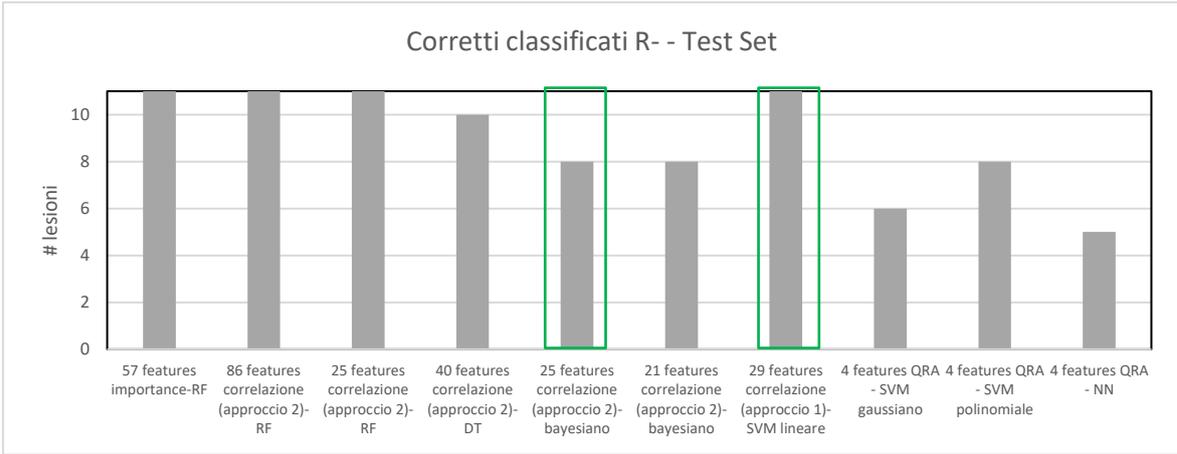
Con il modello di SVM si ottengono percentuali di corretti classificati pari a 57 % (17/30), 67 % (20/30) e 58 % (18/31) sul Training Set, 57 % (49/86), 52 % (34/65) e 100% (11/11) sul Test Set e 30 % (18/59), 60 % (30/50) e 41 % (14/34) sul Validation Set rispettivamente sulle classi R+, S ed R-.

Con il modello di bayesiano si ottengono percentuali di corretti classificati pari a 47 % (14/30), 83 % (25/30) e 58 % (18/31) sul Training Set, 48 % (41/86), 61 % (40/65) e 73 % (8/11) sul Test Set e 49 % (29/59), 36 % (18/50) e 29 % (10/34) sul Validation Set rispettivamente sulle classi R+, S ed R-.

Confronto tra le prestazioni ottenute



Confronto tra le prestazioni ottenute



Confronto tra le prestazioni ottenute

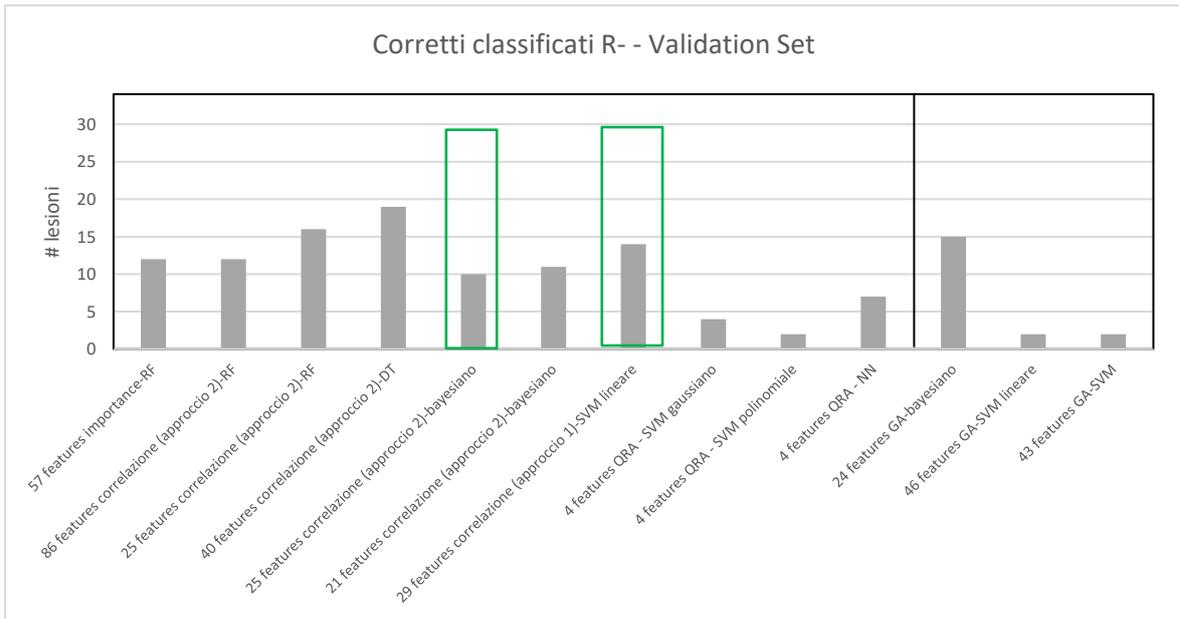
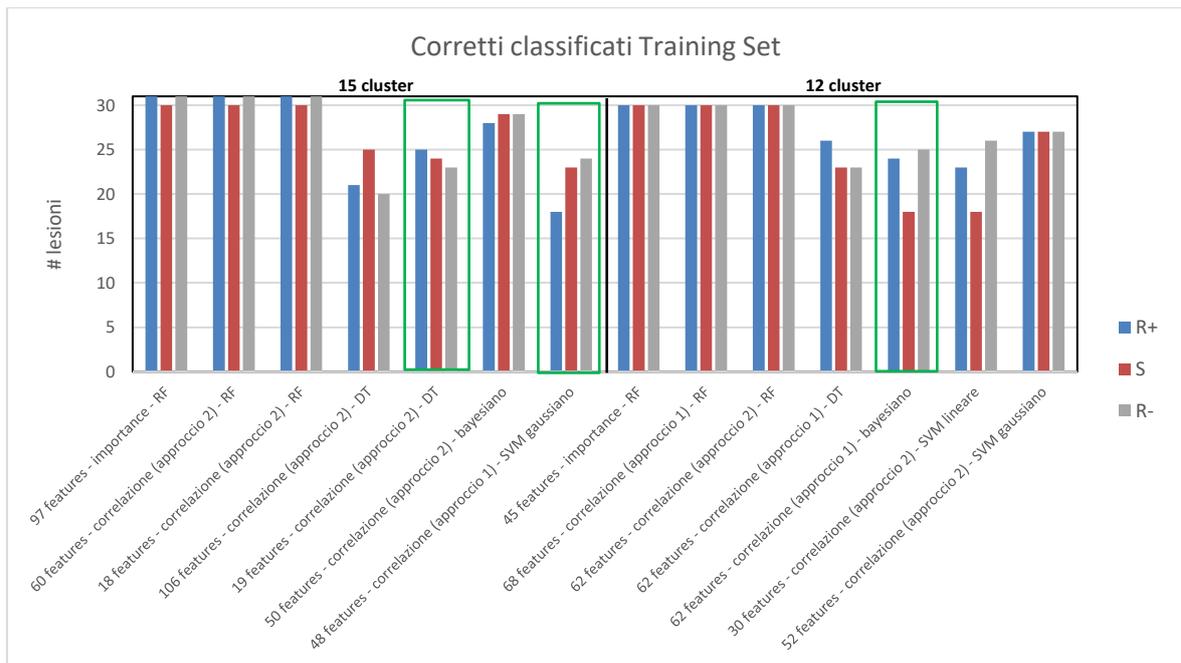
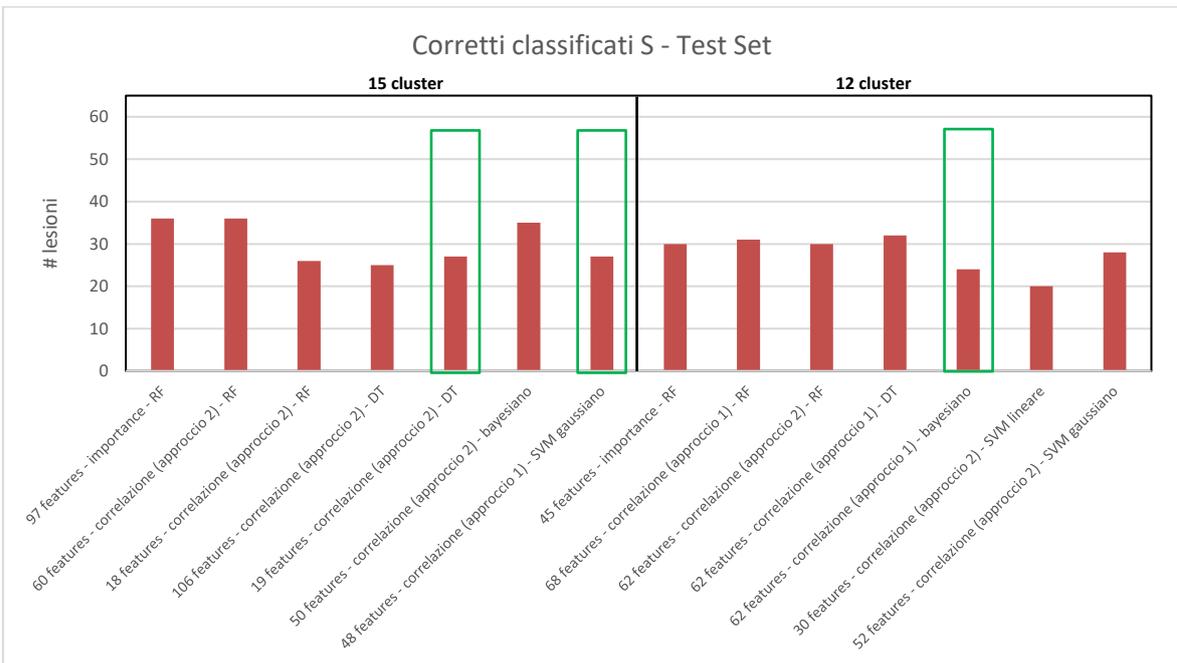
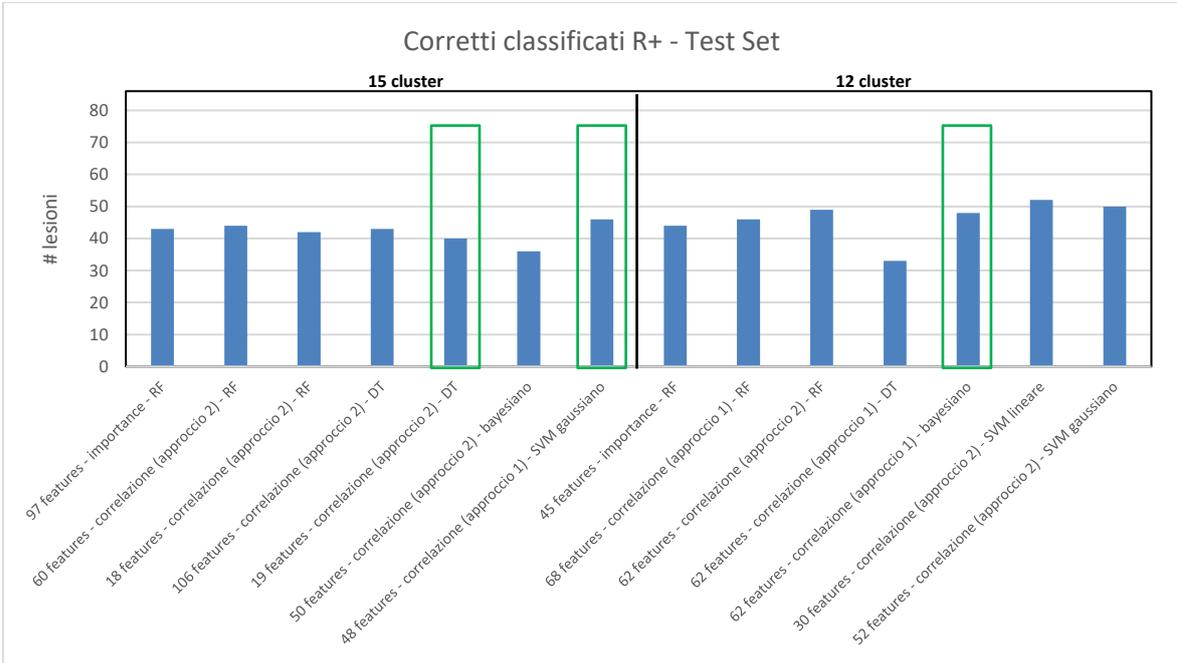


Figura 48: confronto del n° di corretti classificati ottenuti con i modelli migliori individuati nei capitoli Algoritmi genetici, Random Forest, Feature selection basata su Correlazione e QuickReduct Algorithm

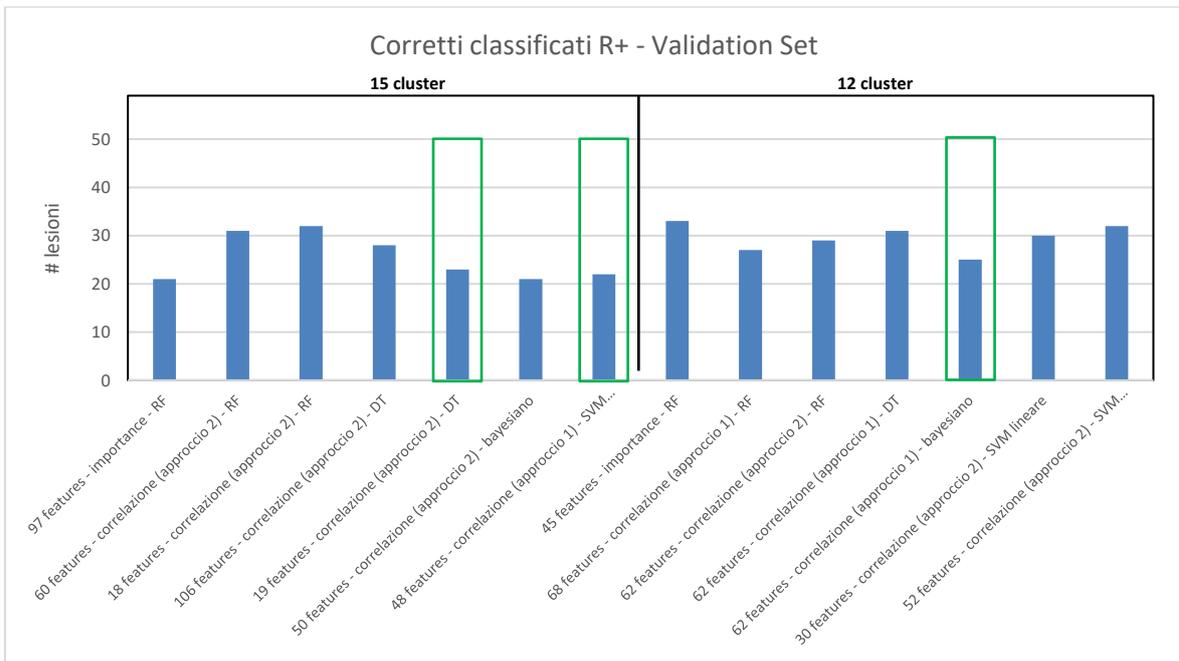
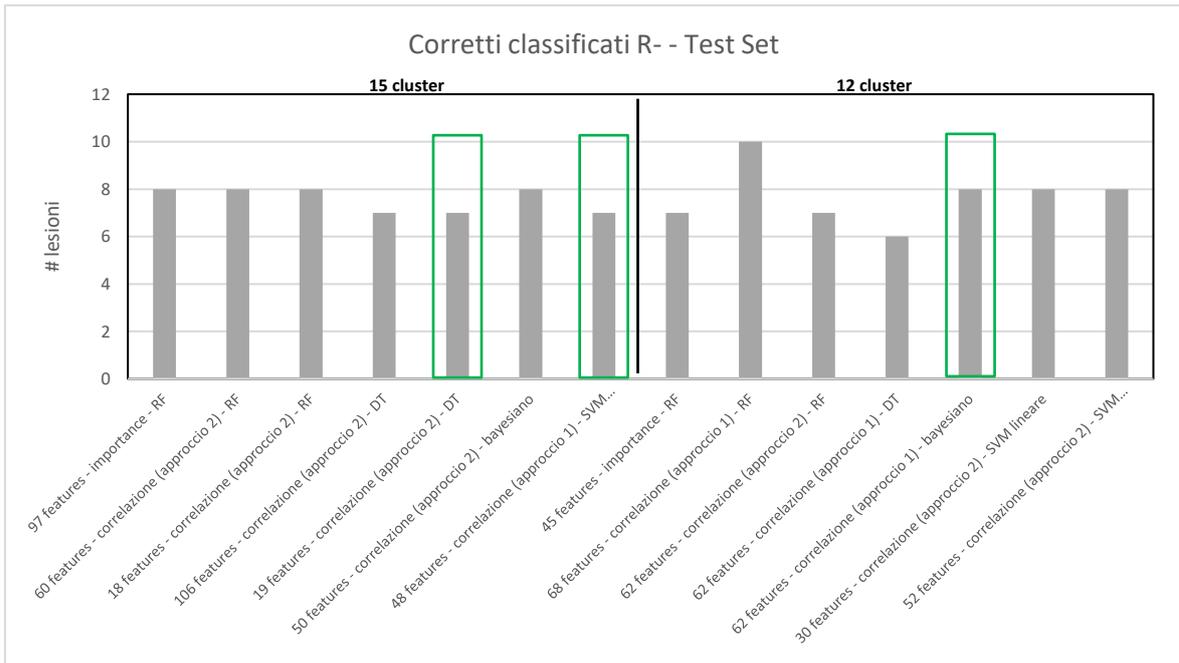
In figura 49 sono messe a confronto le prestazioni relative alle tre classi dei modelli migliori allenati per classificare le lesioni nei cluster (TRS_SOM_2 e TRS_SOM_3) da cui si osserva che la clusterizzazione delle lesioni non ha portato a miglioramenti nelle prestazioni. Con tutti i modelli di RF, anche in questo caso, si osserva il fenomeno di overfitting poiché le prestazioni sul Training Set risultano essere pari al 100% di corretti classificati su tutte e 3 le classi ma si riducono poi su Test Set e Validation Set. Si ottengono prestazioni confrontabili su Test Set e Validation Set con i diversi modelli confrontati. Sono stati individuati come migliori, in verde in figura 49, quei modelli in cui il fenomeno di overfitting è minore.



Confronto tra le prestazioni ottenute



Confronto tra le prestazioni ottenute



Confronto tra le prestazioni ottenute

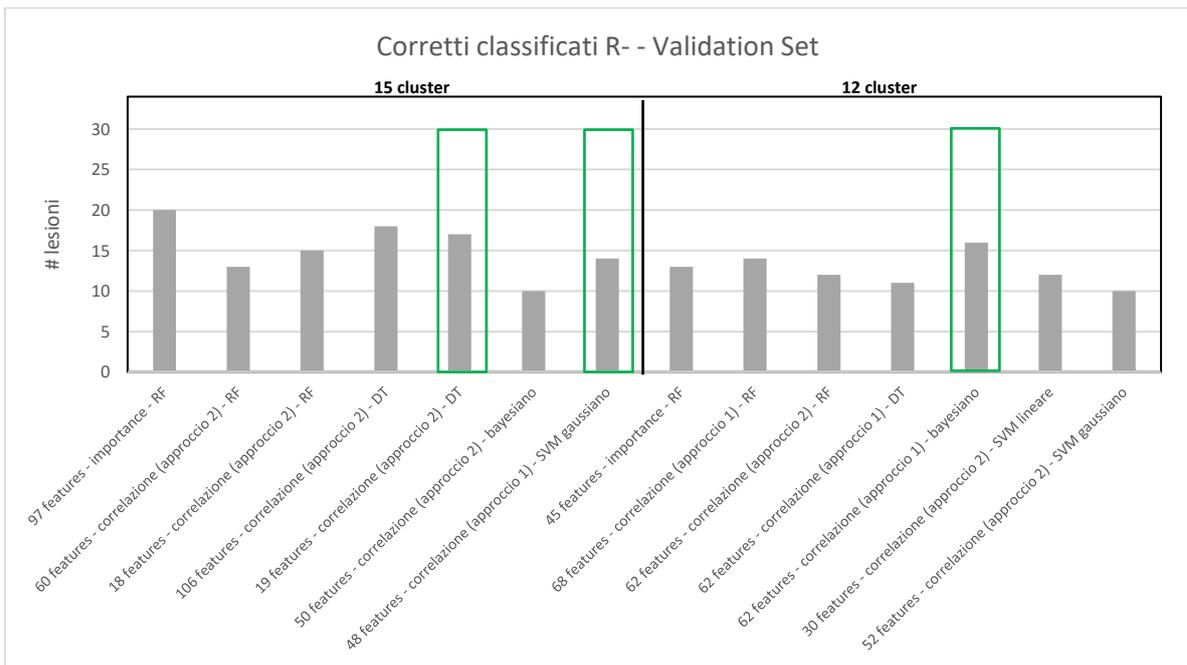
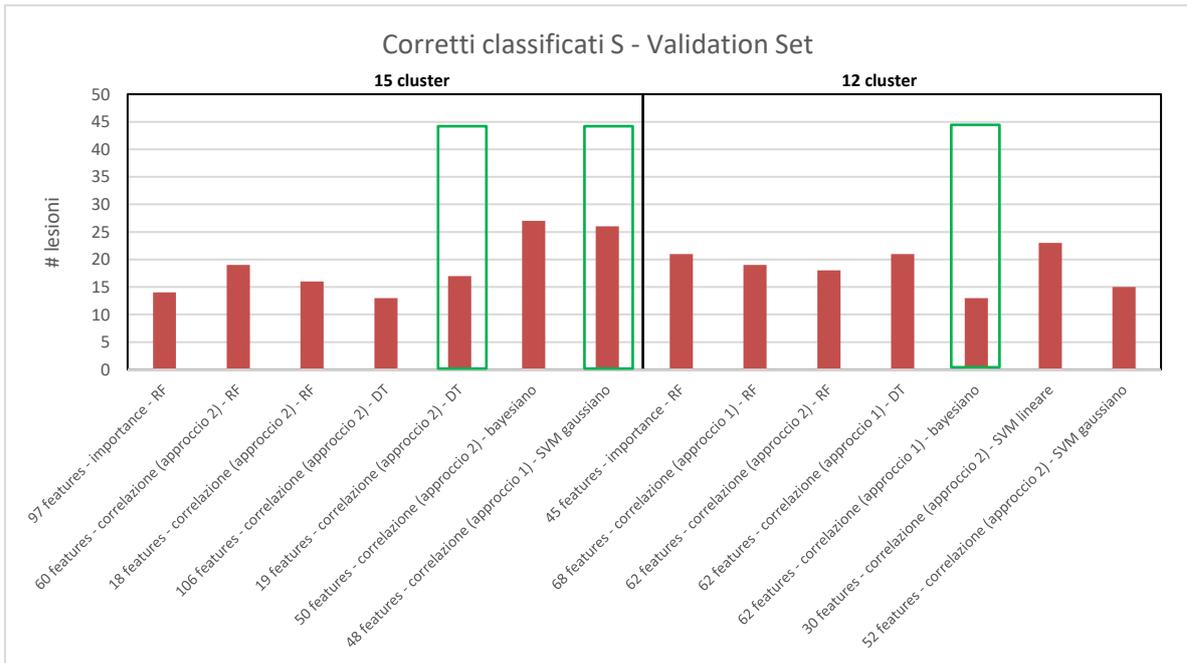


Figura 49: confronto del n° di corretti classificati relativi alle tre classi ottenuti con i modelli migliori allenati per classificare le lesioni nei cluster (TRS_SOM_2 e TRS_SOM_3)

Confrontando le prestazioni sono stati individuati i migliori modelli per ogni tipo di classificatore.

I modelli di RF fornivano prestazioni confrontabili sui 3 dataset che risultano essere in media pari a 52% (45/86), 49% (32/65), 75% (9/12) rispettivamente sulle classi R+, S ed R- sul

Test Set e 44% (26/59), 38% (19/50) e 41% (14/34) rispettivamente sulle classi R+, S ed R- sul Validation Set.

È stato individuato come migliore DT il modello allenato con 23-19 features (features selection basata sulla correlazione; 15 cluster) poiché forniva le migliori prestazioni sulle 3 classi su Test Set e Validation Set: le prestazioni si attestano intorno all'80% di corretti classificati sulle 3 classi sul Training Set ma si riducono al 47% (40/85), 41% (27/65) e 64% (7/11) sul Test Set e a 39% (23/59), 34% (17/50), 50% (17/34) sul Validation Set rispettivamente sulle classi R+, S ed R-.

È stato individuato come miglior bayesiano, il modello allenato con 62 features ($p\text{-value} \leq 0.05$, 12 cluster) che forniva prestazioni superiori al 50% di corretti classificati sulle 3 classi sul Training Set e una minore diminuzione delle prestazioni su Test Set e Validation Set: le prestazioni si attestano intorno all'80% (24-25/30) di corretti classificati sulle classi R+ ed R- e sono pari al 60% (18/30) sulla classe S sul Training Set ma si riducono al 56% (48/85), 37% (24/65) e 73% (8/11) sul Test Set e a 42% (25/59), 26% (13/50), 47% (16/34) sul Validation Set rispettivamente sulle classi R+, S ed R-.

È stato individuato come migliore SVM, il modello gaussiano allenato con 48 features ($p\text{-value} \leq 0.05$, 15 cluster) che permette di ottenere prestazioni pari a 58% (18/31), 77% (23/30) e 77% (24/31) rispettivamente sulle classi R+, S ed R- sul Training Set, pari a 54% (46/85), 41% (27/65) e 64% (7/11) rispettivamente sulle classi R+, S ed R- sul Test Set e pari a 37% (22/59), 52% (26/50), 41% (14/34) rispettivamente sulle classi R+, S ed R- sul Validation Set.

Riconcettualizzazione del problema: data mining

Erano disponibili, per le 396 lesioni fornite, le evoluzioni temporali dei loro diametri relative a follow up (FU) successivi al primo. Con queste informazioni sono state valutate le risposte a lungo termine di ciascuna lesione utilizzando i seguenti criteri:

- CR (complete response) se la lesione scompariva,
- PD (progressive disease) se tra la fine della terapia e il nadir (diametro minimo raggiunto dalla lesione nei diversi FU) c'era un aumento del diametro di almeno 20% (a prescindere dalla differenza tra il baseline e la fine del trattamento). La variazione del diametro è stata calcolata come $\frac{\text{diametro}_{\text{minimo}} - \text{diametro}_{\text{finale}}}{\text{diametro}_{\text{minimo}}}$,
- PR (partial response) se il diametro minimo corrispondeva al diametro relativo all'ultimo FU e la riduzione di diametro era maggiore del 30%,
- quando non si verificava nessuna delle situazioni precedenti la lesioni è stata classificata come stabile.

È stato osservato come circa il 27% delle lesioni, che risultavano responder al primo FU, a lungo termine andavano in progressione (non responder). I clinici hanno inoltre suggerito che, siccome prima o poi tutte le lesioni progrediscono, risulterebbe essere segno di una buona risposta se le lesioni progrediscono dopo 8 mesi.

Sono state escluse 144 lesioni su 396 (85 responder e 59 stabili a lungo termine) perché non avevano un trattamento di almeno 8 mesi (se la lesione risultava stabile o responder a lungo termine). È stato quindi valutato il tempo libero da progressioni per le lesioni (TTP) e le metastasi sono state assegnate alle classi R-, R+ e S usando i seguenti criteri:

- R- (0): se $TTP < 8$ e la risposta a lungo termine era PD,
- R+ (1): se $TTP \geq 8$ e la risposta a lungo termine era CR/PR/PD,
- S (2): se $TTP \geq 8$ e la risposta a lungo termine era stabile.

Visti quindi i risultati ottenuti cercando di predire la risposta al primo FU, è stato deciso di effettuare una nuova concettualizzazione del problema al fine di valutare se, con le features estratte dalle TC al baseline, fosse possibile predire la risposta delle lesioni a lungo termine.

In tabella [18](#) sono riportate le numerosità delle lesioni appartenenti alle nuove classi individuate.

	Candiolo	Niguarda	Totale
Pazienti	15	46	61
Lesioni	74	178	252
Lesioni R-	37	77	114
Lesioni R+	33	95	128
Lesioni S	4	6	10

Tabella 18: caratteristiche dei dati a disposizione.

Per effettuare delle valutazioni preliminari sono state allenate delle reti SOM al fine di analizzare se fosse possibile individuare dei cluster che rispecchiassero la nuova classificazione a lungo termine delle lesioni.

È stata quindi effettuata una clusterizzazione delle lesioni tramite reti SOM di dimensioni 4x4 e 5x5 allenate sul Construction Set e valutate su Construction Set e Validation Set per identificare la presenza di lesioni con risposta a lungo termine differente in uno stesso cluster.

Le reti SOM utilizzate per la fase di clustering sono state allenate sul Construction Set con tutte le 107 features e con il Construction Set costituito dalle sole features significativamente correlate ($p\text{-value} < 0.05$) con la risposta a lungo termine delle lesioni (R-, S, R+).

Risultati

Le prime prove di clusterizzazione sono state effettuate allenando le reti SOM con lo stesso Construction Set utilizzato per le precedenti prove di classificazione in cui è stata sostituita la classe con la risposta a lungo termine (Constr_1). Il Construction Set ottenuto era costituito da 42 pazienti del Niguarda (170 x 111) e il Validation Set (Valid_1) era costituito dai 15 pazienti di Candiolo e 4 del Niguarda (82 x 111). In figura 50 è riportato il numero di lesioni appartenenti a ciascuna classe nei 2 dataset.

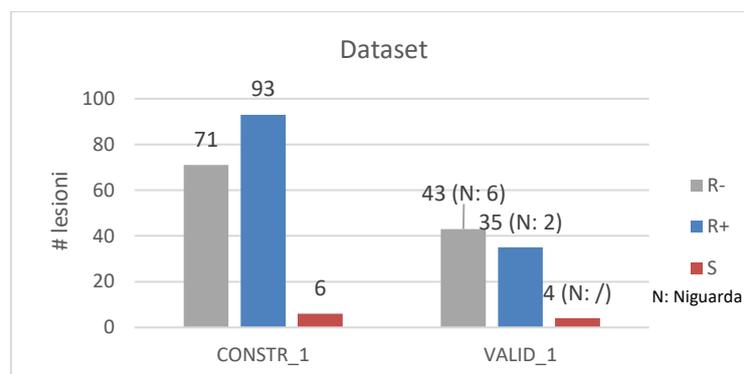


Figura 50: numerosità delle lesioni appartenenti a ciascuna classe nei 2 dataset

Nelle figure 51, 52, 53 e 54 sono riportati i dendrogrammi dei pesi e le mappe relative all'organizzazione delle diverse reti SOM allenate.

Nelle tabelle 19, 20, 21 e 22 sono riportati i neuroni appartenenti a ciascun cluster individuato (prima colonna) e la numerosità delle lesioni di Construction Set (seconda, terza e quarta colonna) e Validation Set (quinta, sesta e settima colonna) appartenenti alle 3 classi in ciascun cluster.

In figura 94 in appendice sono riportati i valori di p-value della correlazione delle 107 features nel Constr_1 in ordine crescente. Si nota come siano presenti 46 features con $p\text{-value} < 0.05$.

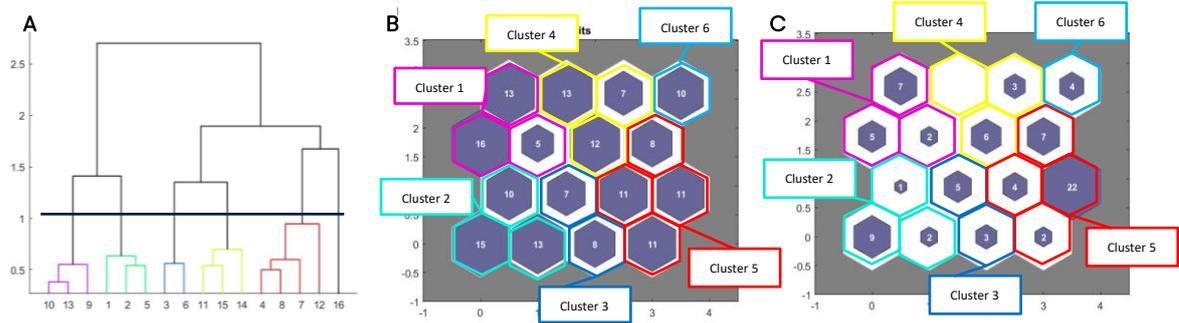


Figura 51: dendrogramma dei pesi della rete SOM 4x4 allenata con le lesioni del Constr_1 (107 features) con il relativo taglio effettuato per clusterizzare le lesioni (Figura A) e mappe con l'organizzazione della rete SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) ottenuto valutando la rete su Constr_1 (Figura B) e Valid_1 (Figura C).

Neuroni	Constr_1			Valid_1		
	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S
3, 6	6	9	0	5	3	0
11, 14, 15	12	18	2	3	5	1
9, 10, 13	12	21	1	6	7	1
1, 2, 5	18	19	1	6	6	0
16	2	8	0	3	1	0
4, 7, 8, 12	21	18	2	20	13	2

Tabella 19: tabella con il numero di lesioni, relativo a ciascuna classe, appartenenti a ciascun cluster ottenuto valutando la rete 4x4 su Constr_1 e Valid_1.

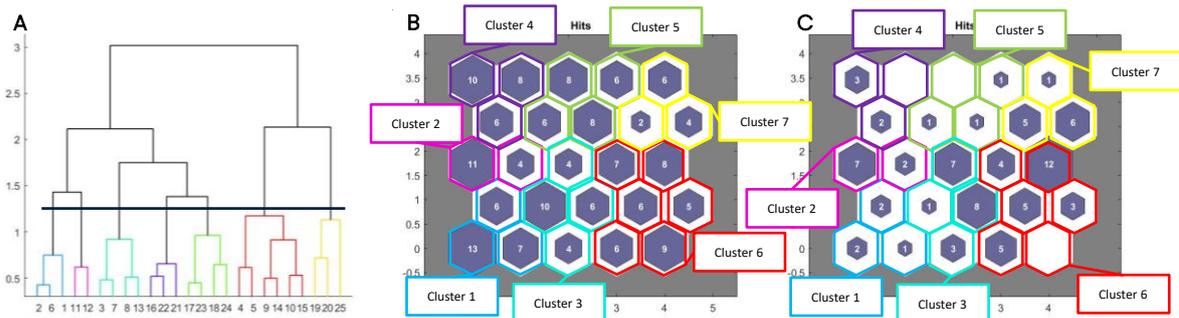


Figura 52: dendrogramma dei pesi della rete SOM 5x5 allenata con le lesioni del Constr_1 (107 features) con il relativo taglio effettuato per clusterizzare le lesioni. (Figura A) e mappe con l'organizzazione della rete SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) ottenute valutando la rete su Construction Set (Figura B) e Validation Set (Figura C).

Neuroni	Constr_1			Valid_1		
	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S
16,21,22	9	13	2	1	3	1
17,18,23, 24	15	12	1	2	0	1
11,12	8	7	0	8	1	0
1,2,6	12	13	1	0	5	0
3,7,8,13	5	19	0	6	13	0
19,20,25	2	10	0	7	4	1
4,5,9,10,14,15	20	19	2	19	9	1

Tabella 20: tabella con il numero di lesioni, relativo a ciascuna classe, appartenenti a ciascun cluster ottenuto valutando la rete 5x5 su Constr_1 e Valid_1.

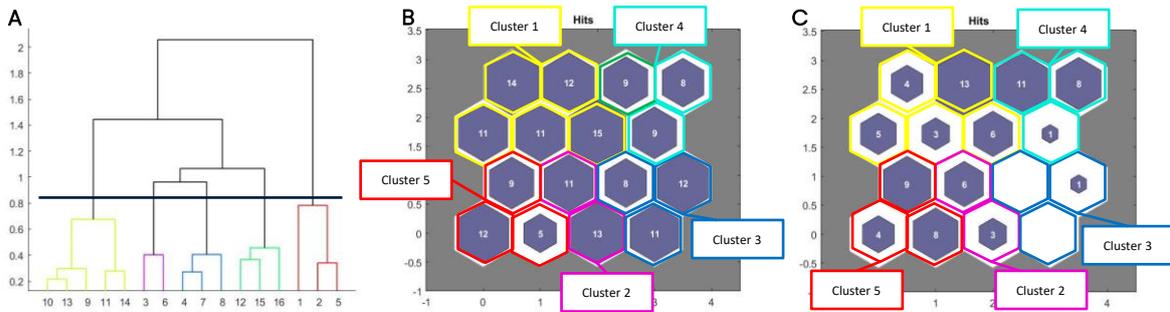


Figura 53: dendrogramma dei pesi della rete SOM 4x4 allenata con le lesioni del Constr_1 (46 features) con il relativo taglio effettuato per clusterizzare le lesioni. (Figura A) e mappe con l'organizzazione della rete SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) ottenute valutando la rete su Construction Set (Figura B) e Validation Set (Figura C).

Neuroni	Constr_1			Valid_1		
	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S
3, 6	7	15	2	5	4	0
4, 7, 8	22	8	1	1	0	0
12, 15, 16	14	12	0	11	9	0
9, 10, 11, 13, 14	22	39	2	17	13	1
1, 2, 5	6	19	1	9	9	3

Tabella 21: tabella con il numero di lesioni, relativo a ciascuna classe, appartenenti a ciascun cluster ottenuto valutando la rete 4x4 su Constr_1 e Valid_1.

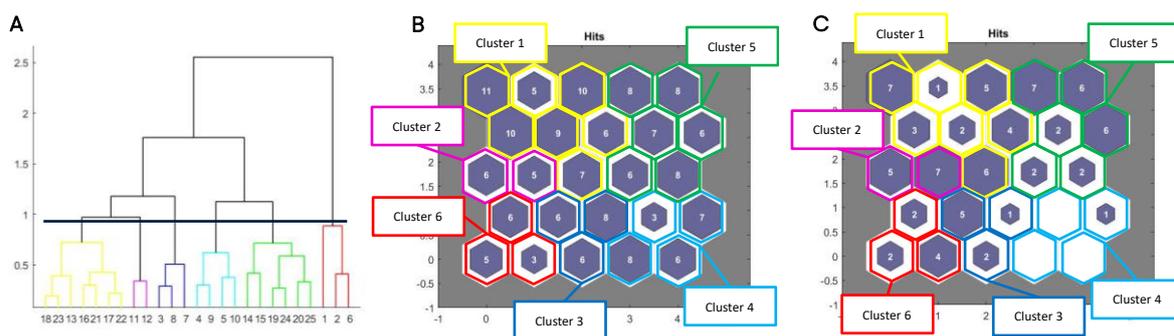


Figura 54: dendrogramma dei pesi della rete SOM 5x5 allenata con le lesioni del Constr_1 (46 features) con il relativo taglio effettuato per clusterizzare le lesioni. (Figura A) e mappe con l’organizzazione della rete SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) ottenute valutando la rete su Construction Set (Figura B) e Validation Set (Figura C).

Neuroni	Constr_1			Valid_1		
	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S
11,12	4	7	0	3	8	1
13,16, 17,18,21,22,23	19	36	3	16	11	1
4,5,9,10	17	6	1	1	0	0
14,15,19,20,24,25	23	19	1	14	11	0
3,7,8	6	14	0	4	3	1
1,2,6	2	11	1	5	2	1

Tabella 22: tabella con il numero di lesioni, relativo a ciascuna classe, appartenenti a ciascun cluster ottenuto valutando la rete 5x5 su Constr_1 e Valid_1.

È stato osservato che, quando si valutava la rete sul Validation Set, la distribuzione delle vittorie non era omogenea nei diversi neuroni. Il Validation Set risultava essere più piccolo rispetto al Construction Set ma in figura 51, ad esempio, si osserva come la maggior parte dei neuroni abbiano poche vittorie mentre un neurone presentava 22 vittorie. Questo potrebbe indicare come l’approccio di costruire un Construction Set costituito unicamente da pazienti provenienti dall’ospedale Niguarda non sia la strada ottimale.

Sono state quindi allenare altre reti SOM con un Construction Set con le stesse dimensioni del precedente (42 pazienti su 61) ma costituito anche da pazienti di Candiolo. I pazienti da inserire nel Construction Set sono stati estratti in modo che la proporzione di lesioni appartenenti a ciascuna classe risultasse simile in Construction Set (Constr_2) e Validation Set (Valid_2). Il Construction Set ottenuto era costituito da 32 pazienti dell’ospedale Niguarda e 10 dell’ospedale di Candiolo (160 x 111) e il Validation Set era costituito da 14 pazienti dell’ospedale Niguarda da 5 dell’ospedale di Candiolo (92 x 111). Il Construction Set è stato normalizzato con l’approccio del min-max scaling e il Validation Set è stato normalizzato con gli stessi minimi e massimi individuati con il Construction Set.

In figura 55 è riportata la proporzione di lesioni appartenenti a ciascuna classe (R-, S, R+) risulta simile nei 2 dataset.

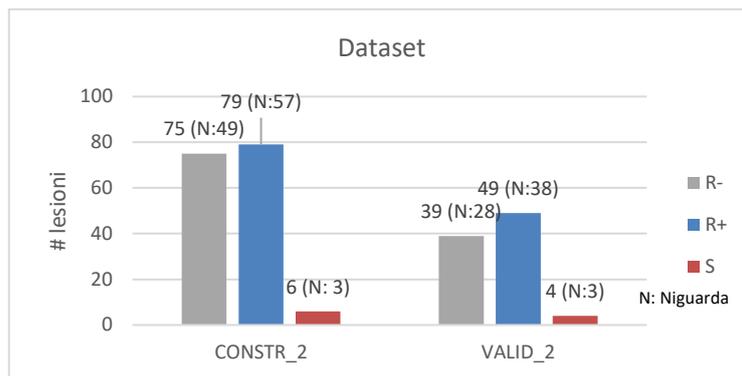


Figura 55: numerosità delle lesioni appartenenti a ciascuna classe nei 2 dataset

Nelle figure 56, 57, 58, 59 sono riportati dendrogrammi dei pesi e le mappe relative all'organizzazione delle diverse reti SOM allenate.

Nelle tabelle 23, 24, 25, 26 sono riportati i neuroni appartenenti a ciascun cluster individuato (prima colonna) e la numerosità delle lesioni di Construction Set e Validation Set appartenenti alle 3 classi in ciascun cluster i cui neuroni sono risultati essere i vincitori (seconda, terza e quarta colonna).

In figura 95 in appendice sono riportati i valori di p-value della correlazione delle 107 features nel Constr_2 in ordine crescente. Si nota come siano presenti 42 features con $p\text{-value} < 0.05$.

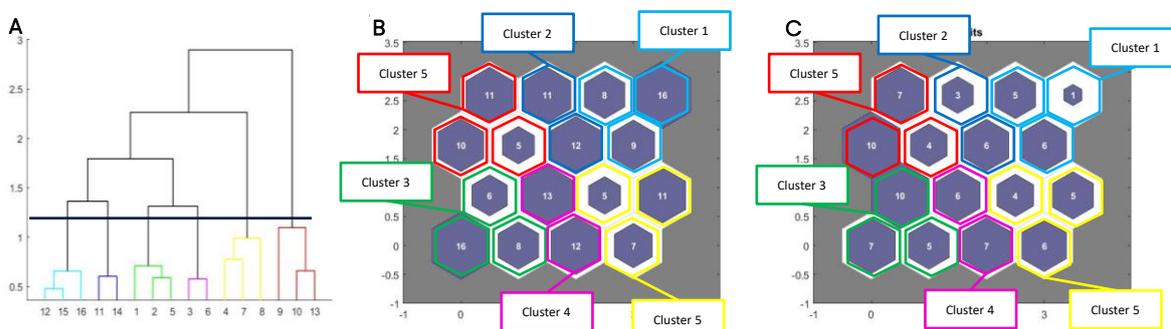


Figura 56: dendrogramma dei pesi della rete SOM 4x4 allenata con le lesioni del Constr_2 (107 features) con il relativo taglio effettuato per clusterizzare le lesioni. (Figura A) e mappe con l'organizzazione della rete SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) ottenute valutando la rete su Constr_2 (Figura B) e Valid_2 (Figura C).

Neuroni	Constr_1			Valid_1		
	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S
3, 6	5	17	3	7	5	1
1, 2, 5	14	15	1	6	15	1
11, 14	14	9	0	4	4	1
12, 15, 16	18	14	1	4	8	0
4, 7, 8	9	13	1	6	8	1
9, 10, 13	15	11	0	12	9	0

Tabella 23: tabella con il numero di lesioni, relativo a ciascuna classe, appartenenti a ciascun cluster ottenuto valutando la rete 4x4 su Constr_2 e Valid_2.

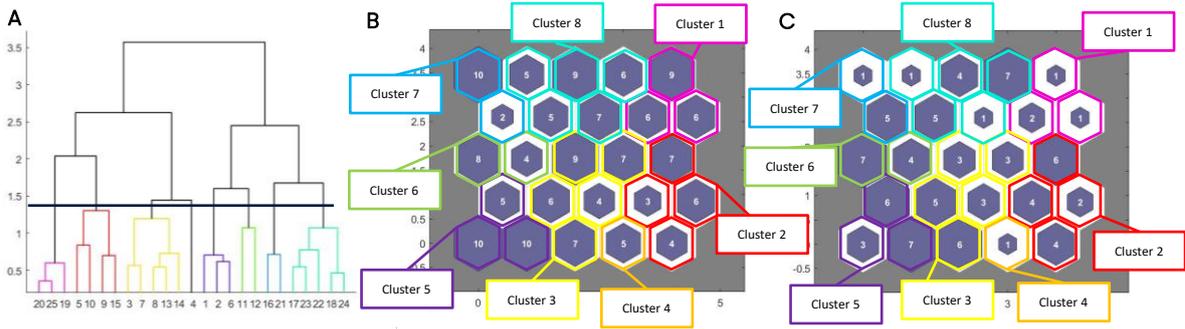


Figura 57: dendrogramma dei pesi della rete SOM 5x5 allenata con le lesioni del Constr_2 (107 features) con il relativo taglio effettuato per clusterizzare le lesioni. (Figura A) e mappe con l'organizzazione della rete SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) ottenute valutando la rete su Constr_2 (Figura B) e Valid_2 (Figura C).

Neuroni	Constr_2			Valid_2		
	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S
4	0	4	1	0	1	0
3, 7, 8, 13, 14	8	22	3	12	7	1
1, 2, 6	13	11	1	4	11	1
11, 12	7	5	0	5	6	0
16, 21	6	6	0	2	4	0
17, 18, 22, 23, 24	21	11	0	10	7	1
19, 20, 25	12	9	0	0	4	0
5, 9, 10, 15	8	11	1	6	9	1

Tabella 24: tabella con il numero di lesioni, relativo a ciascuna classe, appartenenti a ciascun cluster ottenuto valutando la rete 5x5 su Constr_2 e Valid_2.

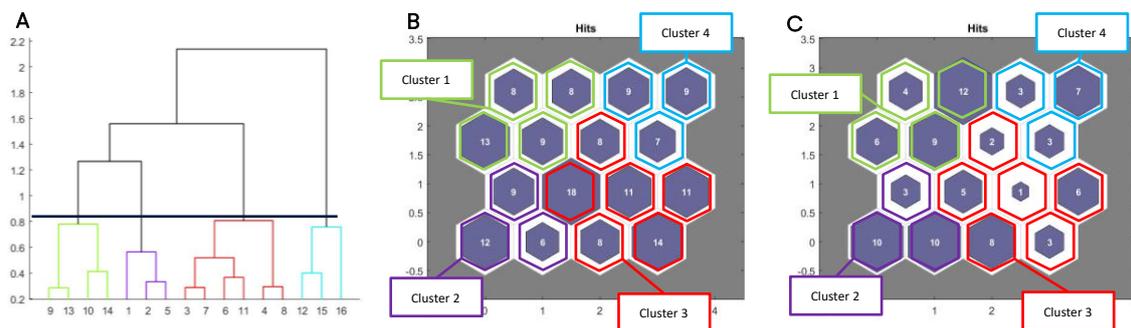


Figura 58: dendrogramma dei pesi della rete SOM 4x4 allenata con le lesioni del Constr_2 (42 features) con il relativo taglio effettuato per clusterizzare le lesioni. (Figura A) e mappe con l'organizzazione della rete SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) ottenute valutando la rete su Constr_2 (Figura B) e Valid_2 (Figura C).

Neuroni	Constr_2			Valid_2		
	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S
1, 2, 5	15	12	0	9	13	1
9, 10, 13, 14	28	10	0	14	16	1
3, 4, 6, 7, 8, 11	24	41	5	10	14	1
12, 15, 16	8	16	1	6	6	1

Tabella 25: tabella con il numero di lesioni, relativo a ciascuna classe, appartenenti a ciascun cluster ottenuto valutando la rete 4x4 su Constr_2 e Valid_2.

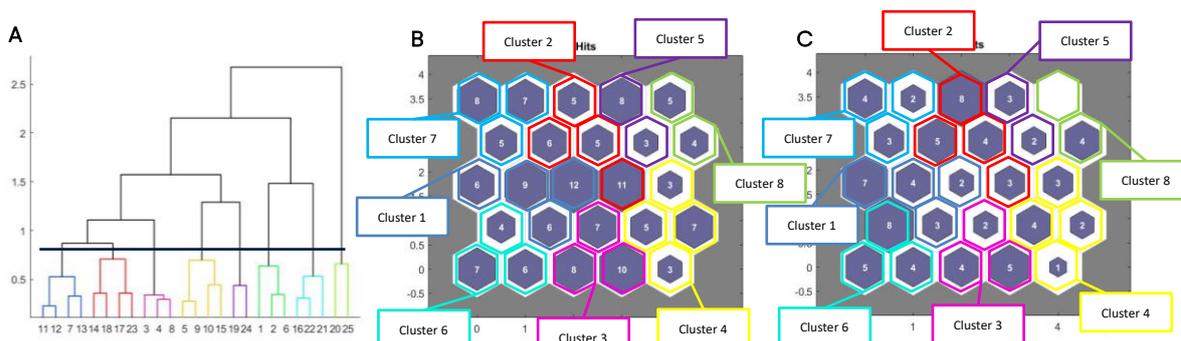


Figura 59: dendrogramma dei pesi della rete SOM 5x5 allenata con le lesioni del Constr_2 (42 features) con il relativo taglio effettuato per clusterizzare le lesioni. (Figura A) e mappe con l'organizzazione della rete SOM e il numero di vittorie di ciascun neurone (il primo neurone è in basso a sinistra) ottenute valutando la rete su Constr_2 (Figura B) e Valid_2 (Figura C).

Neuroni	Constr_2			Valid_2		
	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S	N° lesioni cl R-	N° lesioni cl R+	N° lesioni cl S
7, 11, 12, 13,	18	15	0	6	9	1
14, 17, 18, 23	12	13	2	9	10	1
3, 4, 8	11	12	2	3	8	0
19, 24	5	6	0	3	2	0
5, 9, 10, 15	1	16	1	5	4	1
16, 21, 22	14	6	0	4	4	1
1, 2, 6	11	6	0	8	9	0
20, 25	3	5	1	1	3	0

Tabella 26: tabella con il numero di lesioni, relativo a ciascuna classe, appartenenti a ciascun cluster ottenuto valutando la rete 5x5 su Constr_2 e Valid_2.

Con l'utilizzo del Construction Set costituito da pazienti provenienti da entrambi i centri è stata osservata una maggiore omogeneità nella distribuzione delle vittorie nel momento in cui si valuta la reti SOM sul Validation Set. Con entrambi i Construction Set, tuttavia, è stato osservato che uno stesso cluster era costituito da neuroni che risultavano essere «vincitori» per lesioni appartenenti a classi diverse. Un cluster che riconosceva lesioni per la maggior parte appartenenti ad una classe, inoltre, riconosceva spesso lesioni appartenenti ad un'altra classe sul Validation Set.

Queste ulteriori prove confermano la complessità nel predire la risposta alla terapia delle lesioni sfruttando le features estratte dalle TC al baseline.

Discussioni e sviluppi futuri

In questo lavoro di tesi sono stati costruiti diversi classificatori con lo scopo di predire la risposta al primo ciclo di chemioterapia di singole metastasi epatiche da CRC in modo da permettere ai clinici di valutare eventuali trattamenti alternativi evitando ai pazienti terapie inutili.

Il lavoro svolto presenta alcuni vantaggi rispetto ad altri studi presenti in letteratura. Primo fra tutti, nello studio sono stati sviluppati algoritmi di machine learning per la predizione della risposta alla terapia di singole metastasi e non della lesione di dimensione maggiore come avviene in molti altri studi [4, 9, 11]. È stata inoltre prevista la presenza di un dataset di Validazione, costituito da lesioni non utilizzate per la fase di costruzione dei classificatori, per valutare le prestazioni dei modelli su dati “nuovi” simulando l’utilizzo del sistema in condizioni reali.

Le prestazioni ottenute con i diversi classificatori, tuttavia, non risultano essere sufficientemente buone per poter utilizzare tali modelli nella pratica clinica in quanto si osserva un calo delle prestazioni dei modelli su Test Set e Validation Set rispetto alle prestazioni ottenute sul Training Set.

Con tutti i modelli di RF è stato osservato il fenomeno di overfitting poiché le prestazioni erano pari al 100% di corretti classificati sul dataset di Training ma diminuivano attestandosi in media attorno a 52% (45/86), 49% (32/65), 75% (9/12) rispettivamente sulle classi R+, S ed R- sul Test Set e 44% (26/59), 38% (19/50) e 41% (14/34) rispettivamente sulle classi R+, S ed R- sul Validation Set.

Questa tendenza è stata osservata, seppur in maniera minore, anche con gli altri modelli allenati. Le prestazioni raggiunte con il miglior modello di DT (allenato con un subset di features individuato effettuando feature selection basata sulla correlazione) si attestavano intorno all’80% di corretti classificati sulle 3 classi sul Training Set ma si riducevano al 47% (40/85), 41% (27/65) e 64% (7/11) sul Test Set e a 39% (23/59), 34% (17/50), 50% (17/34) sul Validation Set rispettivamente sulle classi R+, S ed R-.

Il modello di classificatore bayesiano più performante, allenato anch’esso con un subset di features individuato effettuando feature selection basata sulla correlazione, forniva prestazioni che si attestavano all’80% (24-25/30) di corretti classificati sulle classi R+ ed R- e sono pari al 60% (18/30) sulla classe S sul Training Set ma si riducevano al 56% (48/85), 37% (24/65) e 73% (8/11) sul Test Set e a 42% (25/59), 26% (13/50), 47% (16/34) sul Validation Set rispettivamente sulle classi R+, S ed R-. Con il modello di SVM migliore (SVM gaussiano con C pari a 50 e allenato con un subset di variabili individuato effettuando feature selection basata sulla correlazione) sono state ottenute prestazioni pari a 58% (18/31), 77% (23/30) e 77% (24/31) rispettivamente sulle classi R+, S ed R- sul Training Set, pari a 54% (46/85), 41% (27/65) e 64% (7/11) rispettivamente sulle classi R+, S ed R- sul Test Set e pari a 37% (22/59), 52% (26/50), 41% (14/34) rispettivamente sulle classi R+, S ed R- sul Validation Set.

I modelli, quindi, non risultano essere accurati, soprattutto nel predire quali lesioni non risponderanno alla chemioterapia.

Studi clinici sulla previsione della risposta alla terapia di metastasi epatiche presenti in letteratura hanno ottenuto risultati sub-ottimi, con valori di AUC (Area Under The Curve) compresi tra 0,60 e 0,79 nei dataset di validazione [4, 12 e 34], che però risultano essere migliori rispetto a quelli ottenuti nel lavoro di tesi. Occorre tuttavia sottolineare che in letteratura non sono stati trovati lavori in cui si affrontava un problema a 3 classi, come in questo lavoro di tesi, ma solo studi che trattavano il problema binario (R+/ R-). Inoltre, essendo lo scopo della radiomica quello di identificare pazienti con lesioni R- nei casi in cui la maggior parte delle lesioni sia “responder”, allora i risultati sono comunque un buon punto di partenza. Le lesioni correttamente identificate come R- dai classificatori sarebbero sottoposte a biopsia, esame che potrebbe suggerire la somministrazione di trattamenti alternativi o l'ablazione delle metastasi.

I risultati ottenuti sono quindi un buon punto di partenza per successive prove e approfondimenti.

La feature selection basata su correlazione di Spearman tra features e classe, seppur sia un metodo di ranking che non considera la relazione tra variabili, ha dimostrato come le features avessero un basso coefficiente di correlazione con la classe. Applicando l'algoritmo QRA, è stato inoltre osservato che i subset di features ottenuti contenevano anche features poco correlate con la classe. La risposta alla terapia e le features estratte potrebbero non essere linearmente dipendenti e questo dimostra la complessità del problema e la necessità di risolverlo per mezzo di metodi non lineari.

La prova effettuata che utilizza gli algoritmi genetici come metodo di feature selection, seppur promettente in quanto basata su un metodo che seleziona il subset ottimale di features considerando anche le relazioni tra variabili, non ha portato a buoni risultati visto l'ampio spazio delle soluzioni da esplorare e vista la pesantezza computazionale del metodo che non ha consentito di aumentare la dimensione della popolazione di soluzioni iniziale.

Un possibile modo per migliorare i risultati ottenuti potrebbe essere quello di indagare nuovi metodi di feature selection presenti in letteratura come altri algoritmi meta-euristici basati su popolazione (evolutionary computing) [27 e 28] e metodi basati sull'interaction information [29].

L'applicazione di nuovi metodi di feature selection unita ad un dataset di dimensioni maggiori in termini di numero di lesioni potrebbero quindi contribuire al miglioramento dei risultati di classificazione.

Un altro approccio che merita di essere indagato è relativo alla possibilità di predire la risposta a lungo termine delle lesioni (capitolo [Riconcettualizzazione del problema: data mining](#)) utilizzando le features estratte dalle TC al baseline. Si potrebbero quindi valutare le prestazioni dei modelli già implementati per la predizione della risposta alla terapia al primo ciclo di terapia ma utilizzando come classe target la risposta a lungo termine delle lesioni.

Ad oggi in letteratura si stanno diffondendo numerosi studi che superano la “radiomica tradizionale”, che analizza immagini mediche ad un solo istante temporale, per utilizzare invece la deltaradiomica, che considera la variazione delle features radiomiche durante o dopo il trattamento [30]. Questo approccio presenta un elevato costo in termini di tempo dovuto dalla necessità di avere le segmentazioni manuali delle lesioni in più istanti temporali al fine di poter estrarre le features dalle immagini e ottenere il dataset per l'allenamento dei

sistemi di machine learning. Inoltre, questo nuovo approccio cambierebbe l'obiettivo per cui costruire i modelli di machine learning per la predizione della risposta alla terapia in quanto i pazienti dovrebbero comunque essere sottoposti ad almeno un ciclo di trattamento.

La deltaradiomica, non trascurando il cambiamento del tumore durante il trattamento, ha tuttavia dimostrato risultati promettenti in studi precedenti. Chang et al. hanno osservato che le delta-features estratte da immagini di risonanza pre e post trattamento possono potenzialmente fornire una migliore valutazione del trattamento rispetto alle caratteristiche a tempo singolo per prevedere l'overall survival (OS) di pazienti con gliomi maligni ricorrenti. È stato inoltre osservato come la valutazione del trattamento sia sostanzialmente influenzata dal time-point scelto per il calcolo delle delta-features [31]. Fave et al. hanno osservato come le delta-features estratte da immagini TC pre e post radioterapia possano essere prognostiche per la sopravvivenza globale di pazienti con non-small cell lung cancer [32]. Nasief et al. hanno inoltre osservato come le variazioni nei parametri di imaging date dall'utilizzo di 2 diversi scanner TC non abbiano influenzato l'analisi delle delta-features, essendo queste dei valori relativi [33].

La deltaradiomica potrebbe quindi essere un possibile approccio da utilizzare per la predizione della risposta alla terapia delle metastasi epatiche qualora la "radiomica tradizionale" non portasse a miglioramenti.

In conclusione, il lavoro di tesi svolto risulta essere una fase di un progetto promettente che merita di essere portato avanti con ulteriori prove e valutazioni.

Appendice

GA	Valore della fitness	N° di features selezionate	Features selezionate	Parametri del modello selezionati
Basato su classificatore bayesiano	0.690	24	<p>Shape (Elongation, Flatness, Maximum 2D Diameter Row, Mesh Volum, Minor Axis Length, Surface Volume Ratio); Firstorder (Mean); Glc (Autocorrelation, Cluster Prominence, Imc1, Joint Entropy, MCC); Gldm (Dependence Variance, Small Dependence Low Gray Level Emphasis); Glr (Gray Level Non Uniformity, Long Run Low Gray Level Emphasis, Low Gray Level Run Emphasis, Run Entropy, Run Length Non Uniformity); Glszm (Gray Level Variance, Large Area High Gray Level Emphasis, Low Gray Level Zone Emphasis, Small Area Low Gray Level Emphasis); Ngtdm (Busyness)</p>	/
Basato su SVM	0.680	46	<p>Shape (Flatness, Least Axis Length, Maximum 3D Diameter, Surface Volume Ratio); Firstorder (90 Percentile, Energy, Interquartile Range, Mean, Range); Glc (Cluster Prominence, Cluster Shade, Difference Entropy, Difference Variance, Id, Joint Energy, MCC, Sum Average, Sum Entropy); Gldm (Dependence Variance, Gray Level Variance, High Gray Level Emphasis, Small Dependence Emphasis, Small Dependence High Gray Level Emphasis); Glr (Gray Level Non Uniformity, Gray Level Non Uniformity Normalized, Long Run High Gray Level Emphasis, Low Gray Level Run Emphasis, Run Entropy, Run Length Non Uniformity, Run Length Non Uniformity Normalized, Run Variance, Short Run Low Gray Level Emphasis); Glszm (Gray Level Variance, High Gray Level Zone 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 Variance); Ngtdm (Busyness, Coarseness, Contrast, Strength)</p>	C:50 Kernel: lineare

Appendice

Basato su SVM	0.689	43	<p>Shape (Flatness, Major Axis Length, Maximum 2D Diameter Column, Maximum 2D Diameter Row, Maximum 3D Diameter, Surface Area, Voxel Volume); First order (90 Percentile, Energy, Interquartile Range, Maximum, Mean, Minimum, Range, Robust Mean Absolute Deviation, Root Mean Square, Total Energy, Uniformity, Variance); GlcM (Cluster Prominence, Cluster Shade, Contrast, Difference Average, Difference Entropy, Idmn, Imc1, MCC, Maximum Probability, Sum Entropy, Sum Squares); Gldm (Dependence Variance, Gray Level Non Uniformity, 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); GlrIm (Gray Level Variance, Long Run Low Gray Level Emphasis); Glszm (Large Area Emphasis, Low Gray Level Zone Emphasis); Ngtdm (Contrast)</p>	C:100 Kernel: lineare
------------------	-------	----	---	-----------------------------

Tabella 27: Caratteristiche delle soluzioni individuate dal GA

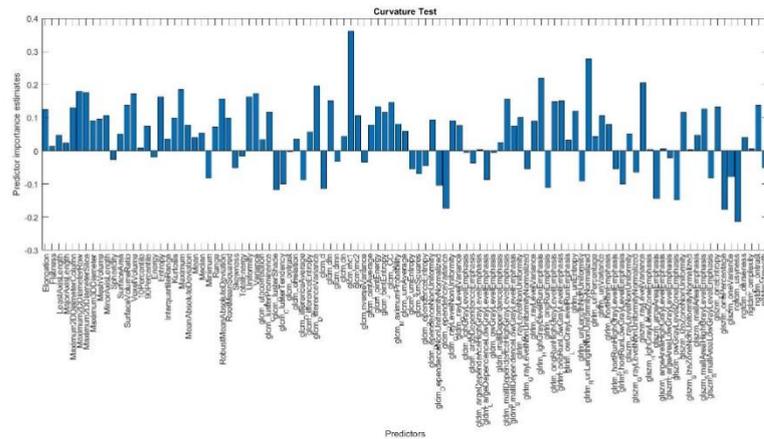


Figura 60: importanza di ciascuna delle 107 features ottenuta allenando un RF con le 107 features e NumPredictorToSample pari a 11 ($\sim\sqrt{107}$)

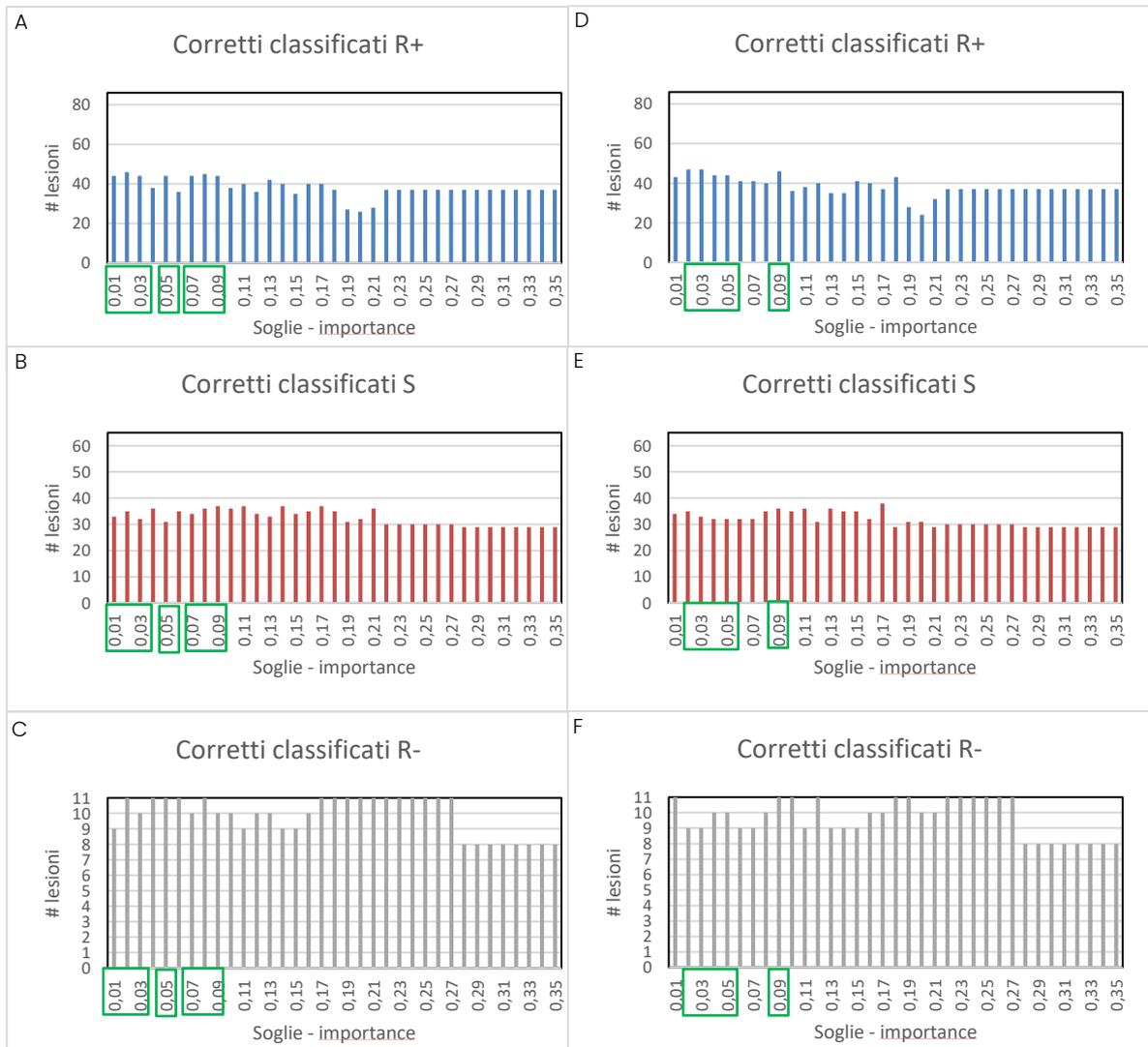


Figura 61: n° di corretti classificati ottenuti per ognuna delle 3 classi con i modelli di RF allenati con un n° di features scelto impostando diverse soglie sull'importanza (Figure A-B-C) e sull'importanza (Figure D-E-F) ottenuta con il modello allenato con 107 features e NumPredictorToSample pari a $\sqrt{n^{\circ} \text{ features}}$ (importanza varia da 0.01 a 0.35 a passi di 0.01). In verde sono indicati i modelli che fornivano buone e non sbilanciate numerosità di corretti classificati sulle 3 classi nel Test Set.

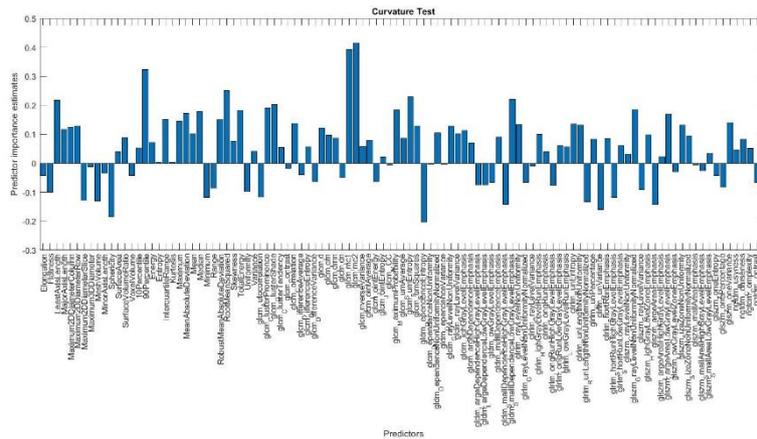


Figura 62: importanza di ciascuna delle 107 features ottenuta allenando un RF con le 107 features e $NumPredictorToSample$ pari a 54 (107/2)

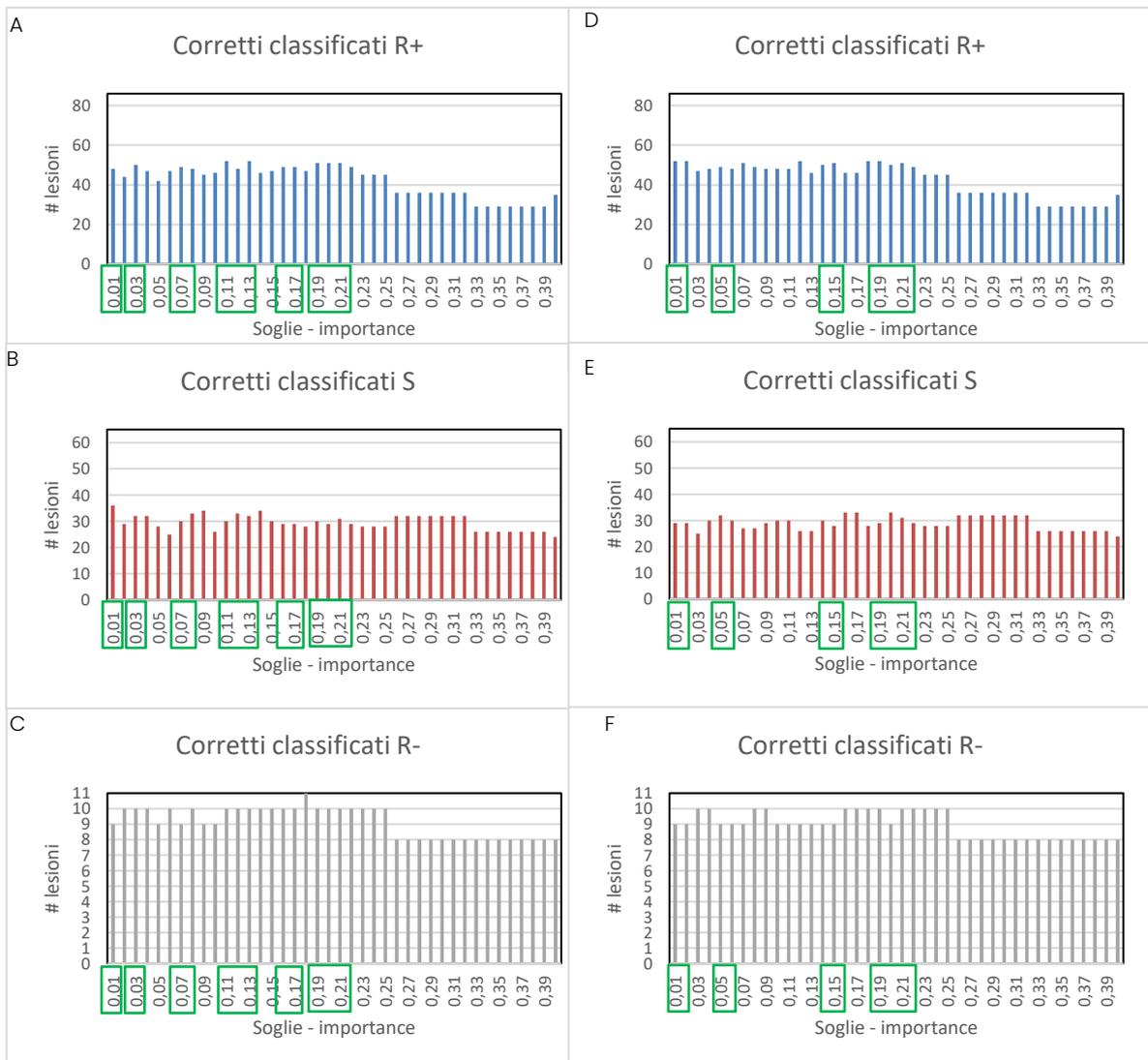


Figura 63: n° di corretti classificati ottenuti per ognuna delle 3 classi con i modelli di RF allenati con un n° di features scelto impostando diverse soglie sull'importanza (Figure A-B-C) e sull'importanza (Figure D-E-F) ottenuta con il modello allenato con 107 features e $NumPredictorToSample$ pari a $(n^{\circ} \text{ features}) / 2$ (soglia varia da 0.01 a 0.40 a passi di 0.01). In verde sono indicati i modelli che fornivano buone e non sbilanciate numerosità di corretti classificati sulle 3 classi nel Test Set.

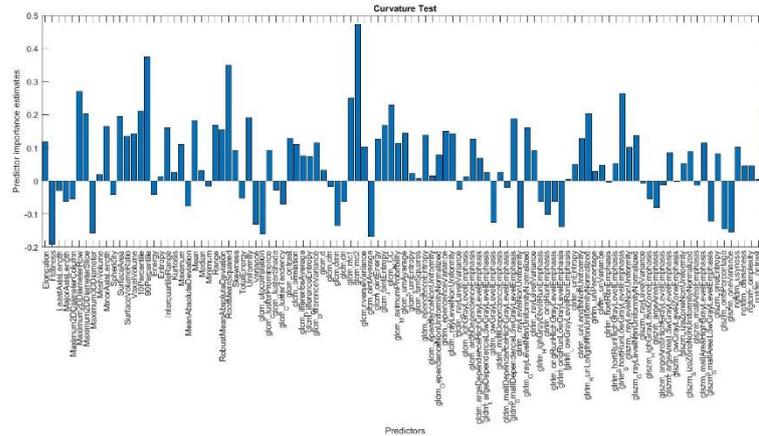


Figura 64: importanza di ciascuna delle 107 features ottenuta allenando un RF con le 107 features e NumPredictorToSample pari a $80^{3/4} \cdot 107$

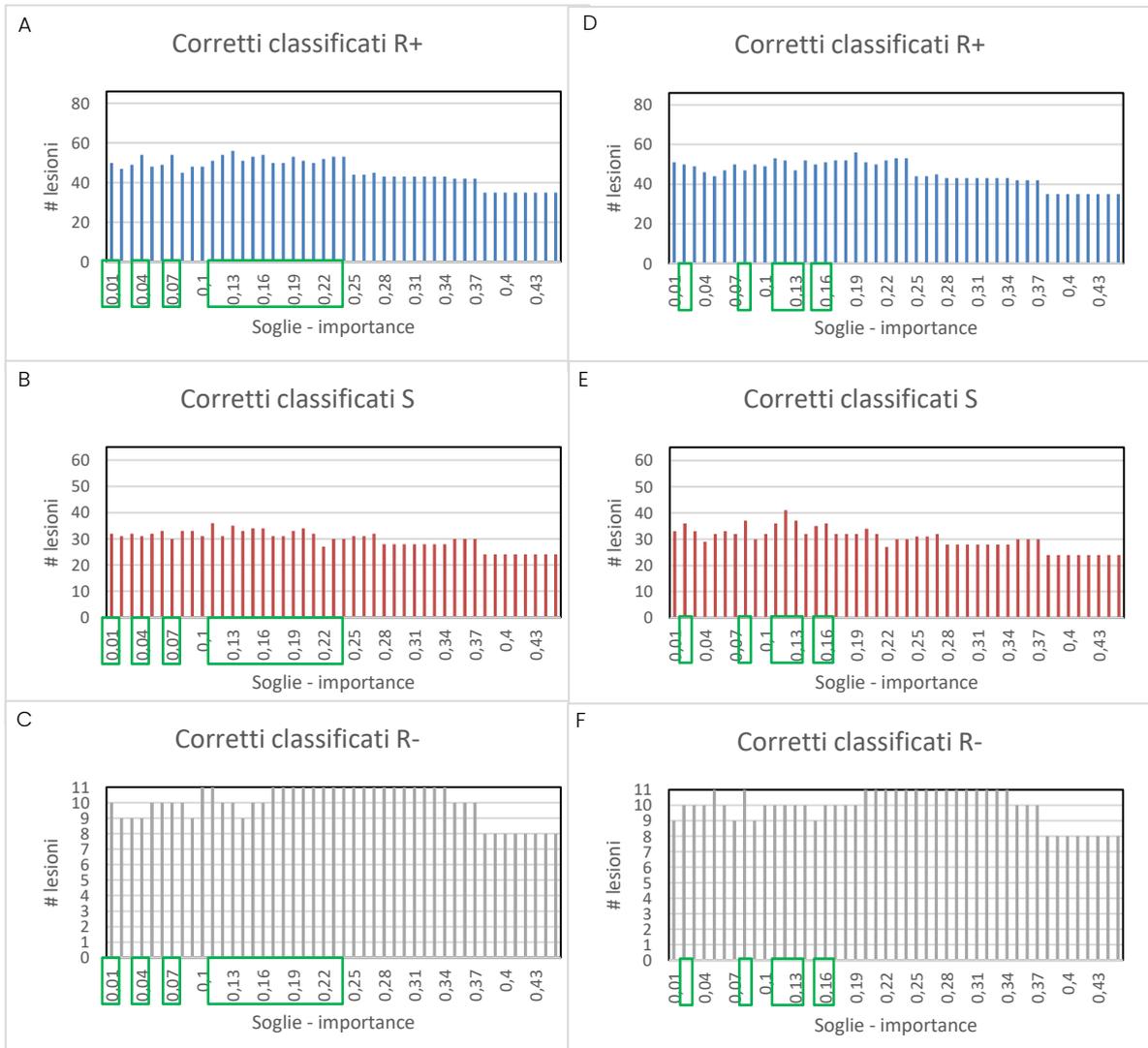
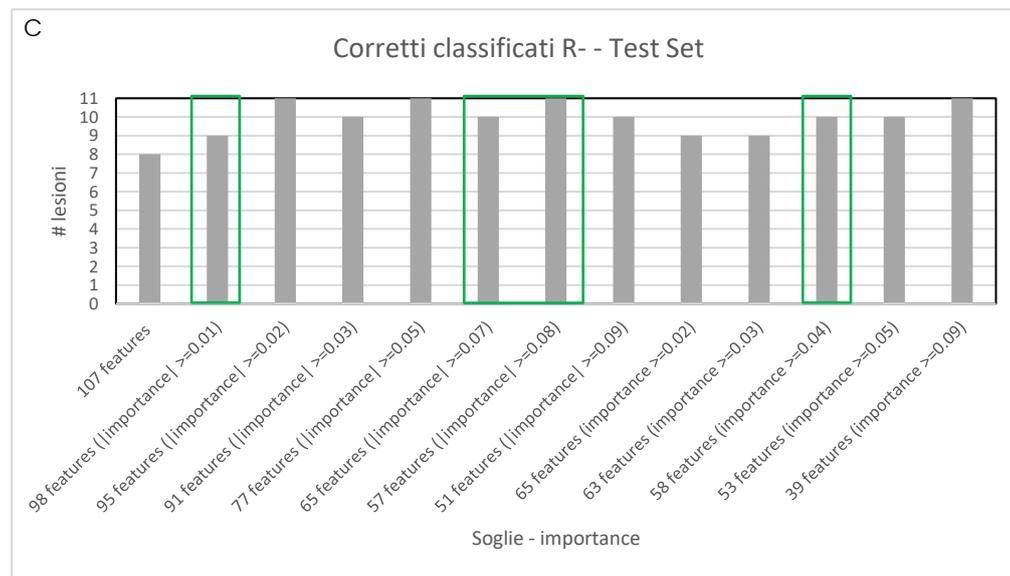
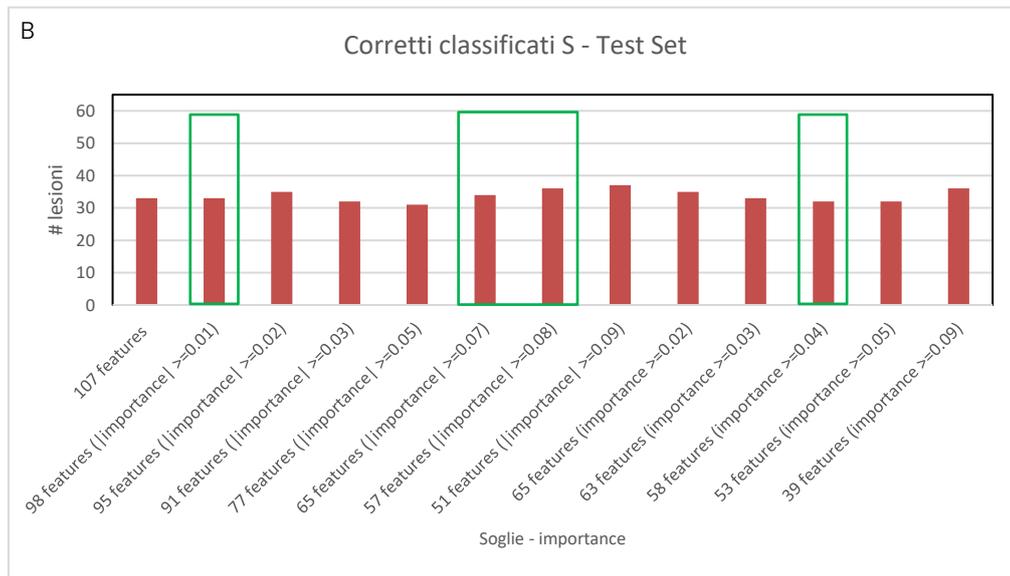
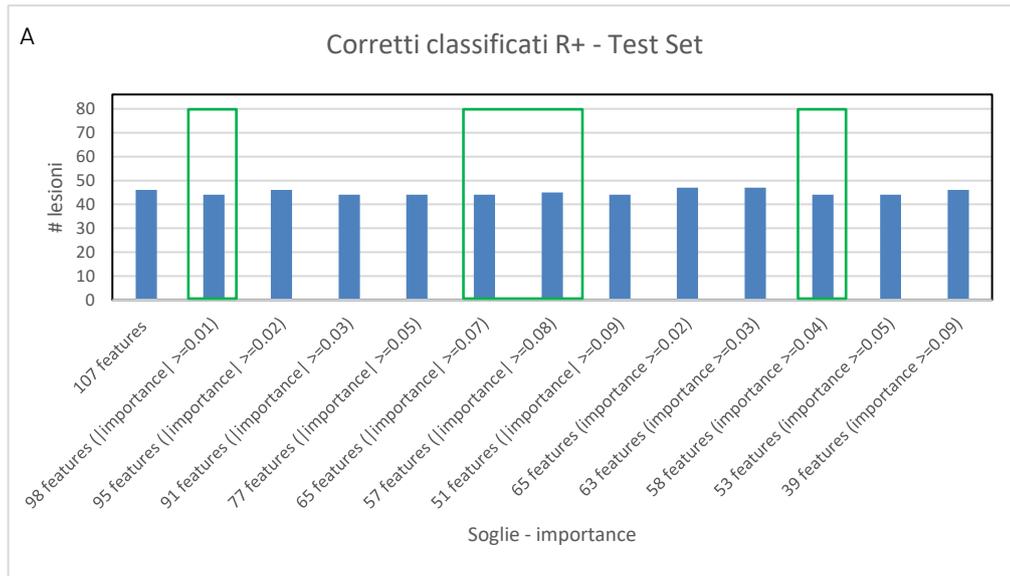
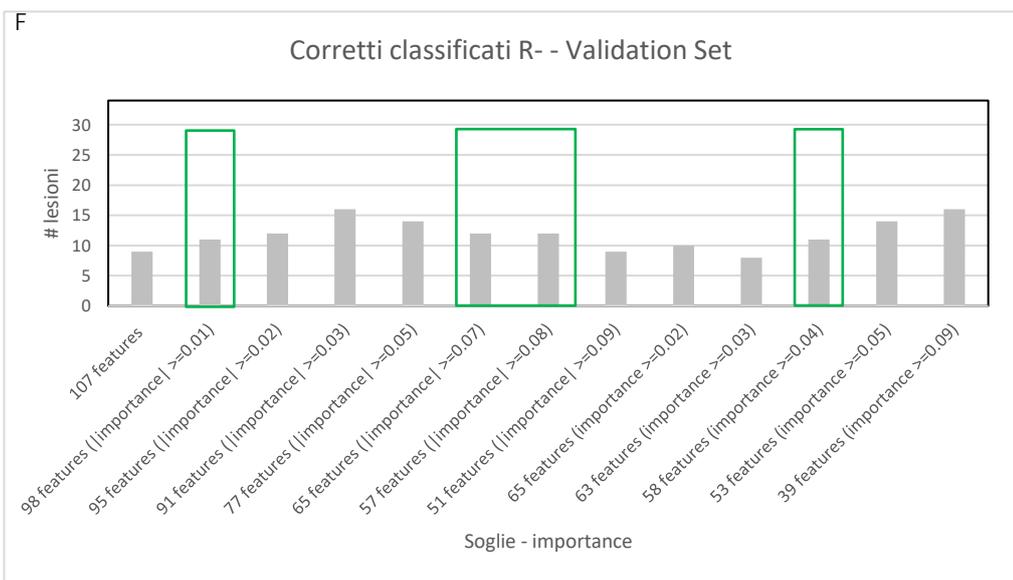
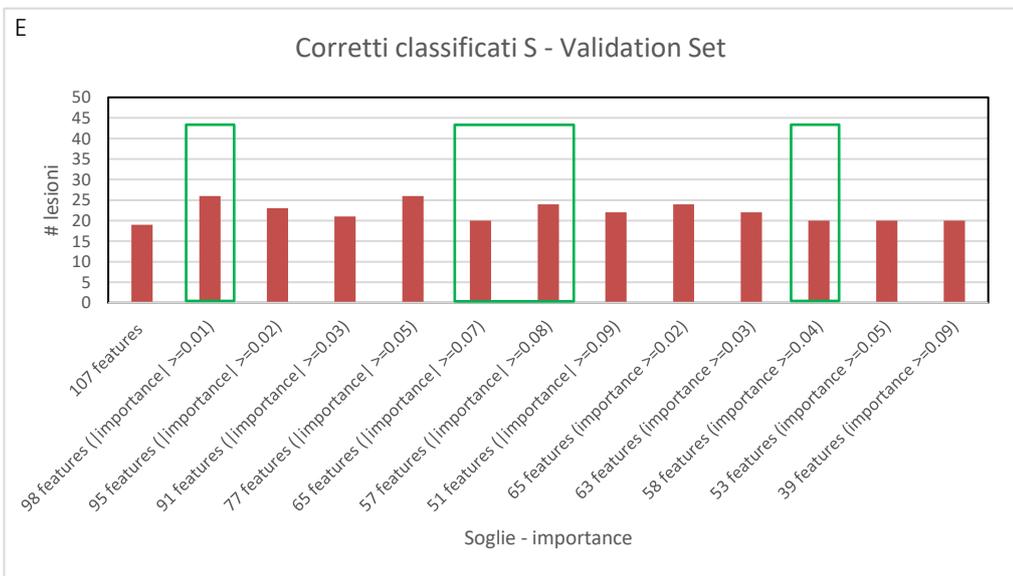
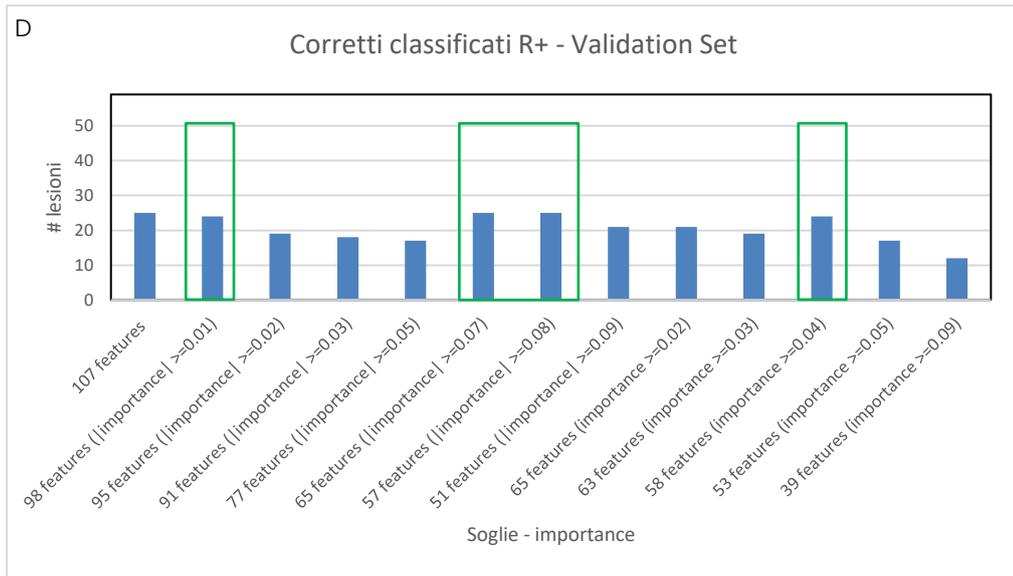
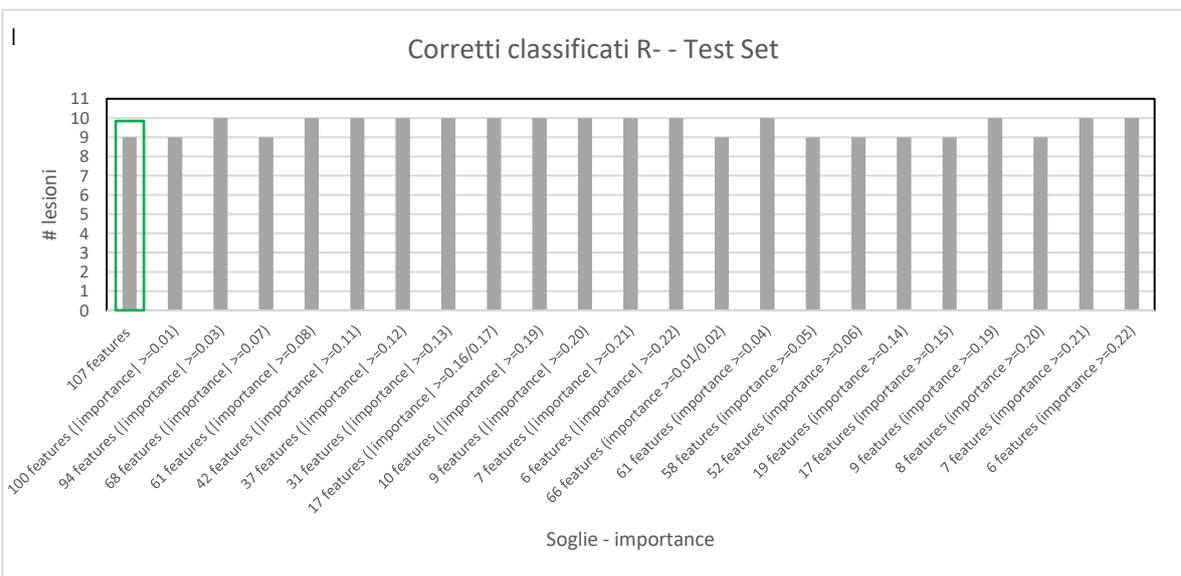
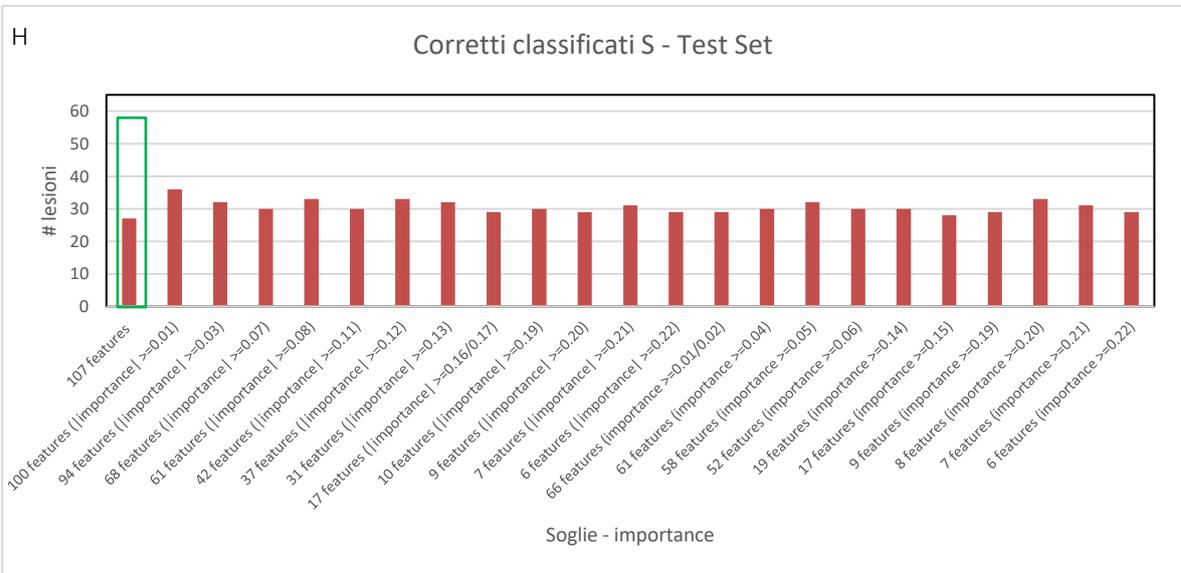
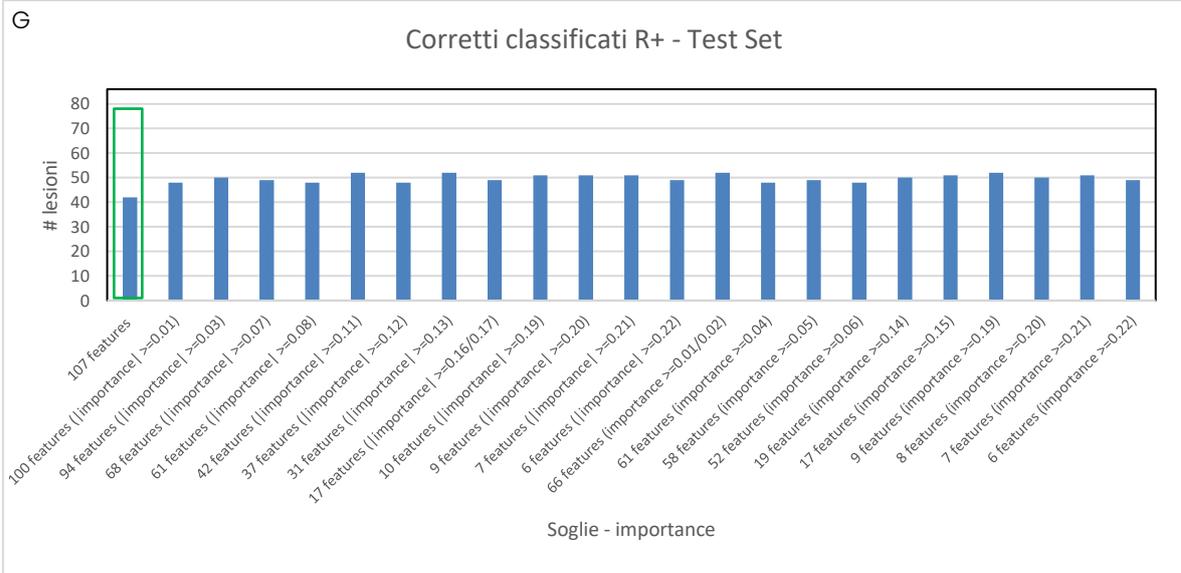


Figura 65: n° di corretti classificati ottenuti per ognuna delle 3 classi con i modelli di RF allenati con un n° di features scelto impostando diverse soglie sull'importance (Figure A-B-C) e sull'importance (Figure D-E-F) ottenuta sul modello allenato con 107 features e NumPredictorToSample pari a $3/4 * n° features$ (soglia varia da 0.01 a 0.45 a passi di 0.01). In verde sono indicati i modelli che fornivano buone e non sbilanciate numerosità di corretti classificati sulle 3 classi nel Test Set

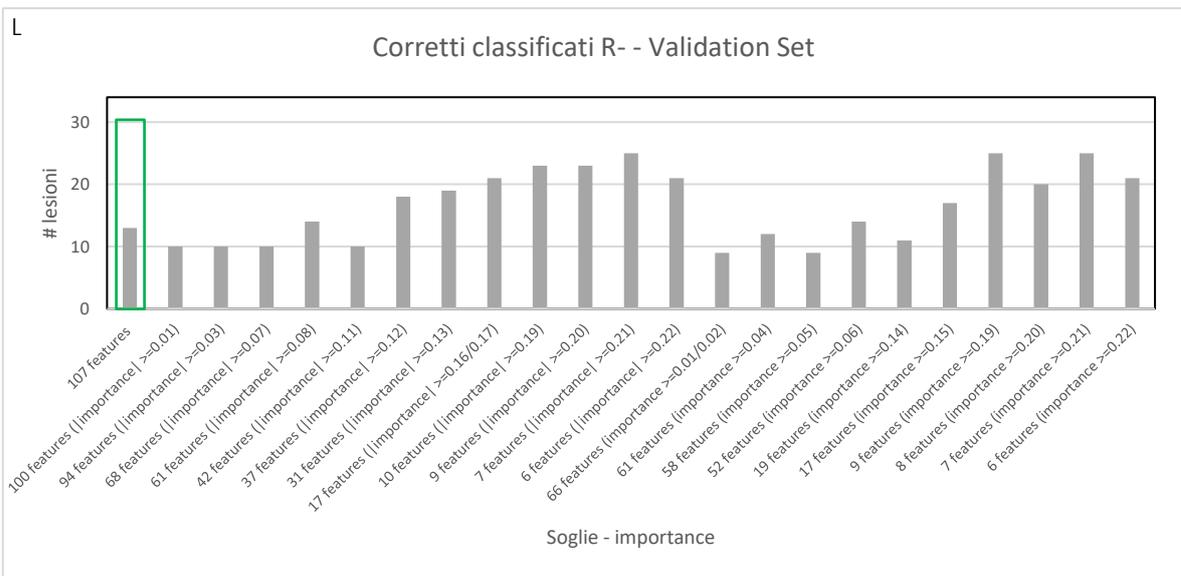
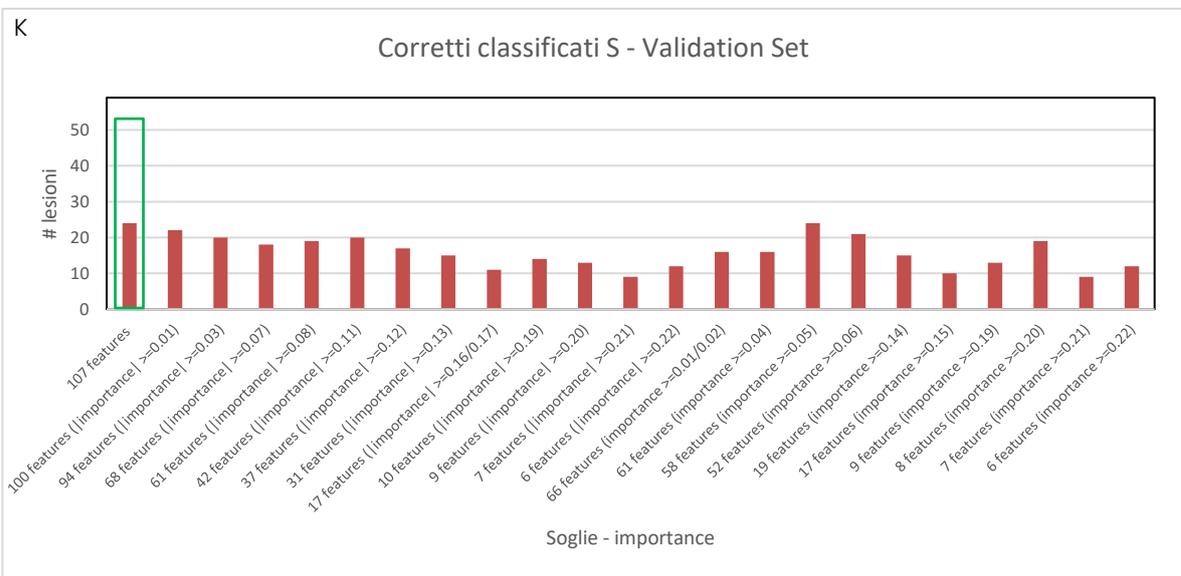
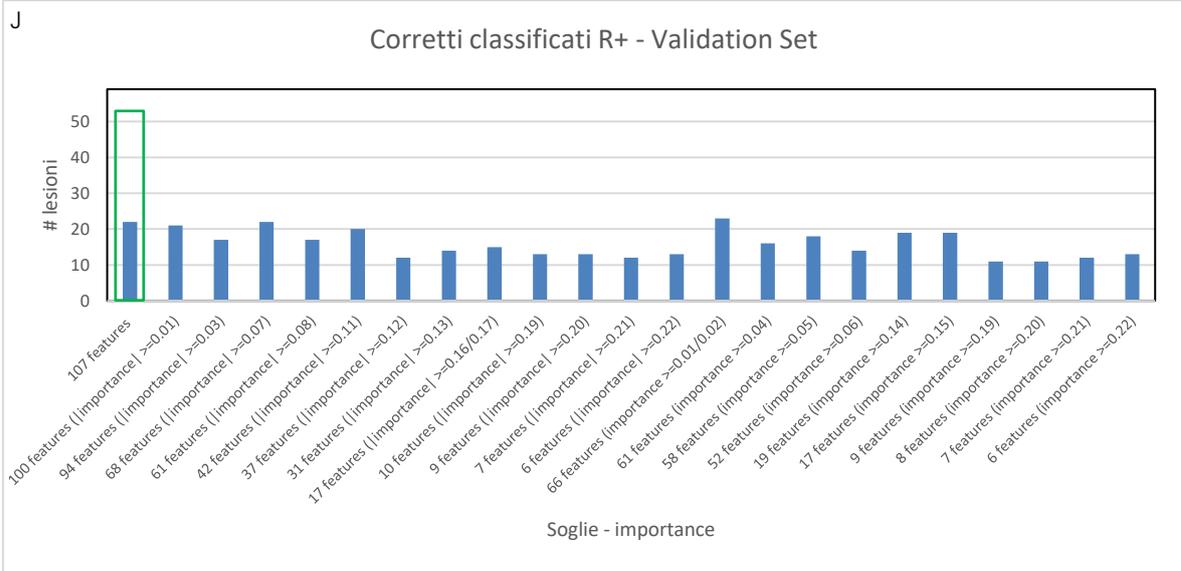
Appendice

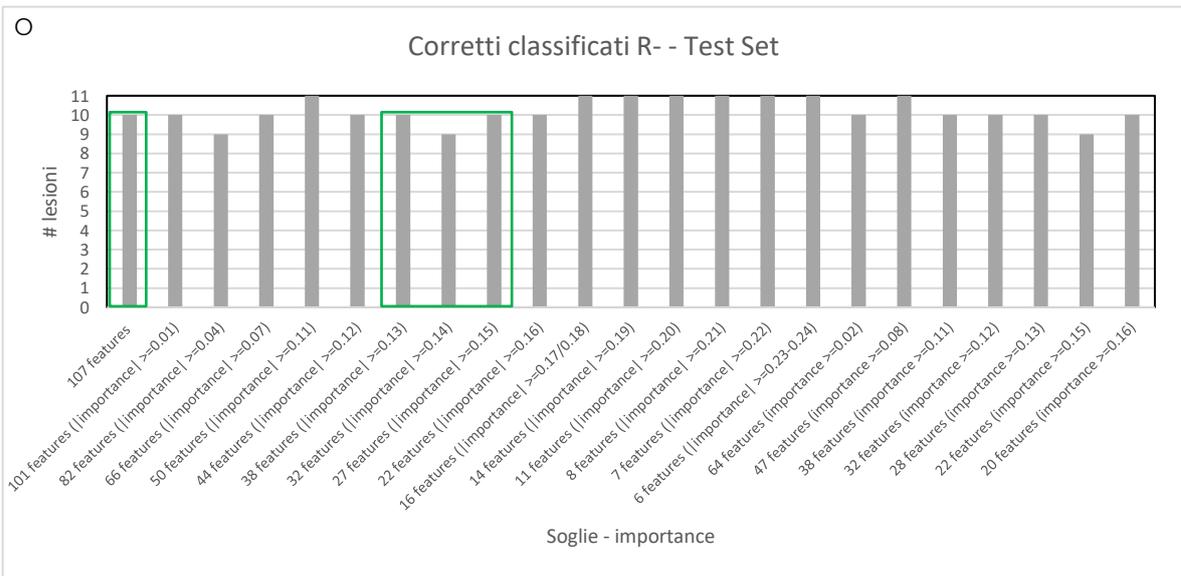
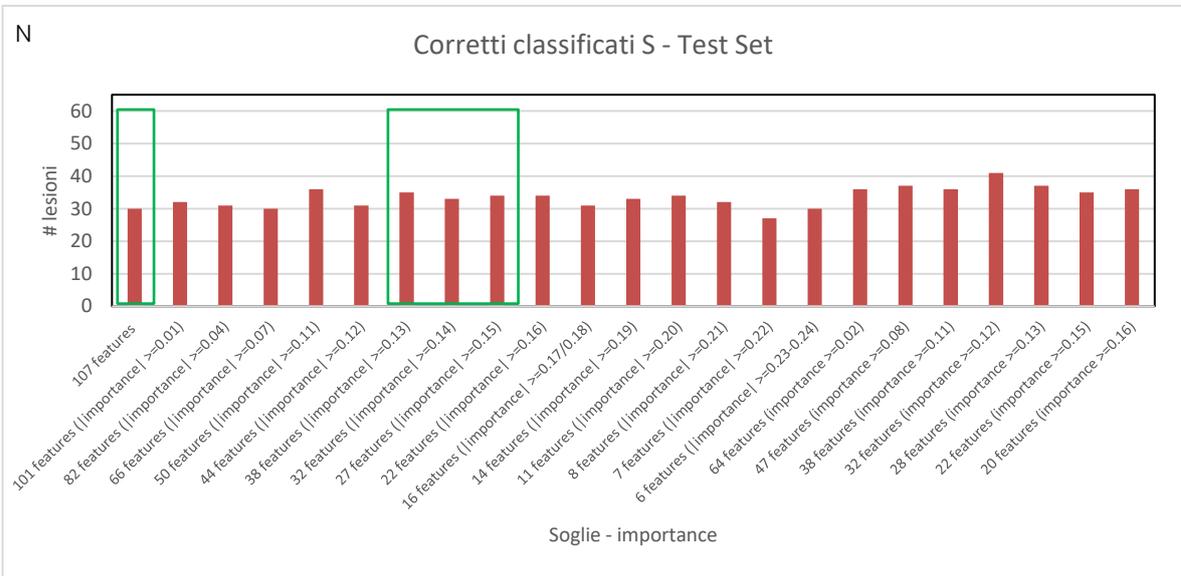
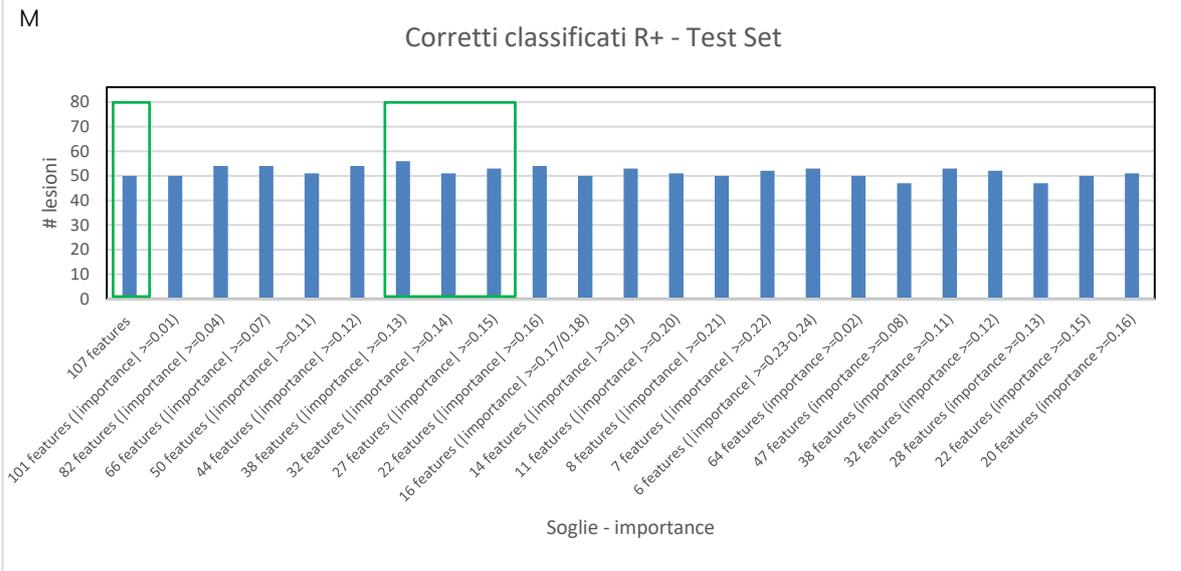






Appendice





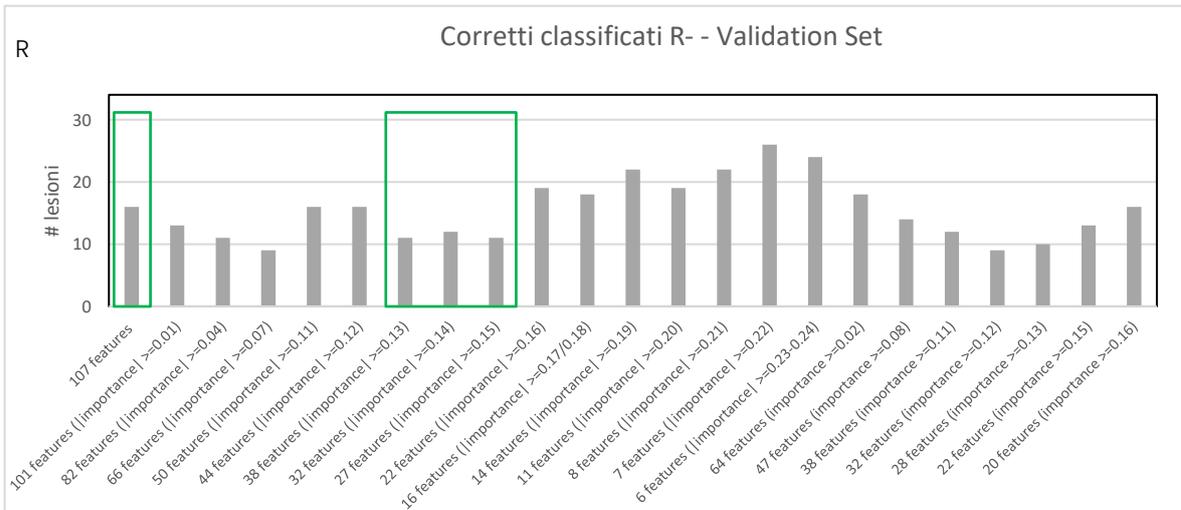
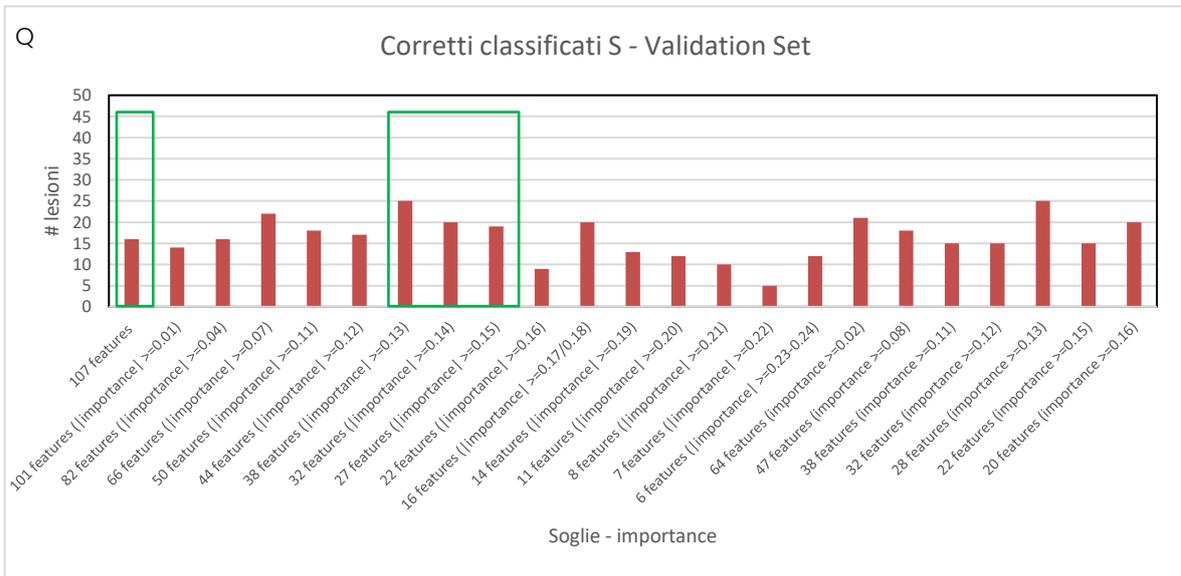
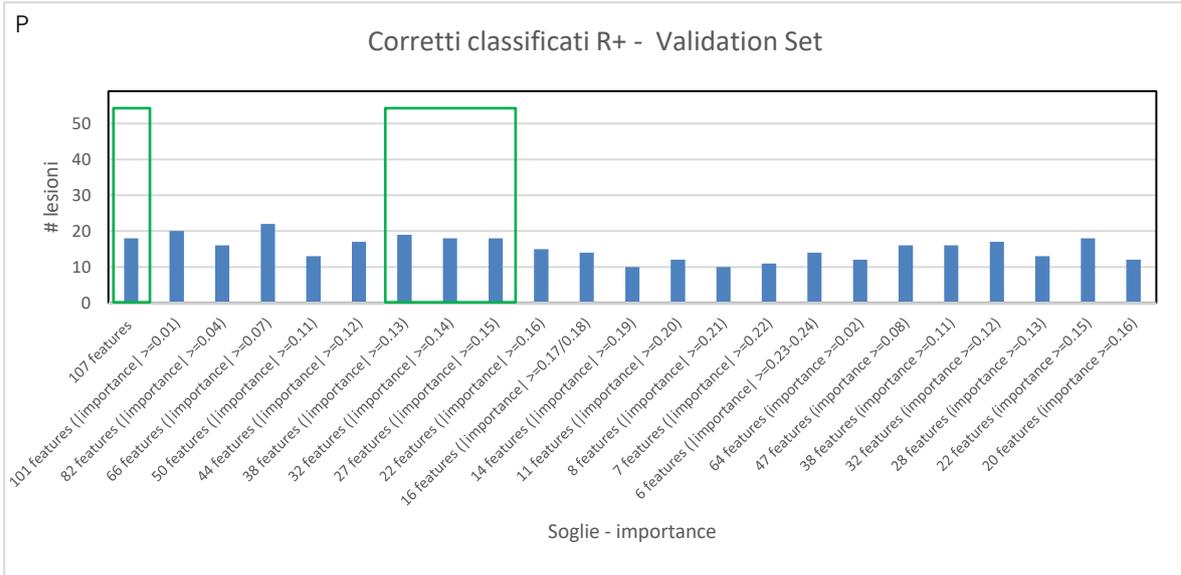


Figura 66: n° di corretti classificati ottenuti per ognuna delle 3 classi sul Test Set e Validation Set con i modelli migliori individuati per ogni NumPredictorsToSamples (Figure A, B, C, D, E, F: $\sqrt{n^\circ features}$; Figure G, H, I, J, K, L,: $n^\circ features/2$; Figure M, N, O, P, Q, R: $\frac{3}{4} * n^\circ features$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	16	5	2
	S	8	23	6
	R-	6	2	23

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	40	12	1
	S	27	34	0
	R-	19	19	10

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare ($C=1$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	22	3	3
	S	7	26	3
	R-	1	1	25

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	45	13	2
	S	26	36	0
	R-	15	16	9

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare ($C=10$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	26	2	3
	S	4	28	1
	R-	0	0	27

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	43	18	1
	S	26	29	0
	R-	17	18	10

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare ($C=50$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	28	2	2
	S	2	28	0
	R-	0	0	29

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	50	21	1
	S	20	27	1
	R-	16	17	9

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare ($C=100$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	0	0	0
	S	0	0	0
	R-	30	30	31

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	0	0	0
	S	0	0	0
	R-	86	65	11

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, $C=1$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	0	0	0
	S	0	0	0
	R-	30	30	31

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	0	0	0
	S	0	0	0
	R-	86	65	11

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, $C=10$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	5	0	1
	S	16	25	12
	R-	9	5	18

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	10	1	0
	S	60	45	5
	R-	16	19	6

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, C=50)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	10	0	1
	S	11	26	9
	R-	9	4	21

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	23	6	0
	S	43	40	2
	R-	20	19	9

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, C=100)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	0	0	0
	S	0	0	0
	R-	30	30	31

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	0	0	0
	S	2	0	0
	R-	84	65	11

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=1)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	10	4	1
	S	10	20	8
	R-	10	6	22

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	31	12	1
	S	37	33	1
	R-	18	20	9

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=10)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	15	4	2
	S	10	24	6
	R-	5	2	23

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	39	12	1
	S	23	35	0
	R-	24	18	10

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=50)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	21	3	2
	S	8	25	6
	R-	1	2	23

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	42	12	1
	S	24	34	0
	R-	20	19	10

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=100)

Figura 67: confusion matrix relative alle prestazioni su Training Set e Test Set dei diversi modelli di SVM allenati con 107 features per individuare le strutture di SVM più performanti da utilizzare per effettuare feature selection basata sulla correlazione

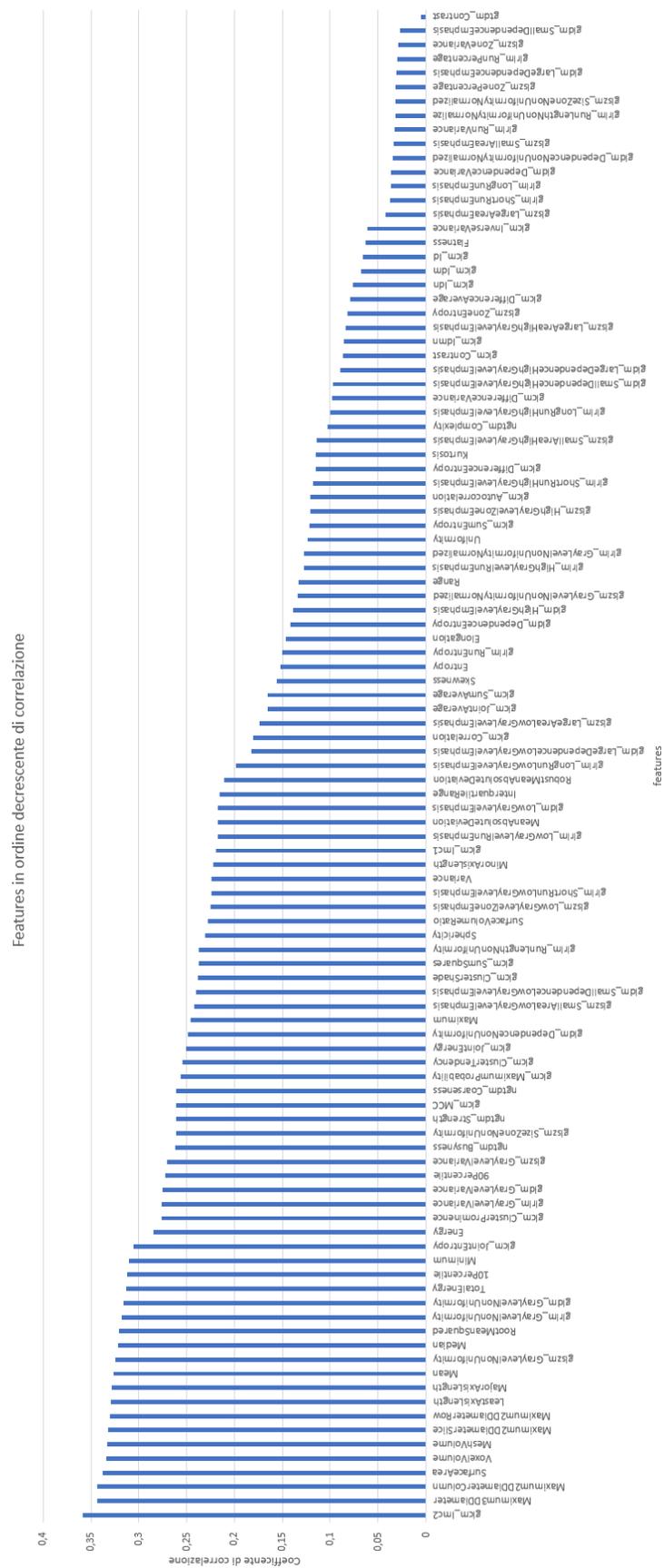


Figura 68: grafico relativo al coefficiente di correlazione di Spearman di ognuna delle 107 features con la classe nel Training Set.

Appendice

Feature	Nome Feature	# di valori distinti iniziali presenti nel Construction Set	# di insiemi individuati usando ChiMerge sul Construction Set
1	Elongation	253	19
2	Flatness	253	11
3	Least Axis Length	253	15
4	Major Axis Length	253	14
5	Maximum 2D Diameter Column	248	16
6	Maximum 2D Diameter Row	249	12
7	Maximum 2D Diameter Slice	249	21
8	Maximum 3D Diameter	253	5
9	Mesh Volume	253	6
10	Minor Axis Length	253	12
11	Sphericity	253	12
12	Surface Area	253	17
13	Surface Volume Ratio	253	19
14	Voxel Volume	252	10
15	10 Percentile	92	6
16	90 Percentile	95	9
17	Energy	253	9
18	Entropy	253	10
19	Interquartile Range	62	13
20	Kurtosis	253	17
21	Maximum	84	10
22	Mean Absolute Deviation	253	17
23	Mean	253	17
24	Median	81	6
25	Minimum	79	8
26	Range	80	9
27	Robust Mean Absolute Deviation	253	9
28	Root Mean Squared	253	13
29	Skewness	253	13
30	TotalEnergy	253	12
31	Uniformity	253	14
32	Variance	253	15
33	glcm Autocorrelation	253	8
34	glcm Cluster Prominence	253	20
35	glcm Cluster Shade	253	10
36	glcm Cluster Tendency	253	11
37	glcm Contrast	253	12
38	glcm Correlation	253	13
39	glcm Difference Average	253	10
40	glcm Difference Entropy	253	12
41	glcm Difference Variance	253	8
42	glcm Id	253	8
43	glcm Idm	253	12
44	glcm Idmn	253	11
45	glcm Idn	253	12
46	glcm Imc1	253	19
47	glcm Imc2	253	23
48	glcm Inverse Variance	253	11
49	glcm Joint Average	253	13
50	glcm Joint Energy	253	18
51	glcm Joint Entropy	253	23

Appendice

52	glcm MCC	253	20
53	glcm Maximum Probability	253	24
54	glcm Sum Average	253	13
55	glcm Sum Entropy	253	18
56	glcm Sum Squares	253	8
57	gldm Dependence Entropy	253	11
58	gldm Dependence Non Uniformity	253	8
59	gldm Dependence Non Uniformity Normalized	253	22
60	gldm Dependence Variance	253	17
61	gldm Gray Level Non Uniformity	253	10
62	gldm Gray Level Variance	253	9
63	gldm High Gray Level Emphasis	253	15
64	gldm Large Dependence Emphasis	253	14
65	gldm Large Dependence High Gray Level Emphasis	253	3
66	gldm Large Dependence Low Gray Level Emphasis	253	19
67	gldm Low Gray Level Emphasis	253	19
68	gldm Small Dependence Emphasis	253	9
69	gldm Small Dependence High Gray Level Emphasis	253	18
70	gldm Small Dependence Low Gray Level Emphasis	253	21
71	glrlm Gray Level Non Uniformity	253	7
72	glrlm Gray Level Non Uniformity Normalized	253	16
73	glrlm Gray Level Variance	253	14
74	glrlm High Gray Level Run Emphasis	253	18
75	glrlm Long Run Emphasis	253	18
76	glrlm Long Run High Gray Level Emphasis	253	17
77	glrlm Long Run Low Gray Level Emphasis	253	7
78	glrlm Low Gray Level Run Emphasis	253	10
79	glrlm Run Entropy	253	20
80	glrlm Run Length Non Uniformity	253	11
81	glrlm Run Length Non Uniformity Normalized	253	13
82	glrlm Run Percentage	253	17
83	glrlm Run Variance	253	17
84	glrlm Short Run Emphasis	253	13
85	glrlm Short Run High Gray Level Emphasis	253	21
86	glrlm Short Run Low Gray Level Emphasis	253	13

Appendice

87	glszm Gray Level Non Uniformity	253	7
88	glszm Gray Level Non Uniformity Normalized	253	10
89	glszm Gray Level Variance	253	12
90	glszm High Gray Level Zone Emphasis	253	13
91	glszm Large Area Emphasis	253	11
92	glszm Large Area High Gray Level Emphasis	253	23
93	glszm Large Area Low Gray Level Emphasis	253	6
94	glszm Low Gray Level Zone Emphasis	253	24
95	glszm Size Zone Non Uniformity	253	17
96	glszm Size Zone Non Uniformity Normalized	253	13
97	glszm Small Area Emphasis	253	14
98	glszm Small Area High Gray Level Emphasis	253	13
99	glszm Small Area Low Gray Level Emphasis	253	15
100	glszm Zone Entropy	253	11
101	glszm Zone Percentage	253	13
102	glszm Zone Variance	253	12
103	glszm Low Gray Level Zone Emphasis	253	12
104	ngtdm Busyness	253	15
105	ngtdm Coarseness	253	7
106	ngtdm Complexity	253	11
107	ngtdm Contrast	253	14

Tabella 28: Risultati della discretizzazione del Constraction Set con ChiMerge.

N° soluzione	N ° Features selezionate	Features
1	4	Maximum 2D Diameter Column, Mean Absolute Deviation, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
2	4	Flatness, glcm Maximum Probability, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis
3	4	Mean Absolute Deviation, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis, ngtdm Contrast
4	4	Mean Absolute Deviation, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis, ngtdm Complexity
5	4	Mean Absolute Deviation, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis, ngtdm Coarseness
6	4	Mean Absolute Deviation, glcm Maximum Probability, glszm Large Area High Gray Level Emphasis, glszm Low Gray Level Zone Emphasis

Appendice

7	4	Mean Absolute Deviation, glcm Maximum Probability, glszm High Gray Level Zone Emphasis, glszm Low Gray Level Zone Emphasis
8	4	Mean Absolute Deviation, glcm Maximum Probability, glrlm Short Run Low Gray Level Emphasis, glszm Low Gray Level Zone Emphasis
9	4	Mean Absolute Deviation, glcm Maximum Probability, glrlm Short Run High Gray Level Emphasis, glszm Low Gray Level Zone Emphasis
10	4	Mean Absolute Deviation, glcm Maximum Probability, glrlm Run Variance, glszm Low Gray Level Zone Emphasis
11	4	Mean Absolute Deviation, glcm Maximum Probability, glrlm Run Percentage, glszm Low Gray Level Zone Emphasis
12	4	Mean Absolute Deviation, glcm Maximum Probability, glrlm Run Entropy, glszm Low Gray Level Zone Emphasis
13	4	Mean Absolute Deviation, glcm Maximum Probability, glrlm Long Run High Gray Level Emphasis, glszm Low Gray Level Zone Emphasis
14	4	Mean Absolute Deviation, glcm Maximum Probability, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis
15	4	Mean Absolute Deviation, glcm Maximum Probability, glrlm High Gray Level Run Emphasis, glszm Low Gray Level Zone Emphasis
16	4	Mean Absolute Deviation, glcm Maximum Probability, gldm Low Gray Level Emphasis, glszm Low Gray Level Zone Emphasis
17	4	Mean Absolute Deviation, glcm Maximum Probability, gldm Large Dependence Emphasis, glszm Low Gray Level Zone Emphasis
18	4	Mean Absolute Deviation, glcm Maximum Probability, gldm High Gray Level Emphasis, glszm Low Gray Level Zone Emphasis
19	4	Mean Absolute Deviation, glcm Maximum Probability, gldm Dependence Non Uniformity Normalized, glszm Low Gray Level Zone Emphasis
20	4	Mean Absolute Deviation, glcm Maximum Probability, glcm Sum Entropy, glszm Low Gray Level Zone Emphasis
21	4	Mean Absolute Deviation, glcm Maximum Probability, glcm Sum Average, glszm Low Gray Level Zone Emphasis
22	4	Mean Absolute Deviation, glcm Joint Average, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
23	4	Mean Absolute Deviation, glcm Imc2, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
24	4	Mean Absolute Deviation, glcm Imc1, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
25	4	Mean Absolute Deviation, glcm Idmn, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
26	4	Mean Absolute Deviation, glcm Difference Variance, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
27	4	Mean Absolute Deviation, glcm Difference Entropy, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
28	4	Mean Absolute Deviation, glcm Difference Average, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
29	4	Mean Absolute Deviation, glcm Contrast, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
30	4	Mean Absolute Deviation, Uniformity, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
31	4	Mean Absolute Deviation, Root Mean Squared, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis

Appendice

32	4	Mean Absolute Deviation, Mean, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
33	4	Kurtosis, Mean Absolute Deviation, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
34	4	10 Percentile, Mean Absolute Deviation, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
35	4	Surface Area, Mean Absolute Deviation, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
36	4	Maximum 2D Diameter Slice, Mean Absolute Deviation, glcm Maximum Probability, glszm Low Gray Level Zone Emphasis
37	4	glcm Maximum Probability, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis, ngtdm Coarseness
38	4	glcm Maximum Probability, glrlm Gray Level Variance, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis
39	4	glcm Maximum Probability, gldm Large Dependence Low Gray Level Emphasis, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis
40	4	glcm Maximum Probability, glcm Sum Entropy, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis
41	4	glcm MCC, glcm Maximum Probability, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis
42	4	glcm Cluster Prominence, glcm Maximum Probability, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis
43	4	Robust Mean Absolute Deviation, glcm Maximum Probability, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis
44	4	Interquartile Range, glcm Maximum Probability, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis
45	4	90 Percentile, glcm Maximum Probability, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis
46	4	Sphericity, glcm Maximum Probability, glrlm Long Run Emphasis, glszm Low Gray Level Zone Emphasis

Tabella 29: sottoinsiemi di features individuati dall'algorithmo di feature selection QRA.

Dataset con valori continui

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	8 ± 5	2 ± 2	1 ± 1
	S	21 ± 6	26 ± 4	24 ± 6
	R-	1 ± 2	2 ± 3	6 ± 6

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	12 ± 9	5 ± 4	1 ± 1
	S	69 ± 13	54 ± 11	8 ± 3
	R-	5 ± 8	6 ± 9	2 ± 3

Tabella relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con un layer nascosto costituito da 5 neuroni

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	12 ± 5	3 ± 5	2 ± 3
	S	16 ± 5	24 ± 5	19 ± 6
	R-	2 ± 2	3 ± 3	10 ± 6

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	20 ± 13	9 ± 9	1 ± 1
	S	58 ± 14	45 ± 11	6 ± 3
	R-	8 ± 9	11 ± 9	4 ± 2

Tabella relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con un layer nascosto costituito da 10 neuroni

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	15 ± 5	4 ± 4	1 ± 2
	S	13 ± 5	22 ± 5	16 ± 6
	R-	2 ± 3	4 ± 4	14 ± 6

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	23 ± 10	10 ± 6	1 ± 1
	S	52 ± 11	41 ± 9	6 ± 2
	R-	11 ± 10	14 ± 9	4 ± 2

Tabella relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con un layer nascosto costituito da 20 neuroni

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	9 ± 6	3 ± 3	1 ± 3
	S	19 ± 6	25 ± 4	21 ± 7
	R-	2 ± 3	2 ± 4	9 ± 7

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	15 ± 11	6 ± 6	1 ± 1
	S	64 ± 15	51 ± 11	7 ± 3
	R-	7 ± 9	8 ± 9	3 ± 3

Tabella relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 2 layer nascosti costituito rispettivamente da 10 e 5 neuroni.

Appendice

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	10 ± 6	3 ± 4	1 ± 3
	S	18 ± 6	25 ± 5	21 ± 6
	R-	2 ± 2	2 ± 3	9 ± 6

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	16 ± 12	7 ± 8	1 ± 1
	S	63 ± 14	50 ± 11	7 ± 2
	R-	7 ± 8	8 ± 8	3 ± 2

Tabelle relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 2 layer nascosti costituito rispettivamente da 10 e 10 neuroni.

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	12 ± 6	4 ± 5	2 ± 4
	S	17 ± 6	24 ± 5	19 ± 6
	R-	1 ± 2	2 ± 3	10 ± 7

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	18 ± 13	9 ± 10	1 ± 1
	S	60 ± 15	47 ± 12	7 ± 2
	R-	8 ± 7	9 ± 8	3 ± 2

Tabelle relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 2 layer nascosti costituito rispettivamente da 20 e 10 neuroni.

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	6 ± 6	2 ± 4	1 ± 3
	S	23 ± 7	27 ± 5	26 ± 6
	R-	1 ± 2	1 ± 2	4 ± 5

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	9 ± 14	5 ± 9	0 ± 1
	S	74 ± 16	56 ± 12	9 ± 2
	R-	3 ± 5	4 ± 6	2 ± 2

Tabelle relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 3 layer nascosti costituito rispettivamente da 10, 5 e 3 neuroni

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	12 ± 6	4 ± 5	1 ± 4
	S	17 ± 6	24 ± 5	20 ± 7
	R-	1 ± 2	2 ± 2	10 ± 7

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	19 ± 15	9 ± 10	1 ± 1
	S	60 ± 16	48 ± 11	7 ± 3
	R-	7 ± 7	8 ± 8	3 ± 2

Tabelle relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 3 layer nascosti costituito rispettivamente da 10, 10 e 10 neuroni

Appendice

TRS		Classe vera			TS		Classe vera		
		R+	S	R-			R+	S	R-
Classe predetta	R+	9 ± 7	3 ± 4	1 ± 3	Classe predetta	R+	15 ± 12	7 ± 8	1 ± 1
	S	20 ± 7	25 ± 5	22 ± 8		S	64 ± 16	50 ± 12	7 ± 3
	R-	1 ± 3	2 ± 4	8 ± 8		R-	7 ± 10	8 ± 10	3 ± 3

Tabella relative alla media \pm deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 3 layer nascosti costituito rispettivamente da 20, 10 e 5 neuroni

Figura 71: relative alla media \pm deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set (con valori continui e normalizzati) delle reti neurali addestrate con le soluzioni del QRA

Dataset con valori discretizzati con CHiMerge

TRS		Classe vera			TS		Classe vera		
		R+	S	R-			R+	S	R-
Classe predetta	R+	8 ± 5	3 ± 2	1 ± 2	Classe predetta	R+	15 ± 10	7 ± 5	0 ± 1
	S	21 ± 5	24 ± 4	24 ± 5		S	64 ± 13	50 ± 9	9 ± 2
	R-	1 ± 3	3 ± 3	6 ± 6		R-	7 ± 8	8 ± 8	2 ± 2

Tabella relative alla media \pm deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con un layer nascosto costituito da 5 neuroni

TRS		Classe vera			TS		Classe vera		
		R+	S	R-			R+	S	R-
Classe predetta	R+	12 ± 5	5 ± 3	2 ± 3	Classe predetta	R+	23 ± 10	11 ± 6	1 ± 1
	S	16 ± 5	21 ± 4	16 ± 6		S	51 ± 12	40 ± 9	6 ± 2
	R-	2 ± 2	4 ± 3	13 ± 7		R-	12 ± 8	14 ± 7	3 ± 2

Tabella relative alla media \pm deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con un layer nascosto costituito da 10 neuroni

TRS		Classe vera			TS		Classe vera		
		R+	S	R-			R+	S	R-
Classe predetta	R+	16 ± 5	6 ± 4	3 ± 3	Classe predetta	R+	26 ± 9	14 ± 6	2 ± 1
	S	12 ± 4	20 ± 5	12 ± 4		S	43 ± 10	33 ± 7	5 ± 2
	R-	2 ± 3	4 ± 3	16 ± 5		R-	17 ± 8	18 ± 7	4 ± 2

Tabella relative alla media \pm deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con un layer nascosto costituito da 20 neuroni

Appendice

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	8 ± 6	3 ± 3	2 ± 2
	S	20 ± 6	24 ± 4	21 ± 7
	R-	2 ± 3	3 ± 3	8 ± 7

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	16 ± 12	8 ± 6	1 ± 1
	S	61 ± 15	48 ± 11	8 ± 2
	R-	9 ± 8	10 ± 9	2 ± 2

Tabella relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 2 layer nascosti costituito rispettivamente da 10 e 5 neuroni

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	13 ± 6	4 ± 3	2 ± 3
	S	15 ± 6	23 ± 4	17 ± 6
	R-	2 ± 2	3 ± 3	12 ± 7

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	21 ± 11	11 ± 7	2 ± 1
	S	53 ± 13	41 ± 9	6 ± 2
	R-	12 ± 8	13 ± 8	3 ± 2

Tabella relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 2 layer nascosti costituito rispettivamente da 10 e 10 neuroni

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	13 ± 6	5 ± 4	2 ± 3
	S	14 ± 5	20 ± 6	14 ± 6
	R-	3 ± 4	5 ± 4	15 ± 7

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	23 ± 11	13 ± 7	2 ± 1
	S	47 ± 12	35 ± 9	6 ± 2
	R-	16 ± 11	17 ± 10	3 ± 2

Tabella relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 2 layer nascosti costituito rispettivamente da 20 e 10 neuroni

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	6 ± 7	2 ± 4	1 ± 3
	S	23 ± 7	26 ± 5	24 ± 8
	R-	1 ± 3	2 ± 4	6 ± 8

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	11 ± 13	6 ± 8	1 ± 2
	S	68 ± 18	52 ± 14	9 ± 3
	R-	7 ± 12	7 ± 10	1 ± 2

Tabella relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 3 layer nascosti costituito rispettivamente da 10, 5 e 3 neuroni

Appendice

TRS		Classe vera			TS		Classe vera		
		R+	S	R-			R+	S	R-
Classe predetta	R+	12 ± 7	4 ± 3	2 ± 4	Classe predetta	R+	21 ± 12	11 ± 8	1 ± 2
	S	16 ± 7	23 ± 4	18 ± 7		S	53 ± 15	42 ± 12	7 ± 3
	R-	2 ± 3	3 ± 3	11 ± 7		R-	12 ± 10	12 ± 9	3 ± 2

Tabelle relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 3 layer nascosti costituito rispettivamente da 10, 10 e 10 neuroni

TRS		Classe vera			TS		Classe vera		
		R+	S	R-			R+	S	R-
Classe predetta	R+	9 ± 6	3 ± 3	2 ± 3	Classe predetta	R+	15 ± 12	8 ± 7	1 ± 2
	S	19 ± 7	24 ± 5	20 ± 7		S	61 ± 16	47 ± 11	7 ± 3
	R-	2 ± 4	3 ± 4	9 ± 7		R-	10 ± 11	10 ± 9	3 ± 2

Tabelle relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set della rete neurale con 3 layer nascosti costituito rispettivamente da 20, 10 e 5 neuroni

Figura 72: tabelle relative alla media ± deviazione standard delle 9 celle delle confusion matrix ottenute dalle 100 ripetizioni delle fasi di inizializzazione, training e verifica su Training Set e Test Set (con valori discretizzati) delle reti neurali allenate con le soluzioni del QRA

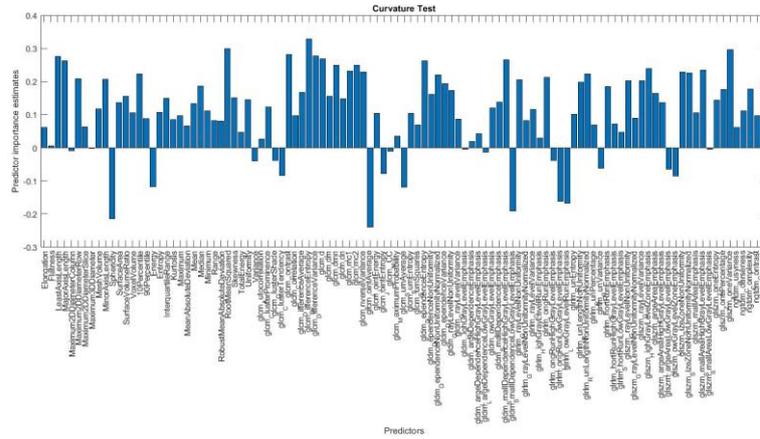


Figura 74: importanza di ciascuna delle 107 features ottenuta allenando un RF con le 107 features e NumPredictorToSample pari a 11 ($\sim\sqrt{107}$)

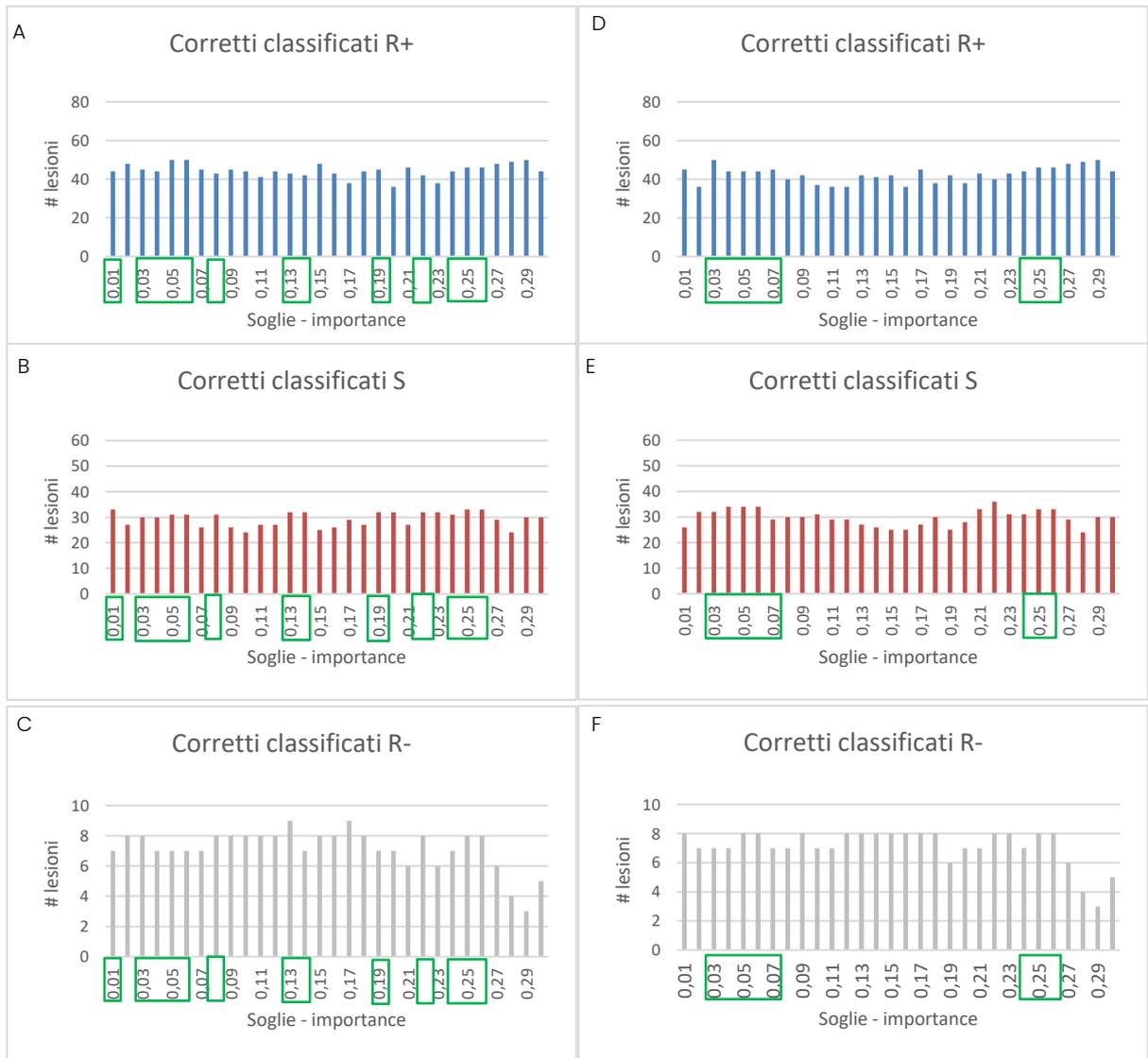


Figura 75: n° di corretti classificati ottenuti per ognuna delle 3 classi con i modelli di RF allenati con TRS_SOM_2 e un n° di features scelto impostando diverse soglie sull'importanza (Figure A-B-C) e sull'importance (Figure D-E-F) ottenuta con il modello allenato con 107 features e NumPredictorToSample pari a $\sqrt{n^{\circ} \text{ features}}$ (importance varia da 0.01 a 0.30 a passi di 0.01). In verde sono indicati i modelli che forniscono buone e non sbilanciate numerosità di corretti classificati sulle 3 classi nel Test Set.

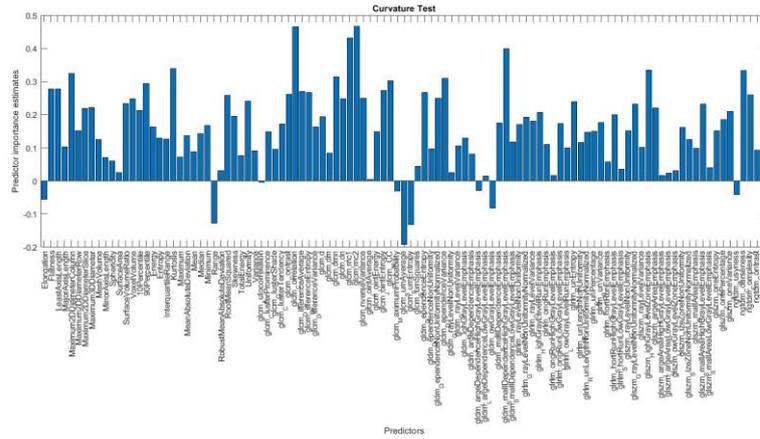


Figura 76: importanza di ciascuna delle 107 features ottenuta allenando un RF con le 107 features e NumPredictorToSample pari a 54 ($(1/2 \cdot 107)$)

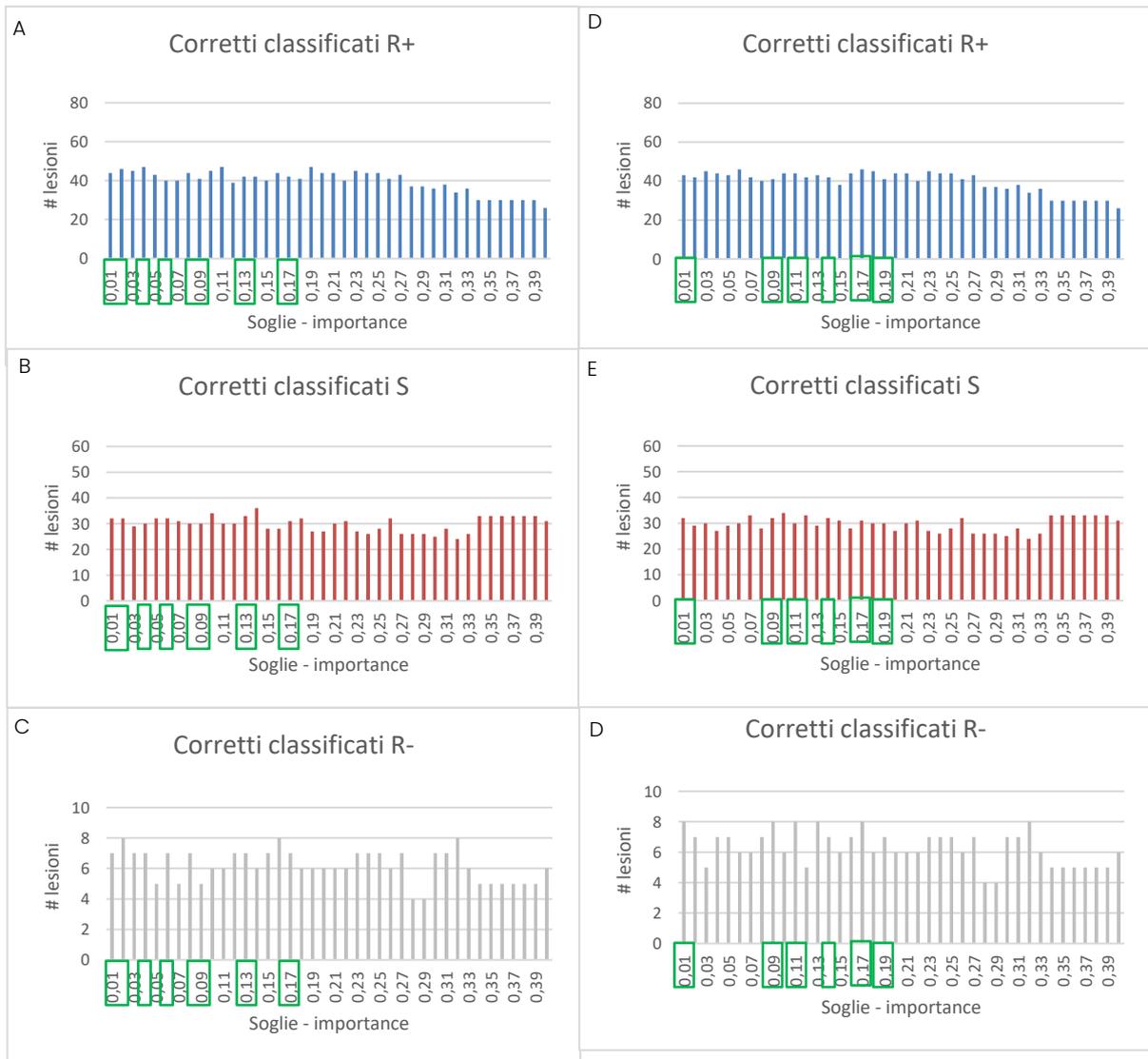


Figura 77: n° di corretti classificati ottenuti per ognuna delle 3 classi con i modelli di RF allenati con TRS_SOM_2 e con un n° di features scelto impostando diverse soglie sull'importanza (Figure A-B-C) e sull'importanza (Figure D-E-F) ottenuta con il modello allenato con 107 features e NumPredictorToSample pari a $(n^{\circ} \text{ features}) / 2$ (soglia varia da 0.01 a 0.40 a passi di 0.01). In verde sono indicati i modelli che forniscono buone e non sbilanciate numerosità di corretti classificati sulle 3 classi nel Test Set

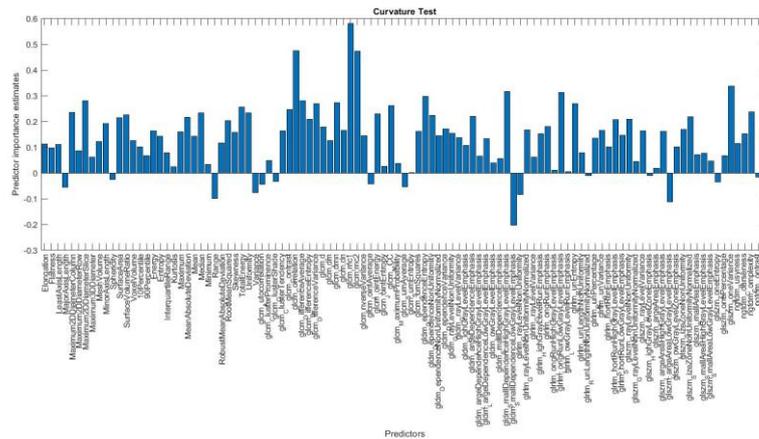


Figura 78: importanza di ciascuna delle 107 features ottenuta allenando un RF con le 107 features e $NumPredictorToSample$ pari a $80 \left(\frac{3}{4} * 107\right)$

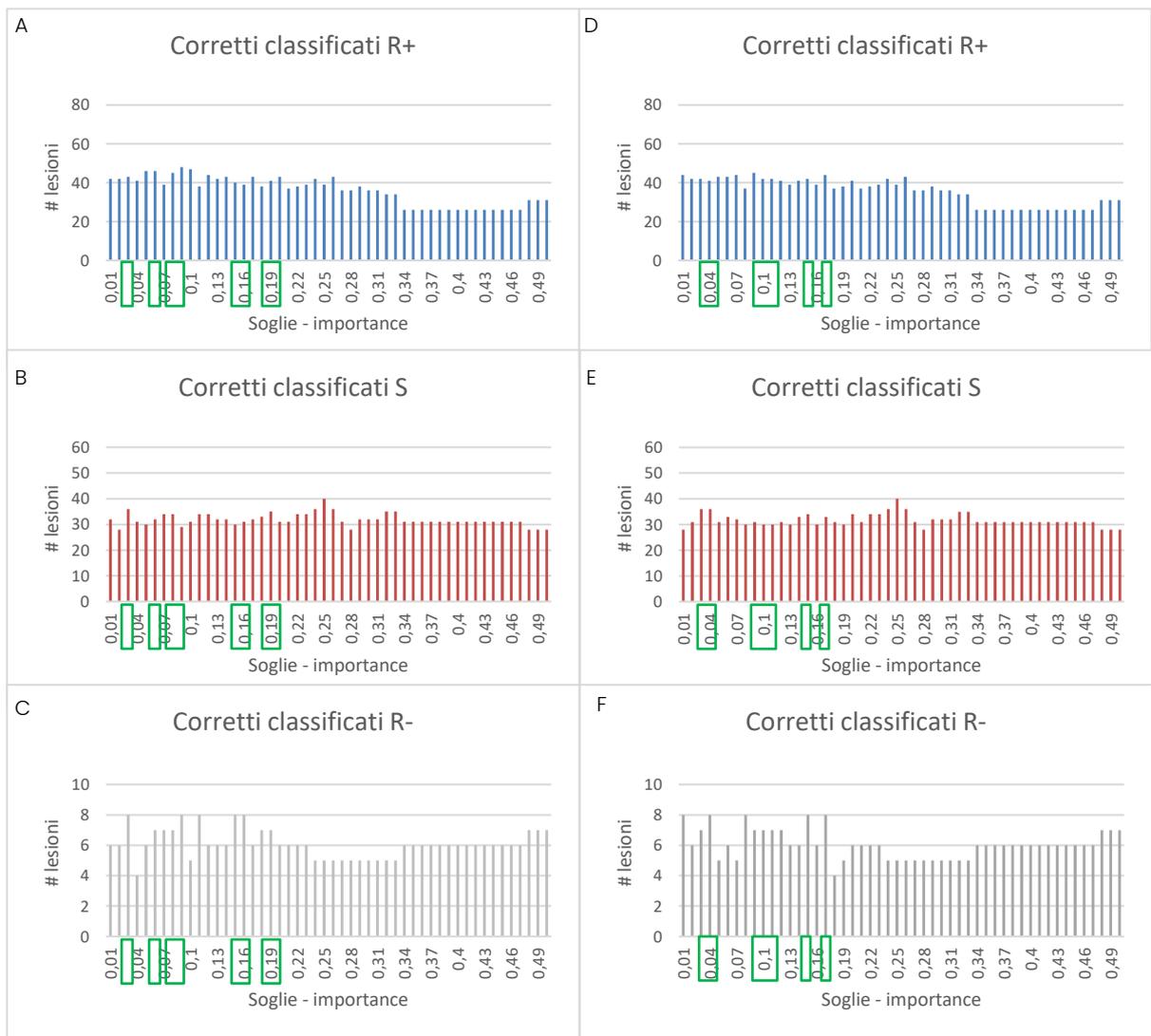
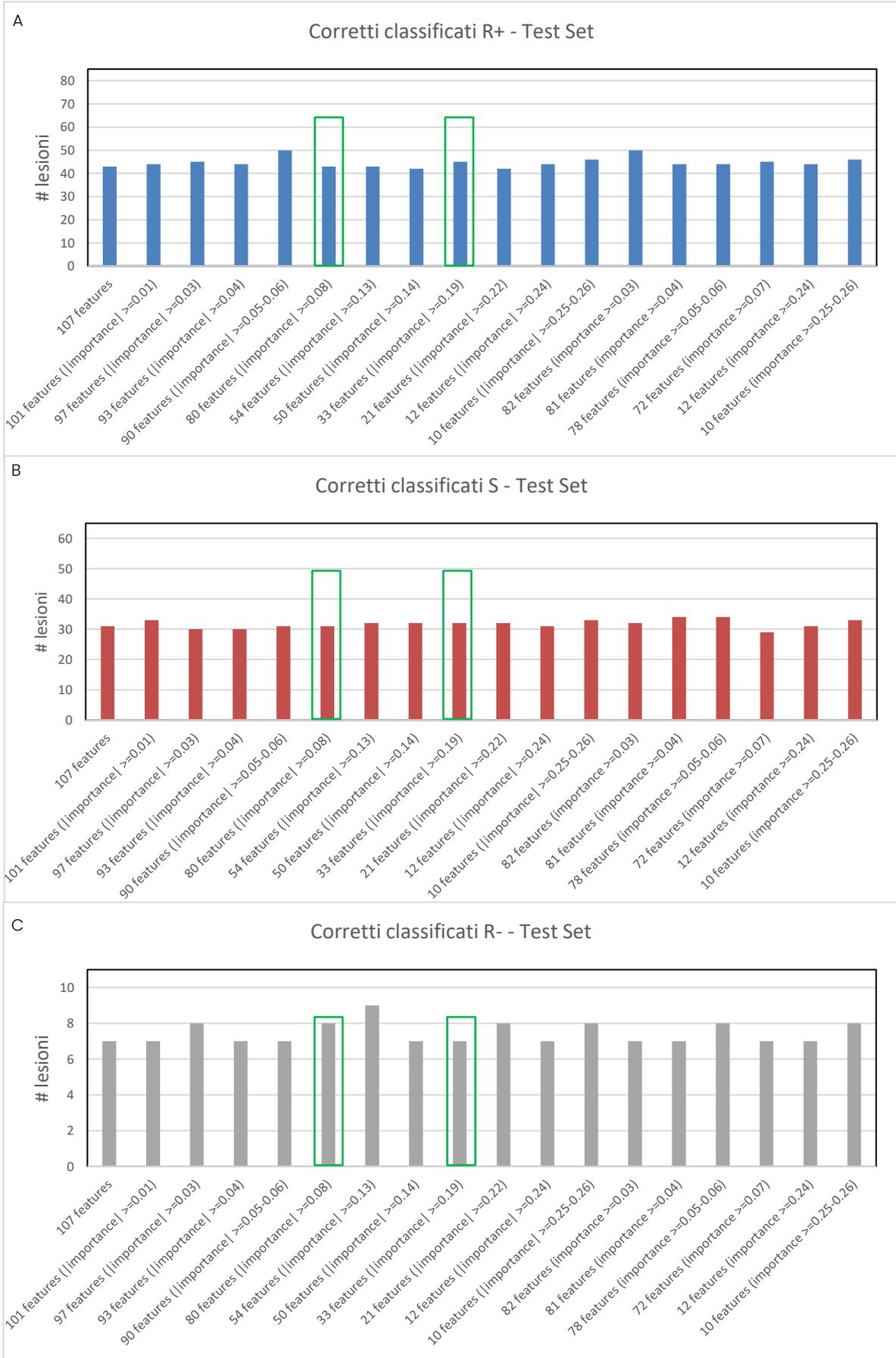
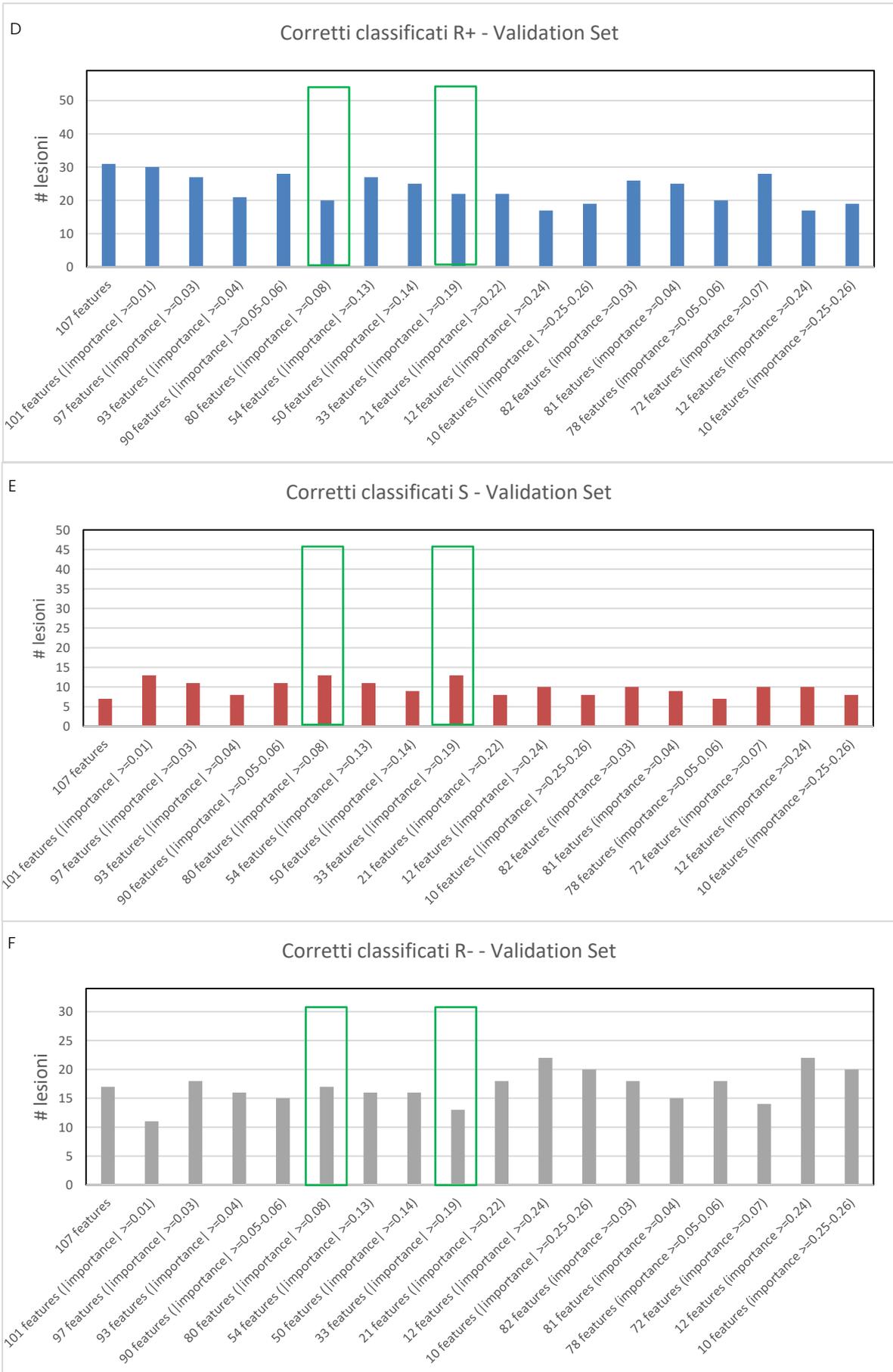


Figura 79: n° di corretti classificati ottenuti per ognuna delle 3 classi con i modelli di RF allenati con TRS_SOM_2 e un n° di features scelto impostando diverse soglie sull'importanza (Figure A-B-C) e sull'importanza (Figure D-E-F) ottenuta con il modello allenato con 107 features e $NumPredictorToSample$ pari a $\frac{3}{4} * n^{\circ} features$ (soglia varia da 0.01 a 0.45 a passi di 0.01). In verde sono indicati i modelli che forniscono buone e non sbilanciate numerosità di corretti classificati sulle 3 classi nel Test Set

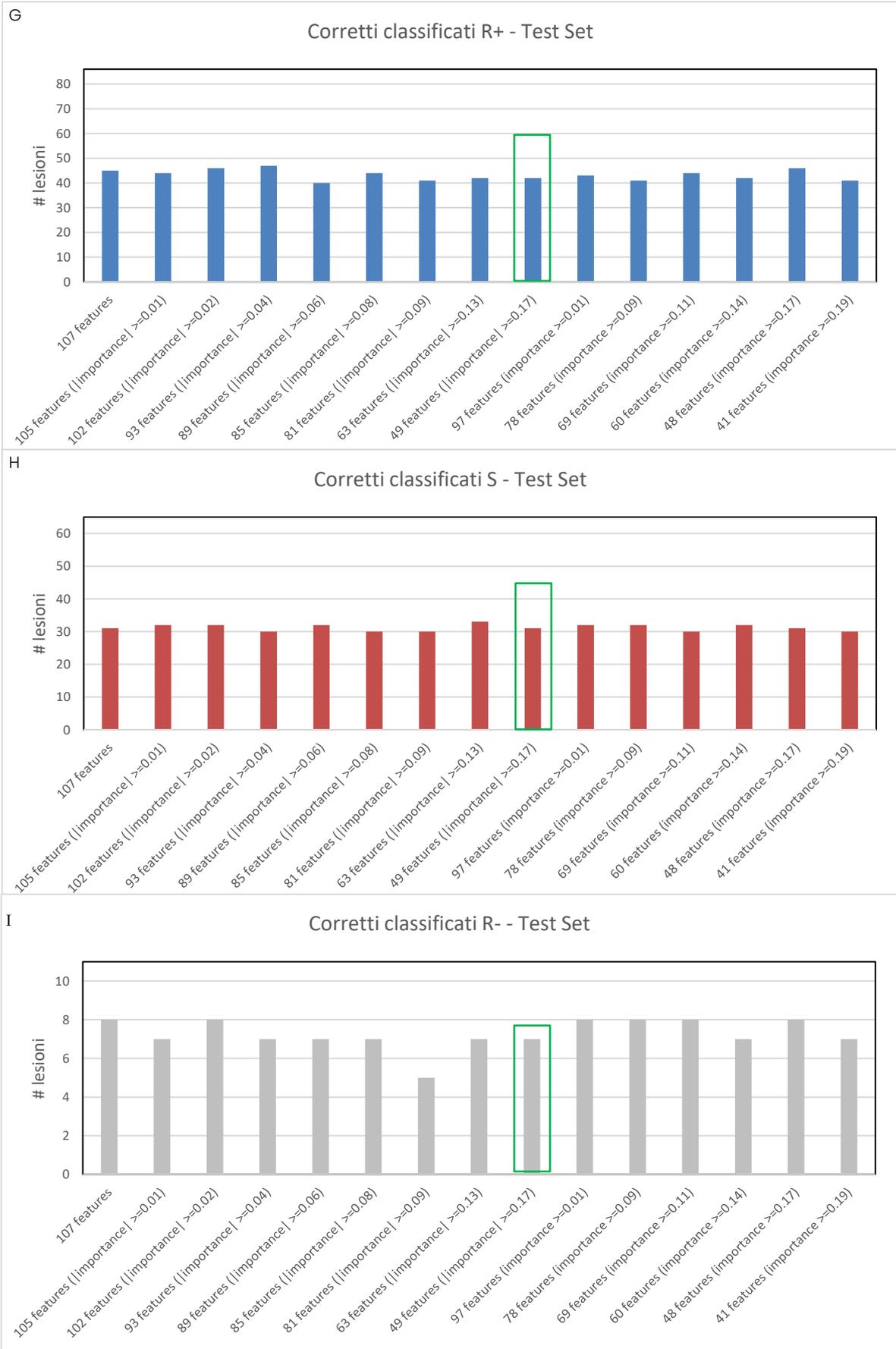
Appendice



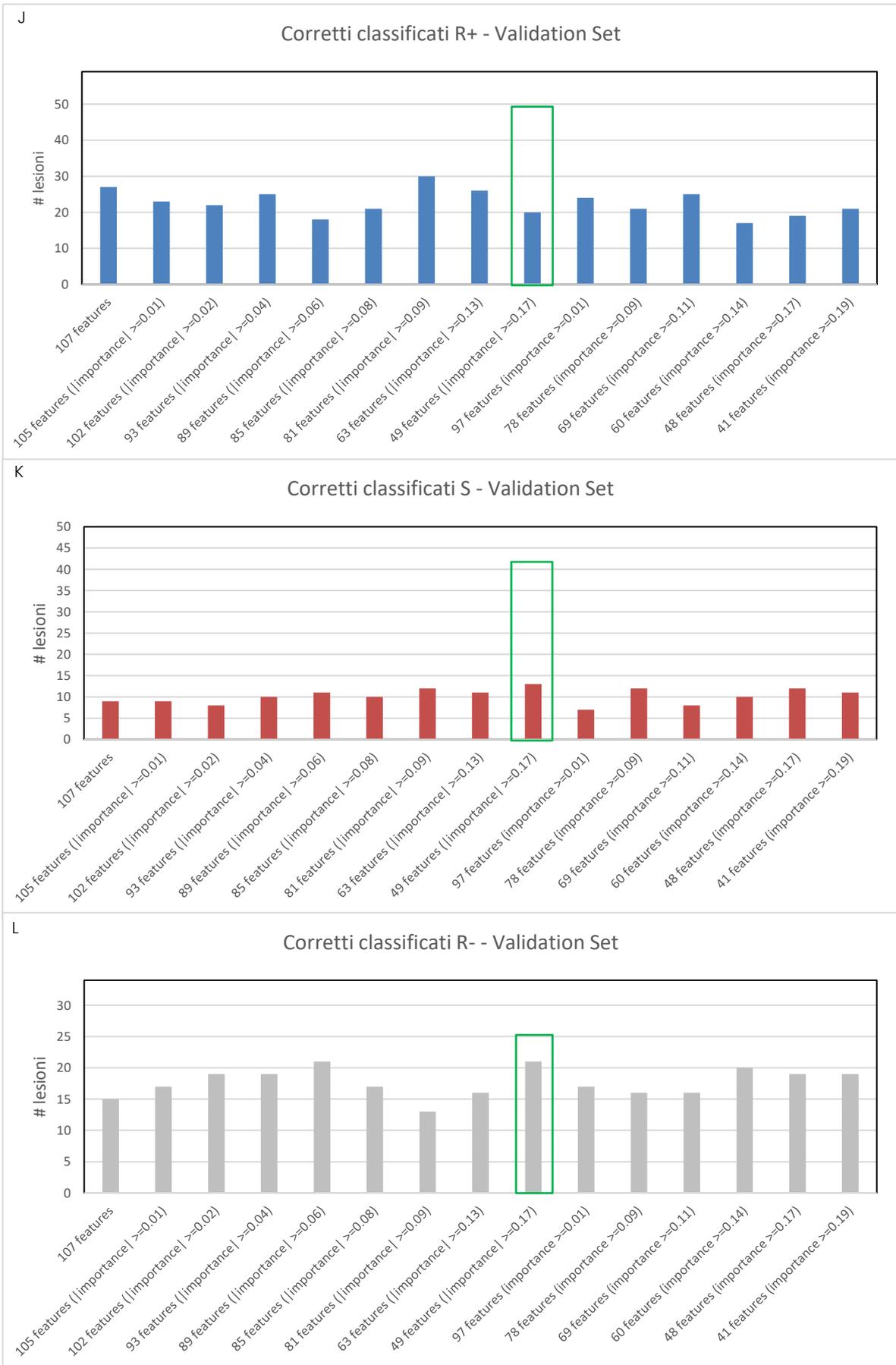
Appendice



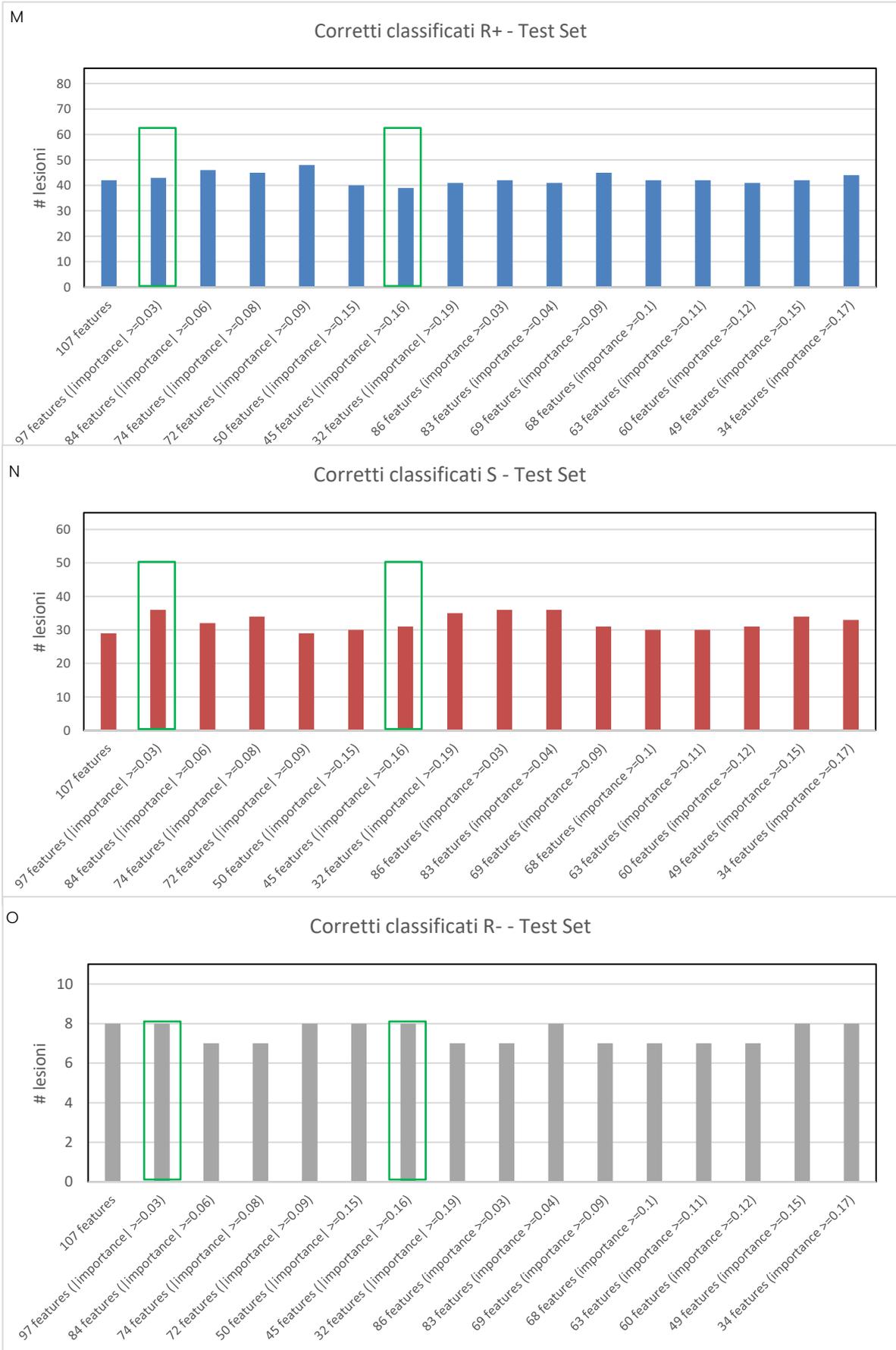
Appendice



Appendice



Appendice



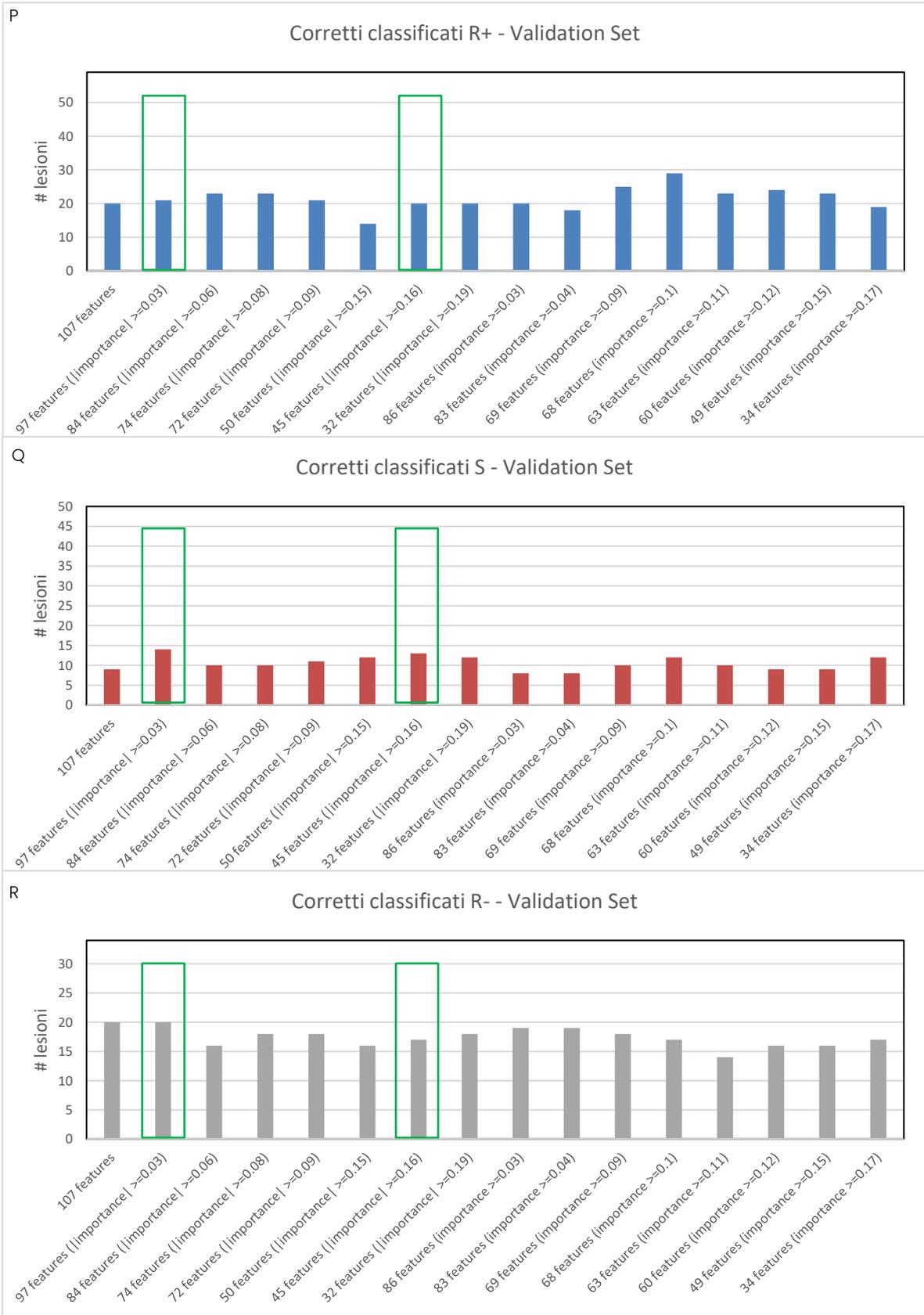


Figura 80: n° di corretti classificati ottenuti per ognuna delle 3 classi sul Test Set e Validation Set con i modelli migliori individuati per ogni NumPredictorsToSamples (Figure A, B, C, D, E, F: $\sqrt{n^\circ \text{features}}$; Figure G, H, I, J, K, L: $n^\circ \text{features}/2$; Figure M, N, O, P, Q, R: $3/4 * n^\circ \text{features}$) allenati sul TRS_SOM_2

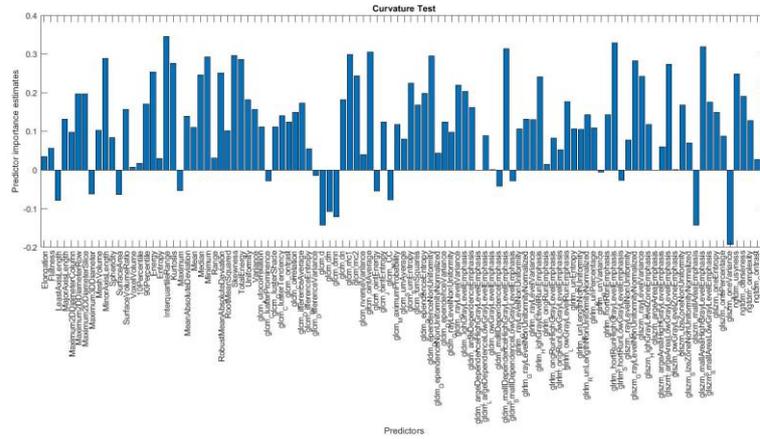


Figura 81: importanza di ciascuna delle 107 features ottenuta allenando un RF con le 107 features e $NumPredictorToSample$ pari a 11 ($\sim\sqrt{107}$)

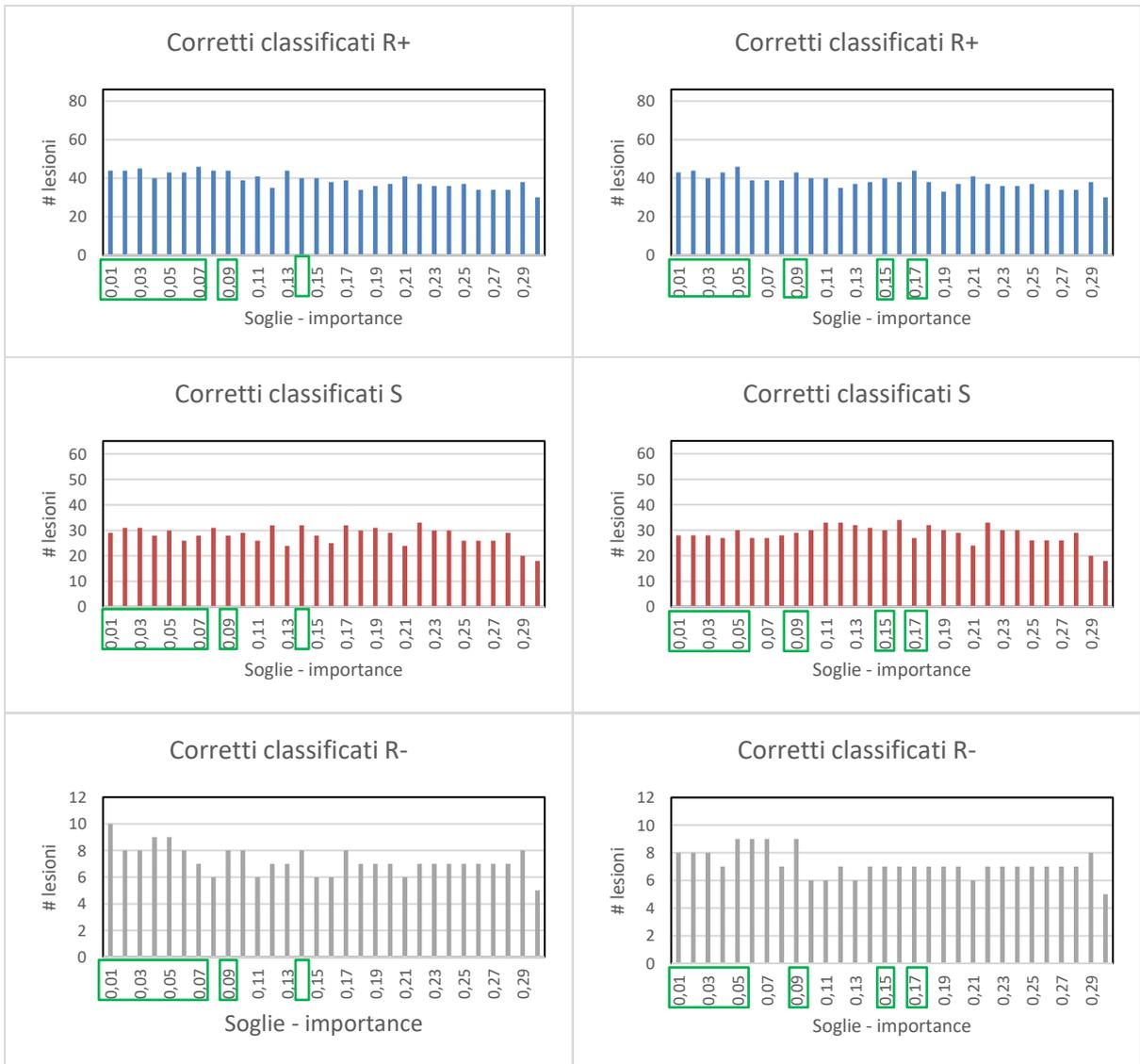


Figura 82: n° di corretti classificati ottenuti per ognuna delle 3 classi con i modelli di RF allenati con TRS_SOM_3 e un n° di features scelto impostando diverse soglie sull'importanza (Figure A-B-C) e sull'importance (Figure D-E-F) ottenuta sul modello allenato con 107 features e $NumPredictorToSample$ pari a $\sqrt{n^\circ features}$ (importance varia da 0.01 a 0.30 a passi di 0.01). In verde sono indicate i modelli che forniscono buone e non sbilanciate numerosità di corretti classificati sulle 3 classi nel Test Set.

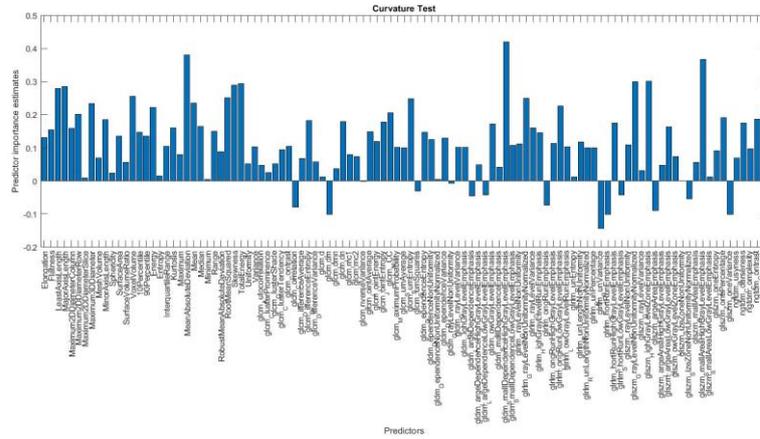
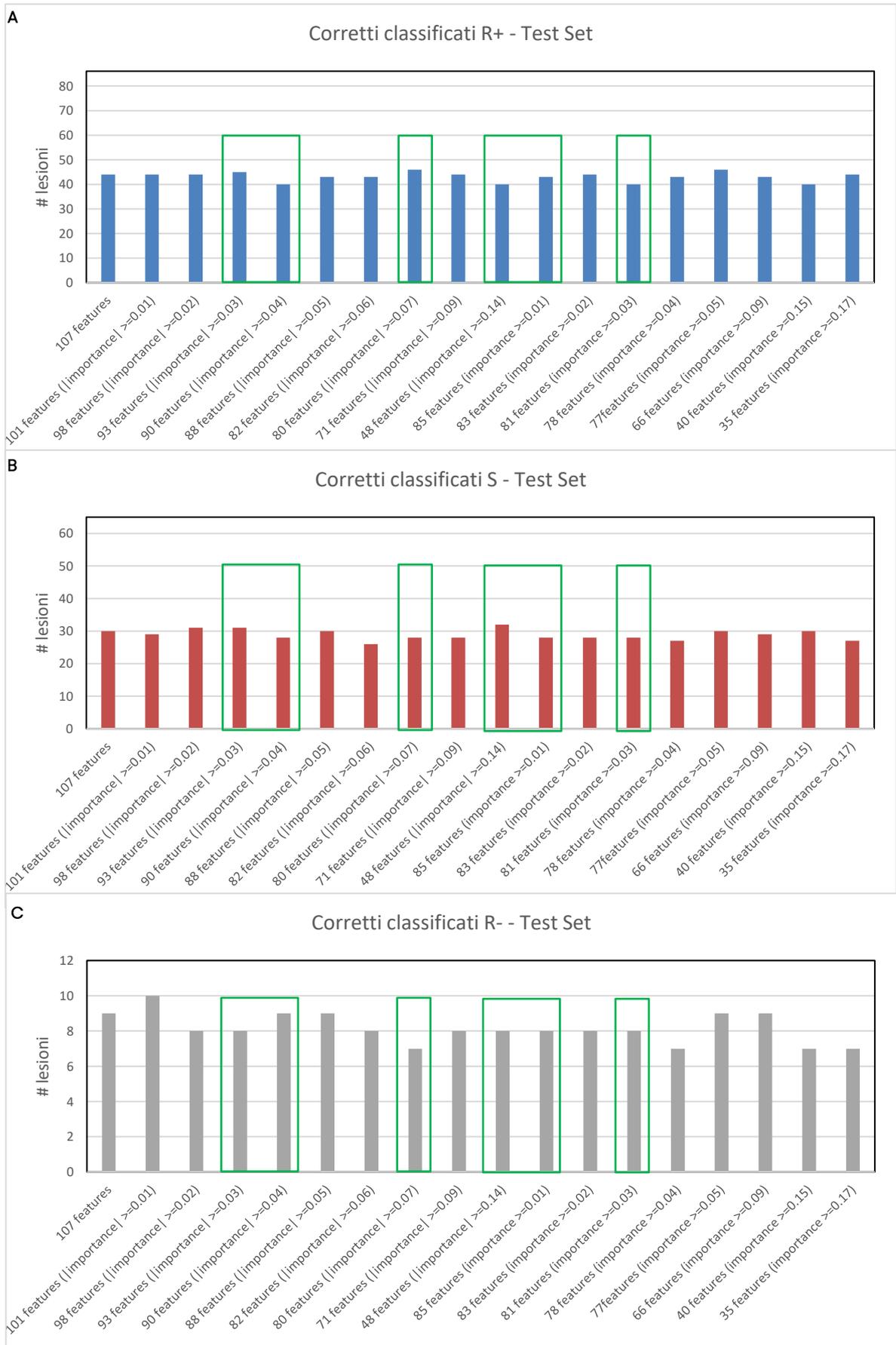
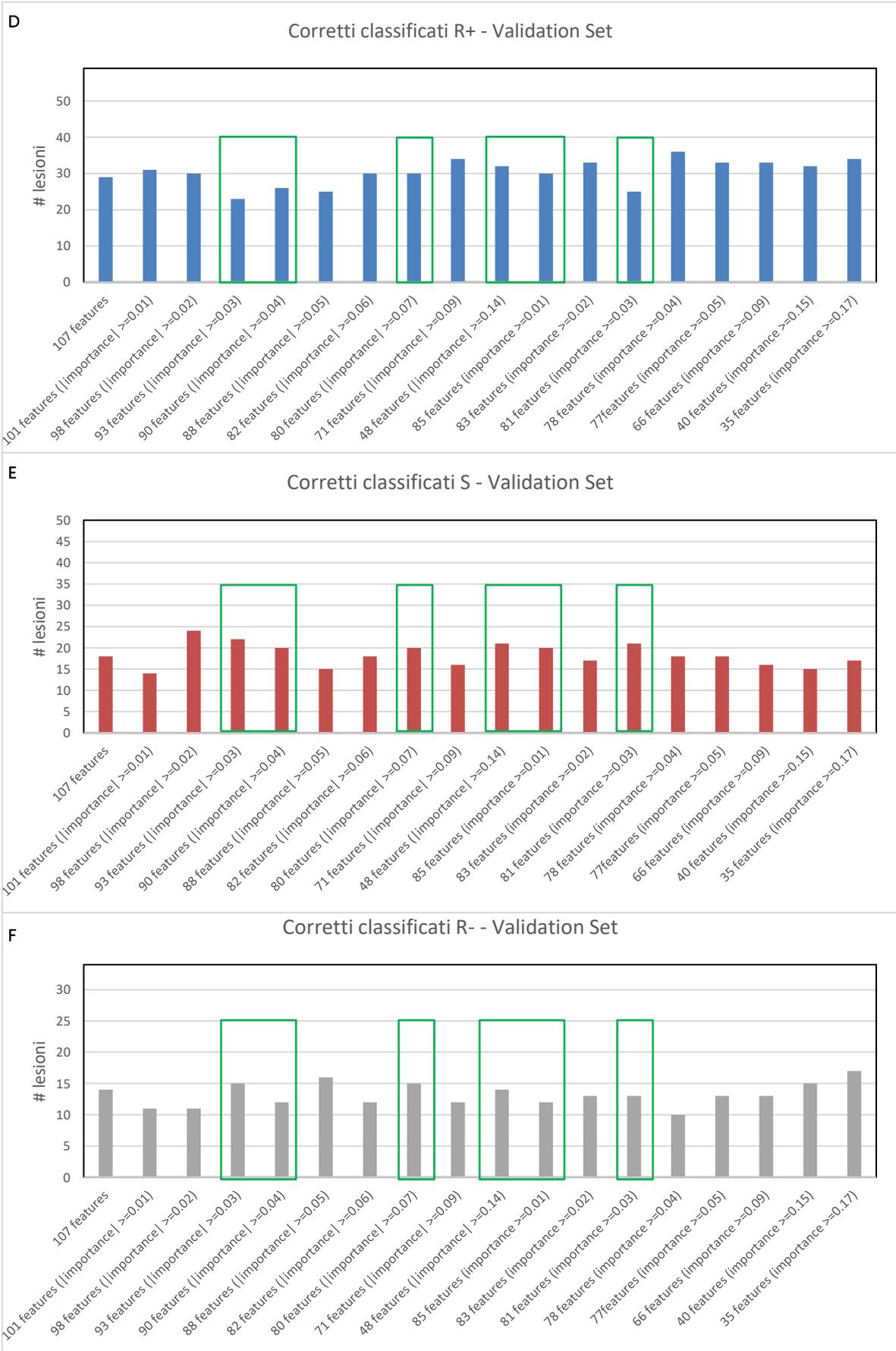


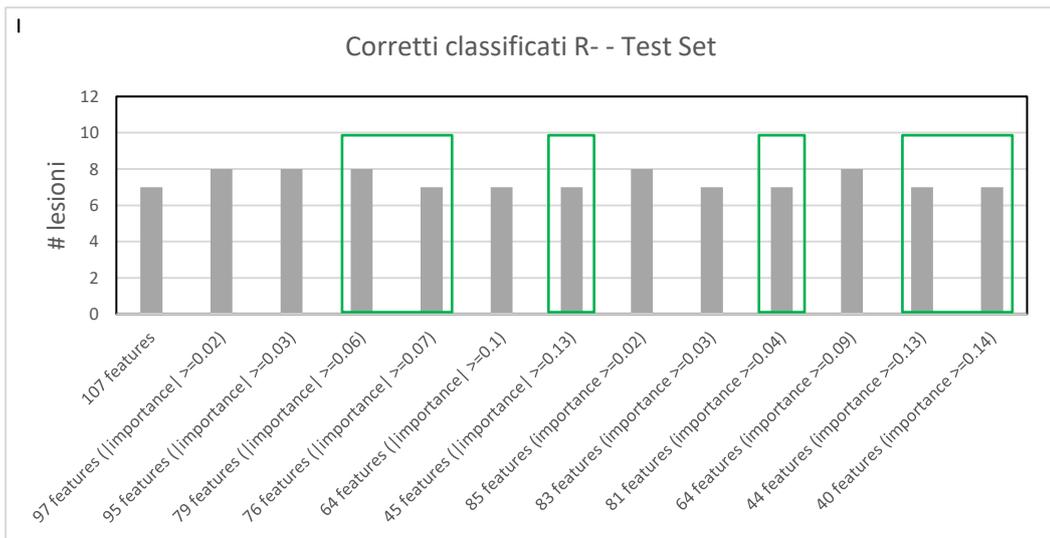
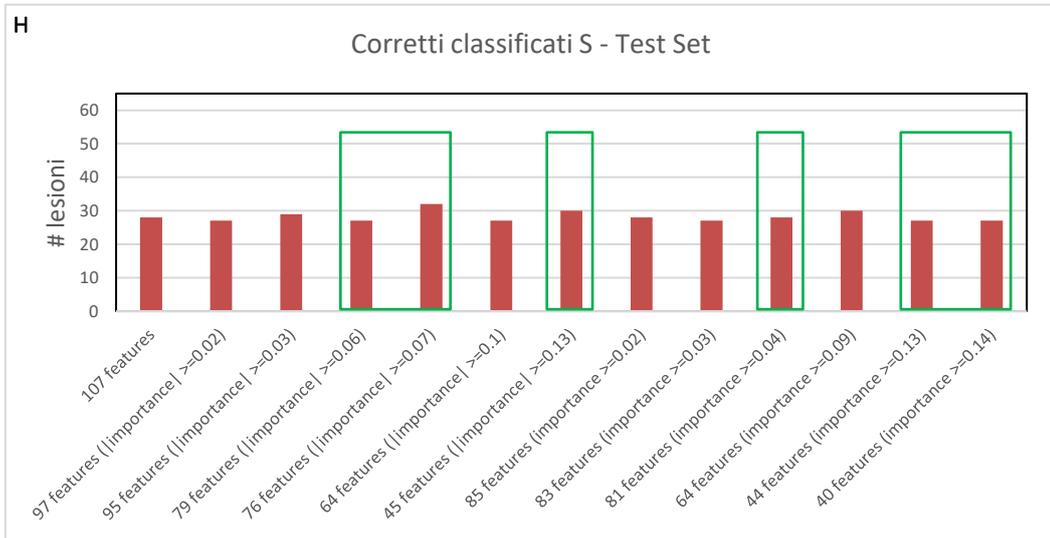
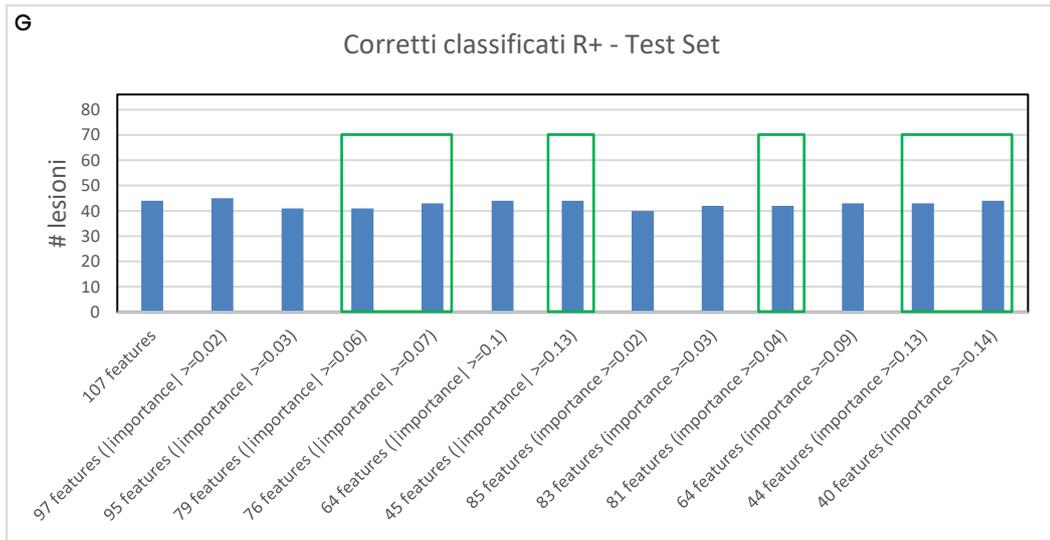
Figura 83: importanza di ciascuna delle 107 features ottenuta allenando un RF con le 107 features e $NumPredictorToSample$ pari a $54 (\frac{1}{2} * 107)$

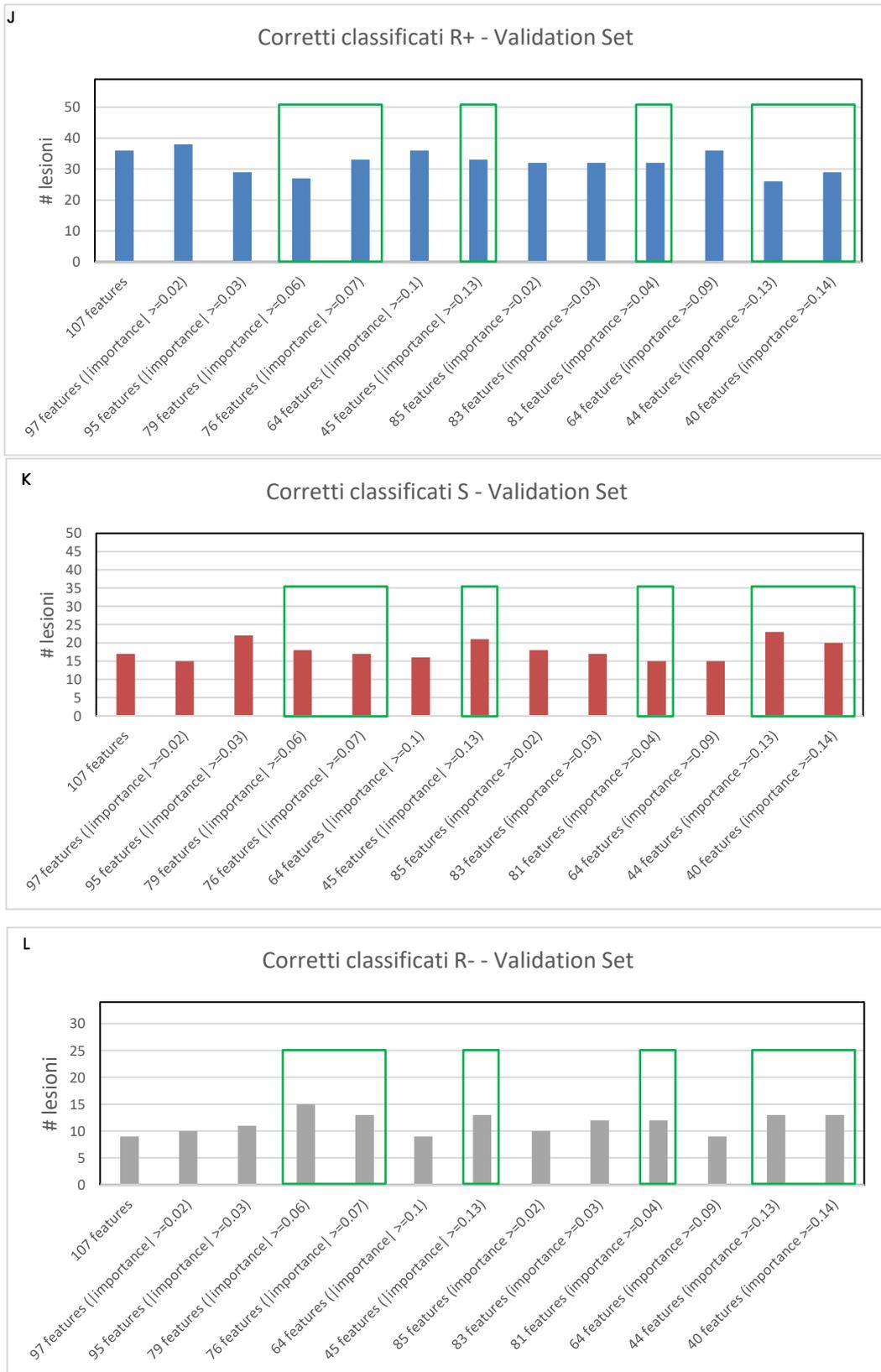


Figura 84: n° di corretti classificati ottenuti per ognuna delle 3 classi con i modelli di RF allenati con TRS_SOM_3 e un n° di features scelto impostando diverse soglie sull'importanza (Figure A-B-C) e sull'importance (Figure D-E-F) ottenuta sul modello allenato con 107 features e $NumPredictorToSample$ pari a $\frac{n^{\circ} features}{2}$ (importance varia da 0.01 a 0.30 a passi di 0.01). In verde sono indicate i modelli che forniscono buone e non sbilanciate numerosità di corretti classificati sulle 3 classi nel Test Set.

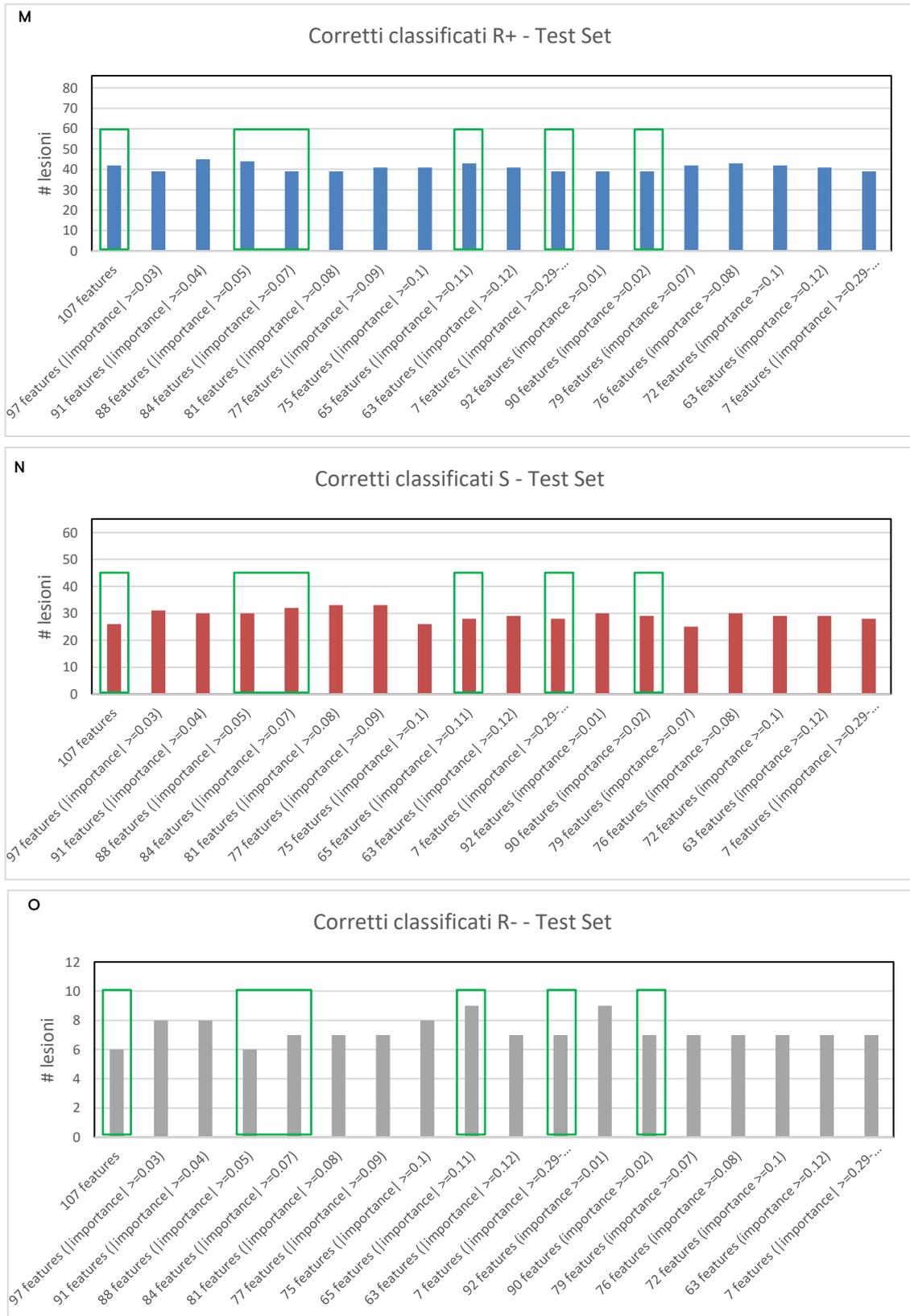








Appendice



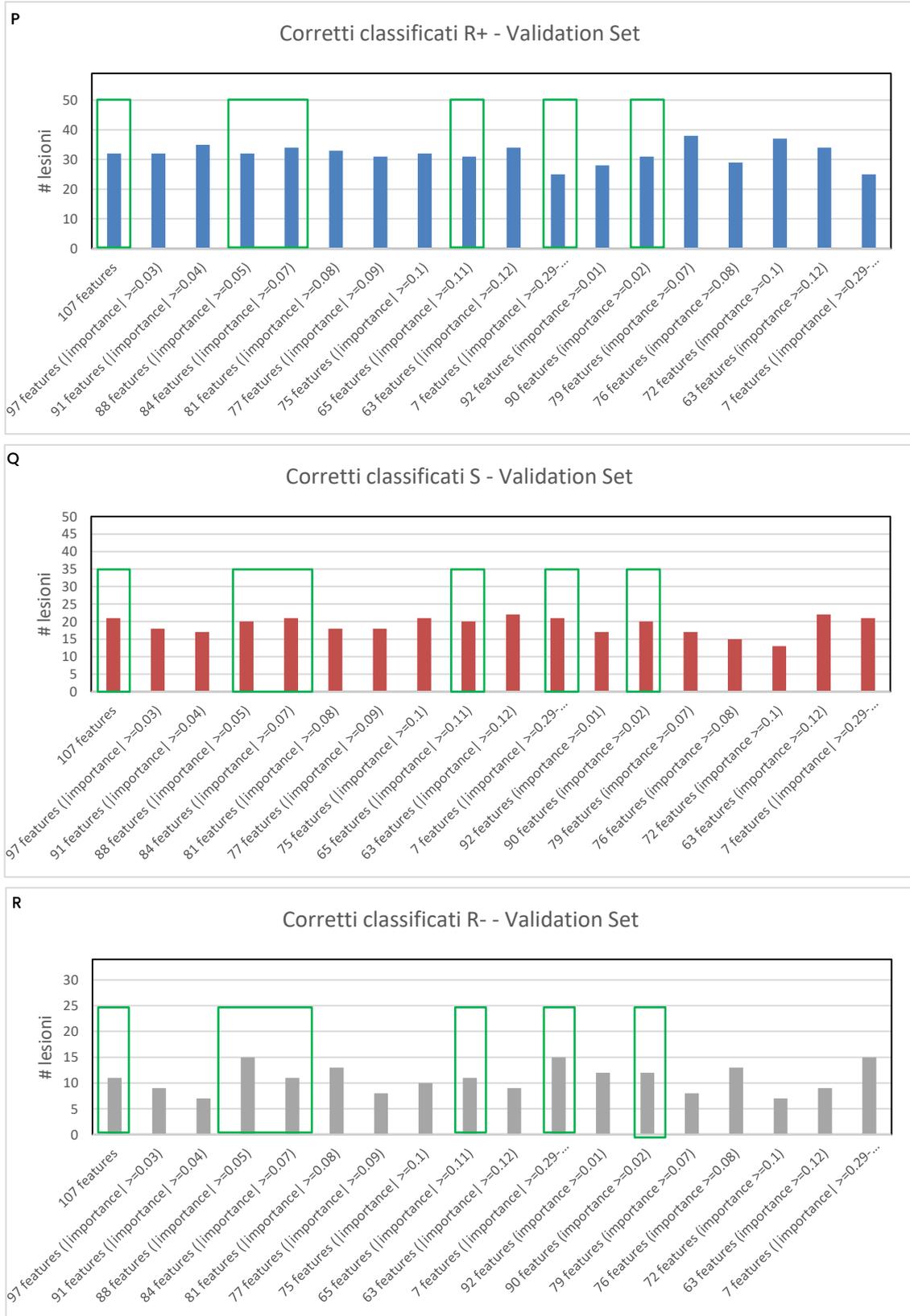


Figura 87: n° di corretti classificati ottenuti per ognuna delle 3 classi sul Test Set e Validation Set con i modelli migliori individuati per ogni *NumPredictorsToSamples* (Figure A, B, C, D, E, F: $\sqrt{n^\circ features}$; Figure G, H, I, J, K, L, M, N, O, P, Q, R: $\frac{3}{4} * n^\circ features$) allenati sul TRS_SOM_3.

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	23	4	2
	S	5	25	5
	R-	3	1	24

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	43	22	2
	S	25	31	3
	R-	17	12	6

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare ($C=1$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	30	0	1
	S	1	30	0
	R-	0	0	30

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	54	27	2
	S	19	28	3
	R-	12	10	6

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare ($C=10$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	31	0	0
	S	0	30	0
	R-	0	0	31

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	52	27	4
	S	20	28	2
	R-	13	10	5

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare ($C=50$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	31	0	0
	S	0	30	0
	R-	0	0	31

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	52	27	4
	S	20	28	2
	R-	13	10	5

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare ($C=100$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	31	30	31
	S	0	0	0
	R-	0	0	0

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	86	65	11
	S	0	0	0
	R-	0	0	0

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, $C=1$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	31	30	31
	S	0	0	0
	R-	0	0	0

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	86	65	11
	S	0	0	0
	R-	0	0	0

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, $C=10$)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	22	17	15
	S	9	13	16
	R-	0	0	0

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	61	41	6
	S	24	24	5
	R-	0	0	0

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, C=50)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	17	11	9
	S	14	18	19
	R-	0	1	3

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	46	26	5
	S	39	38	5
	R-	0	1	1

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, C=100)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	31	29	30
	S	0	1	1
	R-	0	0	0

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	83	64	11
	S	2	1	0
	R-	0	0	0

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=1)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	14	4	3
	S	16	23	18
	R-	1	3	10

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	35	17	2
	S	45	41	4
	R-	5	7	5

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=10)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	23	4	2
	S	5	25	5
	R-	3	1	24

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	43	22	2
	S	25	31	3
	R-	17	12	6

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=50)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	25	3	2
	S	4	24	2
	R-	2	3	27

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	21	21	17
	S	33	21	11
	R-	5	8	6

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=100)

Figura 88: confusion matrix relative alle prestazioni su Training Set e Test Set dei diversi modelli di SVM allenati su TRS_SOM_2 con 107 features per individuare le strutture di SVM più performanti da utilizzare per effettuare feature selection basata sulla correlazione.

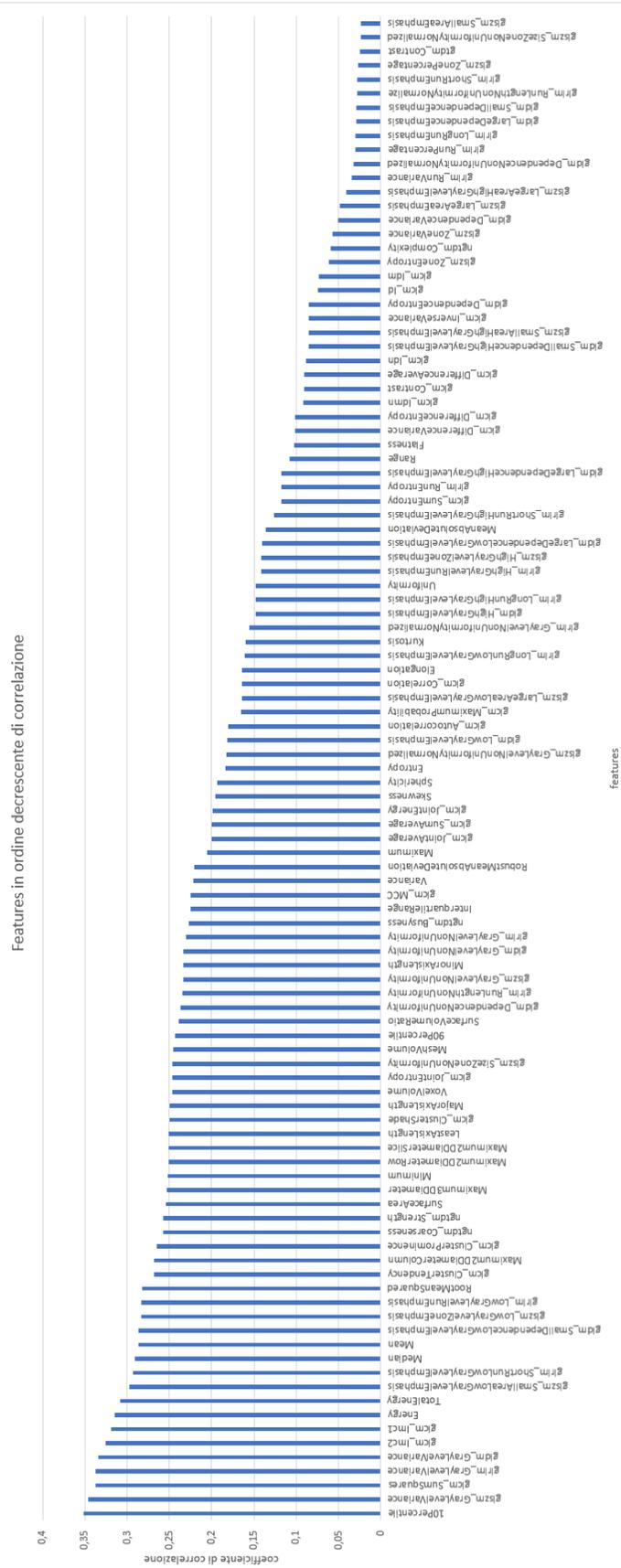


Figura 89: grafico relativo al coefficiente di correlazione di Spearman di ognuna delle 107 features con la classe nel TRS_SOM_2.

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	26	6	3
	S	2	20	2
	R-	2	4	25

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	49	27	3
	S	17	25	2
	R-	20	13	7

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare (C=1)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	29	0	0
	S	1	30	0
	R-	0	0	30

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	51	29	2
	S	26	25	4
	R-	9	11	6

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare (C=10)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	30	0	0
	S	0	30	0
	R-	0	0	30

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	52	28	1
	S	25	26	5
	R-	9	11	6

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare (C=50)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	30	0	0
	S	0	30	0
	R-	0	0	30

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	52	28	1
	S	25	26	5
	R-	9	11	6

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM lineare (C=100)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	30	30	30
	S	0	0	0
	R-	0	0	0

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	86	65	12
	S	0	0	0
	R-	0	0	0

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, C=1)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	30	30	30
	S	0	0	0
	R-	0	0	0

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	86	65	12
	S	0	0	0
	R-	0	0	0

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, C=10)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	26	20	15
	S	0	1	0
	R-	4	9	15

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	67	45	7
	S	0	0	0
	R-	19	20	5

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, C=50)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	25	12	8
	S	1	7	5
	R-	4	11	17

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	59	38	4
	S	7	8	1
	R-	20	19	7

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM polinomiale (grado 3, C=100)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	30	30	28
	S	0	0	0
	R-	0	0	2

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	81	63	12
	S	0	0	0
	R-	5	2	0

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=1)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	22	8	7
	S	4	14	7
	R-	4	8	16

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	54	32	5
	S	14	20	1
	R-	18	13	6

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=10)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	26	7	3
	S	2	19	2
	R-	2	4	25

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	51	28	3
	S	18	25	2
	R-	17	12	7

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=50)

TRS		Classe vera		
		R+	S	R-
Classe predetta	R+	27	5	1
	S	2	25	2
	R-	1	0	27

TS		Classe vera		
		R+	S	R-
Classe predetta	R+	50	30	3
	S	22	27	2
	R-	14	8	7

Confusion matrix relative alle prestazioni su Training Set e Test Set del SVM gaussiano (C=100)

Figura 91: confusion matrix relative alle prestazioni su Training Set e Test Set dei diversi modelli di SVM allenati su TRS_SOM_3 con 107 features per individuare le strutture di SVM più performanti da utilizzare per effettuare feature selection basata sulla correlazione.

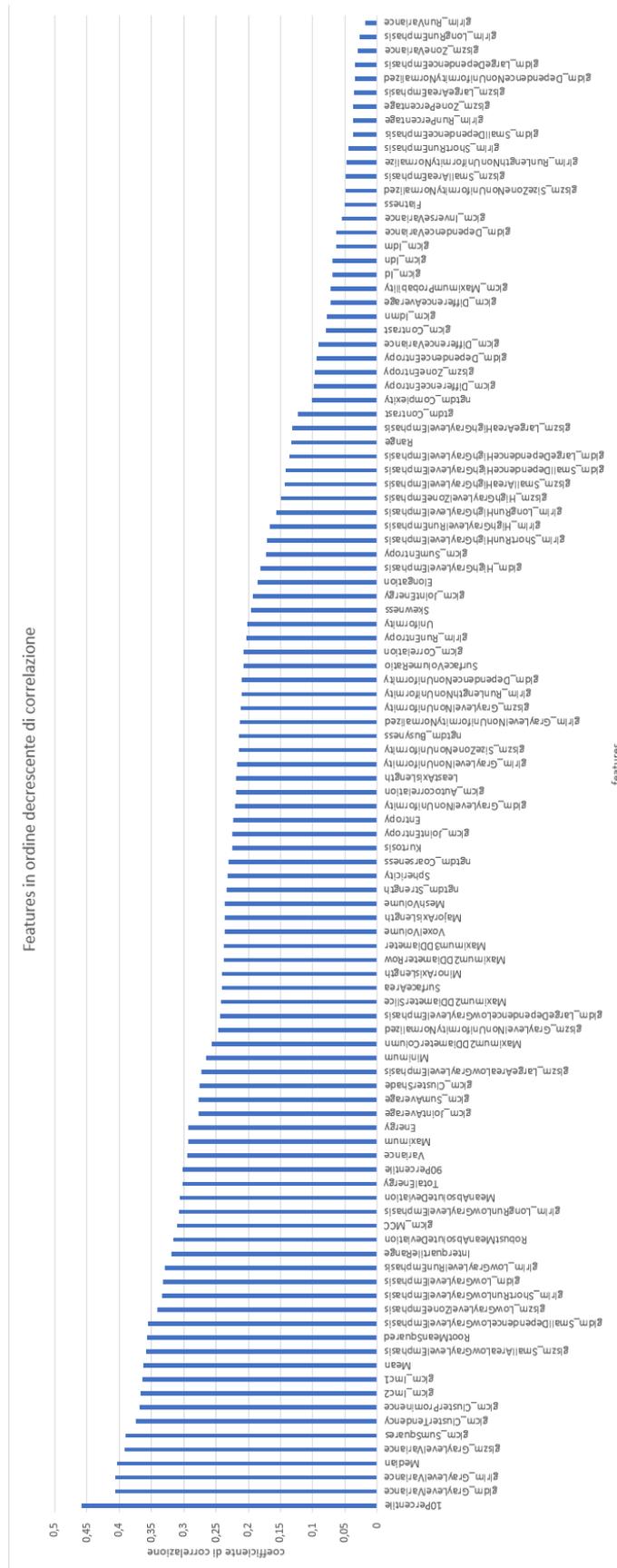


Figura 92: grafico relativo al coefficiente di correlazione di Spearman di ognuna delle 107 features con la classe nel TRS_SOM_3.

Appendice



Figura 95: grafico relativo al p-value del coefficiente di correlazione di Spearman di ognuna delle 107 features con la classe nel Constr_2. In rosso sono indicate le features con p-value < 0.05

Bibliografia

1. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, & F. Bray, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, *CA: A Cancer Journal for Clinicians* **71**, 209–249 (2021)
2. <https://www.airc.it/cancro/informazioni-tumori/guida-ai-tumori/tumore-colon-retto>, ultimo accesso: settembre, 2021
3. E. Van Cutsem, A. Cervantes, B. Nordlinger & D. Arnold, Metastatic Colorectal Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-up, *Annals of Oncology* **25-3**, iii1-ii9 (2014)
4. S. J. Ahn, L. H. Kim, S. J. Park, & J. K. Han. Prediction of the Therapeutic Response after FOLFOX and FOLFIRI Treatment for Patients with Liver Metastasis from Colorectal Cancer Using Computerized CT Texture Analysis, *European Journal of Radiology* **85**10, 1867-1874 (2016)
5. Pyradiomics: <https://pyradiomics.readthedocs.io/en/latest/>, ultimo accesso: settembre, 2021
6. R. J. Gillies, P. E. Kinahan & H. Hricak, Radiomics: Images Are More than Pictures, They Are Data, *Radiology* **278-2**, 563-577 (2016)
7. V. Giannini, S. Rosati, A. Defeudis, G. Balestra, L. Vassallo, G. Cappello, S. Mazzetti, C. De Mattia, F. Rizzetto, A. Torresin, A. Sartore-Bianchi, S. Siena, A. Vanzulli, F. Leone, V. Zagonel, S. Marsoni & Daniele Regge, Radiomics Predicts Response of Individual HER2-amplified Colorectal Cancer Liver Metastases in Patients Treated with HER2-targeted Therapy, *International Journal of Cancer* **147-11**, 3215-3223 (2020)
8. R. Klaassen, R. T. H. M Larue, B. Mearadji, S. O Van Der Woude, J. Stoker, P. Lambin, & H. W. M Van Laarhoven, Feasibility of CT Radiomics to Predict Treatment Response of Individual Liver Metastases in Esophagogastric Cancer Patients, *PLoS One* **13-11** (2018)
9. S. X. Rao, D. MJ Lambregts, R. S. Schnerr, R. CJ Beckers, M. Maas, F. Albarello, R. G. Riedl, C. HC Dejong, M. H Martens, L. A Heijnen, W. H Backes, G. L Beets, M-S. Zeng & R. GH Beets-Tan, CT Texture Analysis in Colorectal Liver Metastases: A Better Way than Size and Volume Measurements to Assess Response to Chemotherapy?, *United European Gastroenterology Journal* **4-2**, 257-263 (2016)
10. M. Ravanelli, G. M. Agazzi, E. Tononcelli, E. Roca, P. Cabassa, G. Baiocchi, A. Berruti, R. Maroldi & D. Farina, Texture features of colorectal liver metastases on pretreatment contrast-enhanced CT may predict response and prognosis in patients treated with bevacizumab-containing chemotherapy: a pilot study including

- comparison with standard chemotherapy, *La Radiologia medica* **124**-9, 877–886 (2019)
11. A. Dohan, B. Gallix, B. Guiu, K. Le Malicot, C. Reinhold, P. Soyer, J. Bennouna, F. Ghiringhelli, E. Barbier, V. Boige, J. Taieb, O. Bouché, E. François, J. M. Phelip, C. Borel, R. Faroux, J. F. Seitz, S. Jacquot, M. Ben Abdelghani, F. Khemissa-Akouz, D. Genet, J. L. Jouve, Y. Rinaldi, F. Desseigne, P. Texereau, E. Suc, C. Lepage, T. Aparicio & C. Hoeffel, Early evaluation using a radiomic signature of unresectable hepatic metastases to predict outcome in patients with colorectal cancer treated with FOLFIRI and bevacizumab, *Gut* **69**-3, 531–539 (2020)
 12. R. Nakanishi, E. Oki, H. Hasuda, E. Sano, Y. Miyashita, A. Sakai, N. Koga, N. Kuriyama, K. Nonaka, Y. Fujimoto, T. Jogo, K. Hokonohara, Q. Hu & Y. Hisamatsu, K. Ando, K., Kimura, Y., Yoshizumi, T., & M. Mori, Radiomics Texture Analysis for the Identification of Colorectal Liver Metastases Sensitive to First-Line Oxaliplatin-Based Chemotherapy, *Annals of surgical oncology* **28**-6, 2975–2985 (2021)
 13. A. Maaref, F. Perdigon Romero, E. Montagnon, M. Cerny, B. Nguyen, F. Vandembroucke, G. Soucy, S. Turcotte, A. Tang & Samuel Kadoury, Predicting the Response to FOLFOX-Based Chemotherapy Regimen from Untreated Liver Metastases on Baseline CT: A Deep Neural Network Approach, *Journal of Digital Imaging* **33**-4, 937-945 (2020)
 14. RECIST: <https://recist.eortc.org/>, ultimo accesso: settembre, 2021
 15. Statistics and Machine Learning Toolbox™: https://it.mathworks.com/help/stats/getting-started-12.html?s_tid=CRUX_lftnav, ultimo accesso: ottobre, 2021
 16. LIBSVM: <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>, ultimo accesso: maggio, 2021
 17. S. Thukral, G. Munjal, Review: a combine approach of feature selection and different classification technique in breast gene data, *International journal of engineering sciences & research technology* vol 6-4, 170-177 (2017).
 18. L. Zheng, F. Chao, N. Mac Parthaláin, D. Zhang & Q. Shen, Feature Grouping and Selection: A Graph-based Approach, *Information Sciences* vol 546, 1256-272 (2021).
 19. R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, & J. Saeed, A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction, *Journal of Applied Science and Technology Trends*, vol. 1- 2, 56 - 70, (2020).
 20. S. Rosati, G. Balestra, V. Giannini, S. Mazzetti, F. Russo & D. Regge, ChiMerge Discretization Method: Impact on a Computer Aided Diagnosis System for Prostate Cancer in MRI, *IEEE International Symposium on Medical Measurements and Applications Proceedings*, 297-302 (2015).
 21. R. Kerber, Chimerge: discretization of numeric attributes, *Proceedings of Association for the advancement of artificial intelligence*, 123-128 (1992).

22. R. Jensen & Q. Shen, *Computational Intelligence and Feature Selection Rough and Fuzzy Approaches* (Hoboken, Wiley IEEE, 2008).
23. S. Surekha, A comparative study of various rough set theory algorithms for feature selection, *International journal of engineering sciences & research technology* vol 6-4, 21-30 (2017).
24. L. Breiman, Random Forests, *Machine Learning* 45, 5–32 (2001)
25. L. Breiman, J. H. Friedman, R. A. Olshen & C. J. Stone, *Classification and Regression Trees* (Boca Raton, FL: Chapman & Hall, 1984).
26. T. Kohonen, Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, 43, 59–69 (1982).
27. A. S. Das, B. Sudeep Padhi, Binary Harmony Search Based Feature Selection and Data Classification Technique, *International journal of innovative research in technology*, vol 6-7, 245-252 (2019).
28. G. I. Sayed, G. Khoriba, & M.H. Haggag, A novel chaotic salp swarm algorithm for global optimization and feature selection. *Applied Intelligence* vol **48**, 3462–3481 (2018)
29. E. S. Hosseini & M. H. Moattar, Evolutionary Feature Subsets Selection Based on Interaction Information for High Dimensional Imbalanced Data Classification, *Applied Soft Computing Journal* vol **82**, 105581-105599 (2019)
30. K. Bera, V. Velche, & A. Madabhushi, Novel Quantitative Imaging for Predicting Response to Therapy: Techniques and Clinical Applications, *American Society of Clinical Oncology*, vol **38**, 1008-1018 (2018).
31. Y. Chang, K. Lafata, W. Sun, C. Wang, Z. Chang, J. P. Kirkpatrick, F. Yin, An investigation of machine learning methods in delta-radiomics feature analysis. *PLOS ONE* **14**-12, (2019)
32. X. Fave, L. Zhang, J. Yang, D. Mackin, P. Balter, D. Gomez, D. Followill, A. Kyle Jones, F. Stingo, Z. Liao, R. Mohan & L. Court, Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. *Scientific Reports* vol **7**, 588 (2017).
33. H. Nasief, C. Zheng, D. Schott, W. Hall, S. Tsai, B. Erickson & X. A. Li, A machine learning based delta-radiomics process for early prediction of treatment response of pancreatic cancer. *npj Precision Oncology* 3, 25 (2019)
34. J. Wei, J. Cheng, D. Gu, F. Chai, N. Hong, Y. Wang, J. Tian, Deep learning-based radiomics predicts response to chemotherapy in colorectal liver metastases, *Medical Physics*, 48(1), 513-522 (2021).