

POLITECNICO DI TORINO

Corso di Laurea Magistrale

in Ingegneria Biomedica

Tesi di Laurea Magistrale

“Costruzione di un classificatore per la stima del rischio di frattura in pazienti osteoporotici”



Relatrice

Prof Gabriella Balestra

Candidato

Tommaso Picciafuoco

Anno accademico 2019/2020

Indice

Abstract.....	1
1 Primo capitolo: introduzione	2
1.1 Sistema scheletrico.....	2
1.1.1 Ciclo vitale dell'osso.....	3
1.2 L'osteoporosi.....	4
1.2.1 Osteoporosi postmenopausale-senile	4
1.2.2 Osteoporosi secondaria.....	5
1.2.3 Statistiche	6
1.3 Diagnosi.....	7
1.3.1 Principio di funzionamento della DEXA	9
1.4 Acquisizione dati	10
1.4.1 Acquisizione lombare	10
1.4.2 Acquisizione femorale.....	12
1.5 Fattori di rischio frattura.....	13
1.6 La ditta "Tecnologie Avanzate" e scopo dello studio	14
1.6.1 Trabecular Bone Score	15
1.6.2 Bone Strain Index	17
1.6.3 Hip Structural Analysis.....	19
1.7 Il Machine Learning nella classificazione.....	19
2. Secondo capitolo: materiali e metodi.....	22
2.1 Dati	22
2.2 Preprocessing	22
2.2.1 Dendrogramma.....	24
2.3 Feature selection.....	27
2.3.1 Genetic algorithm (GA).....	27
2.3.2 Codifica della soluzione	29
2.3.3 Popolazione iniziale	29
2.3.4 Funzione di fitness.....	30
2.3.5 Operatori di selezione.....	33
2.3.6 Operatori di riproduzione	36
2.3.7 Criterio di terminazione.....	38
2.4 Divisione dataset.....	38

2.5 Reti Neurali.....	40
2.5.1 Neurone Artificiale	41
2.5.2 Funzione di attivazione.....	42
2.5.3 Geometria del neurone	44
2.5.4 Apprendimento del neurone artificiale	44
2.5.5 Feedforward network	46
2.5.6 Algoritmo di apprendimento della rete	48
2.5.7 Criteri di terminazione.....	50
2.5.8 Valutazione performance della rete	50
3. Risultati.....	57
3.1 Feature selection.....	57
3.1.1 Costruzione dataset bilanciato.....	57
3.1.2 Algoritmo Genetico.....	58
3.1.2.1 Soluzione Feature Selection con le prime 14 “bloccate”	59
3.1.3 Considerazioni feature selection	62
3.2 Divisione training, test e validation set.....	65
3.2.1 Soluzione n°1 con fitness pari a 0.0920	66
3.2.2 Soluzione n°2 con fitness pari a 0.1061	67
3.2.3 Soluzione n°3 con fitness pari a 0.0883	68
3.2.4 Soluzione n°4 con fitness pari a 0.1002	69
3.2.5 Soluzione n°5 con le feature utilizzate in clinica.....	70
3.2.6 Considerazioni clustering.....	71
3.3 Reti Neurali.....	72
3.3.1 Valutazione delle prestazioni.....	72
3.3.2 Considerazioni sulle reti neurali	88
4. Conclusioni	90
Riferimenti bibliografici.....	92
Appendice A	94

Abstract

Lo scopo dello studio è di valutare la possibilità di stimare il rischio frattura in pazienti osteoporotici attraverso lo sviluppo di un classificatore basato su reti neurali. Il rischio di frattura è influenzato da parametri quali T-score, menopausa, età e fratture pregresse. Ad oggi in ambito clinico, per la stima del rischio, è utilizzata una combinazione di parametri tra cui dati relativi al paziente (età, peso, ecc...) ed altri relativi alle principali zone anatomiche soggette a frattura. Lo studio è stato condotto su 364 pazienti di cui 278 a rischio frattura e 86 non a rischio, costruendo un dataset composto da 46 features per ogni paziente. È stata effettuata una features selection utilizzando gli algoritmi genetici e successivamente sono state allenate diverse strutture di reti neurali. La valutazione delle prestazioni è stata effettuata utilizzando il valore AUC. Il metodo proposto ha evidenziato un'accuratezza nella stima del rischio di frattura, una specificità e sensibilità di circa l'80%, a differenza di un'accuratezza del 65% per la classificazione effettuata mediante i soli parametri utilizzati in clinica. A fronte di queste considerazioni, questo lavoro di tesi dimostra come utilizzando tecniche di machine learning sia possibile ottenere un miglioramento nell'accuratezza della stima del rischio di frattura. Nonostante i risultati siano promettenti, il metodo necessiterebbe di un'ulteriore validazione mediante l'utilizzo di un database più ampio e correttamente bilanciato tra le due classi di pazienti.

1 Primo capitolo: introduzione

1.1 Sistema scheletrico

Il sistema scheletrico [1] è costituito da ossa, articolazioni e dai tessuti ad esse associati come ad esempio cartilagine e legamenti. La struttura ossea costituisce il 18% del peso di un individuo adulto, il restante è composto da 43% muscoli, 25% pelle e grassi, 11% visceri e 3% tessuti nervosi. Gli elementi ossei sono gli organi individuali del sistema scheletrico. Un elemento osseo è costituito da: tessuto osseo, tessuto sanguigno e tessuto nervoso. Alcuni di questi elementi contengono il midollo osseo ed altri invece sono rivestiti, nella zona articolare, da tessuto cartilagineo. Le funzioni del sistema scheletrico e dei suoi componenti sono:

- costituire l'impalcatura del corpo;
- proteggere le strutture vitali;
- fornire aree di ancoraggio ai muscoli e ai legamenti;
- prendere parte alla formazione del sangue attraverso il midollo osseo;
- prendere parte al controllo del contenuto di calcio e fosforo all'interno del corpo.

La funzione del singolo elemento osseo esercita un'influenza decisiva sulla conformazione interna dell'osso e di conseguenza anche sulla configurazione esterna. A livello macroscopico vi sono due tipi di tessuto osseo: osso compatto o corticale, osso spongioso o trabecolare. L'osso corticale è un materiale con densità pari a 2 kg/dm^3 e costituisce: l'intera parte centrale delle ossa lunghe (diafisi), il guscio esterno delle loro estremità (epifisi), il guscio esterno delle ossa corte e i tavolati esterni delle ossa piatte. L'osso spongioso è situato nell'epifisi delle ossa lunghe, tra gli strati di osso corticale delle ossa piatte e all'interno di un sottile guscio di osso corticale nelle ossa corte. L'osso spongioso è anche chiamato trabecolare in quanto è composto da tessuto osseo organizzato in modo da formare tante travate, variamente intrecciate tra loro, che prendono il nome di trabecole. La connessione delle trabecole dà all'osso spongioso un aspetto spugnoso e da questa sua morfologia deriva il suo nome. La densità apparente dell'osso spongioso varia tra $0,15$ e 1 kg/dm^3 . Le trabecole sono prevalentemente orientate secondo le direzioni di

trasmissione delle sollecitazioni a cui è sottoposto l'osso. La superficie esterna dell'osso, ad eccezione delle superfici articolari rivestite da cartilagine, è avvolta da una membrana di tessuto connettivo ricca di vasi sanguigni e di fibre nervose: il periostio. Quest'ultimo si compone di due strati: uno strato fibroso più esterno ricco di vasi sanguigni e nervi, ed uno interno formato da cellule di osteoblasti, osteoclasti e cellule progenitrici osteocondrali. Nel periostio le fibre nervose sensitive costituiscono plessi nervosi responsabili della sua sensibilità al dolore in caso di traumi o danni da sovraccarico. Il tessuto osseo è costituito dalla matrice extracellulare e da cellule ossee. Una matrice ossea matura è costituita in peso secco da circa un 35% di materiale organico (collagene e proteoglicani), mentre il restante 65% da materiale inorganico (cristalli di fosfato di calcio chiamati idrossiapatite). Questi sono responsabili delle principali caratteristiche funzionali dell'osso, il collagene fornisce deformabilità alla matrice ossea, mentre i componenti minerali conferiscono alla matrice resistenza a compressione. Invece, le cellule ossee come osteoblasti, osteoclasti e osteociti sono responsabili dell'attività metabolica dell'osso.

1.1.1 Ciclo vitale dell'osso

L'osso è una struttura metabolicamente attiva ed è sottoposto, durante il corso della vita, ad un processo su scala microscopica chiamato bone turnover (ricambio osseo), che permette un ricambio giornaliero della sua matrice extracellulare al fine di garantire l'omeostasi calcica nel sangue e l'integrità strutturale dello scheletro. Esistono tre tipi di cellule chiave che svolgono diversi ruoli omeostatici: gli osteoclasti, gli osteoblasti e gli osteociti. Gli osteoclasti sono responsabili della degradazione dell'osso disfunzionale, mentre la biosintesi del nuovo osso da sostituire è eseguita dagli osteoblasti. Gli osteociti, le cellule ossee più abbondanti, formano una rete tridimensionale interconnessa in tutto il tessuto osseo agendo come meccanosensori che monitorano lo stress meccanico all'interno dei tessuti ossei e reagendo ai cambiamenti sia di quantità che di direzione del carico applicati sulle ossa. Questo processo è fondamentale in quanto permette di regolare i livelli plasmatici di calcio e fosforo, di rinforzare il tessuto osseo in risposta a stimoli appropriati e per la riparazione di microfratture da stress. In condizioni fisiologiche ottimali, il riassorbimento osseo avviene in circa 10 giorni e la formazione ossea dura circa 3 mesi. Fino al 20% dello scheletro può essere sostituito rimodellandolo ogni anno. In condizioni di equilibrio metabolico la quantità di osso riassorbito e di osso nuovo si

equivale, ma quando la quantità di osso assorbito è maggiore rispetto al neoformato si instaura, prima una condizione di osteopenia, e quando la quantità di minerali scende al di sotto di una determinata soglia, una condizione di osteoporosi.

1.2 L'osteoporosi

L'osteoporosi è una malattia sistemica dell'apparato scheletrico dovuta ad una diminuzione della densità ossea, con un conseguente aumento del rischio di frattura, in particolar modo di vertebre, femore e polso anche per traumi minimi. Come detto precedentemente, questa situazione si instaura quando si ha uno sbilanciamento a favore del riassorbimento osseo nei confronti della formazione del nuovo osso. Si possono distinguere due diversi tipi di osteoporosi: postmenopausale-senile e secondaria.

1.2.1 Osteoporosi postmenopausale-senile

La maggior parte dei casi di osteoporosi sono del tipo postmenopausale-senile, dove, come dice la parola stessa, le principali cause riguardano l'età e la menopausa. Infatti, le donne sono i soggetti più a rischio, innanzitutto perché hanno uno scheletro più leggero e minuto, e in secondo luogo perché, mediamente dopo i 45 anni, vanno incontro alla menopausa, che causa una notevole riduzione della densità ossea. Questa è dovuta al calo di estrogeni che agiscono sull'assorbimento di calcio. Gli estrogeni svolgono infatti un ruolo centrale nella formazione e nel mantenimento del tessuto osseo, ad esempio:

- promuovono il riassorbimento del calcio a livello renale;
- favoriscono la conversione della vitamina D e il conseguente assorbimento intestinale del calcio stesso;
- incrementano la sintesi di calcitonina.

Una loro carenza, invece, stimola una maggiore attività degli osteoclasti (cellule deputate al disassemblaggio del tessuto osseo) e un aumentato riassorbimento. Esistono poi dei fattori secondari che possono aumentare il rischio di osteoporosi e tra questi troviamo:

- la familiarità, infatti la genetica e l'ereditarietà giocano un ruolo di primo piano nello sviluppo dell'osteoporosi e nella suscettibilità alle fratture;

- l'etnia, l'osteoporosi è più frequente nelle popolazioni caucasiche (europei e nordamericani) e asiatiche, che non in quelle africane e di discendenza ispanica.

Infine, esistono dei fattori che dipendono dallo stile di vita e dalle abitudini nutrizionali, che quindi possono essere modificabili, come per esempio:

- un apporto insufficiente di calcio e vitamina D;
- una dieta povera di frutta e verdura;
- un eccesso di proteine, sodio e caffeina;
- il consumo eccessivo di alcool;
- il fumo;
- la magrezza eccessiva e una costituzione minuta;
- la vita sedentaria.

1.2.2 Osteoporosi secondaria

Rientrano invece nelle osteoporosi secondarie tutte quelle causate da un particolare trattamento farmacologico o da alcune patologie. Per fare degli esempi, tra i principali farmaci che causano una riduzione della densità ossea troviamo i glucocorticoidi e le terapie a blocco ormonale adiuvante. Nelle seguenti tabelle si possono vedere le principali patologie e trattamenti farmacologici che possono causare osteoporosi secondaria.

Classe farmacologica	Possibile meccanismo d'azione
Glucocorticoidi	Inibizione attività osteoblastica
Inibitori dell'aromatasi	Ipogonadismo con alto turnover
Inibitori della pompa protonica	Riduzione assorbimento intestinale di calcio

Tabella 1 Principali trattamenti farmacologici che causano una diminuzione della densità ossea.

Malattie endocrine	Malattie ematologiche
Ipogonadismo	Mastocitosi sistemica
Ipercortisolismo	Anemia falciforme
Iperparatiroidismo	Talassemia
Malattie reumatiche	Malattie neurologiche
Artrite reumatoide	Parkinson
Lupus eritematoso sistemico	Sclerosi multipla

Tabella 2 Principali malattie che causano una riduzione della densità ossea.

1.2.3 Statistiche

Nel corso della vita, circa il 40% della popolazione incorre in una frattura del femore, vertebra o polso, in maggioranza dopo i 65 anni. In Italia, il 23% delle donne oltre i 40 anni e il 14% degli uomini con più di 60 anni è affetto da osteoporosi e questi numeri sono in continua crescita, soprattutto in relazione all'aumento dell'aspettativa di vita. Si stima che in Italia l'osteoporosi colpisca circa 5.000.000 di persone, di cui l'80% sono donne in post menopausa.

Le fratture da fragilità per osteoporosi hanno rilevanti conseguenze, sia in termini di mortalità che di disabilità motoria, con elevati costi sia sanitari sia sociali. La mortalità da frattura del femore è del 5% nel periodo immediatamente successivo all'evento e del 15-25% a un anno. Nel 20% dei casi si ha la perdita definitiva della capacità di camminare autonomamente e solo il 30-40% dei soggetti torna alle condizioni precedenti la frattura [2]. Dopo i 75 anni l'osteoporosi è tra le tre malattie croniche più frequenti insieme all'artrosi e all'ipertensione arteriosa. Per quanto riguarda i costi sanitari (figura 1) dovuti a questa patologia è possibile vedere che sono molto maggiori rispetto a infarti, tumori al seno e ictus. Inoltre, il 20% delle fratture sono delle rifratture e prevenirle significherebbe risparmiare 200 milioni di euro.

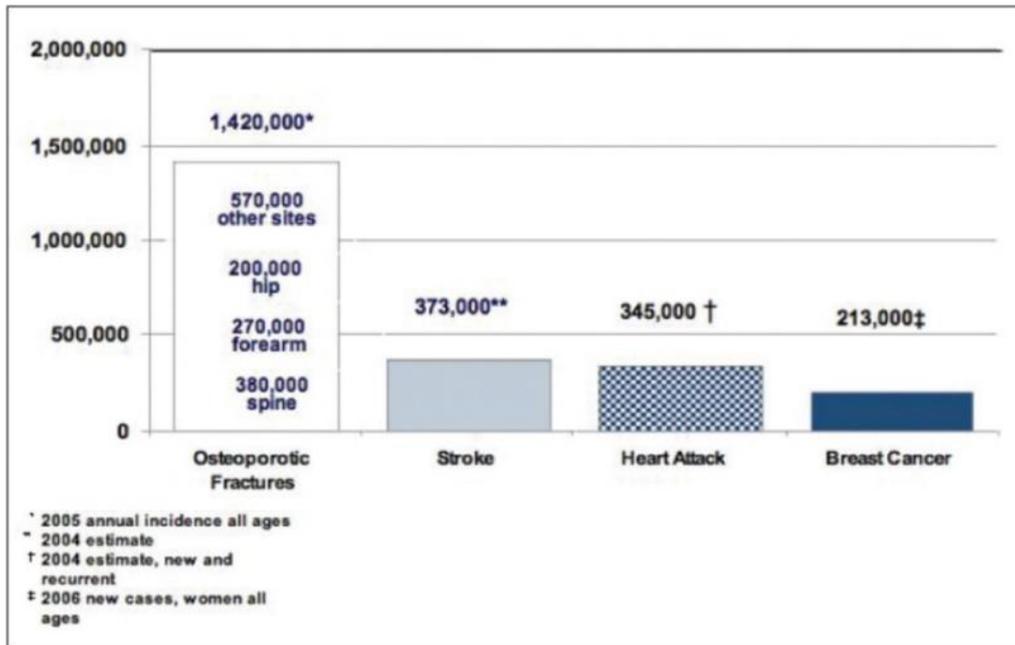


Figura 1 Nella gura è possibile vedere la differenza di costi tra le fratture osteoporotiche, ictus, infarti e tumori al seno [4].

Nel mondo [3] sono 75 milioni le persone colpite da osteoporosi, principalmente in Europa, U.S.A. e Giappone. Nel 2000 sono quasi 4 milioni le fratture da osteoporosi registrate in Europa, di cui quasi 1 milione all'anca, con costi diretti di circa 31,7 miliardi e, secondo alcune stime, nel 2050 saliranno a 76,7 a causa dell'invecchiamento della popolazione. Nel corso della vita le donne hanno 1 probabilità su 6 di subire una frattura all'anca, mentre le probabilità di cancro al seno, l'altra grande minaccia delle donne, è di 1 su 9. Per le donne dopo i 45 anni le giornate di ospedalizzazione per osteoporosi sono superiori a quelle dovute a molte altre malattie come diabete, infarto e cancro al seno.

1.3 Diagnosi

L'indagine densitometrica consente oggi di misurare in modo abbastanza accurato e preciso la massa ossea ed in particolare la sua densità minerale (bone mineral density o BMD) in g/cm² di superficie ossea proiettata. La BMD è responsabile della resistenza meccanica dell'osso per il 60-80%. La BMD è un parametro che restituisce in maniera quantitativa la densità ossea ed è fondamentale per diagnosticare l'osteoporosi. Sappiamo, infatti, che se si sottopone un materiale a compressione la resistenza del materiale al cedimento è proporzionale al quadrato della densità, quindi bastano piccole variazioni di densità per avere un notevole cambiamento alla resistenza a compressione,

stessa cosa per la resistenza allo sforzo di taglio, proporzionale al cubo della densità. Per l'OMS la diagnosi densitometrica di osteoporosi si basa sulla valutazione con tecnica dual-energy x-ray absorptiometry (DEXA o DXA) della densità minerale, raffrontata a quella media di soggetti adulti sani durante il picco di massa ossea. L'unità di misura è rappresentata dalla deviazione standard dal picco medio di massa ossea (T-score). È stato osservato che il rischio di frattura inizia ad aumentare in maniera esponenziale con valori densitometrici di T-score <-2.5 deviazione standard che, secondo l'OMS, rappresenta la soglia per diagnosticare la presenza di osteoporosi. La densitometria ossea rappresenta, quindi, il test diagnostico di osteoporosi e di rischio di frattura. Secondo l'OMS, nell'interpretare i risultati della BMD si adottano le seguenti definizioni:

Stato	T-score
Normale	$-1 < \text{T-score} < +2.5$
Osteopenia	$-2.5 < \text{T-score} < -1$
Osteoporosi	$\text{T-score} < -2.5$
Osteoporosi conclamata	$\text{T-score} < -2.5$ e dalla contemporanea presenza da una o più fratture per fragilità

Tabella 3 Soglie per la diagnosi dell'osteoporosi.

In particolari situazioni, come per donne in menopausa precoce o per uomini al di sotto dei 50 anni, in aggiunta al T-score viene utilizzato il Z-score, definito come la deviazione standard della BMD rispetto a persone della stessa età. In questo caso, la soglia per diagnosticare osteoporosi diventa:

Stato	Z-score
Normale	$-2 < \text{Z-score}$
osteoporosi	$-2 > \text{Z-score}$

Tabella 4 Soglie per la diagnosi dell'osteoporosi in caso di menopausa precoce o uomini sotto i 50 anni.

Dal grafico in figura 2 è possibile vedere che una diminuzione di una sola unità del T-score aumenta esponenzialmente il rischio di frattura. Va precisato che si tratta solo di una

diagnosi densitometrica che può essere tradotta in diagnosi clinica solo dopo una valutazione complessiva di diagnostica differenziale. La soglia diagnostica in T-score, inoltre, non coincide con la soglia terapeutica, poiché altri fattori, scheletrici ed extrascheletrici, condizionano sia il rischio di frattura del singolo soggetto, sia la decisione di intraprendere o meno un trattamento farmacologico.

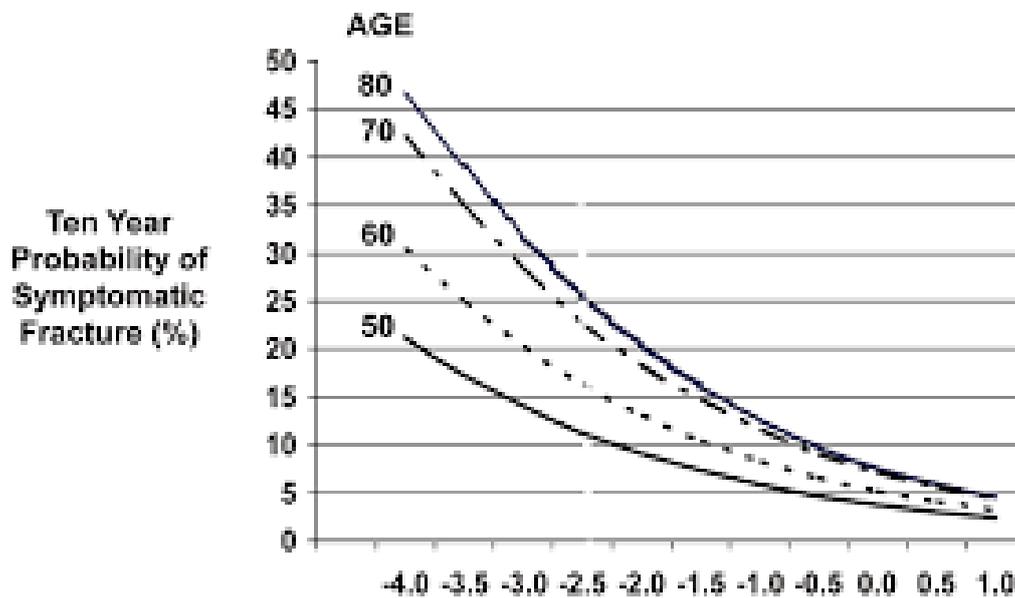


Figura 2 Nel grafico è possibile vedere come diminuisce il rischio di frattura al crescere della BMD, e di conseguenza del T-score [4].

1.3.1 Principio di funzionamento della DEXA

La DEXA [5] è una tecnica densitometrica che utilizza l'assorbimento dei raggi X del distretto scheletrico da esaminare. La misura dell'attenuazione, ad eccezione della Tomografia Computerizzata Quantitativa (QTC), non consente la determinazione del volume nel quale l'osso minerale è distribuito. Infatti, la DEXA è una tecnica proiettiva e si basa sulla rappresentazione bidimensionale della struttura ossea esaminata, simile all'acquisizione della radiologia tradizionale. Le tecniche di densitometria ossea, basate sull'emissione di radiazioni X, hanno sostituito quelle radioisotopiche a singolo (SPA: single photon absorptiometry) e a doppio raggio fotonico (DPA: dual photon absorptiometry), per superare i problemi connessi all'instabilità e al decadimento della sorgente radioattiva. I vantaggi sono importanti: l'aumento della velocità del flusso fotonico e la possibilità di usare un fascio di radiazioni X collimato. Inoltre, la sorgente

radiogena, con una media di corrente al tubo radiogeno di 1mA, produce un flusso fotonico da 500 a 1000 volte più intenso di quello prodotto da una sorgente di 1 curie di Gadolinio 153, usato nella DPA. Altri vantaggi della DXA rispetto alla DPA e alla SPA sono costituiti dalla riduzione dei tempi di scansione, dall'incremento della precisione ed accuratezza, dall'elevata risoluzione spaziale e da una migliore qualità delle immagini. La DXA è stata introdotta nel 1987, è costituita dalla sorgente radiogena, dal sistema che consente la separazione dei due livelli energetici e dall'apparato di rilevazione digitale dei rivelatori. La sorgente a radiazioni X è posizionata sotto il lettino e si muove in maniera sincrona con un sistema di rivelatori posti al di sopra del paziente e questo spiega la definizione di proiezione postero-anteriore (P-A). In base alla modalità di produzione del fascio fotonico, si distinguono:

- sistemi a fascio costante;
- sistemi a fascio pulsato.

I sistemi a fascio costante emettono un fascio radiante costante per intensità e tensione, successivamente filtrato, in modo da ottenere due fasci distinti a differenti bande di energia. Il sistema a fascio pulsato, invece, emette alternativamente un fascio fotonico prodotto, istante per istante, a due differenti tensioni, in modo da generare due distinti fasci fotonici. Durante la scansione del paziente il computer ricostruisce, pixel per pixel, l'immagine della sezione in esame. La distinzione tra osso e tessuti molli viene identificata automaticamente dall'apparecchiatura, sulla base della densità di ciascun pixel.

1.4 Acquisizione dati

I dati sono stati estratti da immagini DXA ottenute da densitometri Hologic modelli Horizon e Discovery. Le immagini derivanti da tali apparecchiature hanno permesso la raccolta di informazioni riguardanti peso/altezza/età del paziente/menopausa, e informazioni sullo stato densitometrico del distretto lombare e femorale.

1.4.1 Acquisizione lombare

Durante una Mineralometria Ossea Computerizzata (MOC) lombare il paziente deve essere posizionato supino. Per minimizzare la curvatura lordotica e rettilineizzare il più possibile la colonna, le gambe devono essere alzate per ottenere un angolo di 90° con la

schiena, che si ottiene utilizzando un supporto per ginocchia. Esistono due diversi tipi di supporto ed è importante per il follow-up utilizzare sempre lo stesso. Per l'acquisizione, la colonna deve essere sempre centrata in modo tale da avere la stessa quantità di parti molli da un lato e dall'altro, accorgimento importante perché la MOC effettui il confronto tra parti molli e tessuto osseo. Per riconoscere e poter numerare correttamente le vertebre, bisogna avere sempre i riferimenti ossei, ovvero includere sempre le ultime coste e le creste iliache. Solitamente le ultime coste stanno ad un'altezza della vertebra T12 e le creste iliache a L5; inoltre le prime tre vertebre L1 L2 L3 hanno una forma a Y mentre L4, un po' più schiacciata, ha forma di H, infine L5 risulta essere molto più schiacciata (figura 3). In aggiunta, ci si può aiutare con i processi trasversi in L3, che con la MOC, sono visibili. Nonostante la vertebra L5 possa essere vista con la MOC, non viene considerata a ni diagnostici, in quanto motivo di un potenziale errore nella misura causato dalle pelvi [5]. Per un corretto follow-up, eccetto casi particolari, si consiglia di ripetere l'esame non prima di 18 mesi, ed è importante simulare le stesse condizioni degli esami precedenti.

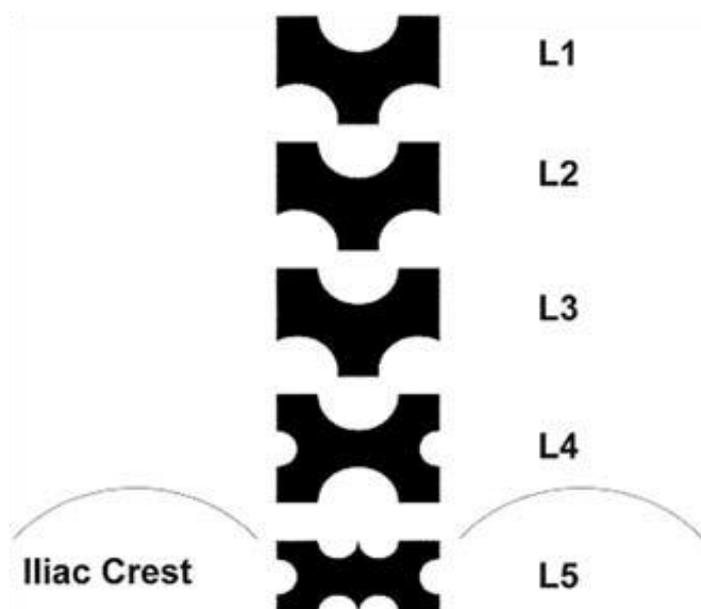


Figura 3 Nell'immagine è possibile vedere la forma delle vertebre lombari da utilizzare come riferimento [5].

Artefatti nella misura:

- errata identificazione della T12 che viene scambiata come L1 che generalmente riduce la BMD del 3.5% con conseguente variazione del T-score;

- in pazienti con scoliosi è importante centrare bene il paziente e identificare le vertebre, è possibile che si ottenga un dato non esattamente attendibile, ma comunque un buon dato per fare diagnosi;
- alterazioni degenerative delle vertebre (40% delle donne dopo i 50 e 85% delle donne dopo i 75), che può portare a valori di T-score tra le vertebre molto diversi tra loro. Le linee guida della società di densitometria internazionale dicono che vanno escluse quelle vertebre con valore di T-score che differisce dalle altre di almeno una unità. Tendenzialmente si escludono le vertebre con un dato migliore (T-score maggiore) a costo di sovrastimare il rischio di frattura del paziente. È necessario mantenere sempre almeno due vertebre altrimenti bisogna cambiare sito della misura (l'esclusione o l'inclusione delle vertebre rientra gli errori più frequenti). La presenza di elementi metallici è facilmente visibile da un'elevata BMD, ovviamente anche in questo caso quelle vertebre verranno escluse dall'esame;
- presenza di osteociti che può causare un significativo aumento della BMD;
- calcificazione aortica, patologia estremamente presente nelle donne in menopausa che può causare un aumento lieve nella misura della BMD;
- sindrome delle faccette articolari che, a differenza della calcificazione aortica, può causare un importante aumento della BMD, specialmente in quelle di grado 2 e 3.

1.4.2 Acquisizione femorale

Le principali regioni di interesse nella diagnosi di osteoporosi sono il collo del femore, il triangolo di Ward, la regione trocanterica e il corpo del femore. Ognuna di queste zone contengono una differente percentuale di osso corticale e trabecolare. Da recenti studi si è visto che non ci sono significative differenze nella misura di BMD tra il femore destro e sinistro, se non fosse possibile misurarle entrambe, un ragionevole metodo nella scelta, potrebbe essere quello di utilizzare la gamba non dominante, in quanto potrebbe contenere una densità ossea minore. Per il posizionamento del paziente si utilizza un device che garantisce un 15-20% di rotazione interna delle gambe. Come riferimento si ha il piccolo trocantere, che permette di vedere se la rotazione è eccessiva o troppo poca (nell'immagine il piccolo trocantere si usa come riferimento quindi non deve scomparire). Nel follow-up si deve garantire sempre lo stesso grado di rotazione, in caso di rotazione

eccessiva meglio continuare e propagare l'errore, ma avere un corretto follow-up. Per il corretto posizionamento, l'inquadratura deve restare un po' al di sotto del piccolo trocantere, comprendere la porzione acetabolare, tutta la testa del femore e spostarsi un po' più all'esterno della regione trocanterica (figura 4). Il collo del femore deve essere posizionato al centro della midline in modo perpendicolare, con la stessa quantità di parti molli da una parte e dall'altra. Molto importante il femore deve essere dritto. Per una questione di riproducibilità, il box del collo del femore deve essere adagiato al trocantere e non deve toccare la branca pubica. Per un corretto follow-up, ad eccezione di casi particolari, si consiglia di ripetere l'esame dopo 18-24 mesi. Sia per la zona femorale che lombare è sconsigliato ripetere l'esame prima di 12 mesi in quanto la precisione dello strumento non consente di misurare le differenze tra due esami. Complessivamente, nel caso in cui ci fosse una classificazione come osteopenia nella zona lombare ed osteoporosi in quella femorale (o viceversa), il paziente viene classificato come osteoporotico.



Figura 4 Nella gura è possibile vedere il corretto posizionamento dell'inquadratura in una acquisizione femorale [5].

1.5 Fattori di rischio frattura

Come detto precedentemente, altri fattori influenzano il rischio di frattura, e vengono utilizzati per una diagnosi complessiva. I fattori di rischio sono molteplici, ma i principali sono:

- Ridotta BMD;
- invecchiamento che causa un rallentamento delle misure protettive, un'alterazione della struttura ossea e un aumento della frequenza di cadute che comportano un aumento del rischio di frattura (a parità di BMD il rischio di frattura è più elevato negli anziani);
- Menopausa precoce;
- Elevato Bone Mass Index (BMI) calcolato come rapporto tra peso espresso in kilogrammi e il quadrato dell'altezza espressa in metri;
- Fratture pregresse che aumentano il rischio di nuove fratture indipendentemente dalla BMD: è dimostrato che pazienti con più di 3 fratture pregresse mostrano un rischio di frattura 10 volte più elevato rispetto a chi non ne ha mai avute;
- Familiarità, che aumenta il rischio di frattura indipendentemente dalla BMD, soprattutto per quanto riguarda le fratture di femore;
- Comorbidità;
- Terapie farmacologiche.

1.6 La ditta “Tecnologie Avanzate” e scopo dello studio

Tecnologie Avanzate nasce nel 1971 con lo scopo di introdurre nel mercato italiano tecnologie medicali innovative nell'ambito della fisica nucleare e diagnostica RX, specializzandosi poi nel corso degli anni in nuovi settori come la radioterapia, neurochirurgia, radiologia e imaging. Inoltre, si impegna nella ricerca e sviluppo di nuove tecnologie e software come algoritmi di contornamento e analisi, tra cui il Bone Strain Index. Quest'ultimo software permette il calcolo automatico delle deformazioni e degli stress di segmenti ossei attraverso il metodo ad elementi finiti. Può essere facilmente installato su una DEXA, e insieme ai principali fattori utilizzati per la diagnosi di osteoporosi e calcolo del rischio di frattura, fornisce una caratterizzazione meccanica completa dell'osso attraverso un indice del rischio di frattura, chiamato Bone Strain Index (BSI), e immagini che permettono di visualizzare la distribuzione delle deformazioni di un determinato distretto corporeo. Scopo di questo studio è proprio quello di valutare come

l'utilizzo di questo parametro incida nella classificazione del rischio di frattura in pazienti con osteoporosi. Come è possibile vedere dal grafico sottostante solo un 30% delle fratture è associata ad un basso valore di BMD, mentre molte delle fratture rientrano in una classificazione di osteopenia, questo a indicare che ci sono altri fattori che influiscono nel rischio di frattura.

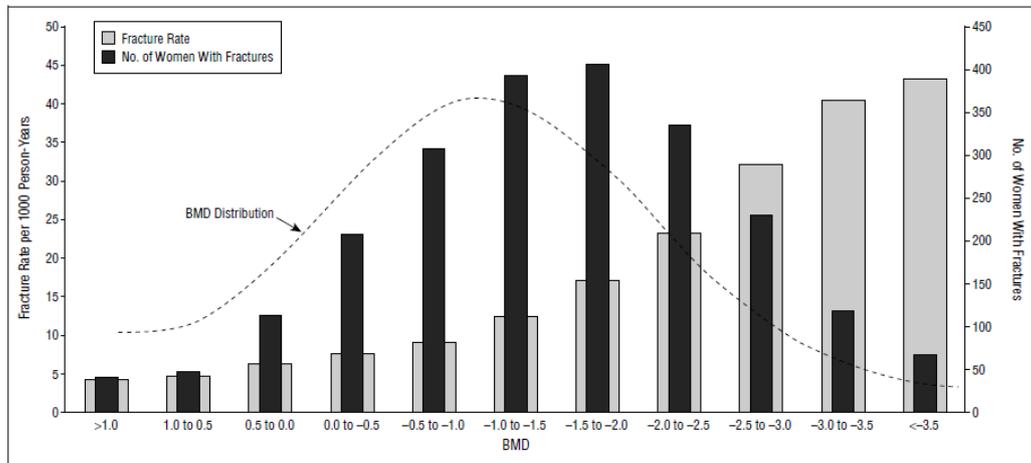


Figura 5 Dall'andamento della distribuzione della BMD si vede che la maggior parte delle fratture si hanno per T-score compresi tra -1 e -2 quindi in zona di osteopenia [6].

Quindi, oltre alla BMD è importante utilizzare altri parametri che ci forniscano informazioni sulla qualità dell'osso.

1.6.1 Trabecular Bone Score

Sappiamo infatti che l'osso è costituito sia da osso compatto, che da osso trabecolare, in particolar modo per quanto riguarda le vertebre che sono costituite per la maggior parte da osso trabecolare perché deve resistere soprattutto a sforzi di compressione. Inoltre, sappiamo che con l'avanzare dell'età, queste subiscono una deformazione dovuta allo schiacciamento. Una semplice misura della BMD non riuscirebbe a rilevare queste deformazioni in quanto la densità ossea non varia, a differenza della texture ossea. Quindi, proprio per misurare la distribuzione spaziale dell'osso, o indice di texture ossea, è stato introdotto un nuovo parametro, il Trabecular Bone Score (TBS). Il TBS è un indice strutturale che valuta le variazioni di livello di grigio dei pixel nell'immagine DXA della colonna lombare, fornendo un indice indiretto della microarchitettura trabecolare. Il TBS

non è un indice diretto della misurazione della microarchitettura ossea, ma piuttosto una visione d'insieme calcolato dalla proiezione della struttura 3D su un piano 2D.

Nella figura sottostante si possono osservare due esempi con stesso volume osseo e volume totale (BV/TV Bone Volume/Total Volume) ma diversa distribuzione spaziale.

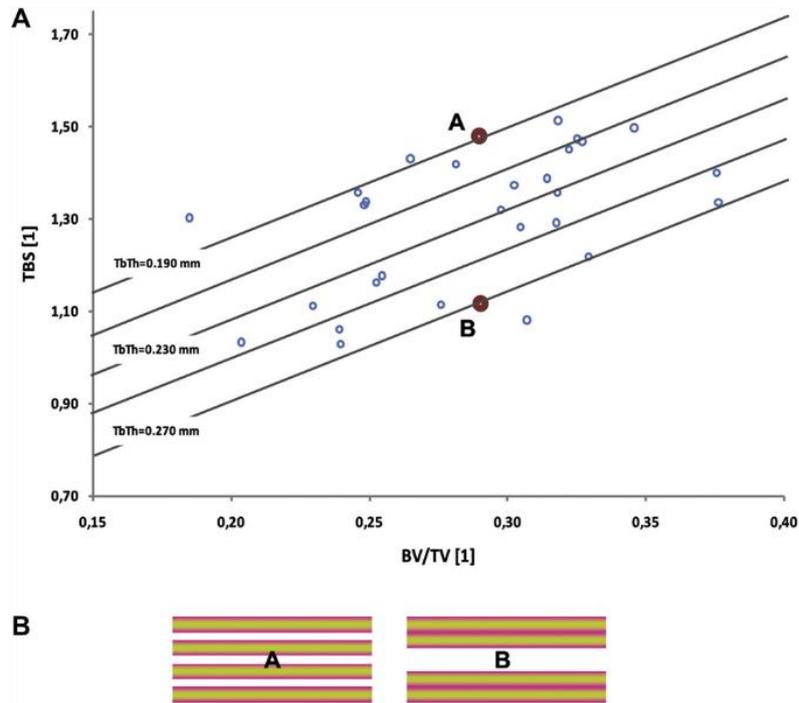


Figura 6 Due esempi con stesso volume ma diversa texture ossea [9].

Questo parametro risulta fondamentale anche per la prevenzione alla frattura in alcune osteoporosi secondarie, come ad esempio per i diabetici, che non mostrano una diminuzione di densità ossea, ma presentano comunque un notevole rischio di frattura. Nell'immagine 7 è possibile vedere un esempio di due vertebre con circa lo stesso BMD ma diverso TBS.

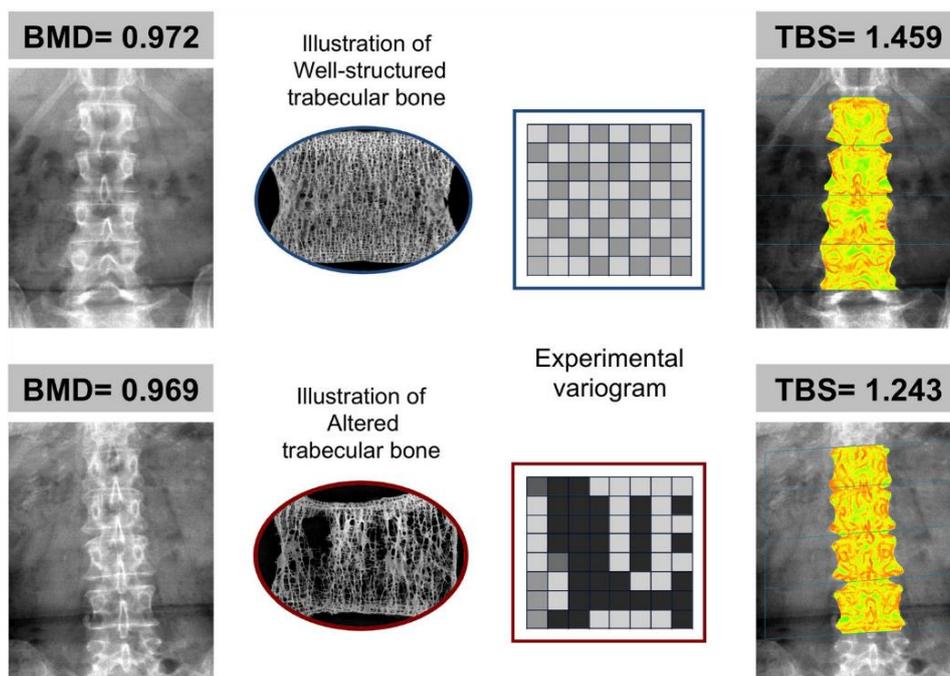


Figura 7 Dalle immagini è possibile vedere due diversi esami con stessa BMD ma TBS diversa [8].

1.6.2 Bone Strain Index

Quando si parla di resistenza meccanica di un materiale è importante fare riferimento non solo alla composizione del materiale, ma anche alla sua geometria e alle condizioni di carico. A questo scopo è stato introdotto un nuovo parametro [9], il Bone Strain Index (BSI), che corrisponde alla deformazione interna calcolata con il metodo degli elementi finiti. È interessante vedere come, in seguito all'applicazione di carichi di diversa intensità, il TBS varia, ma in maniera costante a prescindere dall'intensità del carico applicato, a differenza del BSI che varia in maniera proporzionale, indicando valori di deformazione ossea maggiori per carichi maggiori. In figura 8 si possono osservare le variazioni di questi parametri per i diversi carichi. Come il TBS può essere importante per pazienti diabetici, il BSI risulta molto utile per patologie come la mastocitosi.

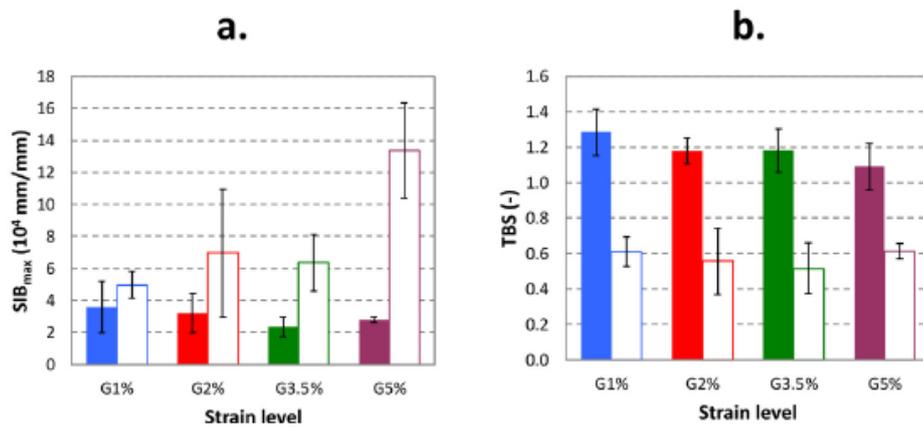


Figura 8 Variazione del BSI e del TBS per i diversi carichi applicati, le barre piene corrispondono al valore prima dell'applicazione del carico mentre quelle vuote dopo l'applicazione [9].

Nell'immagine sottostante è possibile vedere un esempio di DEXA e relativa rappresentazione attraverso il metodo ad elementi finiti, dove nella parte superiore è rappresentata una colonna vertebrale senza fratture, mentre nella parte inferiore una vertebra fratturata.

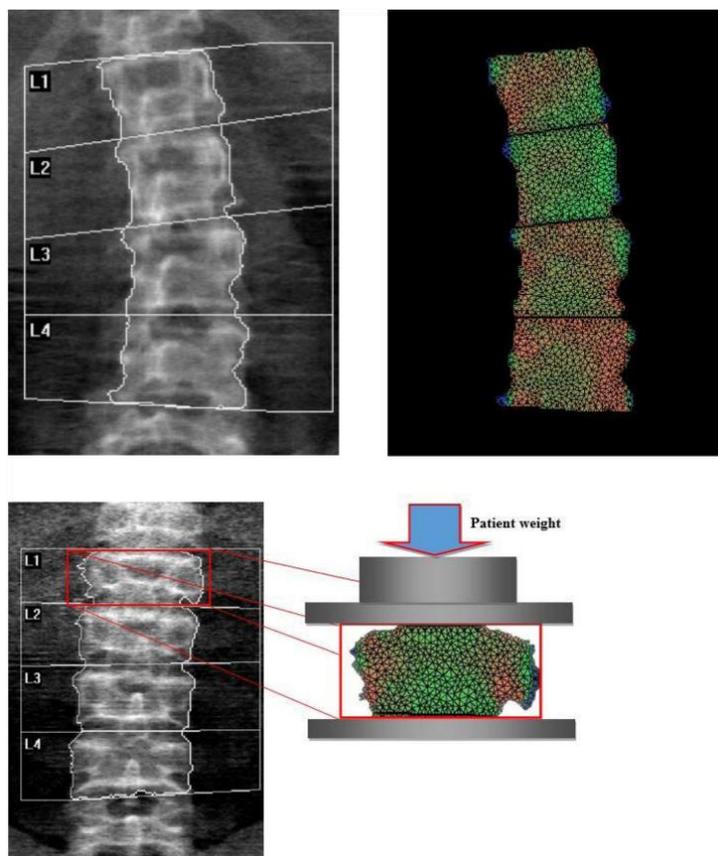


Figura 9 Rappresentazione dell'analisi ad elementi finiti [10].

1.6.3 Hip Structural Analysis

Per quanto riguarda la struttura geometrica del femore, la DEXA restituisce anche la Hip Structural Analysis (HSA), ossia dei parametri relativi alla struttura geometrica associati alla resistenza ossea della zona intertrocanterica, del collo del femore e del corpo del femore. Questi parametri corrispondono a:

- CSA cross section area, relativa alla resistenza a carichi diretti assialmente;
- CSMI crossection moment inertia e Z modulus, relativi alla resistenza a carico flettente;
- BR buckling ratio, relativo stabilita della corteccia sotto carico complessivo;
- Lo spessore dell'osso corticale;
- Diametro interno della corteccia;
- Spessore subperiostale.

Infine, fornisce il NECKANGLE (angolo tra gli assi derivati del collo e del corpo del femore) e l'HAL (Hip Axis Length) distanza dal bordo pelvico al margine esterno del grande trocantere lungo l'asse del collo.

1.7 Il Machine Learning nella classificazione

L'apprendimento attraverso le esperienze è alla base dell'intelligenza umana: imparare da situazioni passate è una delle principali caratteristiche del cervello. Dai primi anni '70 si è cercato di insegnare ad una macchina a ragionare come un uomo, ad elaborare dati come un cervello e, attraverso questi, imparare a riconoscere determinati pattern, con il vantaggio che una macchina riesce ad elaborare una quantità di dati molto maggiore rispetto ad un cervello umano. Questo aspetto è alla base del Machine Learning. Può essere applicato dal campo della finanza, a quello dell'ingegneria e a quello della medicina ad esempio per predire l'andamento di un particolare mercato, o la rottura di un pistone o diagnosticare in tempo una determinata patologia. Classificare correttamente una malattia significa poter intervenire tempestivamente e allo stesso tempo risparmiare tempo e denaro per effettuare esami di approfondimento qualora non ce ne fosse

bisogno. Esistono principalmente due tipi di algoritmi di machine learning definiti come algoritmi di apprendimento supervisionato e non supervisionato.

- Apprendimento supervisionato

Quando si parla di apprendimento supervisionato [12] ci si riferisce a tutti quegli algoritmi dove è conosciuta a priori la classe di appartenenza di ogni osservazione, ad esempio se si tratta di riconoscimento di tumori si è a conoscenza a priori se il tumore è presente o meno. All'interno di questi algoritmi bisogna però fare un'ulteriore distinzione tra classificazione e regressione. Si tratta di classificazione quando il dominio di uscita è discreto, ovvero quando bisogna riconoscere o meno la presenza di una patologia, a differenza della regressione che ha un dominio di uscita continuo, dove ad esempio all'interno di un'immagine bisogna rilevare la profondità di un oggetto, in questo caso non si classifica in vicino o lontano ma si cerca di individuare una profondità in un dominio idealmente infinito.

- Apprendimento non supervisionato

A differenza di quello supervisionato non si è a conoscenza della classe di appartenenza [12] quindi, in questo caso, l'obiettivo è quello di raggruppare tutte le nostre osservazioni del database in sottoinsiemi il più possibile simili tra loro. Il clustering è uno dei principali metodi e prevede la suddivisione del data-set in diversi cluster attraverso una metrica predefinita, facendo in modo che le osservazioni all'interno di un singolo sottoinsieme siano il più possibile simili tra loro. Esistono diverse tecniche di clustering come ad esempio il dendrogramma e il k-means.

DIMENSIONALITY REDUCTION

Quando si hanno a disposizione un elevato numero di dati, la cosa fondamentale da fare, prima di istruire un algoritmo a riconoscere un pattern, è capire se e quali features sono indispensabili al nostro caso. Avere tanti dati non è sempre un vantaggio. Infatti, si possono avere non solo dati inutili o ridondanti, che quindi non farebbero altro che richiedere un costo computazionale elevato, ma anche dati controproducenti ovvero che rischiano di peggiorare le performance del nostro algoritmo. Anche in questo caso, quando si parla di dimensionality reduction, ci si riferisce principalmente a due metodi: feature construction e feature selection.

- Feature construction

La principale caratteristica della feature construction è quella di costruire un nuovo insieme di feature sfruttandone la correlazione tra due o più [11], modificandone quindi il valore, ma mantenendo sempre l'informazione al suo interno. Può risultare vantaggioso quando tutte le feature risultano importanti ai fini della classificazione e quindi il database continua ad essere troppo grande, ma risulta una tecnica di difficile comprensione in quanto le nuove feature sono diverse dalle originali. Le principali tecniche di feature construction sono la Principal Component Analysis (PCA) e la Linear Discriminant Analysis (LDA).

- Feature selection

In questo caso le feature non vengono modificate, ma si cerca di trovare il sottoinsieme che permette di ottenere le prestazioni migliori [11], eliminando quelle ridondanti e quelle dannose ai fini della classificazione. Questa tecnica è di più facile interpretazione in quanto il significato delle feature rimane preservato, ma vengono semplicemente selezionate solo quelle rilevanti. Le feature rilevanti possono essere definite in due modi: forti, ovvero quelle feature che da sole hanno un'elevata importanza per le performance; deboli, ovvero quelle feature che da sole non hanno un gran peso, ma, che insieme ad altre, diventano fondamentali. Anche in questo caso la feature selection può essere del tipo supervisionato o non supervisionato. I due principali modelli di feature selection sono: filter e wrapper. Con il modello filter vengono valutate le performance a prescindere dall'algoritmo di apprendimento, invece con il metodo wrapper la qualità viene valutata durante l'algoritmo di apprendimento. Si possono distinguere tre diversi tipi di ricerca: quella backwards dove si parte con tutto il database e le feature vengono a mano a mano rimosse, quello forwards dove al contrario vengono aggiunte ed infine quello random dove vengono casualmente aggiunte o rimosse alcune feature. In caso di grandi quantità di dati è raccomandato utilizzare dati diversi per la feature selection e per l'allenamento di un algoritmo di apprendimento. La feature selection può essere vista come un problema di ottimizzazione, dove dato un insieme di feature l'obiettivo è trovare il sottoinsieme che restituisce le performance migliori. Tra i principali algoritmi di ottimizzazione troviamo gli evolutionary computation (come il GA Genetic Algorithms) e i computational swarm intelligence (come l'ACO Ant Colony Optimization).

2. Secondo capitolo: materiali e metodi

2.1 Dati

I dati sono stati acquisiti da tre differenti strutture per un totale di 584 pazienti, di cui 45 di sesso maschile e 539 di sesso femminile. Per ogni paziente sono presenti una o più misure della zona lombare e femorale effettuate tramite una DEXA Horizon della Hologic. Queste, per quanto riguarda i pazienti fratturati, possono essere state effettuate sia a frattura già avvenuta che a cavallo della frattura, ovvero di cui si dispongono misure sia prima che dopo l'evento. Per la classificazione, i pazienti con una frattura pregressa sono stati considerati "a rischio", a prescindere dal fatto che si abbiano a disposizione esami prima o dopo la frattura. Invece, per i pazienti senza fratture pregresse, sono stati valutati i T-score delle tre zone anatomiche (lombare, femorale e collo del femore) e sono stati considerati a rischio con un T-score minore di -3 anche solo in una delle aree. Il resto dei pazienti è stato classificato come non a rischio. Del primo centro si hanno a disposizione 5 pazienti maschi e 238 pazienti donne di cui 222 in menopausa e 16 non in menopausa, del secondo 13 pazienti maschi e 138 donne di cui 136 in menopausa e 2 non in menopausa ed infine del terzo 27 pazienti maschi e 163 donne di cui 155 in menopausa e 8 non in menopausa. Di ogni paziente si hanno a disposizione, in aggiunta a tutte le caratteristiche elencate nel capitolo precedente, lo stato del paziente (fratturato o meno) e la zona della frattura (lombare o femorale), l'area delle singole vertebre, l'area delle tre principali zone femorali (collo del femore, intertrocanterica e trocanterica) e la Body Mass Composition (BMC) femorale e lombare, per un totale di 46 feature.

2.2 Preprocessing

Come prima cosa, vista la ridotta numerosità dei pazienti maschili, si è deciso di non considerarli nello studio in quanto il sesso è uno dei principali fattori nella stima del rischio frattura. Questo per cercare di rendere il classificatore quanto più specifico. Stesso ragionamento è stato fatto nei confronti di quelle pazienti non in menopausa, anche questo fattore di fondamentale importanza nell'aumento del rischio frattura. Successivamente, per quanto riguarda l'area delle vertebre, si è deciso di considerare la media delle vertebre presenti nella misura anziché la somma delle aree. Questo a causa della mancanza di alcune vertebre, che, per i motivi espressi nel capitolo precedente, non

sono state considerate. Prendere la somma avrebbe portato ad un errato confronto indipendente dallo stato delle vertebre del paziente. Inoltre, si è deciso di escludere dalle feature il TBS in quanto risultava mancare nella maggior parte delle osservazioni. A questo punto si è passati all'analisi dei dati mancanti, e si è deciso di rimuovere manualmente tutte quelle osservazioni che ne presentavano almeno uno. Questo ha fatto sì che i pazienti disponibili scendessero a 364. Il principale problema dei dati mancanti risiede nei parametri dell'HSA che risultano mancare di frequente. Infine, si è passati alla scelta della misura da tenere in considerazione per ogni paziente. Per quanto riguarda i pazienti a rischio frattura l'obiettivo è quello di prevenire l'evento di conseguenza, dove possibile, si è presa come misura di riferimento quella precedente alla frattura, o, in alternativa, la prima misura disponibile. Al contrario, per quelli non a rischio, si è scelto di prendere come misura di riferimento la prima disponibile in modo tale da posizionarsi in una situazione di "sicurezza". Il dataset così composto risulta essere pronto per la costruzione del classificatore con 364 pazienti di cui 278 a rischio frattura e 86 non a rischio. Questo significativo sbilanciamento tra le due classi potrebbe però causare una polarizzazione nella classificazione a favore della classe più numerosa, di conseguenza, per la feature selection, si è deciso di creare un dataset bilanciato utilizzando tutte le 86 osservazioni dei pazienti non a rischio ed estraendo lo stesso numero di osservazioni dai pazienti a rischio. Per quanto riguarda il metodo per la divisione dei dati, esistono diverse tecniche. Ad esempio, potrebbero essere divisi semplicemente scegliendo in maniera casuale le osservazioni di ogni classe, questo porterebbe però ad una rappresentazione non uniforme dei dati a disposizione. Infatti, per ogni classe, è possibile suddividere ulteriormente le osservazioni in sottoinsiemi, o cluster, simili tra loro, in modo tale poi da poter selezionare le osservazioni da ognuno di questi cluster e fare in modo quindi che ogni sottoinsieme sia rappresentato nel nuovo dataset. Questo metodo prende il nome di clustering e consiste nel raggruppamento di individui di una popolazione al fine di scoprirne la struttura dei dati. In un certo senso, vorremmo che gli individui all'interno di un gruppo fossero vicini o simili tra loro, ma dissimili dagli individui di altri gruppi. In letteratura [17] vengono proposti diversi metodi, ma i più diffusi sono:

- Di tipo gerarchico, che si basano su raggruppamenti tramite una matrice di dissomiglianza;
- A partizioni rapide, che richiedono a priori una suddivisione del dataset in partizioni;

- Modelli misti, che esprimono la funzione di densità di probabilità come somma delle densità dei componenti;
- Metodi della somma dei quadrati, che minimizzano l'errore della somma dei quadrati;
- Clustering spettrale, che utilizzano gli autovettori di un grafico Laplaciano.

Nel nostro caso si è scelto un metodo di tipo gerarchico, nello specifico il dendrogramma, che a differenza degli altri risulta essere di tipo deterministico.

2.2.1 Dendrogramma

Le procedure di clustering gerarchico sono i mezzi più comunemente utilizzati per riepilogare la struttura dei dati. Un albero gerarchico è un insieme nidificato di partizioni rappresentato da un diagramma ad albero o da un dendrogramma (figura 18) [17]. Sezionare un albero a un determinato livello produce una partizione in N gruppi disgiunti. Un esempio di una classificazione gerarchica è la classificazione del regno animale dove ogni specie appartiene a una serie di gruppi nidificati di dimensioni crescenti con un numero decrescente di caratteristiche comuni. Nel produrre un diagramma ad albero, è necessario ordinare i punti in modo che i rami non si incrocino. Questo ordinamento è in qualche modo arbitrario, inoltre, non altera la struttura dell'albero ma solo il suo aspetto. C'è un valore numerico associato ad ogni posizione sull'albero dove si uniscono i rami e questo è una misura della distanza o della dissomiglianza tra i due cluster uniti.

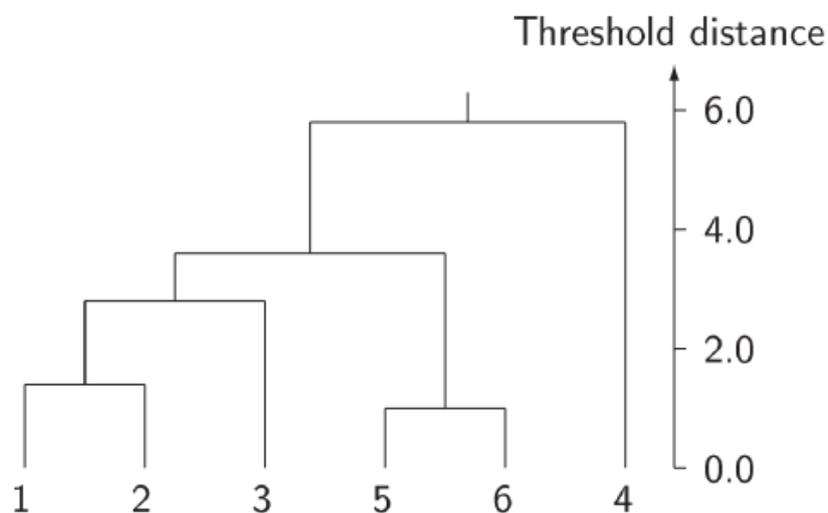


Figura 10 Diagramma ad albero o dendrogramma.

Diverse misure delle distanze tra i cluster danno origine a diverse strutture gerarchiche. Sezionare un albero suddivide i dati in un determinato numero di cluster. Esistono diversi algoritmi per trovare un albero gerarchico, ad esempio un algoritmo agglomerativo [17] inizia con n sotto-cluster, ciascuno contenente un singolo dato e in ciascuna fase unisce i due gruppi più simili per formare un nuovo cluster, riducendo così il numero di cluster di uno. L'algoritmo procede fino a quando tutti i dati rientrano in un singolo cluster. Un algoritmo di divisione, invece, opera al contrario, dividendo successivamente i gruppi a partire da un singolo gruppo e continuando fino a quando non ci sono n gruppi, ciascuno di un singolo individuo. Generalmente, gli algoritmi di divisione sono inefficienti dal punto di vista computazionale. In letteratura [17] vengono proposti diversi metodi che variano a seconda della metrica utilizzata per la misura della similarità, i più diffusi sono:

- Single-link
- Complete-link
- Somma dei quadrati
- Distanza dal centroide
- Distanza dalla mediana
- Media del gruppo

2.3.5.1 Metodo complete-link

Nel metodo complete-link, la distanza tra due gruppi A e B è la distanza tra i due punti più lontani [17], ognuno preso da ciascun gruppo, ovvero:

$$d_{AB} = \max_{i \in A, j \in B} d_{ij}$$

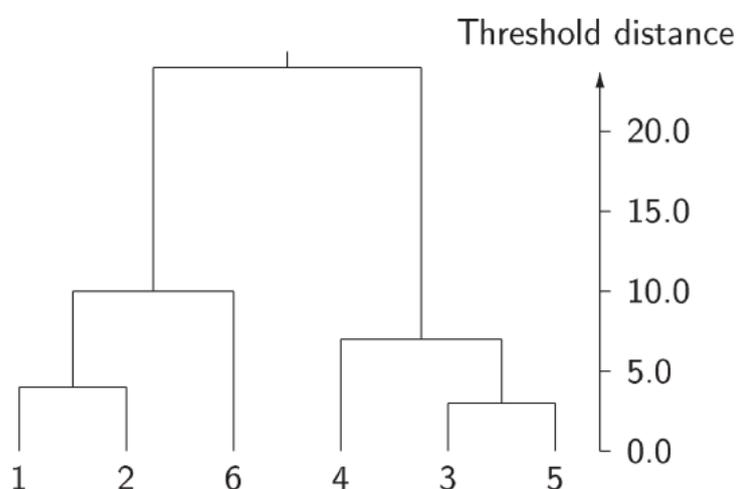
Di seguito è riportato un esempio di come lavora questo algoritmo. Nella situazione di partenza si dispone di 6 elementi con la corrispondente la matrice di dissimilarità, dove vengono espresse le distanze tra ogni singolo elemento che compone un cluster a sé stante.

	1	2	3	4	5	6
1	0	4	13	24	12	8
2		0	10	22	11	10
3			0	7	3	9
4				0	6	18
5					0	8.5
6						0

A questo punto i primi elementi che vengono uniti a formare un cluster sono gli elementi 3 e 5 in quanto sono i più simili. Per la costruzione della nuova matrice di dissimilarità entra in gioco la metrica scelta di conseguenza:

	1	2	(3, 5)	4	6
1	0	4	13	24	8
2		0	11	22	10
(3, 5)			0	7	9
4				0	18
6					0

Dove la distanza tra i nuovi cluster viene definita come la distanza massima tra i punti più lontani di ogni cluster. È per questo motivo che la matrice di dissimilarità varia di metodo in metodo, in quanto è strettamente legata alla metrica utilizzata. Una volta uniti tutti i cluster si ottiene il seguente risultato:



Da questo grafico è possibile vedere come un maggiore "salto" (ovvero una maggiore distanza tra due cluster) implica una maggiore differenza tra le osservazioni appartenenti

ai due cluster. Una volta ottenuto il dendrogramma, l'ultima scelta ricade nel punto di taglio, ovvero nello scegliere a che altezza tagliare l'albero e di conseguenza quanti e quali cluster selezionare. Per la scelta del taglio non esiste un vero e proprio metodo perché dipende da ogni singolo caso. Tipicamente si consiglia di tagliare dove si ha un "salto" maggiore che corrisponde ad una maggiore dissimilarità tra i cluster.

Una volta suddivisi tutti i pazienti in determinati cluster posso essere estratti random gli 86 pazienti per formare il dataset bilanciato.

2.3 Feature selection

Prima di addestrare un algoritmo a riconoscere un determinato pattern è bene verificare se e quali delle feature che si hanno a disposizione sono utili alla classificazione, allo scopo di selezionare solo quelle fondamentali al nostro problema. Infatti, feature ridondanti o dannose potrebbero non solo rallentare l'algoritmo, ma anche peggiorarne le performance. Come detto nel capitolo precedente, questo può essere considerato come un problema di ottimizzazione, con lo scopo di selezionare il sottoinsieme di feature che permette di ottimizzare le performance dell'algoritmo. A questo scopo si è scelto di utilizzare il Genetic Algorithm (GA) che rientra nell'insieme degli Evolutionary Computation. In particolare, per valutare l'importanza di quelle feature che ad oggi, in clinica, vengono ritenute fondamentali in quanto utilizzate per la stima del rischio frattura, si è scelto di portare avanti parallelamente due strade:

- La prima considerando le feature "fondamentali" come obbligatorie, quindi applicando la feature selection solo alle restanti 32 feature;
- La seconda applicando la feature selection a tutte le feature.

Le feature che ad oggi vengono utilizzate in clinica sono: menopausa, età, altezza, peso, BMI, e per le zone lombare, femorale e del collo del femore il T-score, Z-score e il BSI per un totale di 14 feature.

2.3.1 Genetic algorithm (GA)

Gli algoritmi di Evolutionary Computation [13] (EC) si basano su un processo di evoluzione, dove, attraverso la selezione naturale di una popolazione di individui, permettono di convergere verso la soluzione con una migliore capacità di sopravvivenza. Quindi, un

algoritmo evolutivo (EA) è una ricerca stocastica per una soluzione ottimale a un dato problema. Il processo di ricerca evolutiva è influenzato dalle seguenti principali componenti di un EA:

- La codifica della soluzione;
- L'inizializzazione della popolazione iniziale;
- La funzione di fitness;
- Gli operatori di selezione;
- Gli operatori di riproduzione;
- Il criterio di terminazione.

Nella figura sottostante possiamo vedere il flow chart tipico di un algoritmo genetico.

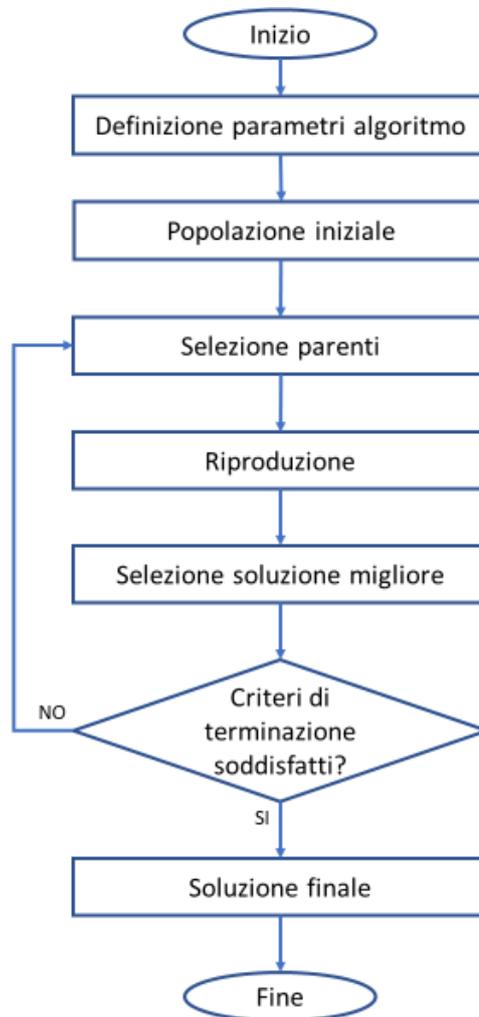


Figura 11 Tipico flow chart di un algoritmo genetico [1]

2.3.2 Codifica della soluzione

La maggior parte degli EA rappresentano soluzioni come vettori di un tipo di dati specifico (un'eccezione è la programmazione genetica (GP) dove gli individui sono rappresentati in un formato ad albero). Una classica rappresentazione di una soluzione dei GA corrisponde in un vettore di lunghezza prefissata, con ogni variabile codificata da una stringa di bit (se la codifica delle variabili ha valori binari la dimensione della soluzione corrisponde alla dimensione del vettore) [13]. Nel nostro caso si è scelto proprio questo tipo di soluzione composta da un vettore di 46 variabili codificate da un bit (0 e 1), dove l'1 indica la presenza di quel determinato gene o feature, al contrario dello 0 che indica l'assenza di quella feature. Di seguito è riportato un esempio di codifica di una soluzione:

Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Codifica	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	1	0	0

Feature	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
Codifica	1	0	0	1	1	1	0	1	0	0	1	1	0	1	0	0	0	1	0	1	0	0	0

Tabella 5 Esempio di rappresentazione di una soluzione codificata con variabile binaria.

2.3.3 Popolazione iniziale

Gli algoritmi evolutivi sono di tipo stocastico [13], di conseguenza il primo step risulta essere proprio l'inizializzazione della popolazione iniziale che solitamente viene impostata random. Questo permette di esplorare tutto lo spazio delle soluzioni. Aumentare la dimensione della popolazione iniziale permette di cercare in uno spazio maggiore ma aumenta notevolmente la complessità ed il costo computazionale dell'algoritmo.

2.3.3.1 Criterio di ammissibilità

Nei nostri due casi sono state inizializzate due popolazioni iniziali diverse in numero e criterio. Nel primo caso, dopo aver fissato le prime 14 feature, sono state generate in maniera random soltanto le ultime 32 con un vincolo legato al numero massimo e minimo di feature selezionabili. Questo vincolo di ammissibilità è stato fissato seguendo semplicemente una regola basata sul buonsenso considerando come ammissibili soluzioni

con un valore minimo di 2 e massimo di 20 feature (per un totale di feature comprese tra 16 e 34). Come dimensione della popolazione è stato scelto poi un valore pari a 1000 soluzioni.

Nel secondo caso la popolazione è stata come prima cosa raddoppiata di dimensione a causa anche dello spazio delle soluzioni nettamente più grande. Poi l'inizializzazione è stata divisa in due parti: la prima creando le prime 1000 soluzioni completamente random con un requisito di ammissibilità di minimo 2 e massimo 34 feature; le seconde 1000 sono state inizializzate facendo in modo che fossero presenti soluzioni che comprendessero sicuramente le prime 14 feature e mantenendo lo stesso criterio di ammissibilità. Questa scelta è stata presa per dare la possibilità all'algoritmo genetico sia di cercare in uno spazio casuale tramite le prime 1000, sia di cercare nell'intorno che comprendesse anche le feature utilizzate in clinica.

2.3.4 Funzione di fitness

La funzione di fitness è quella che ci permette di ottenere un'indicazione diretta sulla "bontà" di una determinata soluzione. Questa rappresenta la funzione obiettivo che descrive il problema di ottimizzazione e solitamente restituisce il valore di fitness assoluto, ma esistono casi in cui si hanno valori di fitness rispetto ad altri cromosomi o rispetto al resto dell'intera popolazione [13]. In base al tipo di problema di ottimizzazione si possono creare diverse funzioni di fitness. Nel nostro caso si è scelto di utilizzare una funzione che restituisse il valore di fitness assoluto, si è scelto quindi di classificare, per ogni soluzione della popolazione, attraverso un algoritmo di k-nearest neighbors (KNN).

2.3.4.1 Il KNN

L'idea che sta alla base della classificazione tramite il k-nearest neighbors è semplice e consiste nell'assegnazione della classe tramite un confronto per similarità, o come dice il nome stesso, in base al vicinato più simile [14]. È un algoritmo piuttosto semplice da un punto di vista concettuale e risulta piuttosto efficace con dataset di piccole dimensioni per questo motivo è stato scelto per valutare la fitness delle soluzioni. Solitamente, per rendere più robusto l'algoritmo, si utilizza un intero vicinato, ovvero un insieme di k vicini, con $k > 1$. È un algoritmo che richiede molta memoria in quanto, per la classificazione, sono necessari molti esempi in fase di esecuzione infatti viene anche chiamata Memory-Based

Classification [14]. Per capire il principio di funzionamento dell' algoritmo si può osservare la figura sottostante dove per assegnare la classe alle nuove osservazioni si valutano i 3 vicini più simili ($k=3$). Nell'esempio è ben visibile che, mentre per una osservazione la classificazione è piuttosto immediata in quanto tutti e 3 i vicini fanno parte della stessa classe, per l'altra non si può dire la stessa cosa.

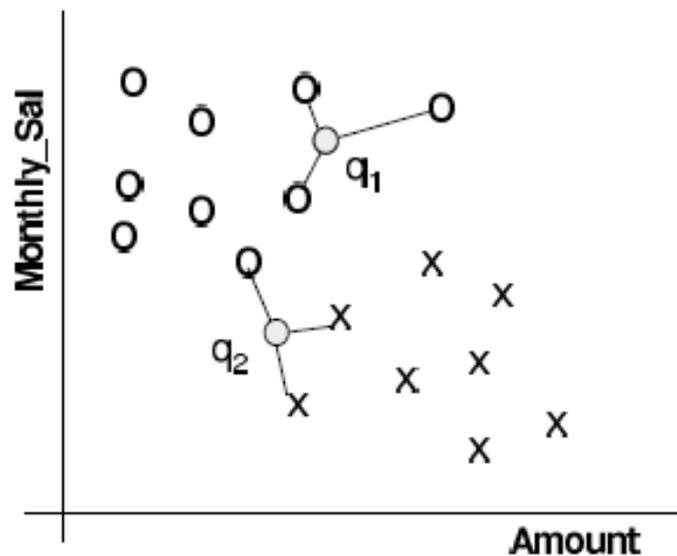


Figura 12 Esempio di classificazione tramite k -nearest neighbors con $k=3$ [14].

Da questo semplice esempio, però, si possono subito dedurre due cose fondamentali: la prima è che per la misura della similarità è necessario scegliere una metrica con la quale confrontare l'osservazione ed il vicinato, mentre, la seconda, è che per classificare in un due classi il numero k di vicini da esplorare deve essere dispari per evitare di incappare in situazioni in cui i vicini delle due classi si equivalgano. Sopponendo quindi di avere un insieme D composto da soluzioni x_i caratterizzate da un insieme di feature $f \in F$ e dover assegnare una nuova osservazione q ad una classe y_i , dovremmo, in primo luogo calcolare la distanza tra q e tutte le osservazioni x_i , nel seguente modo [14]:

$$d(\mathbf{q}, \mathbf{x}_i) = \sum_{f \in F} w_f \delta(\mathbf{q}_f, \mathbf{x}_{if})$$

Come detto precedentemente, una delle principali caratteristiche dell'algoritmo consiste nella scelta della metrica per la misura della distanza. Diversi sono i metodi proposti in letteratura molti dei quali si basano sulla metrica Minkowski [14]:

$$MD_p(\mathbf{q}, \mathbf{x}_i) = \left(\sum_{f \in F} |\mathbf{q}_f - \mathbf{x}_{if}|^p \right)^{\frac{1}{p}}$$

Si può osservare che per valori di $p=1$ la formula equivale alla distanza euclidea, mentre per $p=2$ corrisponde al metodo Manhattan. Valori più grandi di p hanno l'effetto di dare maggior peso agli attributi su cui gli oggetti differiscono maggiormente. Un altro metodo importante corrisponde alla metrica Minkowski per $p=\infty$ che equivale alla distanza di Chebyshev. Nel nostro caso, la scelta della metrica è ricaduta sulla distanza euclidea.

Infine, come ultimo parametro, si è scelto il numero k dei vicini da esaminare. Il potere predittivo dell'algoritmo dipende molto da questo parametro, valori troppo piccoli limiterebbero l'algoritmo ad una ristretta regione senza valutare la distribuzione dell'intero spazio di vicini inoltre aumenterebbe l'impatto della varianza dovuta ad errori casuali. Al contrario, con valori troppo alti si rischierebbe di perdere le informazioni di piccoli dettagli che potrebbero essere rilevanti. Nella pratica, per la scelta di k , si utilizza un valore pari a circa la radice quadrata del numero di osservazioni. Nel nostro caso si è scelto di utilizzare $k=15$.

A questo punto, dover aver costruito il modello utilizzando l'intero dataset, lo si è validato attraverso gli stessi dati. L'accuratezza della predizione (intesa come rapporto tra i corretti classificati e il totale) è poi stata usata per il calcolo del valore di fitness.

$$fitness = 1 - ac$$

$$ac = \frac{\text{corretti classificati}}{\text{totale osservazioni}}$$

Facendo questo, si è notato però che molte soluzioni diverse restituivano lo stesso valore di fitness, per questo motivo si è deciso di introdurre un peso linearmente dipendente al numero di feature selezionate, questo per cercare di favorire, a parità di fitness, soluzioni con un minor numero di feature.

Quindi il valore di fitness da minimizzare risulta essere:

$$fitness = 1 - ac + \frac{peso}{100}$$

Per quanto riguarda il peso, si è fatto in modo che soluzioni con il minimo numero di feature selezionabili avesse un peso pari a zero, mentre soluzioni con il massimo numero di feature selezionabili avesse un peso pari a 1. Il tutto poi diviso 100 per fare in modo che questo non andasse a prevaricare sull'accuratezza della classificazione.

$$peso = \frac{\text{numero feature soluzione}_i - \text{numero minimo feature ammissibili}}{\Delta \text{feature ammissibili}}$$

2.3.5 Operatori di selezione

Prima di introdurre gli operatori di selezione è bene presentare due concetti chiave per la scelta dell'algoritmo da utilizzare, ovvero si tratta di diversità di popolazione e pressione selettiva. I concetti di diversità di popolazione e pressione selettiva sono dei fattori inversamente proporzionali tra di loro e risultano estremamente importanti in un GA. La pressione selettiva [13], indicata anche come tempo di acquisizione, si riferisce al tempo necessario per produrre una popolazione uniforme. È definita come la velocità alla quale la migliore soluzione occuperà l'intera popolazione applicando solo l'operatore di selezione. Un operatore con un'alta pressione selettiva riduce la diversità nella popolazione più rapidamente rispetto agli operatori con una bassa pressione selettiva. Un'alta pressione selettiva può portare a una convergenza prematura, mentre una bassa pressione selettiva può portare a un ristagno nella ricerca di ottimizzazione. La convergenza prematura del GA può verificarsi quando ci sono grandi differenze tra i valori di fitness all'interno di una popolazione, al contrario un ristagno può verificarsi se ci sono piccole differenze tra le fitness delle soluzioni. In questi casi sono state proposte diverse funzioni di ridimensionamento per trovare un equilibrio tra pressione selettiva e diversità della popolazione.

Nei GA l'operatore di selezione consiste in un processo che permette di selezionare un sottoinsieme della popolazione iniziale (genitori) per la riproduzione di una nuova popolazione (figli). Esistono diversi metodi di selezione, i principali sono:

- Tournament selection
- Ranking selection
- Proportionate selection

Tournamen selection

In questo modo le soluzioni vengono selezionate random. Si possono sia prendere le soluzioni in maniera completamente random, oppure si creano dei sottogruppi in maniera random e poi vengono selezionate le migliori all'interno di ciascuno. La selezione di tipo random è quella con il più basso valore di pressione selettiva [13].

Ranking selection

In questo caso, dopo aver calcolato il valore di fitness di ogni soluzione, quest'ultime vengono ordinate partendo dalla soluzione con fitness migliore fino alla peggiore. A questo punto vengono selezionate solo le prime N in base alla dimensione scelta per la popolazione dei genitori. La ranking selection permette di ridurre la pressione selettiva in caso di una elevata varianza nei valori di fitness [13].

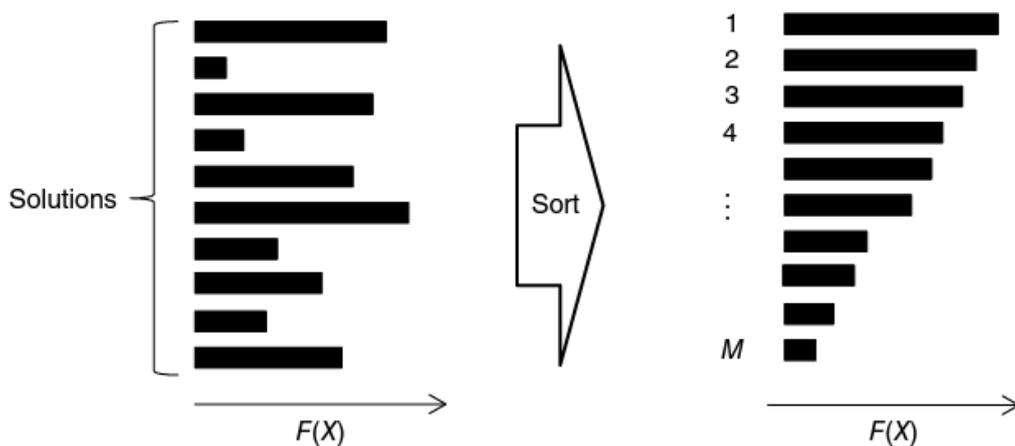


Figura 13 Rappresentazione Ranking Selection

Proportionate selection

Uno degli algoritmi più utilizzati è il proportionate selection dove si utilizza il valore di fitness per calcolare la probabilità di una soluzione di essere scelta attraverso il seguente criterio:

$$P_k = \frac{F(X_k)}{\sum_{j=1}^M F(X_j)}$$

dove $F(X_k)$ e P_k corrispondono rispettivamente al valore di fitness della k -soluzione e alla sua probabilità di essere scelta. Un tipico algoritmo di proporzionate selection è il roulette wheel [13], dove la ruota della roulette è costituita da tanti valori ("fette" della roulette) quante sono le soluzioni. La particolarità di questa roulette consiste che la dimensione della "fetta" della ruota è strettamente legata al valore di fitness, ovvero una più alta probabilità di selezione corrisponde ad una fetta più grande. In questo modo tutte le soluzioni hanno la possibilità di essere scelte, con la differenza che le soluzioni migliori hanno una probabilità maggiore rispetto alle peggiori. Nella figura sottostante si può osservare il principio di funzionamento di questo algoritmo.

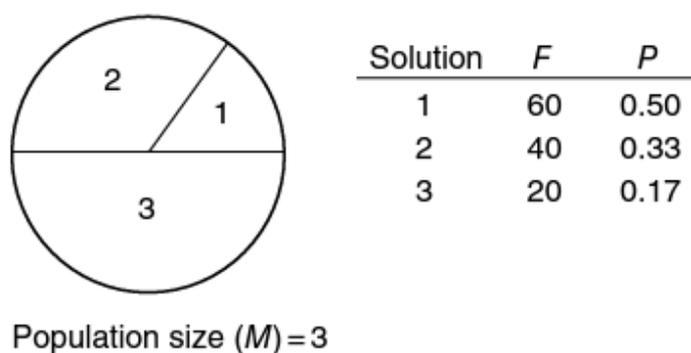


Figura 14 Assegnazione delle probabilità di selezione nel Roulette Wheel

In questo caso ci sono 3 diverse soluzioni alle quali corrispondono 3 diverse probabilità di essere selezionate in base al loro valore di fitness. Ipotizzando di fare girare la ruota della roulette è chiaro come, sebbene tutte le soluzioni possono essere selezionate, quella con una fitness migliore ha una probabilità maggiore rispetto alle altre. La ruota viene fatta girare N volte, dove N corrisponde alla dimensione della popolazione dei genitori, per selezionare N soluzioni.

Nel nostro caso si è scelto di utilizzare come operatore di selezione il proporzionate selection attraverso l'algoritmo roulette wheel, apportando un ridimensionamento lineare ai valori di fitness rispetto al minimo.

$$fitness_{norm\ i} = \frac{1}{1 - fitness_{min} + fitness_i}$$

$$p_{Selezione} = \frac{fitness_{norm\ i}}{\sum_{i=1}^n fitness_i}$$

2.3.6 Operatori di riproduzione

L'algoritmo GA deve generare nuove soluzioni per progredire verso una soluzione ottimale. I genitori fanno figli che costituiscono la totalità o una parte della generazione successiva. Pertanto, la generazione successiva potrebbe essere composta sia da una parte dei genitori che dai figli.

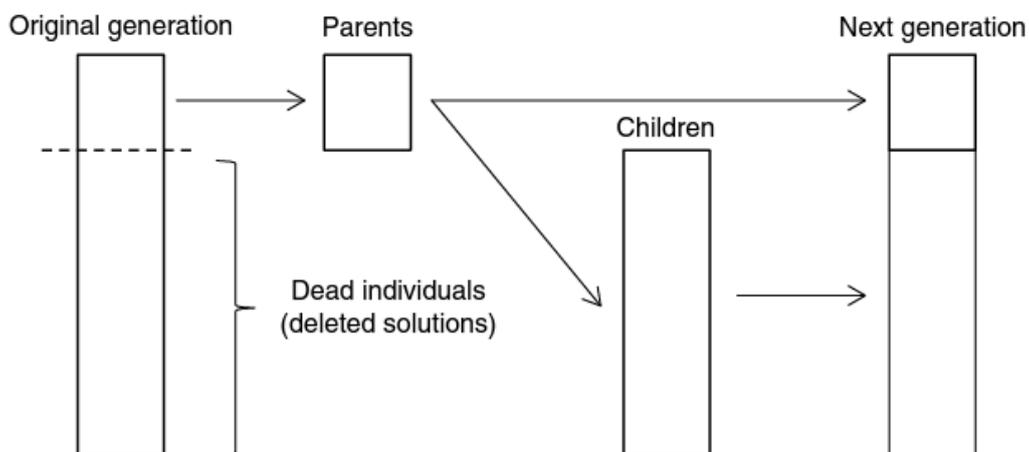


Figura 15 Scelta dei genitori per la costruzione di una nuova popolazione [13].

Una volta selezionato il sottoinsieme di genitori, questi generano figli tramite il parametro di crossover [15]. Non tutti i genitori generano figli, questo avviene con una probabilità che dipende dal valore del crossover. Per ogni soluzione viene generato un numero casuale compreso tra 0 e 1, e se questo valore è minore della probabilità di crossover allora quella determinata soluzione sarà sottoposta al crossover. Una volta generata la prole questa viene sottoposta ad una nuova mutazione attraverso il parametro di mutazione, anche in questo caso con una determinata probabilità come per il crossover.

CROSSOVER

Il crossover avviene tra una coppia di genitori. In questo caso, una volta selezionata la coppia, viene generata la prole come scambio di alcuni geni (sottostringhe delle soluzioni).

Il numero di sottostringhe dipende da un valore, inserito dall'utente, chiamato punto di taglio. In questo modo, ogni soluzione viene dapprima tagliata in tante parti quanti sono i punti di taglio, successivamente vengono scambiate le sottostringhe per la generazione dei nuovi due figli. Solitamente la probabilità di crossover varia tra 0.7 e 0.95 [15], questo per permettere all'algoritmo di convergere verso la soluzione ottimale. Nel nostro caso si è fatta variare la probabilità di crossover tra 0.7 e 0.95 a passo di 0.05 per un totale di 5 combinazioni.

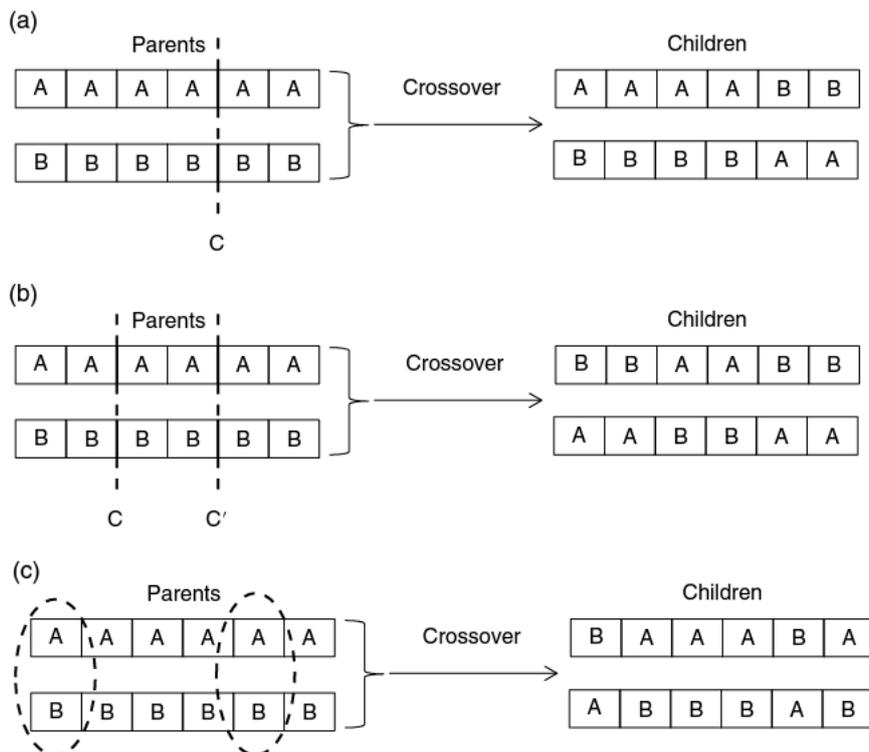


Figura 16 Rappresentazione del crossover [13].

MUTAZIONE

L'operatore di mutazione è importante perché introduce nuovo materiale genetico all'interno della soluzione [16]. Anche in questo caso viene assegnato un valore casuale compreso tra 0 e 1 ad ogni gene della soluzione corrente, viene poi confrontato con la probabilità di mutazione e se minore allora viene applicata la modifica di un gene.

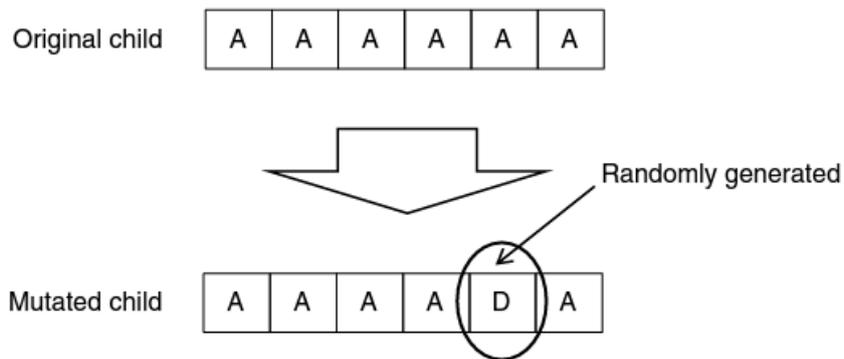


Figura 17 Rappresentazione della mutazione [13].

La probabilità di mutazione è solitamente compresa tra 0.05 e 0.2 [15], questo perché l'obiettivo della mutazione non è quello di far convergere l'algoritmo ma quello di esplorare quanto più possibile lo spazio delle soluzioni modificando anche un singolo gene della soluzione. Nel nostro caso si è fatta variare la probabilità di mutazione tra 0.1 e 0.2 a passo di 0.02 per un totale di 6 combinazioni.

2.3.7 Criterio di terminazione

Il criterio di terminazione ha un ruolo fondamentale nell'algoritmo in quanto permette di capire quando si è raggiunta la soluzione ottimale. Tra i principali troviamo ad esempio il numero massimo di iterazioni e il numero di iterazioni dopo il quale, se non si ha un miglioramento della funzione di fitness, l'algoritmo interrompe la ricerca. Nel nostro caso si sono scelte, come condizioni di stop, un numero massimo di iterazioni pari a 5000 e un numero massimo di iterazioni senza miglioramento pari a 50.

2.4 Divisione dataset

Una volta ottenuto il subset ottimale di feature si è passati alla costruzione e all'allenamento del classificatore. Per fare questo sono necessari tre diversi subset di dati:

- Il training set, utilizzato per allenare la rete;
- Il test set, utilizzato per testare la rete appena allenata;
- Il validation set, utilizzato alla fine del lavoro per validare la rete scelta.

Nel nostro caso si dispone di un dataset non molto numeroso e piuttosto sbilanciato, infatti si hanno 278 pazienti a rischio frattura contro i 86 non a rischio, questo obbliga a fare delle scelte per la suddivisione di questi set di dati. Le osservazioni del validation set non vengono utilizzate per la costruzione del classificatore, ma solamente in fase finale per la validazione, questo implica che la dimensione di questa matrice debba essere piuttosto ridotta. Nel nostro caso si è scelto di comporre il validation set di 10 osservazioni per classe per un totale di 20 osservazioni. A questo punto restano 268 pazienti a rischio e 76 non a rischio. Per quanto riguarda il training set è bene che questo sia bilanciato, ovvero che siano presenti osservazioni di entrambe le classi in egual misura. Il problema principale risiede nella classe meno rappresentata e visto che bisogna fare in modo che sia presente in entrambe i set di dati, si è deciso di comporre il training set di 58 pazienti per classe e le restanti nel test set. Nella figura sottostante è rappresentata la suddivisione del dataset.

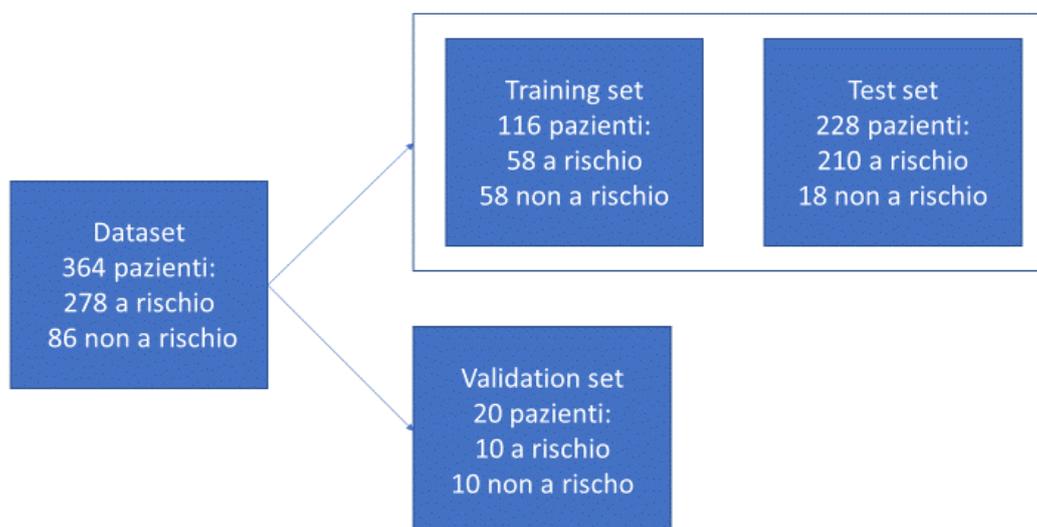


Figura 18 Suddivisione Training, Test e Validation set.

Per la divisione dei dati è stato applicato lo stesso algoritmo di clusterizzazione visto per la feature selection nel capitolo precedente. Ottenuti i singoli cluster e le rispettive osservazioni, è stato possibile costruire i training, test e validation set facendo in modo che ogni cluster sia rappresentato in ognuno dei set di dati.

2.5 Reti Neurali

La capacità di apprendere e memorizzare del cervello umano ha spinto la ricerca nella modellizzazione di sistemi neurali biologici che prendono il nome di Artificial Neural Network (ANN). Gli attuali successi nella modellizzazione neurale riguarda lo svolgimento di compiti specifici, problemi con un singolo obiettivo, a differenza del cervello che riesce a risolvere più problemi contemporaneamente utilizzando più aree. Un neurone artificiale (AN) è una modellizzazione del neurone biologico (BN) [12]. Ogni AN riceve i segnali dall'esterno o da altri neuroni, li elabora e li trasmette ai neuroni ad esso connessi. La trasmissione o l'inibizione del segnale è regolata dai pesi associati ai neuroni, una funzione di attivazione poi, regola il segnale di uscita. Una rete neurale artificiale è composta da più strati di neuroni. Ci sono i layer di input, i layer nascosti e quelli di output. I neuroni di un singolo layer possono essere totalmente o parzialmente connessi ai neuroni dello strato successivo, inoltre, sono possibili anche dei collegamenti ai livelli precedenti. Nella figura sottostante troviamo un esempio di rete neurale artificiale:

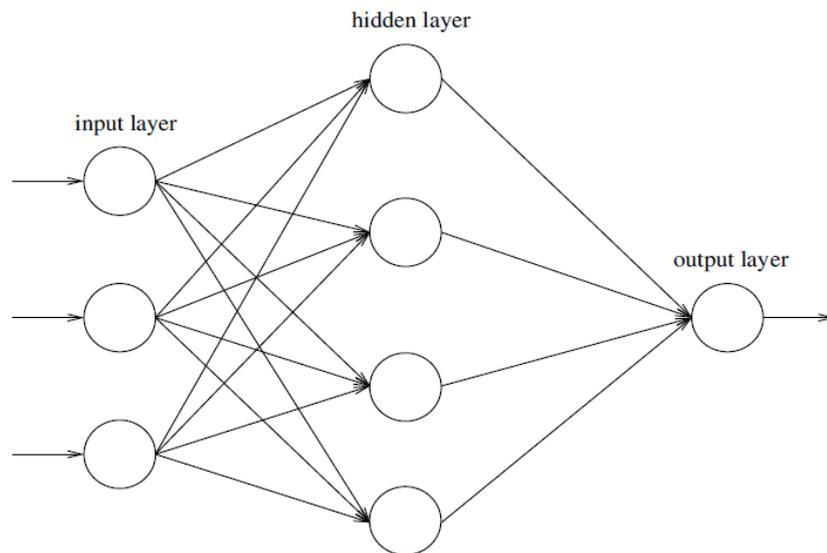


Figura 19 Esempio di struttura di una rete neurale artificiale [12].

In letteratura vengono proposti diversi tipi di reti neurali, tra i principali troviamo:

- Reti single-layer NNs;
- multilayer feedforward NNs;
- temporal NNs;

- self-organizing NNs;
- combinazione di reti neurali supervisionate e non supervisionate.

2.5.1 Neurone Artificiale

Il neurone artificiale (AN), è l'unità funzionale di una rete neurale, ed esegue una mappatura da R^I a un intervallo compreso tipicamente tra $[0;1]$ o tra $[-1;1]$ in base alla funzione di attivazione scelta, dove I indica il numero di segnali di input che arrivano al neurone.

$$f_{AN}: R^I \rightarrow [0,1]$$

o

$$f_{AN}: R^I \rightarrow [-1,1]$$

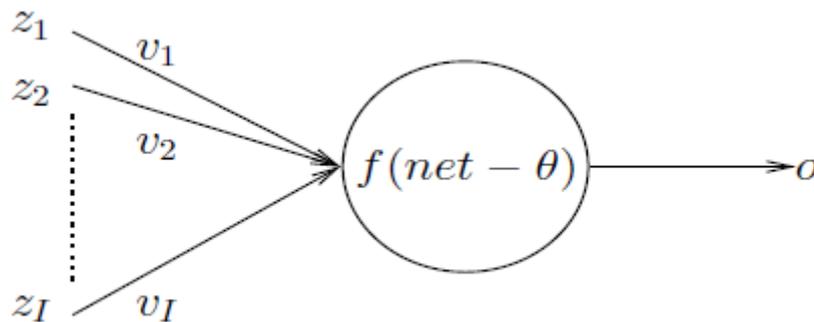


Figura 20 Anatomia di un neurone artificiale [12].

Segnali di input $z = (z_1, z_2, \dots, z_I)$

Pesi $v = (v_1, v_2, \dots, v_I)$

Ad ogni segnale di input è associato un peso che permette di rafforzare o indebolire un determinato input e, attraverso una rete e una funzione di attivazione, viene poi calcolato il segnale di output. Quest'ultimo è fortemente influenzato anche da un parametro θ chiamato anche soglia o bias.

Per quanto riguarda il net input signal, questo viene solitamente calcolato come somma pesata di tutti i segnali di input ed i rispettivi pesi:

$$net = \sum_{i=1}^I z_i v_i$$

I neuroni che utilizzano questa somma pesata vengono chiamati anche “summation unit” (SU). Un’alternativa alla SU è la “product unit” (PU):

$$net = \prod_{i=1}^I z_i^{v_i}$$

Le PU consentono combinazioni di input di ordine superiore, con il vantaggio di una maggiore capacità di informazione.

2.5.2 Funzione di attivazione

Il principale compito della funzione di attivazione è quello di determinare il valore di output una volta ricevuti in ingresso i segnali di input con i rispettivi pesi, e il valore del bias. Tutte le funzioni di attivazione hanno un dominio di input infinito, mentre il dominio di output è differente da funzione a funzione, anche se le più diffuse sono comprese tra $[0,1]$ e $[-1;1]$. Nella pratica le funzioni più utilizzate sono: lineare, gradino, rampa, sigmoide, tangente iperbolica e gaussiana. All’interno di una stessa rete è possibile utilizzare diverse funzioni di attivazione, a seconda delle necessità. Nel nostro caso ne sono state utilizzate due, quella lineare per il layer di output, e quella a tangente iperbolica per il layer di input e quelli nascosti.

Funzione di attivazione lineare

è caratterizzata dalla seguente equazione:

$$f_{AN}(net - \theta) = \lambda(net - \theta)$$

Dove λ rappresenta la pendenza della curva.

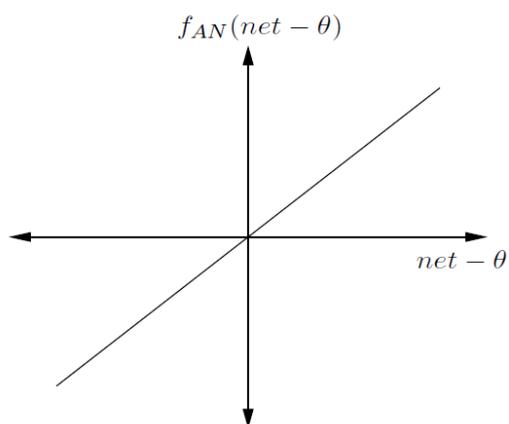


Figura 21 Funzione di attivazione lineare.

Funzione di attivazione tangente iperbolica

È caratterizzata dalle seguenti equazioni:

$$f_{AN}(-\infty) = -1 \quad \text{e} \quad f_{AN}(\infty) = 1$$

$$f_{AN}(net - \theta) = \frac{e^{\lambda(net - \theta)} - e^{-\lambda(net - \theta)}}{e^{\lambda(net - \theta)} + e^{-\lambda(net - \theta)}}$$

Che può essere approssimata a:

$$f_{AN}(net - \theta) = \frac{2}{1 + e^{-\lambda(net - \theta)}} - 1$$

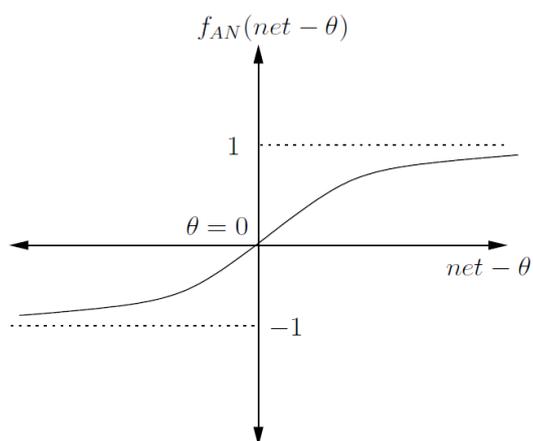


Figura 22 Funzione di attivazione a tangente iperbolica.

2.5.3 Geometria del neurone

I singoli neuroni possono essere utilizzati per risolvere problemi del tipo lineare, questo significa che un neurone può separare lo spazio degli input con vettori che hanno valori o sopra o sotto la soglia. La rete neurale più semplice è composta da un singolo neurone e prende il nome di perceptron [12]. Per problemi che non sono linearmente separabili sono necessari strati di più neuroni. In figura 23 si può vedere come un singolo neurone riesce a dividere lo spazio linearmente (come esempio è stata presa una funzione di attivazione a rampa).

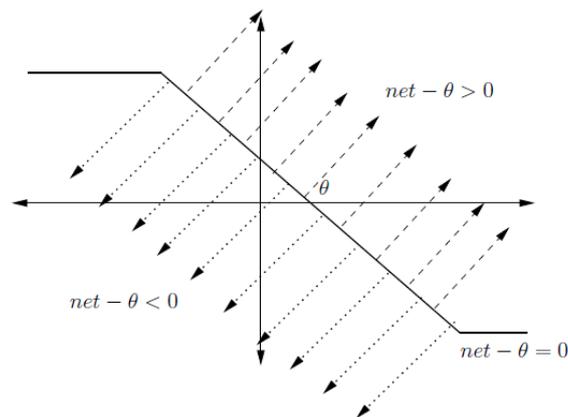


Figura 23 Separazione dello spazio con un singolo neurone [12].

In questo caso l'iperpiano separa i vettori di input con $\sum_i z_i v_i - \theta > 0$ da quelli con $\sum_i z_i v_i - \theta < 0$. A questo punto non resta altro che capire come determinare i vari pesi e la soglia.

2.5.4 Apprendimento del neurone artificiale

In fase di apprendimento, gli unici dati di cui si dispone sono i vettori di input, di conseguenza, attraverso il termine apprendimento, ci si riferisce proprio al metodo attraverso il quale si determinano i valori dei pesi e della soglia. Questi, infatti, vengono aggiornati iterativamente fino a che non viene soddisfatto uno o più criteri. Per semplificare l'apprendimento e l'aggiornamento dei parametri, il vettore di input viene come prima cosa aumentato di dimensione, ovvero si aggiunge un valore z_{i+1} chiamato unità bias. Questo valore è sempre pari a -1, mentre il valore v_{i+1} viene utilizzato per l'aggiornamento della soglia. In questo modo:

$$net = \sum_{i=1}^I z_i v_i - \theta = \sum_{i=1}^I z_i v_i + z_{I+1} v_{I+1} = \sum_{i=1}^{I+1} z_i v_i$$

Dove $\theta = z_{I+1} v_{I+1} = -v_{I+1}$

2.5.4.1 Gradient descent

Uno degli algoritmi più utilizzati per l'aggiornamento dei pesi e della soglia è il gradient descent (GD) [12]. Il GD necessita di una funzione da ottimizzare, ad esempio può essere utilizzata la SSE (Sum Squared Errors) o la MSE (Mean Squared Errors):

$$SSE \quad \varepsilon = \sum_{p=1}^{P_T} (t_p - o_p)^2$$

$$MSE \quad \varepsilon = \frac{1}{P_T} \sum_{p=1}^{P_T} (t_p - o_p)^2$$

Dove t_p e o_p sono rispettivamente la classe reale e quella predetta dall'algoritmo, mentre P_T sono il numero totale dei vettori di input e delle rispettive classi. L'obiettivo del GD è proprio quello di minimizzare ε , ovvero di minimizzare l'errore nella predizione. Questo è possibile calcolando il gradiente di ε e muovendosi in direzione negativa fino a trovare il punto di minimo (figura 24).

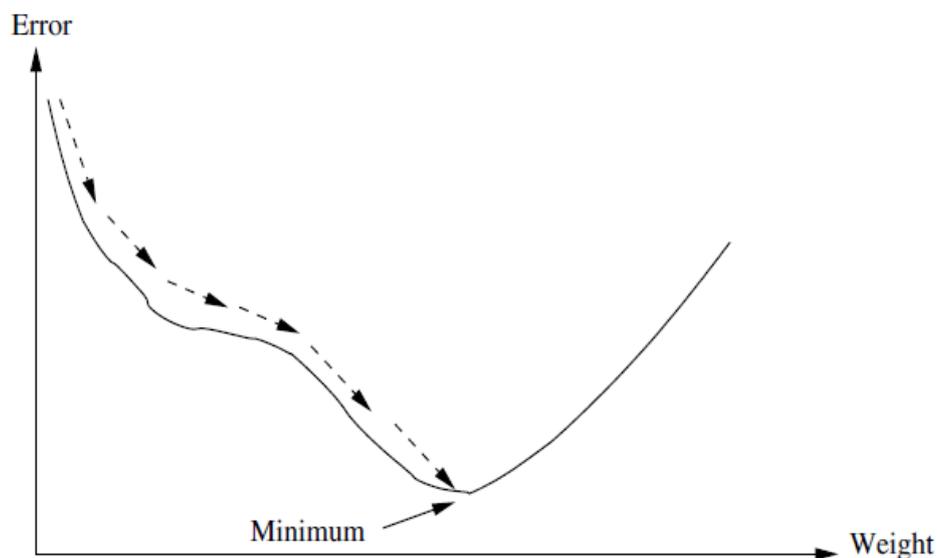


Figura 24 Rappresentazione del Gradient Descent (GD).

Ipotizzando un singolo pattern, i pesi vengono aggiornati utilizzando l'iterazione precedente nel seguente modo:

$$v_i(t) = v_i(t-1) + \Delta v_i(t)$$

Con

$$\Delta v_i(t) = \eta \left(-\frac{\partial \mathcal{E}}{\partial v_i} \right)$$

Dove

$$\left(-\frac{\partial \mathcal{E}}{\partial v_i} \right) = -2(t_p - o_p) \frac{\partial f}{\partial net_p} z_{i,p}$$

E η rappresenta il coefficiente di apprendimento. Questo tipo di algoritmo viene anche chiamato backpropagation in quanto gli errori vengono propagati all'indietro per l'aggiornamento dei pesi e della soglia. Una volta trovati i pesi e la soglia che minimizzano l'errore, la rete risulta essere allenata.

2.5.5 Feedforward network

Un singolo neurone però riesce a separare solamente problemi di tipo lineare, ma per problemi più complicati sono necessarie delle strutture più complesse che permettano di risolvere il problema. Una delle più diffuse è la feedforward network, ovvero una struttura simile alla perceptron vista precedentemente ma formata da più layer, dove i neuroni di un singolo layer ricevono il segnale di input dai layer precedenti e lo propagano a quelli successivi fino al layer di output. Dalla figura sottostante è possibile vedere lo schema di una rete di tipo feedforward.

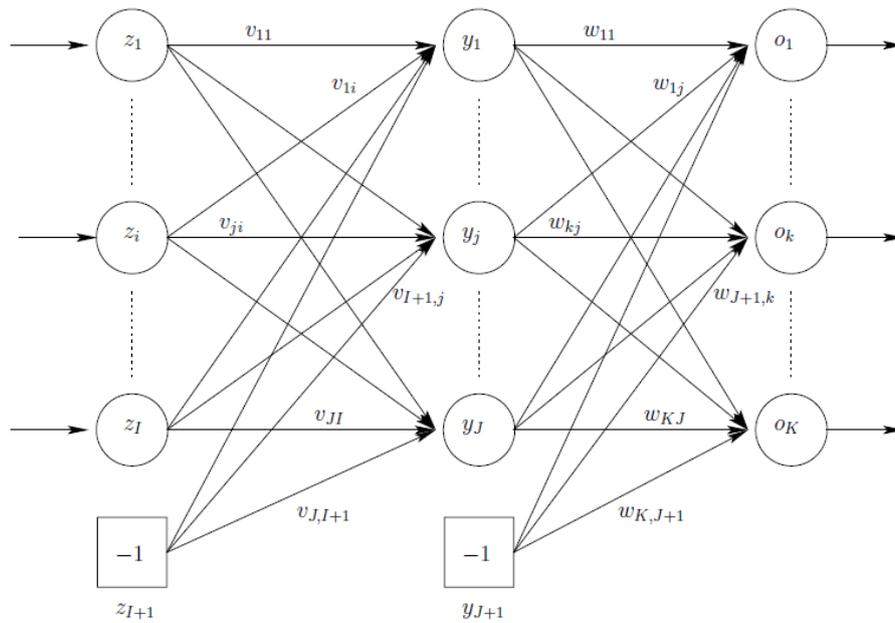


Figura 25 Feedforward network.

Dove ogni output è calcolato come

$$o_{k,p} = f_{ok}(net_{ok,p}) = f_{ok} \left(\sum_{j=1}^{J+1} w_{kj} f_{yj}(net_{yj,p}) \right) = f_{ok} \left(\sum_{j=1}^{J+1} w_{kj} f_{yj} \left(\sum_{i=1}^{I+1} v_{ji} z_{i,p} \right) \right)$$

Dove f_{ok} e f_{yj} sono rispettivamente le funzioni di attivazione dei neuroni di output o_k e dei neuroni nascosti y_j ; w_{kj} è il peso tra l'output o_k e il neurone nascosto y_j ; $z_{i,p}$ è il valore dell'unità di input z_i del pattern di input z_p ; l'input (I+1) e il peso (J+1) sono le unità bias e la soglia del layer successivo.

2.5.5.1 Scelta della struttura

Una delle principali problematiche nell'addestramento di una rete consiste nella scelta della struttura, in particolar modo per quanto riguarda la dimensione ed il numero dei layer nascosti. Infatti, non esiste una regola universale per la scelta della struttura, ma piuttosto delle linee guida che permettono di orientare la decisione. Per quanto riguarda il layer di input infatti, ogni feature in ingresso corrisponde ad un pattern, di conseguenza la dimensione di questo layer corrisponde alla dimensione del dataset. Il layer di output invece dipende dal numero delle classi in cui si vuole allenare l'algoritmo, infatti ogni neurone restituisce un valore che varia all'interno del dominio di uscita della funzione di

attivazione. Nel nostro caso, in cui le classi sono solamente due, è sufficiente un neurone nel layer di output, la criticità sarà poi determinare la soglia che permette di distinguere le due classi. Di conseguenza solo nella dimensione e numero dei layer nascosti si ha la possibilità di scegliere. È chiaro che più piccole sono le strutture e meno complesso sarà l'algoritmo e il costo computazionale, ma al contrario strutture più grandi saranno in grado di risolvere problemi più complessi. Le principali linee guida presenti in letteratura consigliano di:

- iniziare ad allenare reti con 1,2 o massimo 3 layer nascosti;
- utilizzare una struttura dei layer nascosti di tipo piramidale, ovvero con dimensioni via via decrescenti (figura 26);
- fare in modo che il primo layer nascosto non abbia una dimensione maggiore del doppio rispetto al layer di input, questo per cercare di prevenire problemi di overfitting.

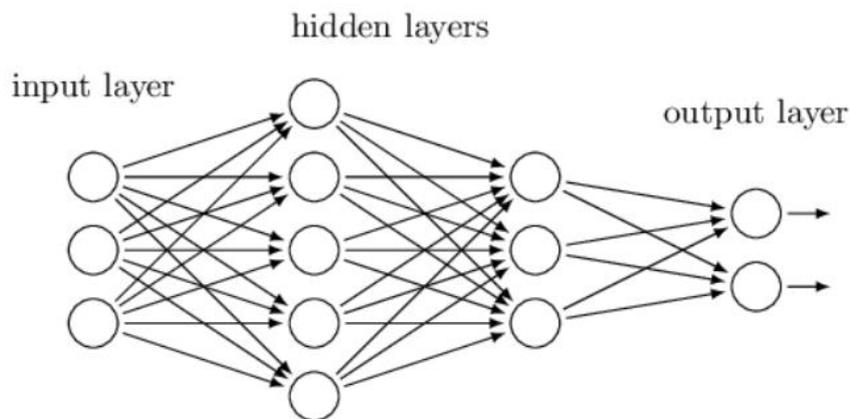


Figura 26 Esempio di struttura di tipo piramidale.

Una volta allenate diverse strutture, la scelta finale dipenderà dalle loro performance.

2.5.6 Algoritmo di apprendimento della rete

L'algoritmo di apprendimento di una rete composta da più strati è molto simile a quello di una struttura perceptron formata da un singolo neurone, allo stesso modo infatti si compone principalmente di due fasi: la prima dove la rete propaga i segnali di input fino al layer di output calcolando i segnali di uscita attraverso i pesi e le funzioni di attivazione,

e una seconda fase dove l'errore viene propagato all'indietro allo scopo di aggiornare i pesi per minimizzare l'errore nella predizione. Molti dei metodi proposti in letteratura sono delle evoluzioni del gradient descent visto precedentemente, allo scopo di ottimizzarne le prestazioni. L'algoritmo utilizzato in questo lavoro è il Levenberg-Marquardt [18], in quanto risulta essere molto efficiente per un limitato numero di pesi (fino ad alcune centinaia). Questo metodo può essere visto come un'ottimizzazione dei metodi del Gradient Descent e quello di Newton.

2.5.6.1 Algoritmo di Levenberg-Marquardt

Supponendo di avere una funzione $V(x)$ da minimizzare rispetto ad un vettore x , il metodo di Newton dice che:

$$\Delta x = -[\nabla^2 V(x)]^{-1} \nabla V(x)$$

Dove $\nabla^2 V(x)$ e $\nabla V(x)$ rappresentano rispettivamente la matrice Hessiana ed il gradiente. A questo punto assumiamo che la funzione $V(x)$ da minimizzare corrisponda proprio al SSE (Sum Squared Errors) visto precedentemente, per cui:

$$V(x) = \sum_{i=1}^N e_i^2(x)$$

La principale caratteristica che differenzia il metodo di Levenberg-Marquardt da quello di Newton consiste nell'approssimazione della matrice Hessiana e del gradiente nel seguente modo [18]:

$$\nabla^2 V(x) = J^T(x)J(x) \quad \text{e} \quad \nabla V(x) = J e(x)$$

Per ottenere infine:

$$\Delta(x) = [J^T(x)J(x) + \mu I]^{-1} J^T(x)e(x)$$

Di conseguenza dover calcolare solo la matrice Jacobiana e non quella Hessiana permette di risparmiare un notevole tempo computazionale. Il parametro μ viene inizializzato e modificato iterativamente tramite un altro parametro β , anch'esso inizializzato dall'utente, dove se $V(x)$ aumenta (ovvero $\Delta(x) > 0$) allora μ viene moltiplicato per β , altrimenti viene diviso per β (tipicamente si assegnano valori di $\mu=0.001$ e di $\beta=10$).

L'algoritmo può quindi essere riassunto principalmente in 5 passi [18]:

1. calcolo dei valori di output e l'SSE attraverso i valori di input;
2. calcolo della matrice Jacobiana;
3. calcolo del $\Delta(x)$
4. ricalcolo dell'SSE attraverso $x + \Delta(x)$, se è minore del precedente allora μ viene ridotta attraverso il parametro β e si ritorna al punto 1, altrimenti μ viene aumentata tramite β e si ritorna al punto 3;
5. l'algoritmo si ritiene essere arrivato a convergenza quando l'SSE è minore di un valore assegnato o quando vengono soddisfatti determinati criteri di terminazione.

2.5.7 Criteri di terminazione

I criteri di terminazione vengono utilizzati dall'algoritmo per capire quando si è arrivati a convergenza, ovvero, per capire quando si sono trovati i pesi e la soglia che permettono di minimizzare l'errore nella predizione. Come per gli algoritmi genetici, anche in questo caso i principali criteri di terminazioni riguardano:

- numero massimo di iterazioni: di solito questo parametro viene impostato piuttosto elevato, ovvero intorno alle 1000 iterazioni;
- numero massimo di iterazioni per i quali non si notano miglioramenti nell'errore di predizione: questo è il principale criterio di terminazione, solitamente si utilizzano 6 iterazioni massime;
- tempo massimo affinché si arrivi a convergenza: tipicamente si imposta un tempo infinito in quanto si prediligono i criteri precedenti.

2.5.8 Valutazione performance della rete

Una volta allenata una determinata rete non resta altro che misurarne le performance. Queste sono infatti fortemente influenzate da diversi parametri come ad esempio l'inizializzazione dei pesi o la struttura della rete. I principali aspetti legati alle performance di una rete sono:

- accuratezza;
- complessità.

Con il termine accuratezza ci si riferisce alla percentuale di corretti classificati rispetto al totale. Nel nostro caso però si parla a tutti gli effetti di un test di screening, ovvero un test effettuato all'interno di una determinata popolazione che permette di identificare il rischio ad una determinata malattia, per poi eventualmente programmare degli esami di approfondimento. In questo tipo di test, oltre all'accuratezza, ci sono altri parametri fondamentali utilizzati per valutare le performance dell'algoritmo come ad esempio:

- sensitività, ovvero la probabilità che il test dia esito positivo per quelle persone affette da quella determinata patologia;
- specificità, ovvero la probabilità che il test dia esito negativo per quelle persone non affette dalla patologia;
- falsi negativi, ovvero quelle persone alle quali il test da esito negativo ma in realtà sono positive alla patologia;
- falsi positivi, ovvero quelle persone alle quali il test da esito positivo ma in realtà non sono affette alla patologia.

Per la valutazione delle performance attraverso questi parametri viene utilizzata una matrice chiamata confusion matrix, dove nelle colonne vengono poste le classi reali delle osservazioni, mentre nelle righe le classi predette. Di conseguenza si otterrà una matrice quadrata di dimensione $n \times n$, dove n è pari al numero di classi. Nel nostro caso quindi si avrà una matrice di dimensioni 2×2 impostata nel seguente modo:

Confusion matrix		Classe reale	
		0	1
Classe predetta	0	TN	FN
	1	FP	TP

Tabella 6 Confusion matrix

Dove:

- TP=true positive;
- FP=false positive;
- TN=true negative;

- FN=false negative.

I parametri utilizzati per la valutazione delle performance vengono calcolati come:

$$\text{Accuratezza } AC = \frac{TN + TP}{TN + FN + TP + FP}$$

$$\text{Sensitività } SE = \frac{TP}{TP + FN}$$

$$\text{Specificità } SP = \frac{TN}{TN + FP}$$

Questi risultano essere fondamentali nella scelta della soglia che permette di distinguere le due classi. Infatti, come detto precedentemente, il neurone di output restituisce un valore in un range all'interno del dominio di uscita della funzione di attivazione, quindi la soglia dovrà essere scelta in modo tale da avere un'accuratezza, sensitività e specificità quanto più possibile elevate. Ad esempio, una soglia troppo bassa tendente allo zero porterebbe ad avere un'altissima sensitività ma una bassissima specificità, al contrario una soglia tendente verso l'uno porterebbe ad avere un'alta specificità contro una bassa sensitività. Nella situazione ideale le distribuzioni di probabilità dei risultati distinguono i pazienti a rischio da quelli non a rischio senza un'overlap delle distribuzioni [19] e di conseguenza la scelta della soglia risulterebbe immediata (figura 27).

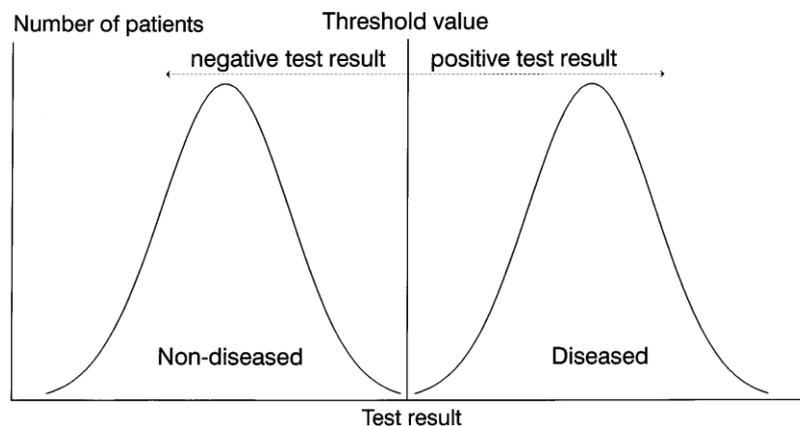


Figura 27 Distribuzione delle probabilità di risultati.

Nella realtà però questa situazione non accade quasi mai, ma le due distribuzioni si sovrappongono, questo rende più difficile la scelta. Come è possibile vedere dalla figura

28 una soglia più bassa permette di ottenere un minor numero di falsi negativi (figura 28.a), al contrario una soglia più alta che permette di agire sui falsi positivi (figura 28.b).

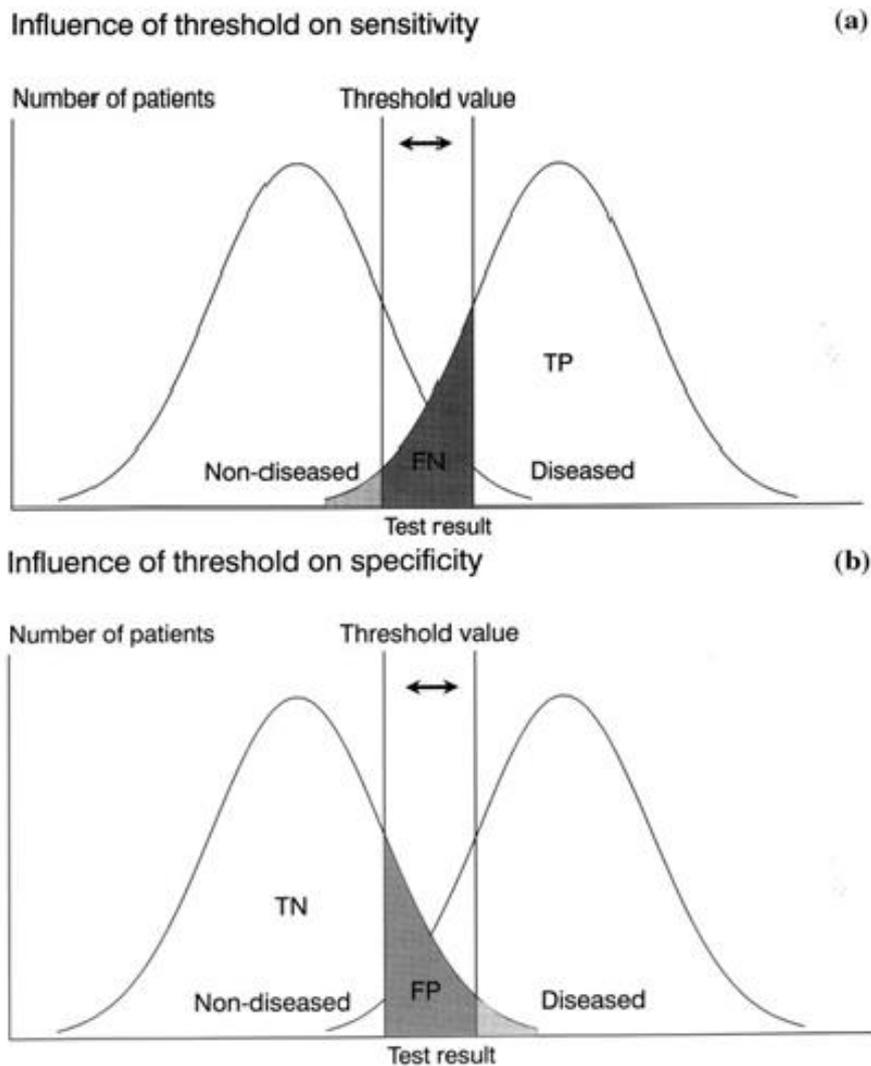


Figura 28 Effetti della variazione della soglia sulla sensibilità e specificità.

L'obiettivo è quello di trovare un valore che permetta di bilanciare tutti i parametri e per fare questo viene creata una curva, chiamata curva ROC (Receiver Operating Characteristic) [19], dove vengono graficate le performance di un classificatore binario al variare della soglia. In questo grafico è possibile trovare il tasso di falsi positivi (identificato come 1-specificità) nell'asse delle ascisse, e la sensibilità nell'asse delle ordinate. Un esempio di curva ROC riferito all'esempio precedente è riportato in figura 30 dove possibile vedere come variano sensibilità e specificità al variare della soglia. In questo caso

la scelta ricadrebbe nella numero 3, ovvero quella che permette di bilanciare la classificazione sia dei pazienti a rischio che di quelli non a rischio.

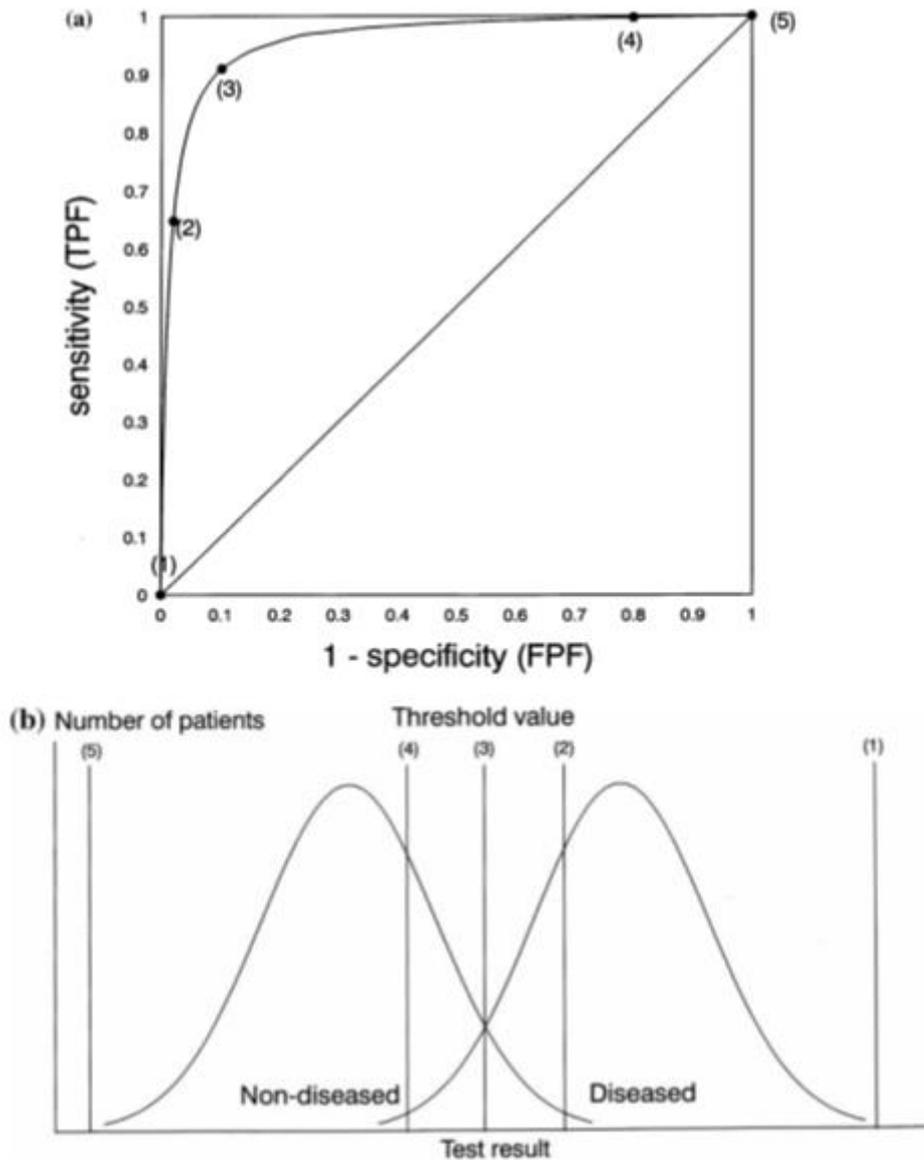


Figura 29 Scelta della soglia attraverso la curva ROC.

Un ulteriore parametro che permette di misurare l'effettiva accuratezza del classificatore è l'area sottesa alla curva (AUC Area Under ROC Curve) [19]. È facilmente visibile come un'area tendente a 1 permetterebbe di scegliere un valore della soglia con degli ottimi valori di sensibilità e specificità (figura 30). Al contrario, valori di area tendenti allo 0 indicano delle pessime performance del classificatore (figura 31). Questo parametro

permette quindi di fare una prima scrematura delle strutture testate, per poi poter selezionare la più performante attraverso la scelta della soglia.

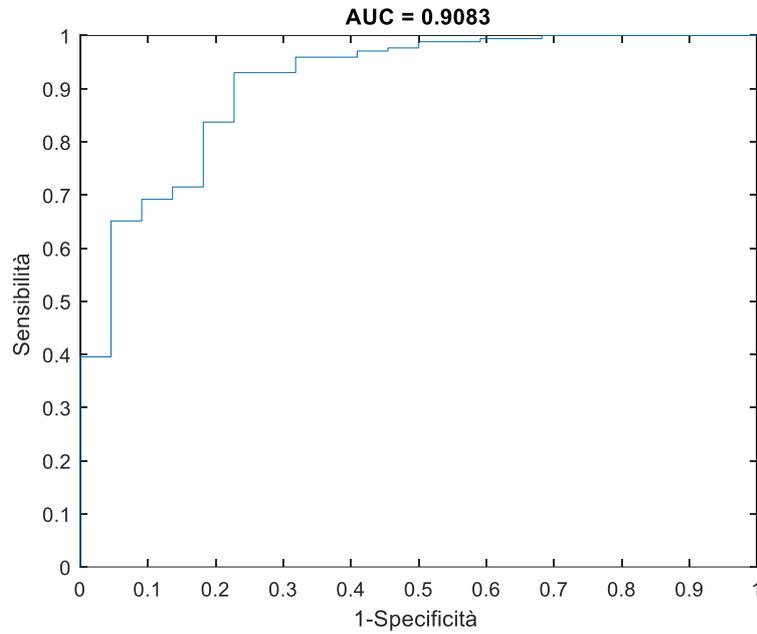


Figura 30 Esempio di curva ROC con area sottesa alla curva pari a 0.9

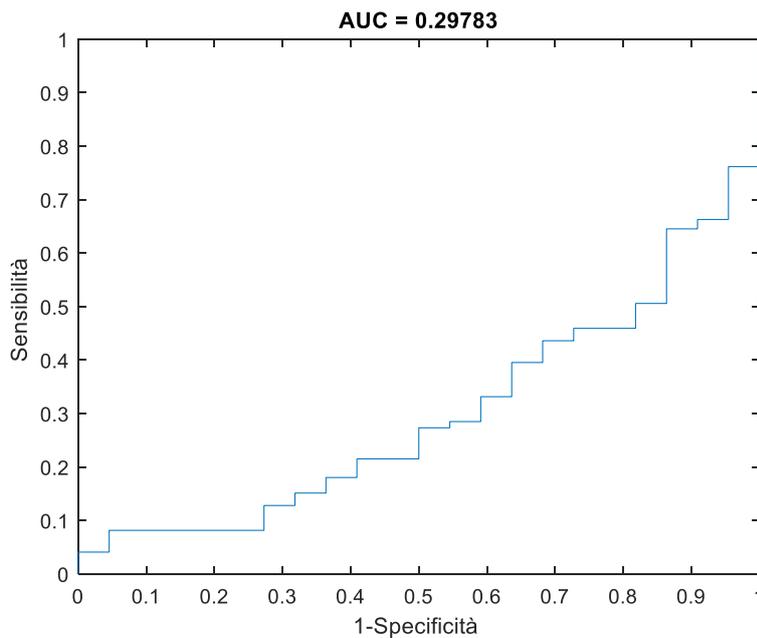


Figura 31 Esempio di curva ROC con area uguale sottesa alla curva pari a 0.29

Per quanto riguarda invece la complessità dell'algoritmo, i principali fattori che la influenzano sono:

- La struttura della rete: più è grande la struttura e maggiori calcoli nell'aggiornamento dei pesi e nel calcolo del valore di output sono richiesti;
- Dimensione del training set: maggiore è la dimensione del training set maggiori sono i pattern presentati in ingresso con un conseguente aumento delle epoche richieste per il calcolo dell'output;
- Complessità dell'algoritmo di ottimizzazione: come visto in precedenza in letteratura vengono proposti diversi algoritmi che permettono di ottimizzare le performance del metodo Gradient Descent, solitamente preso come riferimento.

Altri valori che possono influenzare le performance della rete sono l'inizializzazione dei pesi e il trattamento dei dati in ingresso. Per quanto riguarda l'inizializzazione dei pesi, una pratica molto utilizzata è quella di iniziarli in maniera casuale in un valore compreso tra $[-1,1]$. Come è stato visto per gli algoritmi genetici, quello di inizializzare random è un metodo molto utilizzato per cercare di esplorare tutto lo spazio di ricerca ed evitare di finire in dei minimi locali. Fino ad ora sono stati visti diversi fattori riguardanti esclusivamente le reti, oltre a questi però, è possibile agire anche sui dati in ingresso. Infatti, il trattamento dei dati iniziali influisce nel risultato finale come ad esempio la normalizzazione del dataset. È possibile infatti che elevate differenze tra i dati in ingresso causino dei peggioramenti nelle performance. Per ovviare a questo problema vengono proposti diversi metodi per la normalizzazione dei dati in modo tale da porre tutte le feature in ingresso in un unico range. I metodi più utilizzati sono:

- Min-max scaling, dove ogni feature viene normalizzata rispetto al minimo e al massimo;
- Mean centering scaling, dove ogni feature è centrata intorno al valor medio
- Standardization, dove ogni feature ha una distribuzione standard, ovvero ad ogni feature viene sottratta la media e poi viene divisa per la deviazione standard.

3. Risultati

3.1 Feature selection

L'obiettivo della feature selection è quello di ottenere una soluzione composta da un sottoinsieme di feature iniziali che permettono di ottimizzare le performance della classificazione. Nella situazione ideale la soluzione migliore è una, che si ripete ad ogni iterazione, questo significherebbe che l'algoritmo è robusto e che converge sempre verso il minimo assoluto. Nella realtà non sempre succede, sia perché l'algoritmo potrebbe convergere in minimi relativi sia perché la soluzione migliore potrebbe non essere unica. A questo punto, per le diverse soluzioni trovate, bisognerebbe sceglierne una o, al limite, alcune di queste. Prima di procedere con la feature selection però, visto lo sbilanciamento del dataset originale, si è applicata una clusterizzazione ai dati dei pazienti a rischio frattura per ottenere una matrice bilanciata e, allo stesso tempo, il più possibile rappresentativa di tutti i pazienti a rischio frattura.

3.1.1 Costruzione dataset bilanciato

Per la costruzione di un dataset bilanciato è stata applicata una clusterizzazione mediante dendrogramma e poi, tramite il grafico ad albero, si è scelto il punto di taglio per la scelta del numero di cluster. In figura 32 è possibile vedere come nelle ascisse sono rappresentati i pazienti ordinati in modo tale da non far intrecciare i rami del grafico (l'elevata numerosità dei pazienti non permette di riconoscere il numero identificativo del paziente). Nell'asse delle ordinate è invece rappresentata la misura della distanza tipica della metrica e del metodo scelto. Nel nostro caso si è scelto di utilizzare il metodo complete-link e la metrica della distanza euclidea per la misura della similarità tra i cluster. Una piccola variazione della distanza che causa l'unione di due cluster indica un'elevata similarità tra i due, al contrario un'elevata dissimilarità può essere identificata da un "salto" più marcato.

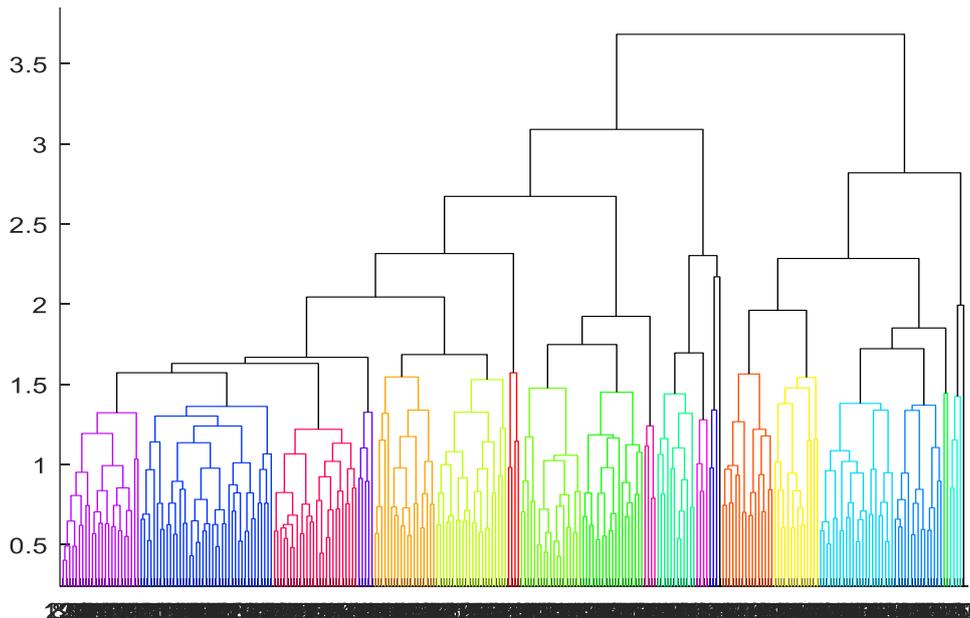


Figura 32 Dendrogramma pazienti con elevato rischio di frattura per la costruzione del dataset bilanciato.

Dal grafico è visibile come il numero dei cluster scelto sia pari a 21. Successivamente sono stati estratti 86 pazienti in maniera iterativa selezionando un paziente alla volta partendo dal primo fino all'ultimo cluster. A questo punto il dataset risulta essere pronto per la feature selection.

3.1.2 Algoritmo Genetico

Dopo aver generato un insieme di popolazioni iniziale composto da 1000 soluzioni ed aver implementato l'algoritmo con i parametri esplicitati, per ogni combinazione dei parametri di crossover e mutazione (30 combinazioni), sono state ripetute 10 iterazioni, ottenendo un totale di 300 soluzioni. Questo per ogni popolazione iniziale. Per la rappresentazione dei risultati sono state create quattro tabelle: le prime tre rappresentano la distribuzione delle soluzioni per le varie combinazioni dei parametri di riproduzione, in particolare per ogni combinazione si hanno le soluzioni uguali, le fitness uguali e le soluzioni simili (ovvero quelle soluzioni che differiscono al massimo per 2 feature); la quarta riferita ai valori di fitness dove si confrontano quante soluzioni con lo stesso valore di fitness si ripetono, il loro numero di feature ed infine il loro valore di fitness senza il peso relativo alla dimensione della soluzione.

3.1.2.1 Soluzione Feature Selection con le prime 14 "bloccate"

Soluzioni uguali		Crossover					
		0.1	0.12	0.14	0.16	0.18	0.2
Mutazione	0.75	0	0	0	2	2	2
	0.8	0	0	0	0	2	0
	0.85	0	0	2	0	0	0
	0.9	2	0	0	0	2	0
	0.95	4	0	3	0	2	0

Tabella 7 Soluzioni uguali.

Fitness uguali		Crossover					
		0.1	0.12	0.14	0.16	0.18	0.2
Mutazione	0.75	3	2	2	3	3	2
	0.8	0	2	3	2	2	3
	0.85	3	2	2	4	3	3
	0.9	2	2	2	2	3	3
	0.95	4	4	3	3	2	3

Tabella 8 Fitness uguali.

Soluzioni simili		Crossover					
		0.1	0.12	0.14	0.16	0.18	0.2
Mutazione	0.75	0	0	0	2	2	2
	0.8	0	0	0	0	2	0
	0.85	0	0	2	0	0	0
	0.9	2	0	0	0	2	0
	0.95	4	0	3	0	2	0

Tabella 9 Soluzioni simili.

N° di ripetizioni	Fitness	N° di feature	1-accuratezza
30	0,0992	28	0,0930
27	0,1046	27	0,0988
27	0,1041	26	0,0988
26	0,0987	27	0,0930
23	0,1026	23	0,0988
19	0,1061	18	0,1047
18	0,0983	26	0,0930
17	0,1036	25	0,0988
14	0,1050	28	0,0988
14	0,0997	29	0,0930
10	0,1031	24	0,0988
9	0,1022	22	0,0988
8	0,1002	30	0,0930
7	0,1006	31	0,0930
5	0,0978	25	0,0930
5	0,1012	20	0,0988
4	0,1017	21	0,0988
4	0,0973	24	0,0930
4	0,0968	23	0,0930
4	0,0954	20	0,0930
3	0,0944	30	0,0872
3	0,1055	29	0,0988
2	0,1060	30	0,0988
2	0,0964	22	0,0930
2	0,0998	17	0,0988
2	0,0948	31	0,0872
2	0,0959	21	0,0930
2	0,1011	32	0,0930
2	0,1003	18	0,0988
1	0,1007	19	0,0988
1	0,0929	27	0,0872
1	0,0939	29	0,0872
1	0,0920	25	0,0872
1	0,0934	28	0,0872

Tabella 10 Confronto tra numero di soluzioni, valori di fitness, numero di feature selezionate e accuratezza.

Da questa prima feature selection è possibile osservare che:

- Una soluzione con 18 feature si ripete per 19 volte con un valore di fitness pari a 0.1061;
- Il miglior valore di fitness pari a 0.0920 si ripete solo 1 volta per una soluzione di 25 feature;
- Il migliore valore di 1-accuratezza pari a 0.0872 si ripete 6 volte per 5 diverse soluzioni con un numero di feature compreso tra 25 e 31;

Da questa prima analisi è possibile dire che, nonostante i valori di fitness siano tutti abbastanza simili, si hanno due interessanti soluzioni: quella che si ripete più stabile, ovvero quella che si ripete per 19 volte; e quella con il miglior valore di fitness. Per questo motivo verranno utilizzate entrambe nell'allenamento delle reti neurali. Di seguito sono riportate le feature selezionate nelle due soluzioni scelte (le prime 14 non sono state riportate in quanto, in questo caso, non soggette alla feature selection):

Soluzione con fitness pari 0.0920 con 25 feature totali

- Area totale del femore
- BMD lombare
- BMD lombare
- BR del collo del femore
- Diametro interno della corteccia del collo del femore
- Spesso subperiostale del collo del femore
- Z modulus del collo del femore
- Diametro interno della corteccia della zona intertrocanterica
- Spesso subperiostale della zona intertrocanterica
- Z modulus della zona intertrocanterica
- HAL

Soluzione con fitness pari a 0.1061 con 18 feature totali

- Area totale del femore
- CSA del collo del femore
- Spessore dell'osso corticale nella zona intertrocanterica
- Spessore subperiostale della zona intertrocanterica

3.1.2.2 Soluzione Feature selection applicata a tutte le feature

Queste soluzioni si riferiscono al caso in cui è stata lasciata la libertà al GA di selezionare o meno tutte le feature, comprese quelle utilizzate in clinica. In questo caso, visto l'aumentare dello spazio delle soluzioni si è deciso di raddoppiare l'insieme delle soluzioni iniziali (per un totale di 2000 soluzioni) facendo in modo che le prime 1000 fossero generate completamente random, e che le seconde 1000 contenessero almeno tutte le feature utilizzate in clinica in modo tale da verificare l'effettiva importanza di quei parametri.

Soluzioni uguali		Crossover					
		0.1	0.12	0.14	0.16	0.18	0.2
Mutazione	0.75	8	9	7	9	8	8
	0.8	10	7	7	7	10	9
	0.85	10	9	9	10	10	8
	0.9	10	10	9	8	9	9
	0.95	10	9	9	8	10	10

Tabella 11 Soluzioni uguali.

Fitness uguali		Crossover					
		0.1	0.12	0.14	0.16	0.18	0.2
Mutazione	0.75	8	9	7	9	8	8
	0.8	10	7	7	7	10	9
	0.85	10	9	9	10	10	8
	0.9	10	10	9	8	9	9
	0.95	10	9	9	8	10	10

Tabella 12 Fitness uguali.

Soluzioni simili		Crossover					
		0.1	0.12	0.14	0.16	0.18	0.2
Mutazione	0.75	8	9	7	9	8	8
	0.8	10	7	7	7	10	9
	0.85	10	9	9	10	10	8
	0.9	10	10	9	8	9	9
	0.95	10	9	9	8	10	10

Tabella 13 Soluzioni simili.

N° di ripetizioni	Fitness	N° di feature	1-accuratezza
266	0,1002	25	0,0930
7	0,0996	23	0,0930
5	0,0990	21	0,0930
4	0,0999	24	0,0930
3	0,0986	20	0,0930
3	0,0980	18	0,0930
2	0,0974	16	0,0930
2	0,0938	23	0,0872
2	0,0993	22	0,0930
1	0,0919	17	0,0872
1	0,0965	13	0,0930
1	0,0883	24	0,0814
1	0,0935	22	0,0872
1	0,0977	17	0,0930
1	0,0941	24	0,0872

Tabella 14 Confronto tra numero di soluzioni, valori di fitness, numero di feature selezionate e accuratezza.

Anche in questo caso sono presenti due interessanti soluzioni: quella con il miglior valore di fitness pari a 0.0883 ed un totale di 24 feature; quella che si ripete per 266 volte con un valore di fitness pari a 0.1002 ed un numero di feature pari a 25. Di conseguenza, come nel caso precedente, verranno considerate entrambe per l'allenamento delle reti neurali. Di seguito sono riportate tutte le feature delle due soluzioni.

Soluzione con fitness pari a 0.0883 con 24 feature totali

- Menopausa
- Altezza
- Peso
- BMI
- T-score, Z-score e BSI della zona lombare
- T-score, Z-score e BSI del femore
- Area totale del femore
- BMD della zona lombare
- CSA, CSMI, spessore subperiostale e Z modulus del collo del femore
- CSMI, diametro interno della corteccia e spessore subperiostale della zona intertrocanterica
- Spessore osso corticale, diametro interno della corteccia e Z modulus dell'albero del femore
- NECKANGLE
- HAL

Soluzione con fitness pari a 0.1002 con 25 feature totali

- Menopausa
- Altezza
- Peso
- Età
- BMI
- T-score, Z-score e BSI della zona lombare
- T-score, Z-score e BSI del femore
- T-score, Z-score e BSI del collo del femore
- Area totale del femore
- BMD della zona lombare e femorale
- CSA, CSMI, diametro interno della corteccia e Z modulus del collo del femore
- Diametro interno della corteccia della zona intertrocanterica
- CSA e spessore subperiostale dell'albero del femore
- HAL

3.1.3 Considerazioni feature selection

Gli algoritmi genetici permettono di risolvere problemi di ottimizzazione partendo da un insieme di soluzioni iniziali casuali, questo per permettere di esplorare tutto lo spazio delle soluzioni, cercando poi, attraverso gli operatori di selezione e riproduzione, di far

convergere l' algoritmo verso la soluzione con un miglior valore di fitness. Questo, risulta quindi essere un valore di fondamentale importanza nell'evoluzione dell' algoritmo e nella scelta finale della soluzione perciò, anche la scelta della funzione di fitness risulta essere decisiva. Osservando i risultati, quello che possiamo dire da questa prima feature selection, è che apparentemente non esiste una soluzione migliore di altre in termini di stabilità e accuratezza. Infatti, si hanno soluzioni con il miglior valore di fitness che si ripetono solo una volta e soluzioni con fitness leggermente peggiore che si ripetono diverse volte per questo motivo si è scelto di portare avanti parallelamente tutte e 4 le soluzioni proposte precedentemente. Un'interessante analisi sarebbe quella di confrontare queste quattro soluzioni per determinare quali e quante soluzioni hanno in comune e capire quali di queste sono più importanti di altre. Di seguito sono riportate le tabelle che contengono le soluzioni scelte.

	Menopause	Height	Weight	Age	BMI	T-score lumbar	Z-score Lumbar	BSI lumbar
S. fitness 0.0920	1	1	1	1	1	1	1	1
S. fitness 0.1061	1	1	1	1	1	1	1	1
S. fitness 0.0883	1	1	1	0	1	1	1	1
S. fitness 0.1002	1	1	1	1	1	1	1	1

	T-score femur	Z-score femur	BSI femur	T-score neck	Z-score neck	BSI neck
S. fitness 0.0920	1	1	1	1	1	1
S. fitness 0.1061	1	1	1	1	1	1
S. fitness 0.0883	1	1	1	0	0	0
S. fitness 0.1002	1	1	1	1	1	1

	Lumbar Avg Area	Femur Area	Neck Area	Lumbar BMC	Lumbar BMD	Femur BMC	Femur BMD	Neck BMC
S. fitness 0.0920	0	1	0	1	1	0	0	0
S. fitness 0.1061	0	1	0	0	0	0	0	0
S. fitness 0.0883	0	1	0	0	1	0	0	0
S. fitness 0.1002	0	1	0	0	1	0	1	0

	Neck BMD	NN BR	NN CSA	NN CSMI	NN Cort Thick	NN EndoCort Width	NN SubPeri Width	NN Z
S. fitness 0.0920	0	1	0	0	0	1	1	1
S. fitness 0.1061	0	0	1	0	0	0	0	0
S. fitness 0.0883	0	0	1	1	0	0	1	1
S. fitness 0.1002	0	0	1	1	0	1	0	1

	IT BR	IT CSA	IT CSMI	IT Cort Thick	IT EndoCort Width	IT SubPeri Width	IT Z	FS BR
S. fitness 0.0920	0	0	0	0	1	1	1	0
S. fitness 0.1061	0	0	0	1	0	1	0	0
S. fitness 0.0883	0	0	1	0	1	1	0	0
S. fitness 0.1002	0	0	0	0	1	0	0	0

	FS CSA	FS CSMI	FS Cort Thick	FS EndoCort Width	FS SubPeri Width	FS Z	NECKANGLE	HAL
S. fitness 0.0920	0	0	0	0	0	0	0	1
S. fitness 0.1061	0	0	0	0	0	0	0	0
S. fitness 0.0883	0	0	1	1	0	1	1	1
S. fitness 0.1002	1	0	0	0	1	0	0	1

Da questo confronto è possibile vedere l'importanza di alcune feature rispetto ad altre. Come ci si aspettava molte delle feature utilizzate in clinica sono presenti in tutte le soluzioni, in particolar modo del T-score della zona lombare e femorale che rappresentano anche la soglia di diagnosi dell'osteoporosi. Al contrario, è possibile osservare anche alcune feature che non sono mai presenti, ad indicare una bassa utilità nella classificazione. Infine, per quanto riguarda quelle presenti solo alcune volte, è possibile affermare che queste siano quelle feature chiamate "deboli", ovvero che magari da sole hanno un basso impatto nelle prestazioni della classificazione, ma in combinazione con altre diventano fondamentali.

Infine, per l'allenamento delle reti si è deciso anche di utilizzare solo le feature utilizzate in clinica, questo per poter confrontare le prestazioni della feature selection e l'utilizzo di altre feature con quella che è oggi la pratica clinica.

3.2 Divisione training, test e validation set

Ottenuti i dataset con le feature selezionate si è passati alla divisione della matrice nel training set, test set e validation set. Per fare questo, come per l'algoritmo genetico, si è praticata prima una clusterizzazione attraverso il dendrogramma, poi, attraverso il grafico ad albero, si è scelto il punto di taglio. Infine, si è fatto in modo che tutti cluster fossero rappresentati in tutti e tre i set di dati.

3.2.1 Soluzione n°1 con fitness pari a 0.0920

In figura è rappresentato il dendrogramma dei pazienti non a rischio frattura. Per questa matrice di dati è stato scelto un numero di cluster pari a 8.

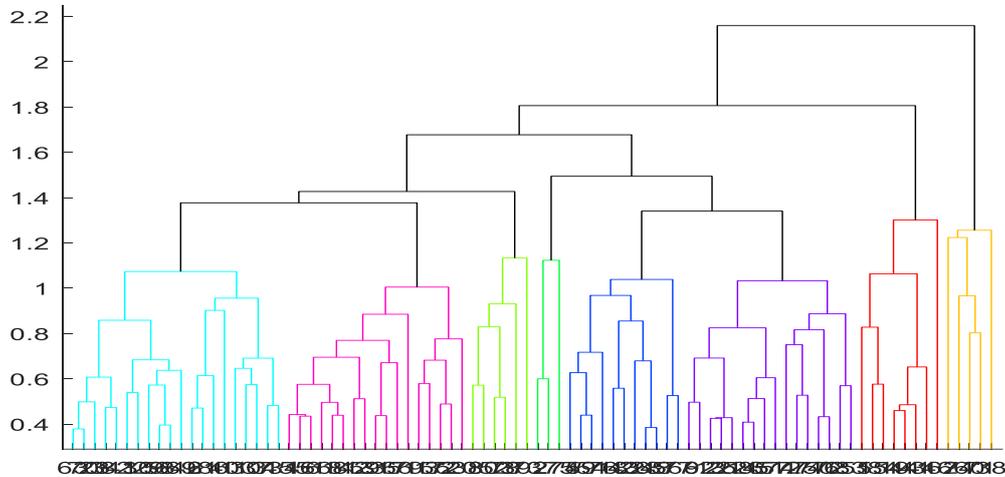


Figura 33 Dendrogramma pazienti non a rischio frattura della soluzione n°1.

Per quanto riguarda invece il dendrogramma dei pazienti a rischio frattura, il punto di taglio scelto porta ad avere 24 cluster. In questo caso si hanno diversi cluster composti da una singola osservazione. Per evitare questo si sarebbero dovuti avere solo 3 cluster ed uno sbilanciamento sproporzionato nella dimensione dei cluster di conseguenza si è optato per un taglio a 24 cluster.

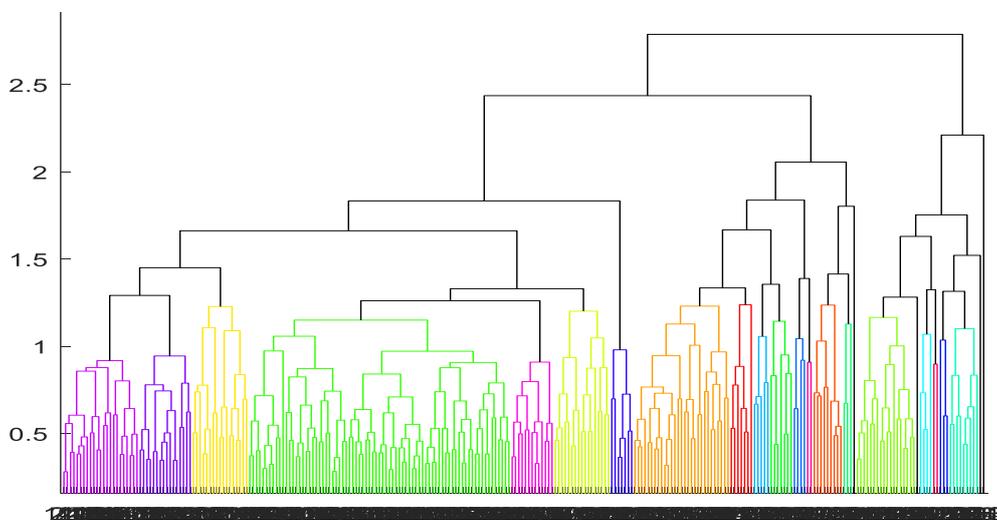


Figura 34 Dendrogramma pazienti a rischio frattura della soluzione n°1.

3.2.3 Soluzione n°3 con fitness pari a 0.0883

Per i pazienti non a rischio il taglio scelto ha portato ad ottenere un numero di cluster pari a 12.

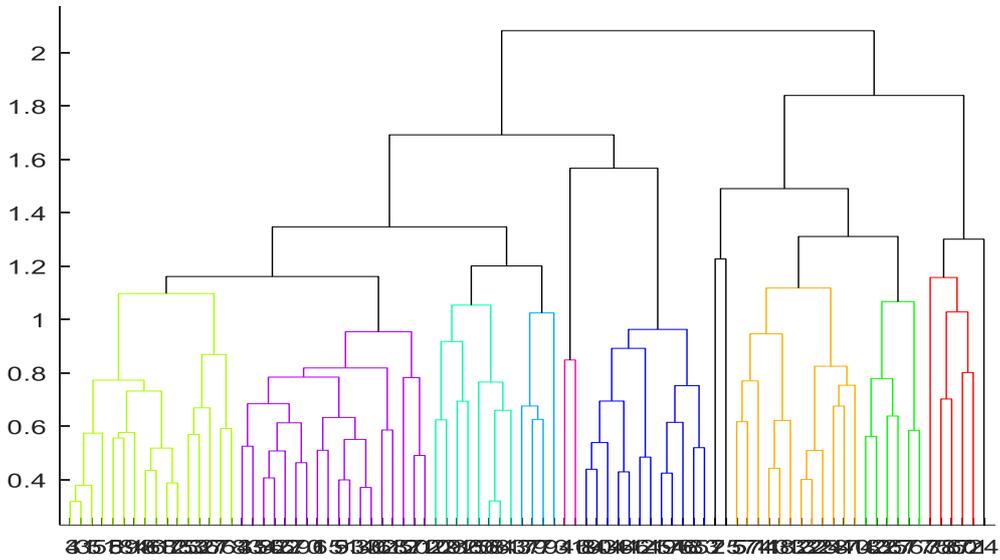


Figura 37 Dendrogramma pazienti non a rischio frattura della soluzione n°3.

Per quanto riguarda invece il dendrogramma dei pazienti a rischio frattura è stato scelto un numero di cluster pari a 19.

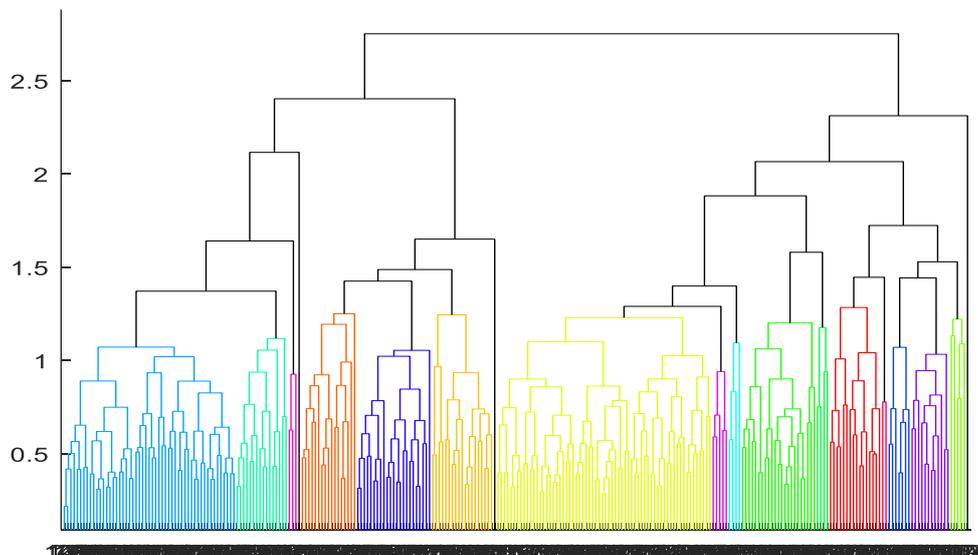


Figura 38 Dendrogramma pazienti a rischio frattura della soluzione n°3.

3.2.4 Soluzione n°4 con fitness pari a 0.1002

Per i pazienti non a rischio il taglio scelto ha portato ad ottenere un numero di cluster pari a 12.

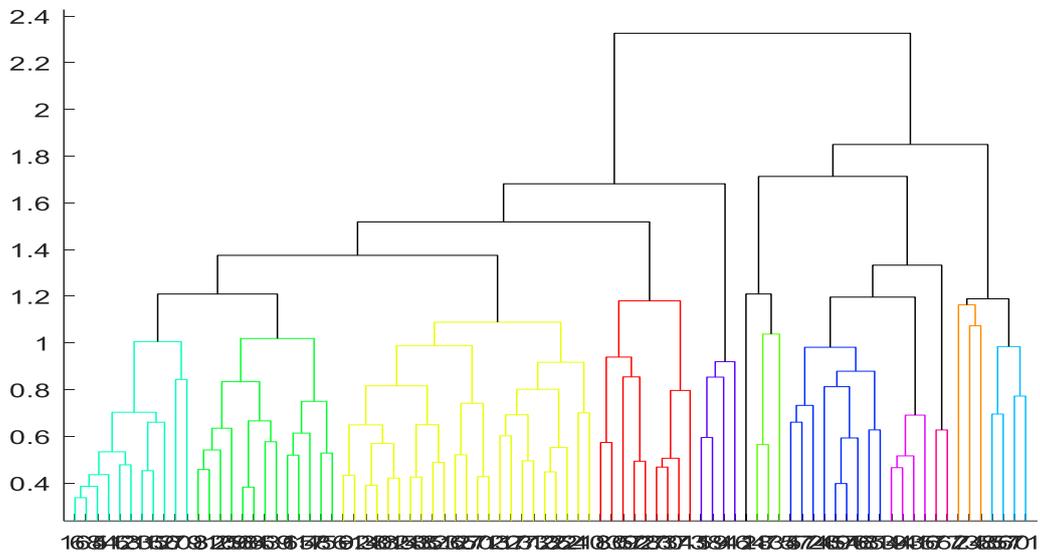


Figura 39 Dendrogramma pazienti non a rischio soluzione n°4.

Per i pazienti a rischio il taglio scelto ha portato ad ottenere un numero di cluster pari a 20.

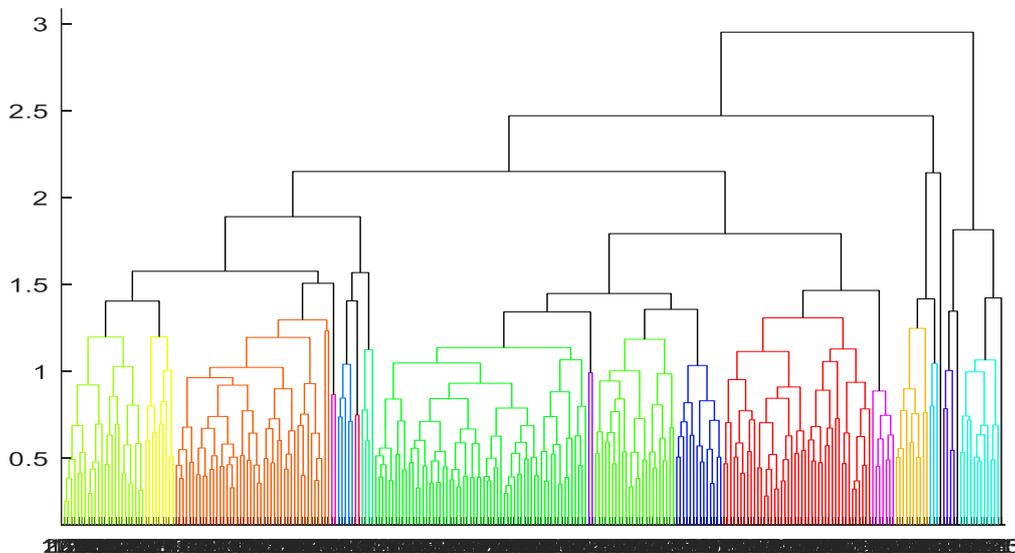


Figura 40 Dendrogramma pazienti a rischio frattura soluzione n°4.

3.2.5 Soluzione n°5 con le feature utilizzate in clinica

Per i pazienti non a rischio il taglio scelto ha portato ad ottenere un numero di cluster pari a 13.

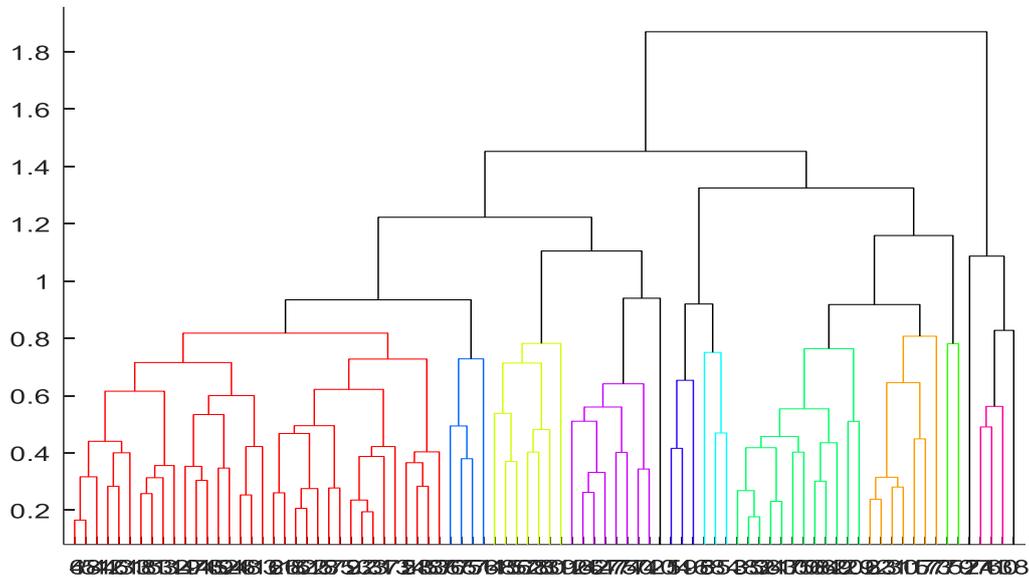


Figura 41 Dendrogramma pazienti non a rischio soluzione n°5.

Per i pazienti a rischio il taglio scelto ha portato ad ottenere un numero di cluster pari a 21.

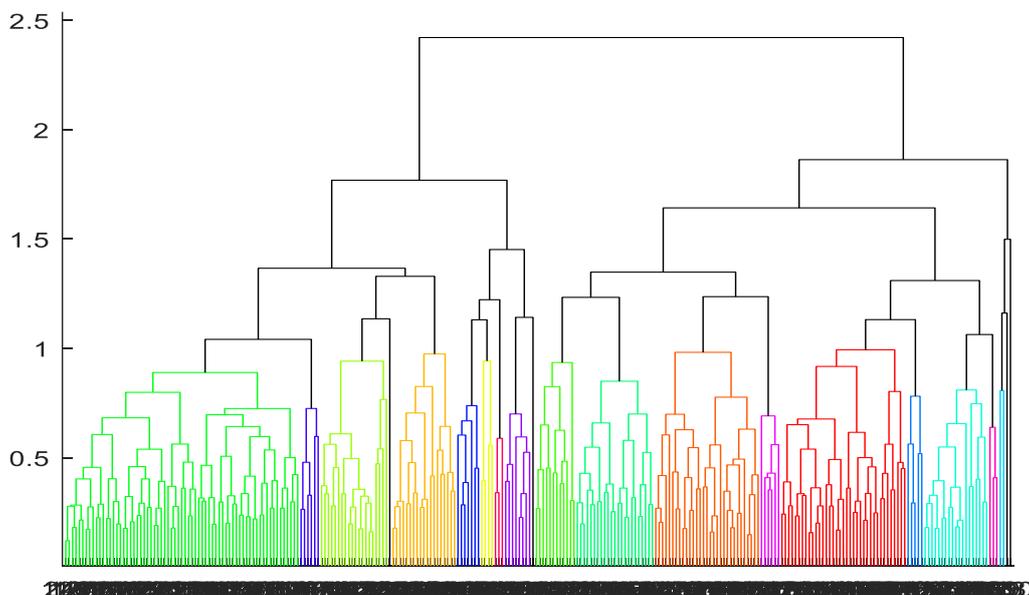


Figura 42 Dendrogramma pazienti a rischio soluzioni n°5.

3.2.6 Considerazioni clustering

Effettuato il clustering si è passati alla divisione del dataset. Da un primo sguardo dei dendrogrammi è facile capire che non tutte le soluzioni hanno lo stesso numero di cluster, ma soprattutto non tutti i cluster hanno la stessa dimensione. Di conseguenza si è dovuto scegliere un criterio con il quale creare i diversi training, test e validation set. La prima criticità la si trova nei cluster composti da una sola osservazione. Inserirla nel validation o nel test set non sarebbe stata la scelta ottimale in quanto la rete non sarebbe stata allenata a riconoscerla, quindi tutte le osservazioni che compongono un cluster a sé stante sono state inserite nel training set. Una seconda osservazione la si può fare pensando che i dati contenuti nel validation set non vengono mai utilizzati per l'allenamento della rete, di conseguenza questi dati dovranno essere presi dai cluster più numerosi in modo tale da permettere comunque alla rete di poter riconoscere bene tutti i tipi di osservazioni. Per questo motivo le 10 osservazioni per classe del validation set sono state prese dai 5 cluster più rappresentati. A questo punto, le osservazioni del training set sono state estratte dal primo all'ultimo cluster iterativamente fino a raggiungere le 58 osservazioni per classe. Le restanti infine, sono state inserite nel test set (18 pazienti non a rischio e 210 a rischio). Ricapitolando:

- Validation set delle prime due soluzioni composto da 2 osservazioni selezionate dai primi 5 cluster più rappresentati per un totale di 10 osservazioni per cluster;
- Osservazioni che compongono un cluster a sé stante inserite nel training set;
- Training set composto selezionando iterativamente osservazioni partendo dal primo fino all'ultimo cluster fino al raggiungimento delle 58 per classe;
- Test set composto dalle restanti osservazioni.

Va notato però che le estrazioni all'interno di ciascun cluster sono state effettuate in maniera random. Di conseguenza, per quanto riguarda le sole creazioni del training e del test set, sono state ripetute per 5 volte in modo tale da non rendere la rete neurale allenata soggetta alla casualità dell'estrazione random. Quindi, tutte le reti verranno allenate utilizzando tutti e 5 i training e test set e la valutazione delle prestazioni verrà effettuata considerando tutti e 5 i risultati.

3.3 Reti Neurali

A questo punto risulta essere tutto pronto per iniziare a testare le diverse reti neurali, non resta altro che scegliere quante e quali strutture allenare, e come inizializzare i pesi. Per quanto riguarda la dimensione si sono allenate le seguenti strutture:

- Con 1 layer nascosto con un numero di neuroni variabile tra il doppio delle feature della soluzione corrente e 10 neuroni a passo di 10;
- Con 2 layers nascosti, con un numero di neuroni del primo layer variabile tra il doppio delle feature della soluzione corrente e 10 neuroni a passo di 10, ed un numero di neuroni del secondo layer pari alla metà del layer precedente;
- Con 3 layer nascosti, con un numero di neuroni del primo layer variabile tra il doppio delle feature della soluzione corrente e 10 neuroni a passo di 10, ed un numero di neuroni del secondo e terzo layer pari alla metà del layer precedente.

Infine, i pesi sono stati inizializzati random in un valore compreso tra -1 e 1. Come per la clusterizzazione, anche in questo caso per ogni struttura l'inizializzazione dei pesi e l'allenamento della rete sono stati effettuati per 5 volte. Di conseguenza ogni struttura verrà allenata per un totale di 25 volte (5 diverse matrici di training e test set, 5 ripetizioni per matrice).

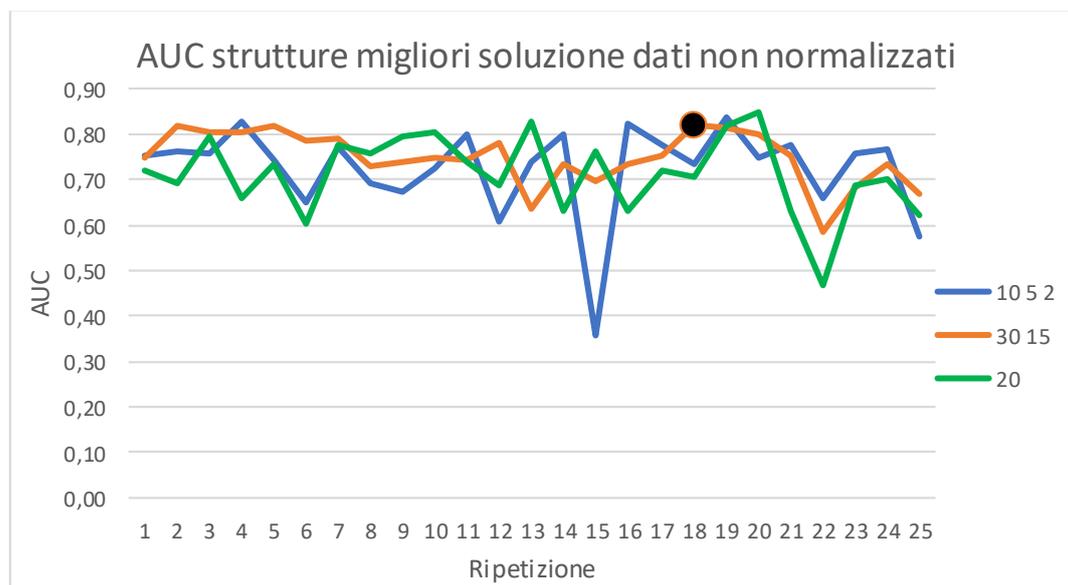
3.3.1 Valutazione delle prestazioni

Per la valutazione delle performance prima sono state allenate le strutture attraverso il training set, poi sono state testate attraverso il test set utilizzando la curva ROC e l'area sottesa alla curva (AUC). Per una valutazione complessiva delle strutture, oltre al valore AUC di ciascuna, sono state calcolate media, mediana e il delta di tutte e 25 le ripetizioni. Infine, per una prima scrematura, è stata creata una tabella per ogni dataset (presente in appendice A) con le tre migliori strutture da 1,2 e 3 layers nascosti, dove per strutture migliori si intendono quelle con un più alto valore della mediana delle 25 ripetizioni. La scelta della mediana piuttosto che della media è dovuta al fatto che quest'ultima è maggiormente influenzata dagli outliers, ovvero da quelle strutture che forniscono valori di AUC molto maggiori o minori rispetto all'andamento medio. Una volta selezionate le strutture migliori, queste sono poi state validate con il validation set attraverso il quale è anche stato scelto il valore della soglia. Infine, per valutare l'impatto della normalizzazione

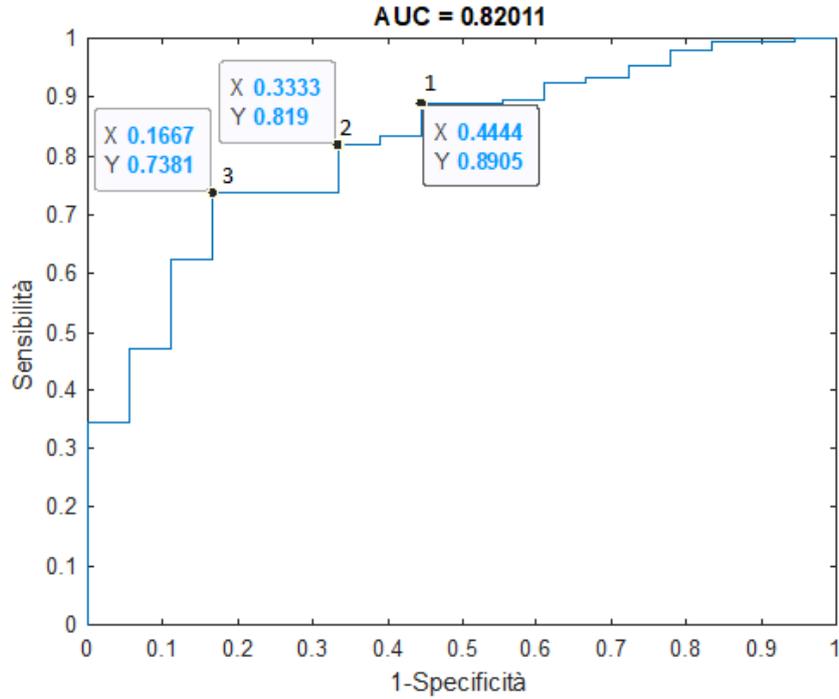
dei dati, tutte le strutture sono state allenate sia con i dati originali che attraverso i dati normalizzati utilizzando il metodo min-max scaling, ovvero facendo in modo che ogni feature fosse compresa nel range tra 0 e 1. Per una migliore visualizzazione dei risultati, di seguito sono riportati dei grafici che rappresentano l'andamento del valore AUC delle 3 strutture migliori per ogni soluzione del GA. Per la scelta della più performante, evidenziata nel grafico tramite un pallino nero, verrà valutato non solo il valore massimo, ma anche l'andamento medio che permette di considerare anche la stabilità della struttura. Per ognuna di queste poi, viene rappresentata la curva ROC corrispondente e l'effetto che ha la scelta della soglia sulle prestazioni mediante l'utilizzo della confusion matrix.

3.3.1.1 Soluzione n°1 con feature bloccate e fitness pari a 0.0920

Risultati soluzione n°1 con i dati non normalizzati:



Struttura con 2 layers nascosti da 30 e 15 neuroni, ripetizione n°18:

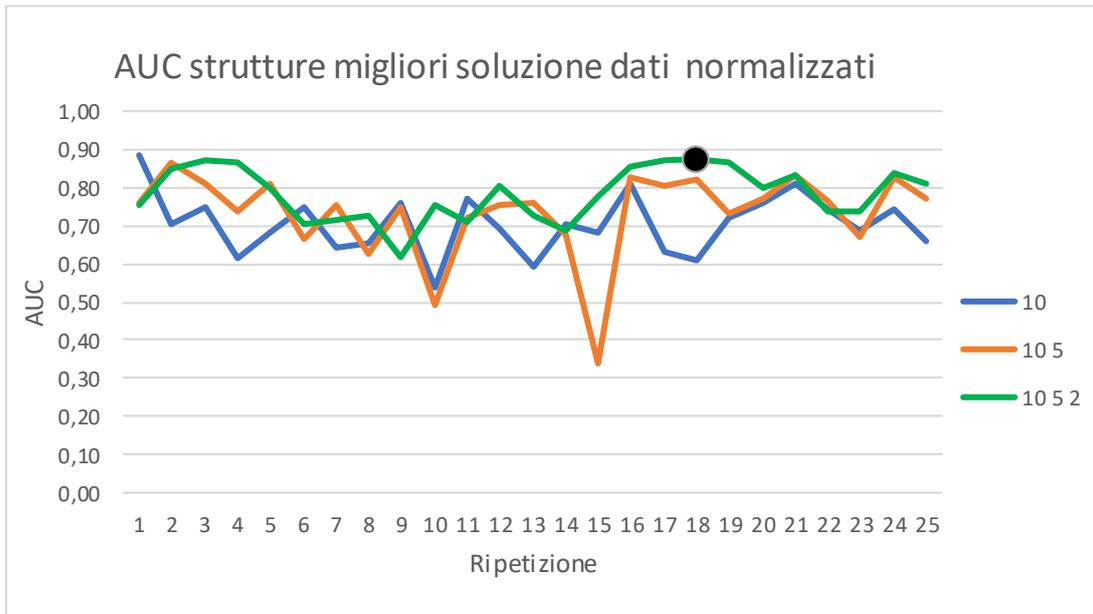


$S(1) = 0.4070$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 6 & 4 \\ 2 & 8 \end{bmatrix}$	70%	80%	60%
Training + test set	$\begin{bmatrix} 48 & 28 \\ 25 & 243 \end{bmatrix}$	84%	90%	63%

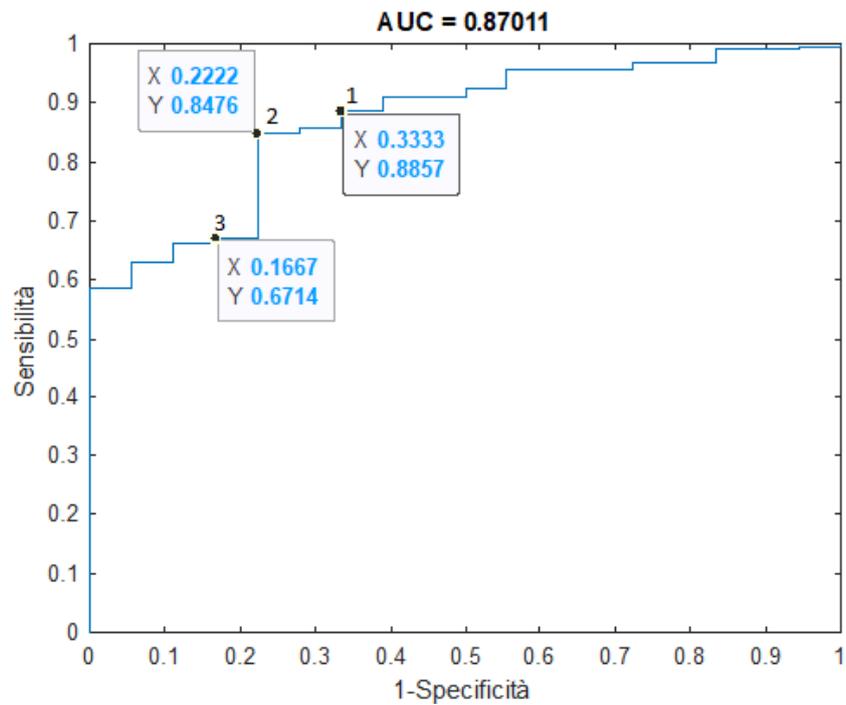
$S(2) = 0.4726$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 7 & 3 \\ 3 & 7 \end{bmatrix}$	70%	70%	70%
Training + test set	$\begin{bmatrix} 56 & 20 \\ 45 & 223 \end{bmatrix}$	81%	83%	73%

$S(3) = 0.5380$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 9 & 1 \\ 3 & 7 \end{bmatrix}$	80%	70%	90%
Training + test set	$\begin{bmatrix} 65 & 11 \\ 67 & 201 \end{bmatrix}$	77%	75%	85%

Risultati soluzione n°1 con i dati normalizzati:



Struttura con 3 layers nascosti da 10 5 2 neuroni, ripetizione n°18:



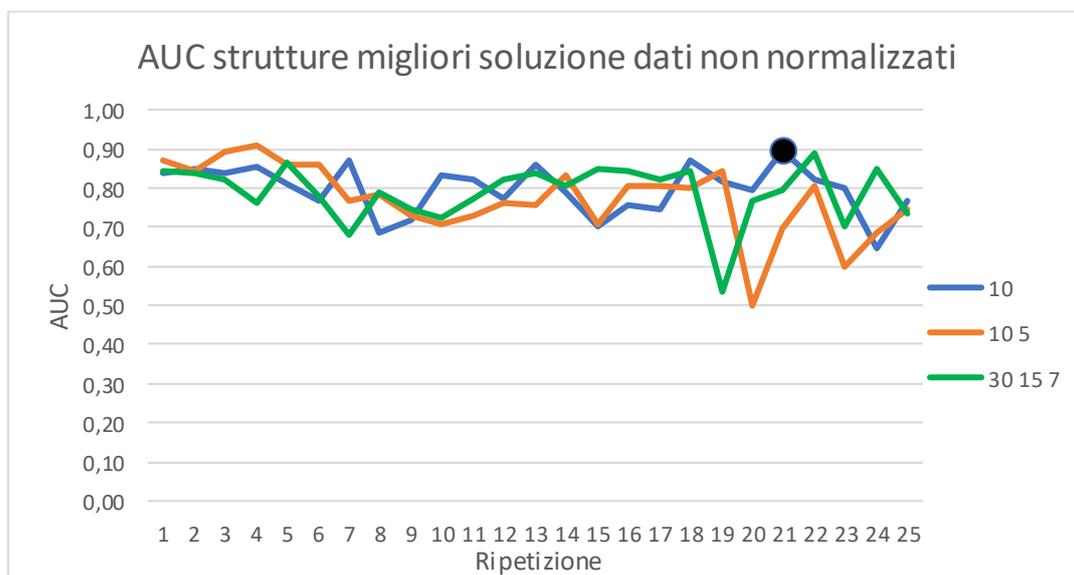
$S(1) = 0.2968$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 6 & 4 \\ 2 & 8 \end{bmatrix}$	70%	80%	60%
Training + test set	$\begin{bmatrix} 49 & 27 \\ 26 & 242 \end{bmatrix}$	84%	90%	64%

$S(2) = 0.3411$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 6 & 4 \\ 2 & 8 \end{bmatrix}$	70%	80%	60%
Training + test set	$\begin{bmatrix} 53 & 23 \\ 36 & 232 \end{bmatrix}$	83%	86%	70%

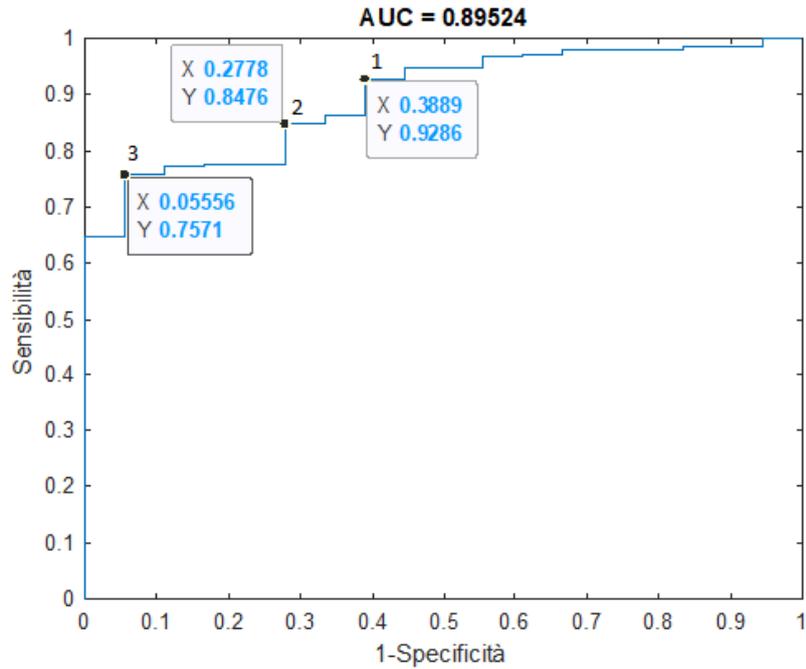
$S(3) = 0.5380$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 7 & 3 \\ 3 & 7 \end{bmatrix}$	70%	70%	70%
Training + test set	$\begin{bmatrix} 66 & 10 \\ 78 & 190 \end{bmatrix}$	74%	70%	86%

3.3.1.2 Soluzione n°2 con feature bloccate e fitness pari a 0.1061

Risultati soluzione n°2 con i dati non normalizzati:



Struttura con uno strato nascosto da 10 neuroni, ripetizione n°21:

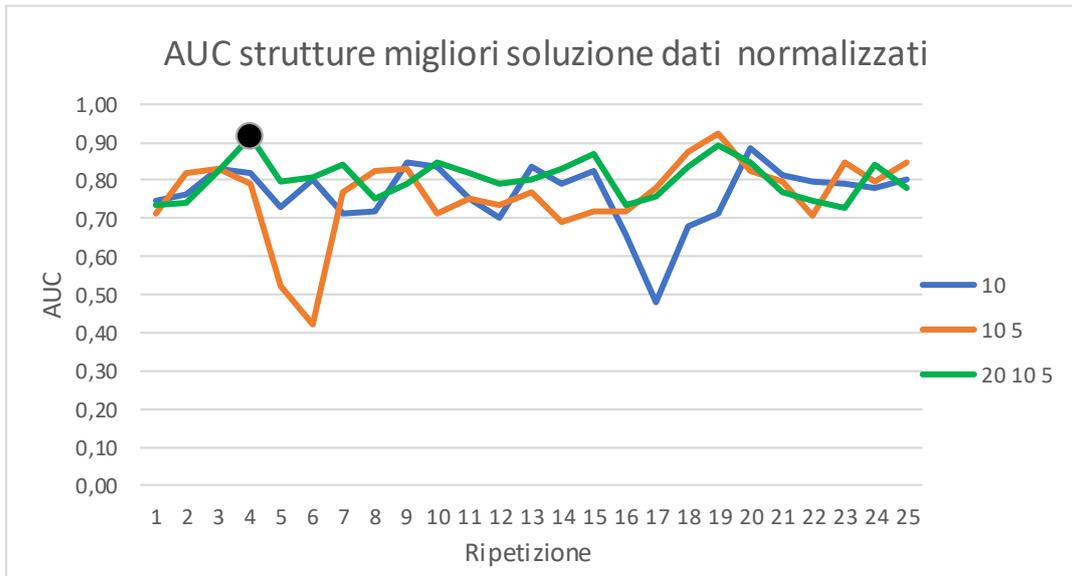


$S(1) = 0.3835$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 6 & 4 \\ 1 & 9 \end{bmatrix}$	75%	90%	60%
Training + test set	$\begin{bmatrix} 39 & 37 \\ 22 & 246 \end{bmatrix}$	83%	91%	51%

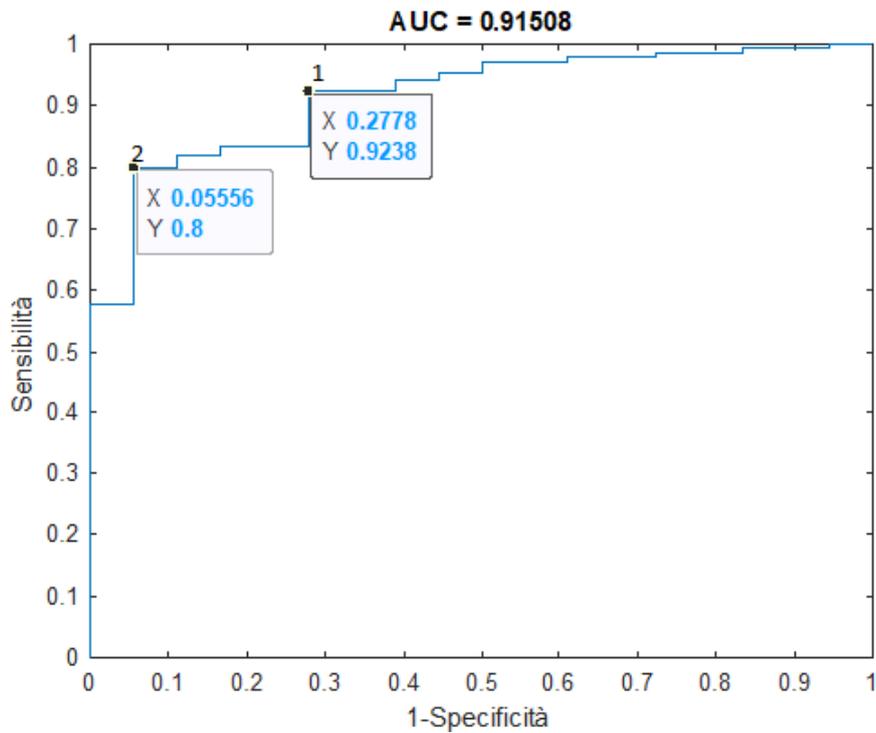
$S(2) = 0.4561$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 7 & 3 \\ 2 & 8 \end{bmatrix}$	75%	80%	70%
Training + test set	$\begin{bmatrix} 54 & 22 \\ 44 & 224 \end{bmatrix}$	80%	83%	71%

$S(3) = 0.5046$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 7 & 3 \\ 3 & 7 \end{bmatrix}$	70%	70%	70%
Training + test set	$\begin{bmatrix} 63 & 13 \\ 64 & 204 \end{bmatrix}$	77%	76%	82%

Risultati soluzione n°2 con i dati normalizzati:



Struttura con 3 layers nascosti da 20 10 e 5 neuroni, ripetizione n°4:

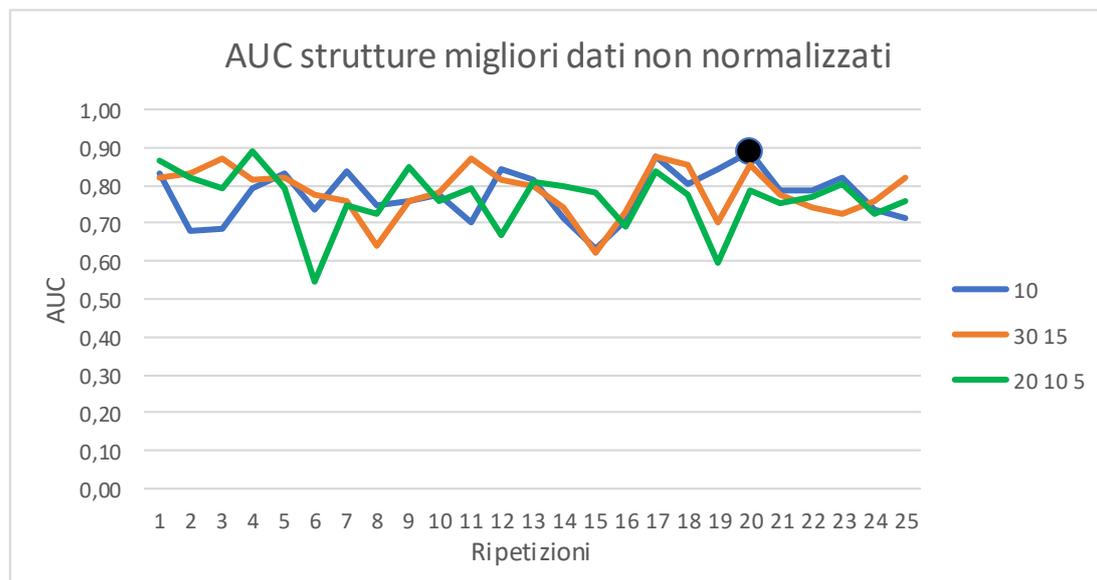


$S(1) = 0.3161$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 8 & 2 \\ 1 & 9 \end{bmatrix}$	85%	90%	80%
Training + test set	$\begin{bmatrix} 45 & 31 \\ 24 & 244 \end{bmatrix}$	84%	91%	59%

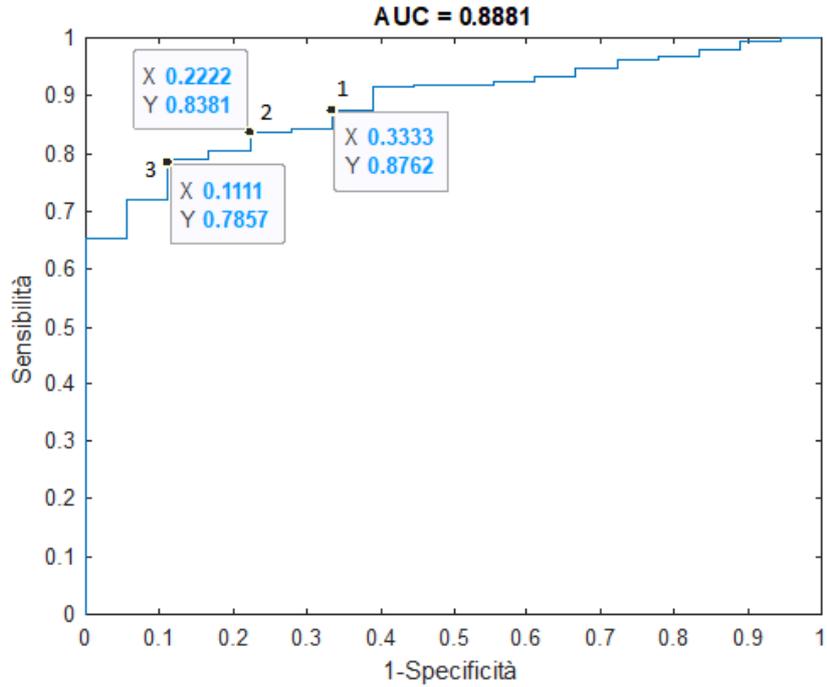
$S(2) = 0.4401$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 7 & 3 \\ 2 & 8 \end{bmatrix}$	80%	80%	80%
Training + test set	$\begin{bmatrix} 58 & 18 \\ 55 & 213 \end{bmatrix}$	79%	79%	76%

3.3.1.3 Soluzione n°3 con le feature libere fitness pari a 0.0883

Risultati soluzione n°3 con i non dati normalizzati



Struttura con un layer nascosto da 10 neuroni, ripetizione n° 20:

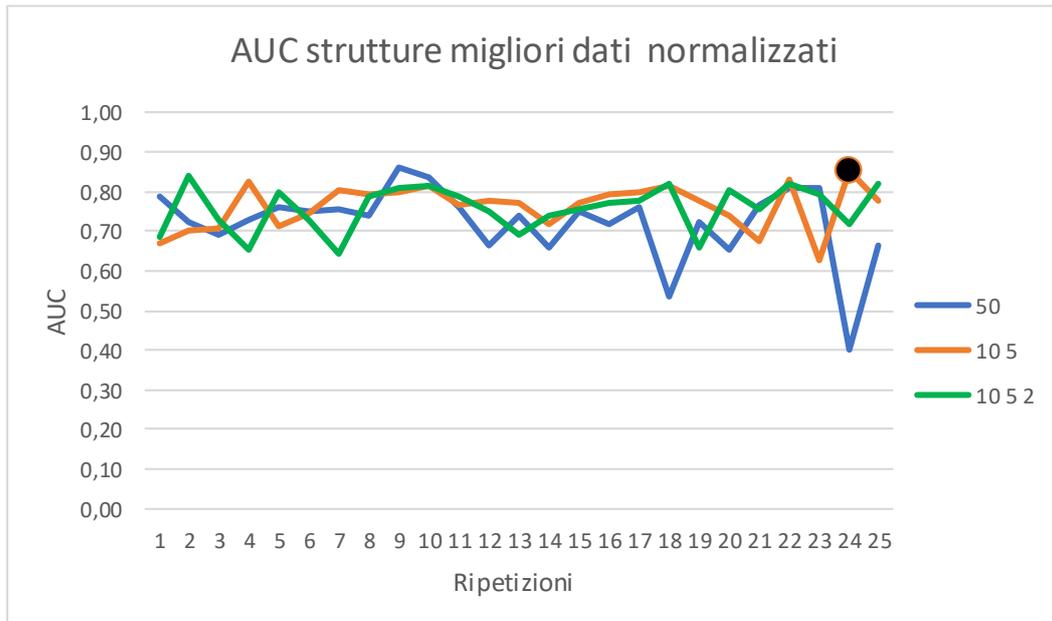


$S(1) = 0.4236$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 6 & 4 \\ 1 & 9 \end{bmatrix}$	75%	90%	60%
Training + test set	$\begin{bmatrix} 53 & 23 \\ 33 & 235 \end{bmatrix}$	83%	87%	69%

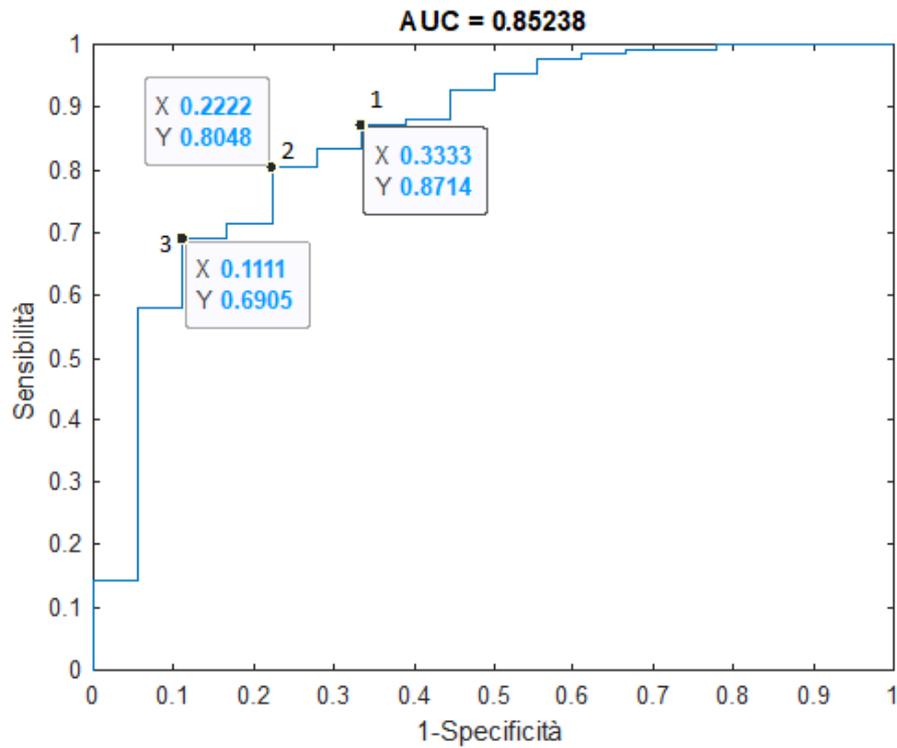
$S(2) = 0.4480$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 7 & 3 \\ 1 & 9 \end{bmatrix}$	80%	90%	70%
Training + test set	$\begin{bmatrix} 57 & 19 \\ 41 & 227 \end{bmatrix}$	82%	84%	75%

$S(3) = 0.4735$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 7 & 3 \\ 1 & 9 \end{bmatrix}$	80%	90%	70%
Training + test set	$\begin{bmatrix} 63 & 13 \\ 54 & 214 \end{bmatrix}$	80%	80%	83%

Risultati soluzione n°3 con i dati normalizzati



Struttura con due layers anscoisti da 10 e 5 neuroni, ripetizione n°24:



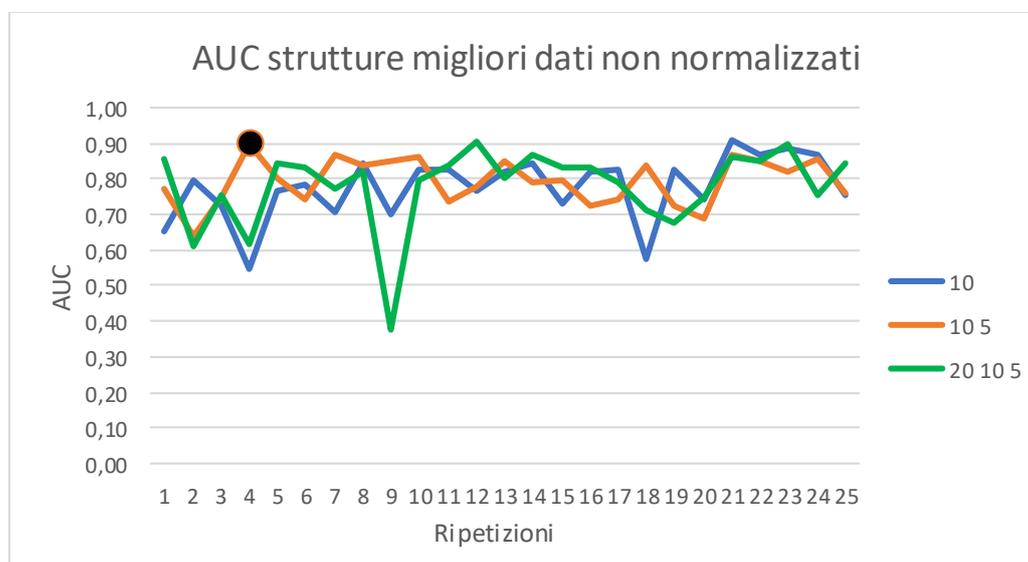
$S(1) = 0.3426$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 6 & 4 \\ 0 & 10 \end{bmatrix}$	80%	100%	60%
Training + test set	$\begin{bmatrix} 41 & 35 \\ 34 & 234 \end{bmatrix}$	80%	87%	54%

$S(2) = 0.3923$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 8 & 2 \\ 1 & 9 \end{bmatrix}$	85%	90%	80%
Training + test set	$\begin{bmatrix} 48 & 28 \\ 50 & 218 \end{bmatrix}$	77%	81%	63%

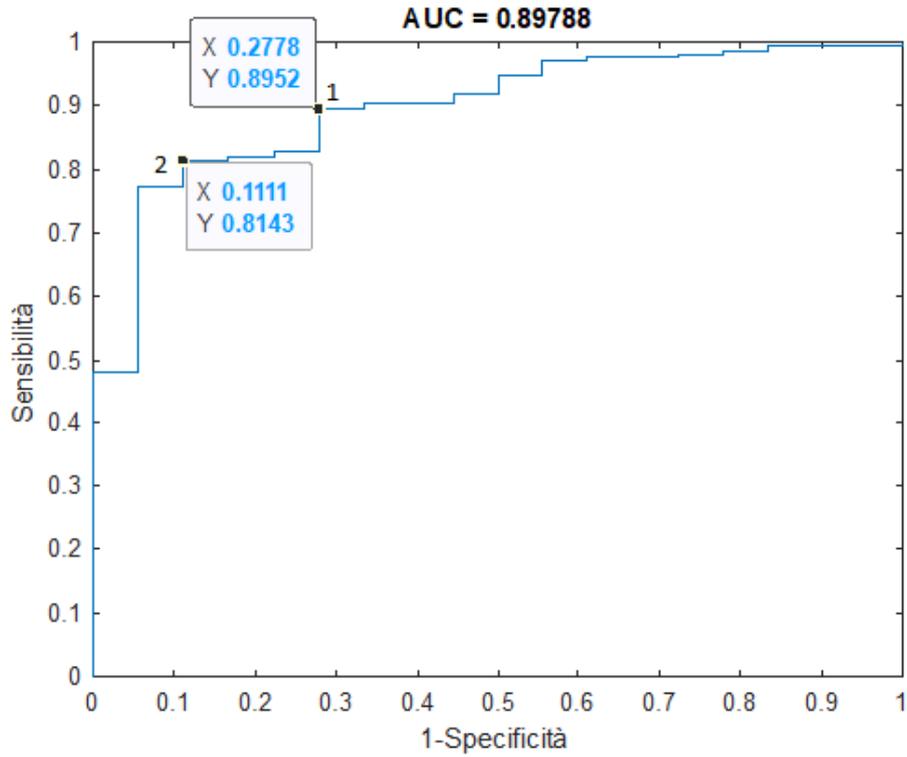
$S(3) = 0.4721$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 9 & 1 \\ 2 & 8 \end{bmatrix}$	85%	80%	90%
Training + test set	$\begin{bmatrix} 65 & 11 \\ 82 & 186 \end{bmatrix}$	72%	69%	85%

3.3.1.4 Soluzione n°4 con le feature libere fitness pari a 0.1002

Risultati soluzione n°4 con i non dati normalizzati



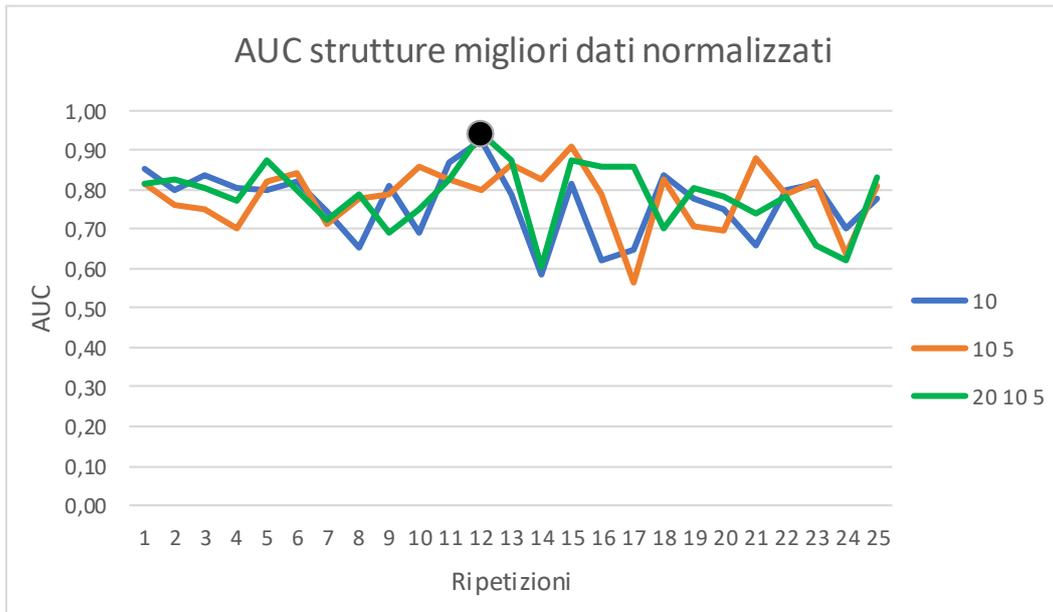
Struttura da due layers nascosta da 10 e 5 neuroni, ripetizione n°4:



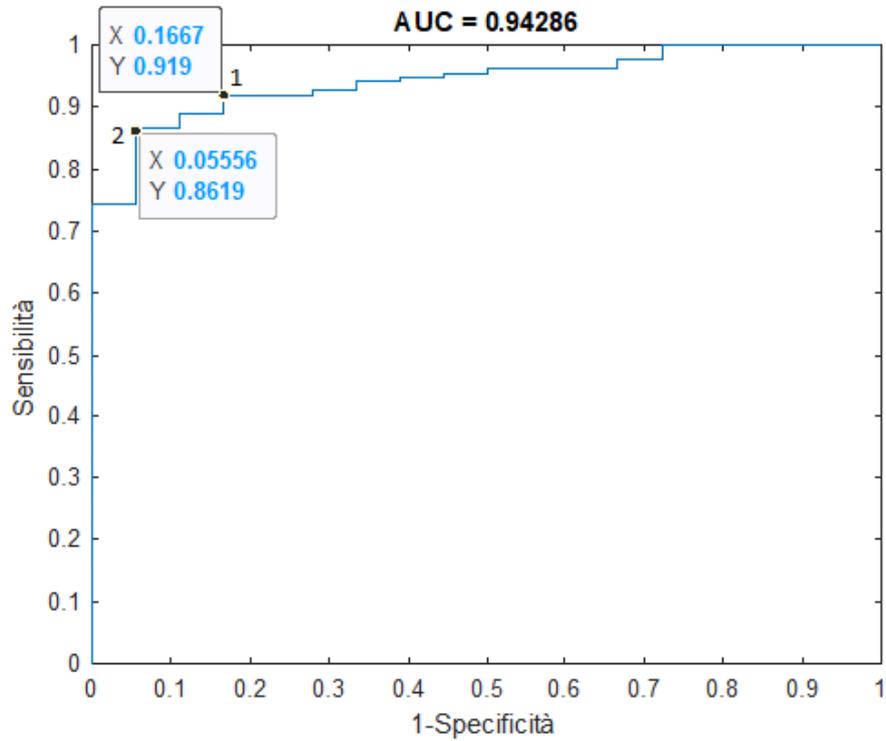
$S(1) = 0.4821$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 7 & 3 \\ 2 & 8 \end{bmatrix}$	75%	80%	70%
Training + test set	$\begin{bmatrix} 45 & 31 \\ 37 & 231 \end{bmatrix}$	80%	86%	59%

$S(2) = 0.5571$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 9 & 1 \\ 4 & 6 \end{bmatrix}$	75%	60%	90%
Training + test set	$\begin{bmatrix} 54 & 22 \\ 56 & 212 \end{bmatrix}$	77%	79%	71%

Risultati soluzione n°4 con i dati normalizzati



Struttura a tre layers nascosti da 20 10 e 5 neuroni, ripetizione n°12:

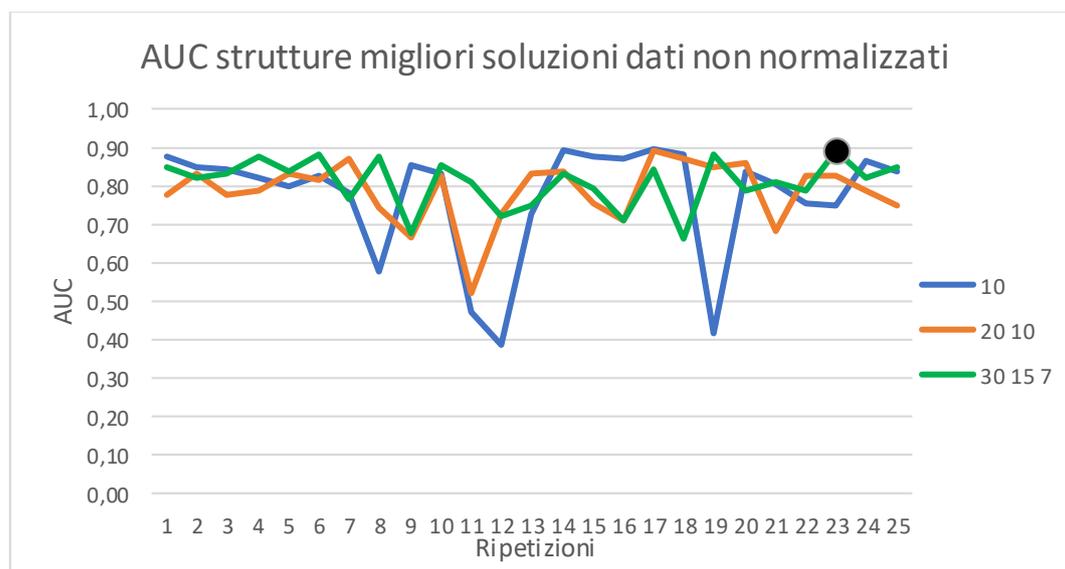


$S(1) = 0.3850$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 2 & 8 \\ 1 & 9 \end{bmatrix}$	55%	90%	20%
Training + test set	$\begin{bmatrix} 40 & 36 \\ 31 & 237 \end{bmatrix}$	80%	88%	52%

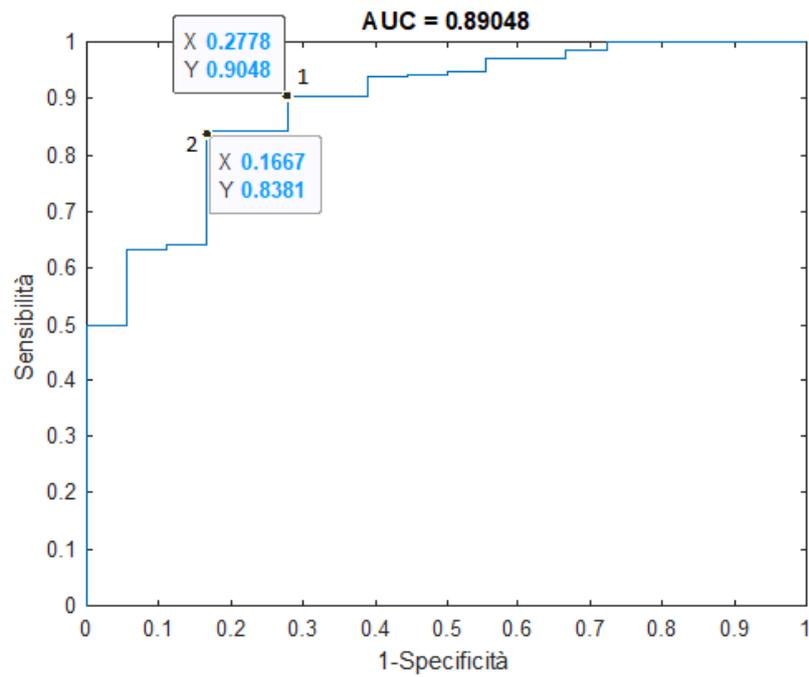
$S(2) = 0.4554$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 5 & 5 \\ 3 & 7 \end{bmatrix}$	60%	70%	50%
Training + test set	$\begin{bmatrix} 50 & 26 \\ 43 & 225 \end{bmatrix}$	80%	84%	66%

3.3.1.5 Soluzione n°5 con le clinical feature

Risultati soluzione n°5 con i dati non normalizzati:



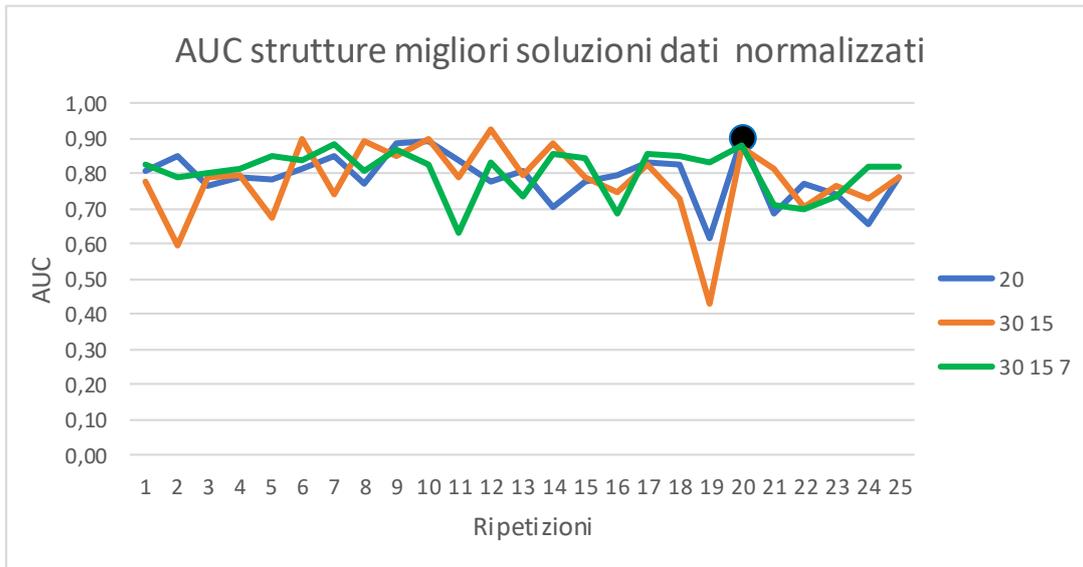
Struttura a 3 layers nascosti da 30 15 e 7 neuroni, ripetizione n°23:



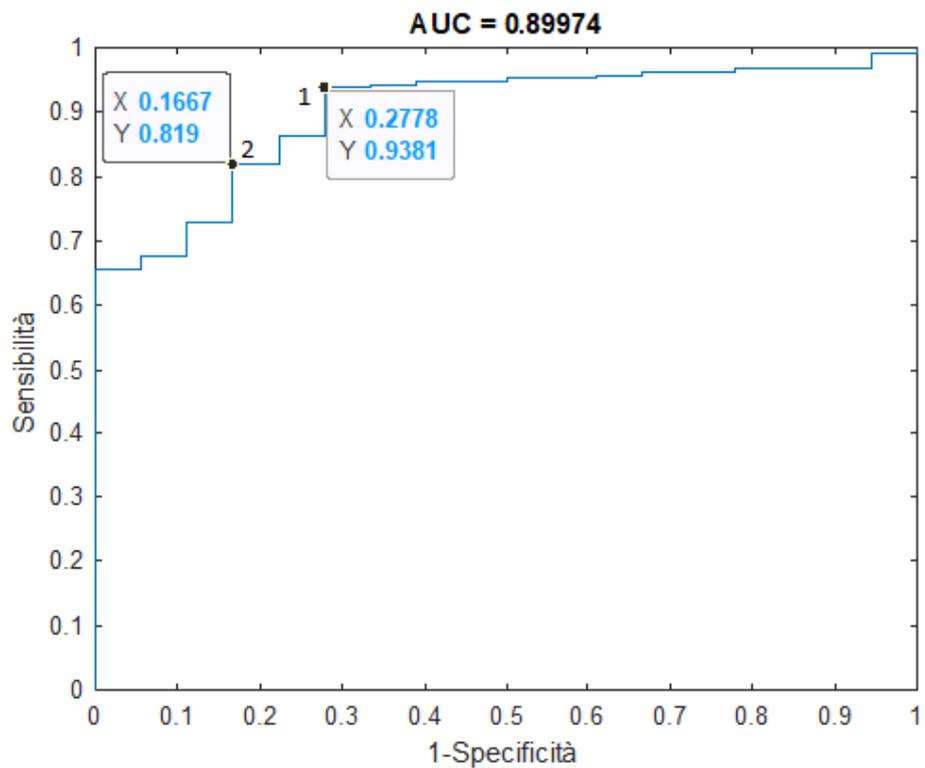
$S(1) = 0.1268$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 4 & 6 \\ 1 & 9 \end{bmatrix}$	65%	90%	40%
Training + test set	$\begin{bmatrix} 45 & 31 \\ 25 & 243 \end{bmatrix}$	83%	90%	59%

$S(2) = 0.1938$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 5 & 5 \\ 3 & 7 \end{bmatrix}$	60%	60%	60%
Training + test set	$\begin{bmatrix} 54 & 22 \\ 38 & 230 \end{bmatrix}$	82%	85%	71%

Risultati soluzione n°5 con i dati normalizzati:



Struttura da un layer nascosto da 20 neuroni, ripetizione n°20:



$S(1) = 0.3399$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 2 & 8 \\ 2 & 8 \end{bmatrix}$	50%	80%	20%
Training + test set	$\begin{bmatrix} 42 & 33 \\ 16 & 252 \end{bmatrix}$	85%	94%	55%

$S(2) = 0.5741$	Confusion matrix	Accuratezza	Sensibilità	Specificità
Validation set	$\begin{bmatrix} 6 & 4 \\ 3 & 7 \end{bmatrix}$	65%	70%	60%
Training + test set	$\begin{bmatrix} 62 & 14 \\ 44 & 224 \end{bmatrix}$	83%	83%	82%

3.3.2 Considerazioni sulle reti neurali

Come si può vedere dai risultati la percentuale di accuratezza di tutte le strutture delle soluzioni ottenute tramite feature selection si aggira tra il 70% e 80%, con la percentuale di sensibilità spesso maggiore rispetto quella della specificità. Vero è che durante un test di screening è molto più importante riuscire a classificare correttamente i pazienti a rischio che non il contrario, ma avere una specificità eccessivamente bassa, del 40/50%, renderebbe quasi nullo l'utilizzo della classificazione, quindi in fase di scelta sarebbe opportuno non solo dare priorità alla sensibilità ma anche ad un equilibrio tra la corretta classificazione delle due classi. Questa problematica può essere dovuta principalmente allo sbilanciamento del dataset a favore dei pazienti a rischio frattura che ha portato ad una polarizzazione del classificatore. La quinta soluzione, ottenuta con le 14 feature utilizzate in clinica, permette di ottenere un'accuratezza, sensitività e specificità di circa il 65%. Questo dimostra come, utilizzando una combinazione di feature diverse dalle sole 14 usate fino ad ora, permette di ottenere delle prestazioni migliori. Un dato da tenere in considerazione è il metodo con cui è stato costruito il validation set. Infatti, quest'ultimo è stato creato selezionando pazienti dai primi 5 cluster più rappresentati, di conseguenza questi dati non descrivono a pieno la capacità di classificazione delle reti. Inoltre, a causa della ridotta dimensione della matrice, l'accuratezza varia con un passo del 5%, mentre sensibilità e specificità del 10%. Per questi motivi, alle prestazioni relative al validation set, sono state affiancate quelle del training e test set. Questo confronto è stato effettuato

per avere la possibilità di una visione più generica delle prestazioni, in particolar modo per controllare la corrispondenza delle prestazioni tra validation set con quelle del training e test set. Nella tabella sottostante sono riportate le prestazioni della struttura migliore di ogni soluzione de GA:

Soluzione	Dataset	Accuratezza	Sensibilità	Specificità
1 fitness=0.0920	validation	70%	70%	70%
	Train+test	81%	83%	73%
2 fitness=0.1061	validation	80%	80%	80%
	Train+test	79%	79%	76%
3 fitness=0.0883	validation	80%	90%	70%
	Train+test	82%	84%	75%
4 fitness=0.1002	validation	75%	80%	70%
	Train+test	80%	86%	59%
5 cliniche	validation	65%	70%	60%
	Train+test	83%	83%	82%

4. Conclusioni

L'osteoporosi è una malattia sistemica dell'apparato scheletrico dovuta ad una diminuzione della densità ossea, con un conseguente aumento del rischio di frattura, in particolar modo di vertebre, femore e polso anche per traumi minimi. Ad oggi l'osteoporosi viene diagnosticata mediante la valutazione del T-score, che indica la deviazione standard della densità minerale ossea (BMD) rispetto il picco medio di massa ossea. Questa viene misurata tramite la tecnica dual-energy x-ray absorptiometry (DEXA). È stato dimostrato che il rischio di frattura aumenta con il diminuire del T-score, ma anche altri parametri incidono particolarmente nella sua valutazione, di conseguenza in clinica, per la valutazione del rischio frattura vengono utilizzate una combinazione di parametri tra cui dati relativi al paziente (età, peso, ecc...) ed altri relativi alle principali zone anatomiche soggette a frattura (zona femorale e lombare). Lo scopo di questo studio è la costruzione di un classificatore per la stima del rischio di frattura complessiva in pazienti osteoporotici, ovvero sia femorale che lombare, e di valutare l'importanza di quelle feature utilizzate in clinica in combinazione con altri parametri ricavati da una DEXA Horizon della Hologic. I dati a disposizione sono composti da un totale di 364 pazienti di cui 278 a rischio frattura e 86 non a rischio ed un totale di 46 feature per ogni paziente. È stata effettuata una prima feature selection mediante Algoritmi Genetici. Per valutare l'incidenza delle 14 feature utilizzate in clinica, la feature selection è stata applicata due volte con due diverse strategie: la prima effettuando la feature selection solo alle restanti 32 feature e mantenendo le prime 14; la seconda effettuando la feature selection a tutte le 46 feature. I risultati di questa prima fase dimostrano che non è possibile determinare una soluzione migliore in assoluto, ma piuttosto soluzioni diverse in numero e tipo di feature che restituiscono la stessa "bontà" in termini di accuratezza e stabilità nella classificazione. Di conseguenza per l'allenamento delle reti neurali sono state selezionate 4 diverse soluzioni del GA. È interessante notare come tre di queste soluzioni presentano almeno tutte le feature utilizzate in clinica, mentre una ne contiene 10 su 14. Inoltre, per un confronto con quella che è ad oggi la pratica clinica, si è deciso di allenare delle reti anche utilizzando solo i 14 parametri. Per ognuna di queste soluzioni sono poi state allenate diverse strutture per dimensione e numero di neuroni, validate poi attraverso la curva ROC e l'area sottesa alla curva (AUC). Attraverso i dati a disposizione i risultati dimostrano che l'utilizzo di algoritmi di machine learning permettono di aumentare

l'accuratezza della stima rispetto l'utilizzo delle sole feature utilizzate ad oggi in clinica. Nonostante i risultati siano promettenti non è possibile determinare una coppia soluzione-struttura migliore in quanto la ridotta numerosità dei pazienti non a rischio non permette di ottenere una soluzione abbastanza robusta da poter affermare in maniera assoluta quale sia la migliore. Questo è infatti uno dei principali limiti di questo lavoro. È possibile però affermare con certezza che si ottengono prestazioni decisamente migliori utilizzando una combinazione di tutte o alcune delle feature utilizzate in clinica e delle altre 32 disponibili. In futuro, sarebbe interessante poter approfondire questo lavoro aumentando il numero di pazienti non a rischio frattura in modo tale da ottenere un dataset più ampio e bilanciato. Inoltre, un altro sviluppo interessante potrebbe essere quello di costruire un classificatore specifico per ogni zona anatomica (lombare e femorale) in modo tale da ottenere non solo una stima più precisa, ma anche la zona del rischio di frattura per poter indirizzare eventuali trattamenti farmacologici.

Riferimenti bibliografici

- [1] Seeley, Stephens, Tate, "Anatomia", 2th edizione, Idelson-Gnocchi.
- [2] Ministero della Salute, <http://www.salute.gov.it>.
- [3] Osservatorio Nazionale Salute della Donna, <https://www.ondaosservatorio.it>.
- [4] Nelson B. Watts, MD, FACP, MACE, John P. Bilezikian, MD, MACE, Pauline M. Camacho, MD, FACE, Susan L. Greenspan, MD, FACP, FACE, Steven T. Harris, MD, FACE, Stephen F. Hodgson, MD, FACP, MACE, Michael Kleerekoper, MD, MACE, Marjorie M. Luckey, MD, FACE, Michael R. McClung, MD, FACP, FACE, Rachel Pessah Pollack, MD, and Steven M. Petak, MD, JD, FACE, FCLM. "American Association Of Clinical Endocrinologists Medical Guidelines For Clinical Practice For The Diagnosis And Treatment Of Postmenopausal Osteoporosis: Executive Summary Of Recommendations". *Endocr Pract.* 2010; 16(6): 1016-1019.
- [5] Sydney Lou Bonnick, 2010, *Bone Densitometry in Clinical Practice, application and interpretation*, 3th edition, Humana Press.
- [6] Ethel S. Siris; Ya-Ting Chen; Thomas A. Abbott; Elizabeth Barrett-Connor; Paul D. Miller; Lois E. Wehren; Marc L. Berger. "Bone Mineral Density Thresholds for Pharmacological Intervention to Prevent Fractures." *Arch Intern Med.* 2004;164(10), May 24, 2004. Web.
- [7] Mohammad J Mirzaali, Flavia Libonati, Davide Ferrario, Luca Rinaudo, Carmelo Messina, Fabio M Ulivieri, Bruno M Cesana, Matteo Strano, and Laura Vergani. "Determinants of Bone Damage: An Ex-vivo Study on Porcine Vertebrae." *PLoS ONE* 13.8 (2018): E0202210. Web.
- [8] Silva, Barbara C, William D Leslie, Heinrich Resch, Olivier Lamy, Olga Lesnyak, Neil Binkley, Eugene V McCloskey, John A Kanis, and John P Bilezikian. "Trabecular Bone Score: A Noninvasive Analytical Method Based Upon the DXA Image." *Journal of Bone and Mineral Research* 29.3 (2014): 518-30. Web.
- [9] Colombo, Chiara, Flavia Libonati, Luca Rinaudo, Martina Bellazzi, Fabio Massimo Ulivieri, and Laura Vergani. "A New Finite Element Based Parameter to Predict Bone Fracture." *PloS One* 14.12 (2019): E0225905. Web.

- [10] Fabio Massimo Ulivieri, Luca Petruccio Piodi, Luca Rinaudo, Paolo Scanagatta, and Bruno Mario Cesana. "Bone Strain Index in the Prediction of Vertebral Fragility Refracture." *European Radiology Experimental* 4.1 (2020): 1-6. Web.
- [11] Lee, John Aldo, Jordan, M, and Verleysen, Michel. *Nonlinear Dimensionality Reduction*. Springer New York, 2007. Web.
- [12] Andries P. Engelbrecht, "Computational Intelligence, an introduction", Second Edition, Wiley
- [13] Bozorg-Haddad, Omid, Loaiciga, Hugo A, and Solgi, Mohammad. "Meta-heuristic and Evolutionary Algorithms for Engineering Optimization." Wiley, 2017. Web
- [14] Cunningham, Padraig & Delany, Sarah. (2007). "k-Nearest neighbour classifiers." *Mult Classif Syst*.
- [15] Melanie Mitchel, "An introduction to Genetic Algorithms", fifth edition, MIT press LTD
- [16] Falco, Ivano & Tarantino, Ernesto & Della Cioppa, Antonio. (2003). I meccanismi di mutazione negli Algoritmi Genetici: verso una migliore imitazione della natura.
- [17] Webb, Andrew R., and Cosey, Keith D. *Statistical Pattern Recognition*. John Wiley & Sons, 2011. Web.
- [18] Hagan, M.T, and M.B Menhaj. "Training Feedforward Networks with the Marquardt Algorithm." *IEEE Transactions on Neural Networks* 5.6 (1994): 989-93. Web.
- [19] Van Erkel, Arian R, and Peter M.Th Pattynama. "Receiver Operating Characteristic (ROC) Analysis: Basic Principles and Applications in Radiology." *European Journal of Radiology* 27.2 (1998): 88-94. Web.

Appendice A

Nelle tabelle seguenti sono riportate le 3 migliori strutture con 1,2 e 3 layers nascosti per ogni soluzione, sia attraverso il dataset originale che normalizzato.

Soluzione n°1 fitness 0.0920 feature bloccate con i dati non normalizzati

Divisione	Ripetizione	20	10	30	30 e 15	10 e 5	20 e 10	10 e 5 e 2	30 e 15 e 7	40 e 20 e 10
Prima divisione training-test	1	0,72	0,74	0,77	0,75	0,69	0,79	0,75	0,79	0,80
	2	0,69	0,75	0,76	0,82	0,80	0,83	0,76	0,76	0,79
	3	0,80	0,79	0,83	0,81	0,77	0,69	0,76	0,78	0,85
	4	0,66	0,63	0,80	0,80	0,77	0,86	0,83	0,77	0,83
	5	0,74	0,71	0,72	0,82	0,76	0,74	0,74	0,69	0,81
Seconda divisione training-test	1	0,60	0,76	0,49	0,79	0,81	0,28	0,65	0,69	0,63
	2	0,78	0,72	0,79	0,79	0,76	0,65	0,77	0,78	0,75
	3	0,76	0,80	0,56	0,73	0,82	0,70	0,69	0,70	0,76
	4	0,79	0,72	0,75	0,74	0,75	0,74	0,67	0,76	0,64
	5	0,80	0,69	0,83	0,75	0,35	0,72	0,73	0,67	0,61
Terza divisione training-test	1	0,74	0,69	0,71	0,74	0,73	0,72	0,80	0,82	0,70
	2	0,69	0,68	0,53	0,78	0,64	0,76	0,61	0,77	0,69
	3	0,83	0,69	0,56	0,63	0,76	0,68	0,74	0,82	0,84
	4	0,63	0,67	0,78	0,74	0,76	0,80	0,80	0,74	0,70
	5	0,76	0,54	0,64	0,69	0,82	0,58	0,36	0,48	0,57
Quarta divisione training-test	1	0,63	0,69	0,74	0,74	0,75	0,74	0,82	0,68	0,68
	2	0,72	0,83	0,70	0,75	0,74	0,86	0,77	0,75	0,81
	3	0,71	0,74	0,72	0,82	0,66	0,75	0,73	0,81	0,84
	4	0,82	0,83	0,61	0,81	0,64	0,75	0,84	0,81	0,82
	5	0,85	0,78	0,75	0,80	0,69	0,81	0,75	0,79	0,72
Quinta divisione training-test	1	0,63	0,55	0,43	0,75	0,78	0,79	0,78	0,71	0,81
	2	0,47	0,60	0,69	0,58	0,66	0,62	0,66	0,64	0,79
	3	0,69	0,57	0,67	0,68	0,80	0,57	0,76	0,69	0,65
	4	0,70	0,74	0,60	0,74	0,74	0,71	0,77	0,52	0,68
	5	0,62	0,72	0,50	0,67	0,70	0,75	0,58	0,63	0,69
	avg	0,71	0,71	0,68	0,75	0,73	0,72	0,72	0,72	0,74
	median	0,72	0,72	0,71	0,75	0,75	0,74	0,75	0,75	0,75
	delta	0,38	0,29	0,40	0,24	0,47	0,58	0,48	0,35	0,29

Soluzione n°1 fitness 0.0920 feature bloccate con i dati normalizzati

Divisione	Ripetizione	40	50	10	10 e 5	20 e 10	30 e 15	10 e 5 e 2	50 e 25 e 12	40 e 20 e 10
Prima divisione training-test	1	0,78	0,79	0,88	0,76	0,67	0,76	0,76	0,84	0,88
	2	0,42	0,78	0,71	0,86	0,85	0,84	0,85	0,59	0,74
	3	0,80	0,55	0,75	0,81	0,77	0,85	0,87	0,84	0,72
	4	0,75	0,71	0,61	0,74	0,80	0,74	0,86	0,87	0,77
	5	0,83	0,82	0,68	0,81	0,72	0,80	0,80	0,84	0,68
Seconda divisione training-test	1	0,62	0,71	0,75	0,67	0,68	0,59	0,70	0,60	0,75
	2	0,83	0,79	0,64	0,76	0,75	0,76	0,71	0,74	0,76
	3	0,76	0,78	0,65	0,63	0,71	0,75	0,72	0,81	0,81
	4	0,74	0,78	0,76	0,75	0,73	0,68	0,62	0,74	0,69
	5	0,70	0,73	0,54	0,49	0,73	0,64	0,75	0,72	0,75
Terza divisione training-test	1	0,72	0,62	0,77	0,72	0,83	0,73	0,71	0,59	0,76
	2	0,55	0,67	0,69	0,75	0,68	0,70	0,81	0,77	0,83
	3	0,62	0,69	0,59	0,76	0,73	0,37	0,73	0,77	0,73
	4	0,75	0,52	0,71	0,68	0,75	0,70	0,69	0,72	0,82
	5	0,72	0,82	0,68	0,34	0,70	0,74	0,78	0,67	0,77
Quarta divisione training-test	1	0,70	0,84	0,81	0,83	0,83	0,78	0,85	0,76	0,83
	2	0,73	0,68	0,63	0,81	0,79	0,76	0,87	0,70	0,80
	3	0,81	0,68	0,61	0,82	0,66	0,78	0,87	0,80	0,80
	4	0,81	0,73	0,72	0,73	0,85	0,75	0,86	0,86	0,73
	5	0,71	0,67	0,76	0,77	0,81	0,66	0,80	0,77	0,74
Quinta divisione training-test	1	0,79	0,82	0,81	0,83	0,82	0,79	0,83	0,83	0,43
	2	0,77	0,63	0,74	0,76	0,80	0,67	0,74	0,84	0,72
	3	0,73	0,73	0,69	0,67	0,76	0,68	0,74	0,69	0,75
	4	0,73	0,68	0,74	0,82	0,73	0,79	0,84	0,72	0,78
	5	0,80	0,78	0,66	0,77	0,80	0,80	0,81	0,65	0,73
	avg	0,73	0,72	0,70	0,73	0,76	0,73	0,78	0,75	0,75
	median	0,74	0,73	0,71	0,76	0,75	0,75	0,80	0,76	0,75
	delta	0,41	0,32	0,35	0,53	0,19	0,48	0,26	0,29	0,44

Soluzione n°2 fitness 0.1061 feature bloccate con i dati non normalizzati

Divisione	Ripetizione	10	40	30	10 e 5	40 e 20	20 e 10	30 e 15 e 7	40 e 20 e 10	10 e 5 e 2
Prima divisione training-test	1	0,84	0,75	0,89	0,87	0,86	0,80	0,84	0,85	0,86
	2	0,85	0,86	0,84	0,84	0,63	0,81	0,84	0,85	0,82
	3	0,84	0,88	0,81	0,89	0,86	0,65	0,82	0,80	0,87
	4	0,85	0,81	0,73	0,91	0,78	0,89	0,76	0,90	0,84
	5	0,81	0,78	0,89	0,86	0,84	0,70	0,87	0,90	0,74
Seconda divisione training-test	1	0,76	0,70	0,78	0,86	0,50	0,56	0,78	0,60	0,78
	2	0,87	0,52	0,67	0,77	0,85	0,75	0,68	0,69	0,74
	3	0,69	0,61	0,65	0,79	0,76	0,72	0,79	0,74	0,84
	4	0,72	0,63	0,83	0,73	0,70	0,79	0,74	0,74	0,80
	5	0,83	0,62	0,69	0,71	0,76	0,79	0,72	0,68	0,35
Terza divisione training-test	1	0,82	0,77	0,79	0,73	0,80	0,87	0,77	0,87	0,79
	2	0,77	0,86	0,82	0,76	0,79	0,76	0,82	0,56	0,81
	3	0,86	0,82	0,85	0,76	0,72	0,77	0,84	0,75	0,83
	4	0,79	0,89	0,68	0,83	0,86	0,78	0,81	0,86	0,75
	5	0,70	0,82	0,86	0,71	0,81	0,88	0,85	0,84	0,77
Quarta divisione training-test	1	0,76	0,62	0,73	0,80	0,67	0,61	0,84	0,81	0,80
	2	0,74	0,55	0,33	0,80	0,67	0,66	0,82	0,79	0,63
	3	0,87	0,81	0,80	0,80	0,65	0,65	0,85	0,83	0,87
	4	0,81	0,55	0,61	0,84	0,67	0,75	0,53	0,78	0,77
	5	0,79	0,80	0,75	0,50	0,69	0,87	0,77	0,68	0,67
Quinta divisione training-test	1	0,90	0,63	0,58	0,70	0,79	0,60	0,79	0,79	0,73
	2	0,82	0,74	0,70	0,80	0,66	0,86	0,89	0,74	0,87
	3	0,80	0,68	0,71	0,60	0,76	0,67	0,70	0,88	0,85
	4	0,65	0,39	0,75	0,69	0,93	0,76	0,85	0,83	0,71
	5	0,77	0,79	0,81	0,74	0,76	0,72	0,74	0,80	0,90
	avg	0,80	0,72	0,74	0,77	0,75	0,75	0,79	0,78	0,78
	median	0,81	0,75	0,75	0,79	0,76	0,76	0,81	0,80	0,80
	delta	0,25	0,50	0,55	0,41	0,43	0,33	0,35	0,34	0,55

Soluzione n°2 fitness 0.1061 feature bloccate con i dati normalizzati

Divisione	Ripetizione	10	20	30	20 e 10	10 e 5	30 e 15	10 e 5 e 2	20 e 10 e 5	30 e 15 e 7
Prima divisione training-test	1	0,74	0,82	0,73	0,86	0,71	0,80	0,78	0,74	0,84
	2	0,76	0,69	0,79	0,87	0,82	0,85	0,87	0,74	0,77
	3	0,83	0,82	0,87	0,76	0,83	0,76	0,88	0,82	0,68
	4	0,82	0,83	0,76	0,82	0,79	0,87	0,78	0,92	0,83
	5	0,73	0,80	0,69	0,42	0,52	0,78	0,74	0,80	0,84
Seconda divisione training-test	1	0,80	0,78	0,78	0,84	0,42	0,76	0,80	0,81	0,69
	2	0,71	0,78	0,60	0,69	0,77	0,69	0,82	0,84	0,75
	3	0,72	0,68	0,74	0,82	0,82	0,68	0,74	0,75	0,79
	4	0,85	0,82	0,79	0,81	0,83	0,79	0,80	0,79	0,80
	5	0,84	0,80	0,73	0,67	0,71	0,64	0,81	0,84	0,85
Terza divisione training-test	1	0,75	0,78	0,64	0,74	0,75	0,64	0,69	0,82	0,50
	2	0,70	0,27	0,83	0,48	0,73	0,79	0,81	0,79	0,81
	3	0,84	0,59	0,89	0,80	0,77	0,66	0,78	0,80	0,81
	4	0,79	0,73	0,76	0,81	0,69	0,78	0,79	0,83	0,77
	5	0,82	0,70	0,76	0,79	0,72	0,74	0,84	0,87	0,85
Quarta divisione training-test	1	0,66	0,70	0,78	0,79	0,72	0,81	0,81	0,74	0,88
	2	0,48	0,81	0,68	0,84	0,78	0,81	0,88	0,76	0,67
	3	0,68	0,83	0,79	0,87	0,88	0,70	0,40	0,83	0,82
	4	0,71	0,85	0,80	0,89	0,92	0,83	0,83	0,89	0,78
	5	0,88	0,84	0,75	0,67	0,83	0,81	0,88	0,85	0,86
Quinta divisione training-test	1	0,81	0,71	0,72	0,81	0,80	0,73	0,86	0,77	0,76
	2	0,79	0,75	0,80	0,79	0,71	0,80	0,85	0,74	0,41
	3	0,79	0,64	0,70	0,79	0,85	0,82	0,65	0,73	0,76
	4	0,78	0,75	0,76	0,79	0,80	0,72	0,84	0,84	0,80
	5	0,80	0,73	0,74	0,82	0,84	0,85	0,80	0,78	0,80
	avg	0,76	0,74	0,76	0,77	0,76	0,76	0,79	0,80	0,77
	median	0,79	0,78	0,76	0,80	0,78	0,78	0,81	0,80	0,80
	delta	0,40	0,57	0,29	0,48	0,50	0,23	0,48	0,19	0,47

Soluzione n°3 con le clinical feature con i dati non normalizzati

Divisione	Ripetizione	10	30	20	20 e 10	10 e 5	30 e 15	10 e 5 e 2	30 e 15 e 7	20 e 10 e 5
Prima divisione training-test	1	0,87	0,82	0,80	0,77	0,79	0,78	0,90	0,85	0,78
	2	0,85	0,71	0,62	0,83	0,86	0,69	0,82	0,82	0,85
	3	0,84	0,84	0,75	0,78	0,83	0,79	0,81	0,83	0,88
	4	0,82	0,85	0,59	0,79	0,88	0,71	0,82	0,87	0,85
	5	0,80	0,82	0,78	0,83	0,74	0,82	0,89	0,84	0,89
Seconda divisione training-test	1	0,83	0,58	0,64	0,81	0,84	0,77	0,88	0,88	0,72
	2	0,78	0,82	0,83	0,87	0,70	0,69	0,88	0,77	0,82
	3	0,58	0,79	0,86	0,74	0,80	0,73	0,36	0,88	0,78
	4	0,85	0,81	0,84	0,67	0,83	0,87	0,87	0,68	0,90
	5	0,83	0,73	0,89	0,83	0,73	0,80	0,76	0,86	0,69
Terza divisione training-test	1	0,47	0,76	0,80	0,52	0,70	0,78	0,76	0,81	0,78
	2	0,39	0,82	0,49	0,73	0,83	0,82	0,68	0,72	0,78
	3	0,72	0,81	0,87	0,83	0,66	0,82	0,82	0,75	0,82
	4	0,89	0,84	0,80	0,84	0,81	0,73	0,79	0,83	0,81
	5	0,87	0,83	0,84	0,76	0,89	0,82	0,89	0,79	0,85
Quarta divisione training-test	1	0,87	0,85	0,85	0,71	0,76	0,89	0,90	0,71	0,57
	2	0,89	0,90	0,81	0,89	0,89	0,81	0,84	0,84	0,83
	3	0,88	0,78	0,89	0,87	0,90	0,86	0,81	0,66	0,89
	4	0,42	0,81	0,77	0,85	0,80	0,77	0,86	0,88	0,71
	5	0,84	0,75	0,87	0,86	0,73	0,89	0,89	0,79	0,86
Quinta divisione training-test	1	0,80	0,84	0,60	0,68	0,87	0,87	0,81	0,81	0,77
	2	0,75	0,72	0,74	0,82	0,85	0,76	0,82	0,79	0,56
	3	0,75	0,84	0,70	0,82	0,81	0,82	0,76	0,89	0,85
	4	0,87	0,82	0,83	0,78	0,86	0,73	0,88	0,82	0,87
	5	0,84	0,76	0,47	0,75	0,80	0,78	0,83	0,85	0,80
	avg	0,77	0,80	0,76	0,79	0,81	0,79	0,81	0,81	0,79
	median	0,83	0,82	0,80	0,81	0,81	0,79	0,82	0,82	0,82
	delta	0,51	0,32	0,41	0,37	0,24	0,20	0,54	0,23	0,34

Soluzione n°3 con le clinical feature con i dati normalizzati

Divisione	Ripetizione	20	30	10	10 e 5	30 e 15	20 e 10	30 e 15 e 7	10 e 5 e 2	20 e 10 e 5
Prima divisione training-test	1	0,81	0,80	0,75	0,73	0,78	0,83	0,82	0,81	0,71
	2	0,85	0,84	0,74	0,73	0,60	0,64	0,79	0,74	0,75
	3	0,76	0,77	0,69	0,84	0,79	0,74	0,80	0,83	0,84
	4	0,79	0,78	0,81	0,77	0,79	0,51	0,81	0,78	0,77
	5	0,78	0,66	0,81	0,77	0,67	0,77	0,85	0,83	0,78
Seconda divisione training-test	1	0,81	0,87	0,76	0,88	0,90	0,86	0,83	0,85	0,85
	2	0,85	0,88	0,74	0,78	0,74	0,87	0,88	0,85	0,78
	3	0,77	0,79	0,80	0,88	0,89	0,86	0,80	0,82	0,88
	4	0,89	0,78	0,77	0,89	0,85	0,87	0,87	0,86	0,84
	5	0,89	0,72	0,77	0,90	0,90	0,82	0,83	0,77	0,80
Terza divisione training-test	1	0,84	0,80	0,86	0,77	0,79	0,85	0,63	0,86	0,78
	2	0,77	0,61	0,79	0,80	0,92	0,24	0,83	0,85	0,85
	3	0,81	0,84	0,82	0,85	0,79	0,89	0,74	0,86	0,72
	4	0,71	0,84	0,86	0,84	0,88	0,86	0,85	0,74	0,91
	5	0,78	0,83	0,84	0,78	0,79	0,80	0,84	0,83	0,90
Quarta divisione training-test	1	0,79	0,74	0,65	0,88	0,74	0,76	0,69	0,84	0,82
	2	0,83	0,71	0,86	0,86	0,82	0,83	0,86	0,78	0,88
	3	0,82	0,81	0,75	0,85	0,73	0,78	0,85	0,74	0,81
	4	0,62	0,84	0,88	0,85	0,43	0,78	0,83	0,80	0,75
	5	0,90	0,71	0,84	0,84	0,87	0,71	0,88	0,85	0,80
Quinta divisione training-test	1	0,69	0,72	0,84	0,66	0,81	0,75	0,71	0,80	0,73
	2	0,77	0,78	0,77	0,72	0,70	0,79	0,70	0,80	0,80
	3	0,74	0,73	0,65	0,75	0,76	0,77	0,73	0,82	0,70
	4	0,66	0,52	0,71	0,77	0,73	0,72	0,82	0,74	0,80
	5	0,79	0,69	0,68	0,72	0,79	0,72	0,82	0,83	0,82
	avg	0,79	0,76	0,78	0,80	0,78	0,76	0,80	0,81	0,80
	median	0,79	0,78	0,77	0,80	0,79	0,78	0,82	0,82	0,80
	delta	0,28	0,36	0,22	0,25	0,50	0,65	0,25	0,12	0,21

Soluzione n°4 con le feature libere fitness pari a 0.0883 con i non dati normalizzati

Divisione	Ripetizione	10	50	20	30 e 15	10 e 5	20 e 10	30 e 15 e 7	20 e 10 e 5	50 e 25 e 12
Prima divisione training-test	1	0,83	0,84	0,86	0,82	0,70	0,87	0,83	0,87	0,56
	2	0,68	0,67	0,65	0,83	0,84	0,53	0,82	0,82	0,79
	3	0,68	0,83	0,80	0,87	0,82	0,87	0,80	0,79	0,81
	4	0,79	0,82	0,81	0,82	0,71	0,67	0,84	0,89	0,83
	5	0,83	0,69	0,76	0,82	0,82	0,69	0,81	0,79	0,77
Seconda divisione training-test	1	0,74	0,59	0,67	0,78	0,66	0,70	0,68	0,55	0,75
	2	0,84	0,76	0,73	0,76	0,67	0,84	0,47	0,74	0,40
	3	0,75	0,76	0,50	0,64	0,87	0,80	0,73	0,72	0,67
	4	0,76	0,78	0,74	0,76	0,83	0,72	0,85	0,85	0,59
	5	0,78	0,87	0,58	0,78	0,74	0,77	0,64	0,76	0,57
Terza divisione training-test	1	0,70	0,67	0,51	0,87	0,79	0,75	0,80	0,79	0,75
	2	0,84	0,64	0,81	0,81	0,81	0,79	0,86	0,67	0,83
	3	0,81	0,75	0,78	0,80	0,77	0,80	0,79	0,81	0,76
	4	0,71	0,87	0,78	0,74	0,80	0,76	0,82	0,79	0,79
	5	0,63	0,88	0,77	0,62	0,81	0,77	0,72	0,78	0,76
Quarta divisione training-test	1	0,71	0,79	0,52	0,73	0,59	0,78	0,79	0,69	0,69
	2	0,88	0,79	0,66	0,88	0,80	0,55	0,81	0,84	0,85
	3	0,81	0,73	0,83	0,85	0,79	0,71	0,62	0,78	0,77
	4	0,84	0,79	0,84	0,70	0,77	0,80	0,78	0,59	0,69
	5	0,89	0,77	0,81	0,85	0,77	0,59	0,84	0,79	0,77
Quinta divisione training-test	1	0,79	0,78	0,74	0,78	0,70	0,76	0,76	0,75	0,77
	2	0,79	0,56	0,67	0,74	0,78	0,79	0,78	0,77	0,78
	3	0,82	0,68	0,76	0,72	0,72	0,81	0,69	0,80	0,77
	4	0,73	0,77	0,79	0,76	0,67	0,67	0,81	0,72	0,70
	5	0,71	0,60	0,74	0,82	0,77	0,69	0,70	0,76	0,66
	avg	0,77	0,75	0,72	0,78	0,76	0,74	0,76	0,76	0,72
	median	0,79	0,77	0,76	0,78	0,77	0,76	0,79	0,78	0,76
	delta	0,26	0,32	0,36	0,25	0,28	0,34	0,39	0,34	0,45

Soluzione n°4 con le feature libere fitness pari a 0.0883 con i dati normalizzati

Divisione	Ripetizione	50	20	40	10 e 5	40 e 20	30 e 15	10 e 5 e 2	20 e 10 e 5	30 e 15 e 7
Prima divisione training-test	1	0,79	0,63	0,75	0,67	0,76	0,69	0,69	0,80	0,78
	2	0,72	0,78	0,63	0,70	0,75	0,75	0,84	0,82	0,65
	3	0,69	0,69	0,72	0,71	0,54	0,64	0,73	0,67	0,75
	4	0,73	0,71	0,70	0,82	0,79	0,67	0,65	0,69	0,68
	5	0,76	0,67	0,65	0,71	0,55	0,73	0,80	0,80	0,72
Seconda divisione training-test	1	0,75	0,74	0,76	0,74	0,69	0,75	0,73	0,78	0,71
	2	0,75	0,56	0,81	0,80	0,67	0,80	0,64	0,84	0,84
	3	0,74	0,74	0,69	0,80	0,82	0,64	0,79	0,82	0,82
	4	0,86	0,87	0,71	0,80	0,70	0,75	0,81	0,79	0,73
	5	0,83	0,81	0,73	0,81	0,70	0,76	0,81	0,77	0,79
Terza divisione training-test	1	0,76	0,64	0,66	0,77	0,78	0,78	0,79	0,60	0,71
	2	0,66	0,68	0,69	0,77	0,74	0,76	0,75	0,77	0,79
	3	0,74	0,79	0,65	0,77	0,80	0,64	0,69	0,75	0,78
	4	0,66	0,70	0,67	0,72	0,71	0,72	0,74	0,61	0,64
	5	0,75	0,74	0,73	0,77	0,74	0,74	0,76	0,61	0,81
Quarta divisione training-test	1	0,72	0,65	0,77	0,79	0,65	0,80	0,77	0,77	0,66
	2	0,76	0,77	0,74	0,80	0,77	0,84	0,78	0,75	0,75
	3	0,54	0,70	0,72	0,81	0,78	0,73	0,82	0,77	0,74
	4	0,72	0,78	0,61	0,78	0,69	0,73	0,66	0,69	0,56
	5	0,65	0,84	0,70	0,74	0,77	0,83	0,81	0,73	0,74
Quinta divisione training-test	1	0,77	0,68	0,70	0,68	0,61	0,79	0,75	0,84	0,75
	2	0,81	0,84	0,76	0,83	0,75	0,73	0,82	0,72	0,74
	3	0,81	0,77	0,78	0,63	0,78	0,72	0,79	0,69	0,83
	4	0,40	0,28	0,78	0,85	0,64	0,69	0,72	0,82	0,71
	5	0,67	0,71	0,71	0,78	0,76	0,75	0,82	0,70	0,80
	avg	0,72	0,71	0,71	0,76	0,72	0,74	0,76	0,74	0,74
	median	0,74	0,71	0,71	0,77	0,74	0,74	0,77	0,77	0,74
	delta	0,46	0,59	0,19	0,23	0,27	0,20	0,20	0,24	0,28

Soluzione n°5 con le feature libere fitness pari a 0.1002 con i dati non normalizzati

Divisione	Ripetizione	10	20	30	10 e 5	50 e 25	30 e 15	20 e 10 e 5	10 e 5 e 2	30 e 15 e 7
Prima divisione training-test	1	0,65	0,80	0,71	0,77	0,85	0,77	0,86	0,78	0,79
	2	0,80	0,71	0,80	0,64	0,81	0,75	0,61	0,84	0,75
	3	0,72	0,76	0,57	0,75	0,76	0,77	0,75	0,77	0,71
	4	0,55	0,82	0,72	0,90	0,76	0,55	0,62	0,76	0,57
	5	0,76	0,81	0,68	0,80	0,72	0,62	0,84	0,72	0,65
Seconda divisione training-test	1	0,78	0,81	0,66	0,74	0,64	0,72	0,83	0,84	0,79
	2	0,71	0,69	0,76	0,87	0,64	0,69	0,77	0,77	0,27
	3	0,84	0,67	0,77	0,84	0,72	0,78	0,82	0,71	0,82
	4	0,70	0,62	0,66	0,85	0,81	0,64	0,38	0,85	0,80
	5	0,82	0,82	0,68	0,86	0,65	0,66	0,80	0,81	0,84
Terza divisione training-test	1	0,83	0,80	0,64	0,74	0,40	0,79	0,84	0,88	0,84
	2	0,77	0,76	0,84	0,78	0,69	0,68	0,90	0,88	0,71
	3	0,82	0,39	0,78	0,85	0,79	0,83	0,80	0,86	0,82
	4	0,84	0,70	0,72	0,79	0,74	0,82	0,87	0,90	0,84
	5	0,73	0,81	0,70	0,80	0,76	0,73	0,83	0,87	0,80
Quarta divisione training-test	1	0,82	0,36	0,50	0,72	0,80	0,78	0,83	0,78	0,81
	2	0,83	0,81	0,75	0,74	0,61	0,78	0,79	0,76	0,82
	3	0,57	0,62	0,74	0,84	0,79	0,80	0,71	0,82	0,81
	4	0,83	0,53	0,74	0,73	0,71	0,78	0,68	0,81	0,77
	5	0,74	0,72	0,64	0,69	0,76	0,75	0,75	0,67	0,37
Quinta divisione training-test	1	0,91	0,86	0,86	0,87	0,58	0,64	0,86	0,88	0,68
	2	0,87	0,80	0,76	0,85	0,76	0,69	0,85	0,86	0,71
	3	0,88	0,75	0,79	0,82	0,84	0,74	0,90	0,78	0,84
	4	0,87	0,52	0,84	0,86	0,73	0,47	0,76	0,79	0,85
	5	0,76	0,85	0,72	0,76	0,77	0,82	0,85	0,85	0,74
	avg	0,78	0,71	0,72	0,79	0,72	0,72	0,78	0,81	0,74
	median	0,80	0,76	0,72	0,80	0,76	0,75	0,82	0,81	0,79
	delta	0,36	0,50	0,36	0,26	0,45	0,36	0,53	0,23	0,58

Soluzione n°5 con le feature libere fitness pari a 0.1002 con i dati normalizzati

Divisione	Ripetizione	10	20	40	10 e 5	20 e 10	30 e 15	10 e 5 e 2	40 e 20 e 10	20 e 10 e 5
Prima divisione training-test	1	0,85	0,83	0,81	0,81	0,62	0,83	0,83	0,80	0,82
	2	0,80	0,81	0,82	0,76	0,82	0,79	0,88	0,84	0,83
	3	0,84	0,86	0,71	0,75	0,83	0,66	0,86	0,78	0,81
	4	0,80	0,78	0,73	0,70	0,83	0,46	0,61	0,83	0,77
	5	0,80	0,73	0,81	0,82	0,75	0,87	0,84	0,86	0,87
Seconda divisione training-test	1	0,82	0,64	0,71	0,84	0,85	0,72	0,84	0,81	0,80
	2	0,74	0,77	0,74	0,71	0,77	0,57	0,76	0,82	0,72
	3	0,66	0,68	0,75	0,78	0,82	0,78	0,77	0,72	0,79
	4	0,81	0,73	0,62	0,79	0,82	0,75	0,81	0,42	0,69
	5	0,69	0,74	0,68	0,86	0,66	0,58	0,85	0,84	0,75
Terza divisione training-test	1	0,87	0,66	0,81	0,82	0,75	0,85	0,77	0,84	0,83
	2	0,92	0,88	0,73	0,80	0,87	0,78	0,88	0,87	0,94
	3	0,79	0,75	0,83	0,86	0,68	0,94	0,85	0,81	0,87
	4	0,58	0,77	0,74	0,82	0,80	0,95	0,86	0,74	0,60
	5	0,81	0,81	0,76	0,91	0,81	0,79	0,84	0,84	0,88
Quarta divisione training-test	1	0,62	0,53	0,78	0,79	0,80	0,64	0,78	0,87	0,86
	2	0,65	0,84	0,72	0,56	0,76	0,73	0,85	0,78	0,86
	3	0,84	0,82	0,79	0,83	0,79	0,83	0,79	0,78	0,70
	4	0,78	0,71	0,72	0,71	0,70	0,80	0,79	0,83	0,80
	5	0,75	0,49	0,76	0,70	0,93	0,69	0,74	0,73	0,78
Quinta divisione training-test	1	0,66	0,62	0,75	0,88	0,89	0,86	0,85	0,82	0,74
	2	0,80	0,71	0,78	0,79	0,73	0,82	0,71	0,89	0,78
	3	0,81	0,80	0,77	0,82	0,60	0,78	0,80	0,82	0,66
	4	0,70	0,76	0,84	0,64	0,71	0,82	0,84	0,76	0,62
	5	0,78	0,83	0,73	0,81	0,84	0,44	0,81	0,79	0,83
	avg	0,77	0,74	0,76	0,78	0,78	0,75	0,81	0,80	0,78
	median	0,80	0,76	0,75	0,80	0,80	0,78	0,83	0,82	0,80
	delta	0,34	0,38	0,22	0,34	0,33	0,51	0,28	0,47	0,34