

UNIVERSITÀ DEGLI STUDI DEL MOLISE

DIPARTIMENTO DI MEDICINA E SCIENZE DELLA SALUTE

“VINCENZO TIBERIO”



Corso di Dottorato in Medicina clinica e Traslazionale

Tesi sperimentale in Radiomica

RADIOMICA E METODI FORMALI: APPLICAZIONI PROSSIME E FUTURE

Candidato

dr. Pasquale GUERRIERO

Matr. 170022

A handwritten signature in black ink, appearing to be 'Pasquale Guerriero'.

Coordinatore del corso di dottorato
Chiar.mo Prof. Giovanni Scapagnini

A handwritten signature in black ink, appearing to be 'Giovanni Scapagnini'.

Relatore

Chiar.mo Prof.

Ferdinando Caranci

A handwritten signature in black ink, appearing to be 'Ferdinando Caranci'.

XXXVI ciclo

INDICE

INTRODUZIONE	1
1 BACKGROUND.....	3
1.1 RADIOMICA.....	3
1.2 INTELLIGENZA ARTIFICIALE.....	13
1.3 SOFTWARE UTILIZZATI	19
1.4 METODI FORMALI.....	25
1.5 TECNICA DEL MODEL CHECKING.....	27
2 AN INNOVATIVE PROSTATE CANCER METHOD TROUGHT FORMAL VERIFICATION.....	31
2.1 STATO DELL'ARTE.....	31
2.2 MATERIALI E METODI.....	33
2.3 SCOPO DELLO STUDIO.....	42
2.4 RISULTATI.....	43
2.5 CONCLUSIONI.....	48
3 EXPLENABILITY OF FORMAL METHODS	49
4 ALTRI LAVORI E SVILUPPI FUTURI.....	79
4.1 LAVORI PUBBLICATI.....	79
4.2 POSSIBILITA' FUTURE.....	80

BIBLIOGRAFIA	82
---------------------------	-----------

INTRODUZIONE

I tempi che stiamo vivendo sono caratterizzati da un lato “dall’invasione” da parte dell’AI in tutti i settori, medicina compresa, e dall’altro dalla netta affermazione che la medicina deve di precisione, basata sul singolo individuo, personalizzata. Questo ha portato me e i colleghi che hanno lavorato con me alla messa in pratica di alcuni progetti, di cui alcuni già pubblicati, atti alla implementazione di software di intelligenza artificiale (o meglio di radiomica) che potessero aiutare il medico radiologo nella diagnosi precoce e non invasiva di patologie a partire da immagini radiologiche. Nella prima parte di questo lavoro di tesi viene appunto descritto un software per la diagnosi automatica del cancro alla prostata e come si è arrivati alla sua implementazione, i risultati ottenuti e i possibili sviluppi futuri. Voglio anticipare non solo il successo nell’implementazione del software, ma anche gli ottimi risultati ottenuti in termini di sensibilità (compresa tra 0.95 e 1) e specificità (pari a 1). Questa ricerca è stata resa possibile grazie allo sviluppo di un modello formale del carcinoma prostatico.

Ci siamo anche accorti, però, che in ambito medico c’è una certa diffidenza verso l’intelligenza artificiale sia da parte dei medici, che ovviamente non sono abituati a vedersi restituito un verdetto da una

macchina senza capire le motivazioni per cui la macchina dia quella risposta, ma anche da parte dei pazienti, che non comprendendo i meccanismi di funzionamento di quelle che vengono viste come “scatole nere”, non si sentono sicuri nell’affidare la propria vita ad un calcolatore elettronico i cui procedimenti di calcolo non sono noti.

Partendo quindi da queste considerazioni, dal lavoro già anticipato prima sul carcinoma della prostata e da lavori analoghi in cui il software era stato implementato e utilizzato, ad esempio, per il Covid-19, abbiamo voluto smembrare e analizzare il sistema dei “metodi formali” per poter fornire una spiegazione razionale dei risultati ottenuti.

Questo ci ha permesso di associare i risultati a valori tangibili, incasellabili e che quindi ci hanno fatto identificare un pattern di riferimento.

Inoltre, per poter favorire il radiologo nella diagnosi, si è voluto anche “geolocalizzare” all’interno della singola slice la malattia, in modo da indirizzare lo sguardo del radiologo che poteva con semplicità ritrovare l’immagine “patologica” e confermare il risultato dato dalla macchina o meno.

1 BACKGROUND

1.1 LA RADIOMICA

La radiomica rappresenta una nuova frontiera della medicina basata sull'estrazione di dati quantitativi dalle immagini radiologiche e sull'utilizzo di questi ultimi per la creazione di sistemi di supporto alle decisioni cliniche.

I dati quantitativi estratti dalle immagini possono essere integrati in modelli morfologici e predittivi multidisciplinari necessari per la gestione del paziente, anche mediante l'utilizzo di metodiche di *machine learning* e di intelligenza artificiale^[1]. Queste caratteristiche morfologico/predittive possono essere suddivise in diversi gruppi: dimensioni e caratteristiche basate sulla morfologia, descrittori delle relazioni tra i *pixel* o *voxel* che compongono l'immagine. Le definizioni matematiche di queste relazioni, chiamate "descrittori", sono indipendenti dalla modalità di *imaging*, ma i loro valori sono dipendenti dal settaggio della macchina radiologica. In genere, questi descrittori mostrano come i valori numerici presenti all'interno delle immagini siano distribuiti spazialmente (e non) all'interno del tessuto studiato.^{[2] [3]}

Partendo dal presupposto che ad ogni intensità di grigio in un'immagine digitale in bianco e nero corrisponda un valore numerico ben preciso, la distribuzione nello spazio dei valori di grigio ottenuti nelle strutture esaminate con queste tecniche è influenzata da diversi fattori, che dipendono dal tipo di metodica utilizzata per ottenere le immagini. Questo tipo di analisi è denominata “*texture analysis*”^[4]^[5].

Nelle immagini TC ad esempio l'intensità delle immagini è proporzionale alla radiopacità dei tessuti: quindi, la distribuzione nello spazio di questo parametro all'interno di una lesione tumorale, può essere influenzata, ad esempio, dalla presenza di necrosi o di calcificazioni; tutti parametri che possono essere correlati con l'aggressività della malattia.

Informazioni di tipo differente possono essere ottenute usando la stessa matematica di base anche analizzando immagini ecografiche^[6], di risonanza magnetica^[7]^[8] o ottenute mediante la PET^[9] e la SPECT^[10]. Mentre le immagini ottenute mediante tecniche di radiodiagnostica hanno come grosso vantaggio l'elevata risoluzione (che consente di campionare tanti *voxel* o *pixel*, fornendo un'elevata numerosità nella stima dei descrittori, che sono variabili di tipo statistico), le tecniche medico-nucleari hanno tendenzialmente una migliore risoluzione di contrasto e consentono di studiare più in

dettaglio i processi biologici a carico delle lesioni sospette; tuttavia la bassa risoluzione spaziale di queste ultime metodiche può costituire un problema nello studio delle lesioni di piccole dimensioni.

In letteratura e in oncologia, spesso i descrittori ottenuti utilizzando queste tecniche sono stati correlati con l'analisi genetica delle lesioni valutate mediante le immagini prima della loro asportazione, rilevando come in molti casi la differente struttura di queste correli con differenti mutazioni a carico delle cellule tumorali che le costituiscono. Tale approccio è denominato radiogenomica.

Un obiettivo chiave della medicina moderna è la "medicina di precisione", la cui finalità consiste nel personalizzare il trattamento sulla base delle caratteristiche specifiche del paziente e della sua malattia. La Radiomica sta rapidamente emergendo come tecnologia della medicina personalizzata ed è attualmente uno dei campi di ricerca di maggior interesse.

Una delle principali applicazioni della Radiomica è l'Oncologia, dal momento che la diagnostica per immagini è in grado di ottenere informazioni e caratterizzare i tessuti in modo non invasivo, a differenza di quanto avviene con biopsie o interventi chirurgici invasivi.

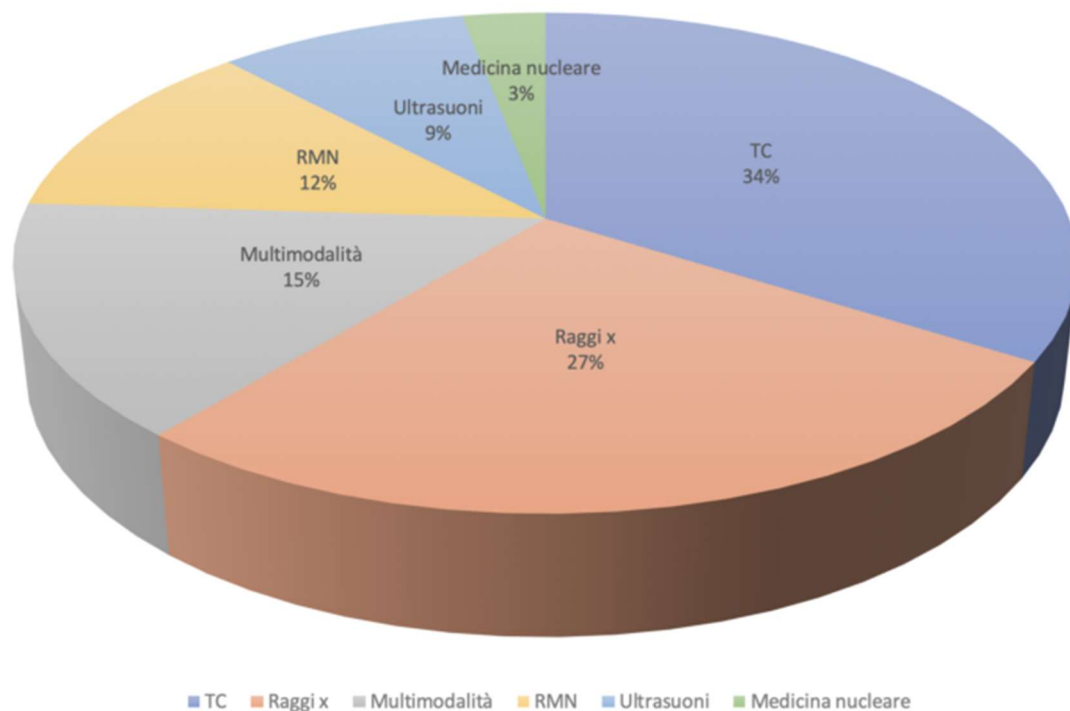


Fig.6 – Rappresentazione delle immagini radiologiche più frequentemente utilizzate per ricavare i dati in radiomica.^[11]

Nel diagramma riportato in figura 6 è possibile capire quali siano le metodiche di *imaging* più utilizzate nella radiomica e si nota una netta prevalenza della TC seguita poi dall' RX e solo in minor parte vengono utilizzate anche le altre tecniche radiologiche di *imaging* .

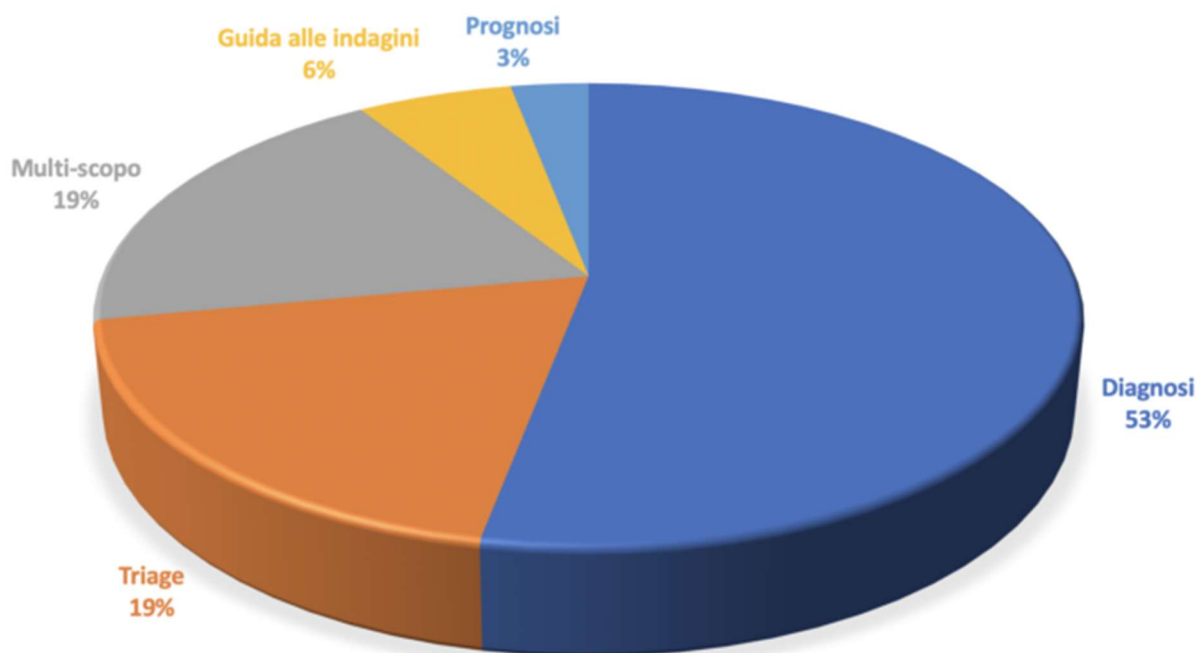


Fig.7 – Rappresentazione degli scopi della radiomica^[11]

Nel diagramma riportato in figura 7 è invece possibile comprendere quale sia l'obiettivo finale della radiomica da parte del radiologo e permette di evidenziare come venga utilizzata maggiormente a scopo diagnostico.

Questo enorme patrimonio di dati numerici che la Radiomica ci permette di ottenere, che non riuscirebbe ad essere elaborato tramite la semplice osservazione visiva delle immagini mediche da parte dell'operatore, va a definire numerose caratteristiche della malattia e

dell'ambiente circostante, relative ad esempio alla forma, al volume ed alla struttura tissutale.

Con tali tecniche, è possibile studiare l'eventuale correlazione fra i dati ottenuti dalle immagini e le caratteristiche molecolari e genomiche della malattia, con l'obiettivo finale di estrarre direttamente dalle immagini indicazioni sull'aggressività della malattia, sulle terapie più indicate e sulla loro risposta.

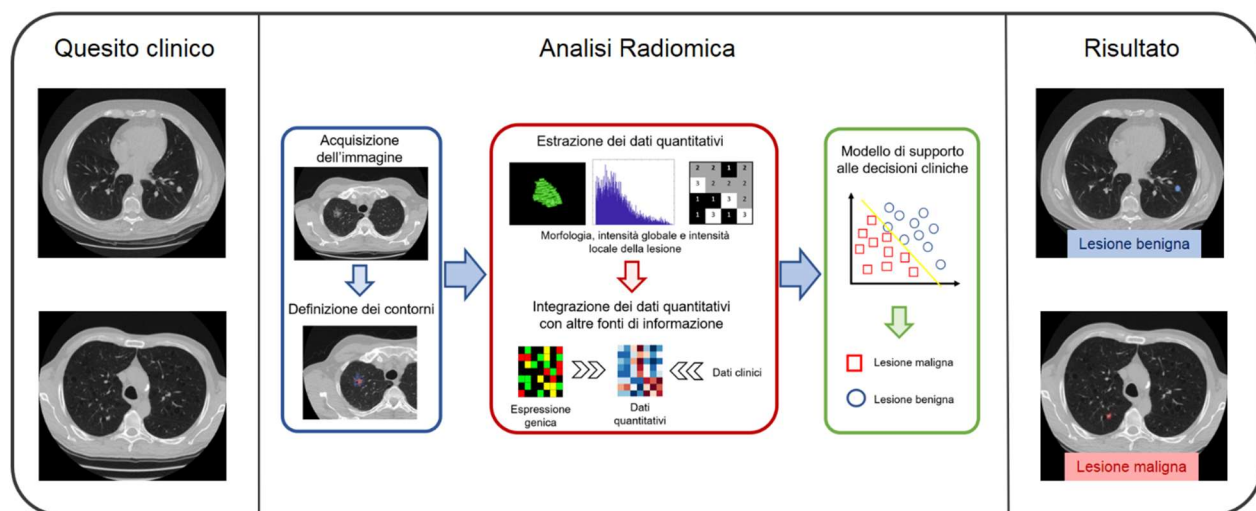


Fig.8 – Rappresentazione del processo completo di un'analisi radiomica, a partire dall'immagine fino alla diagnosi automatizzata.

In figura 8 è rappresentato il processo di Radiomica: i parametri quantitativi calcolati a partire dalle immagini cliniche vengono analizzati in combinazione a caratteristiche biologiche, genetiche e cliniche del paziente ed utilizzati per creare dei modelli di supporto alle decisioni cliniche^[12].

L'obiettivo è quello di raccogliere tutti i dati derivanti dalle indagini di *imaging* di ogni singolo paziente e di osservare come, tali dati, possono condizionare l'effetto delle terapie o predire l'evoluzione della malattia.

Il principale vantaggio è quello di comunicare ai pazienti, che non risponderanno alla terapia, di dover intraprendere un trattamento diverso da quello pianificato, senza aspettare di osservare in vivo la mancata risposta, riducendone anche i costi sanitari.

Nasce così la necessità di realizzare un classificatore, ossia un sistema capace di effettuare delle diagnosi a partire da immagini mediche. Per valutare la bontà di un classificatore in termini di capacità di classificare, si introducono quattro metriche:

- *True Positive* (TP): indica i pazienti malati classificati correttamente come malati; dove la "T" di "*True*" è il *Gold standard*, in questo caso il medico che è a conoscenza della verità, mentre la "P" di "*Positive*" è il risultato del classificatore;
- *True Negative* (TN): indica i pazienti sani correttamente classificati come sani;
- *False Positive* (FP): indica i pazienti sani erroneamente classificati come malati; in tale caso si ha una classificazione errata;

- *False Negative* (FN): indica i pazienti malati erroneamente classificati come sani ed anche in questo caso il risultato del classificatore è errato.

L'errore di classificazione può essere calcolato utilizzando la *Confusion Matrix* (presente nella Figura 9), ovvero una matrice costituita da due righe e due colonne, sommando tutti gli elementi al di fuori della diagonale principale.

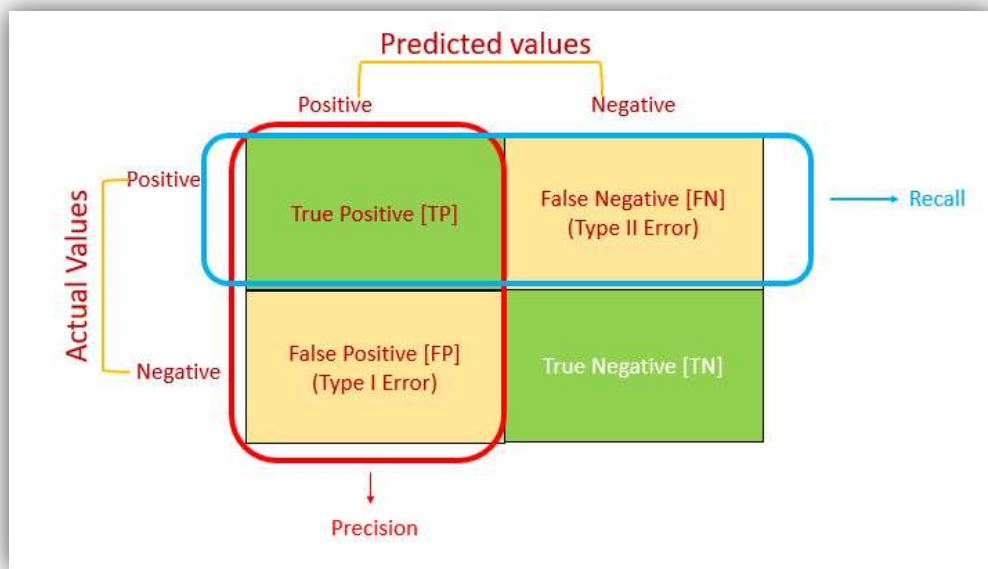


Fig.9 – *Confusion Matrix*.

In campo medico si preferisce avere i *False Positive* più elevati rispetto ai *False Negative*, in quanto per i primi, si pensa che il paziente sia malato quando realmente non lo è, quindi è un messaggio

falsamente allarmante che non provoca danni al paziente. Viceversa, per i *False Negative* si ha un messaggio falsamente rassicurante poiché si pensa che la malattia sia assente, quando in realtà il paziente è malato. In tale circostanza non si mettono in atto tutte quelle procedure e terapie atte a rilevare la malattia e, in alcuni casi, ritardare una diagnosi potrebbe compromettere la vita del paziente preso in esame.

A partire dalle metriche precedentemente descritte, è possibile ricavare quattro indici utili in questa fase di classificazione:

- **Accuracy:** rappresenta il numero di pattern correttamente classificati. In genere si preferisce esprimerlo in percentuale, indicando così la percentuale di classificazioni corrette.

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

Tanto più tale valore è vicino ad 1, tanto più efficiente sarà il classificatore.

- **Precision:** è un indice di correttezza che valuta i pazienti correttamente classificati come positivi sul numero complessivo di pazienti positivi.

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

- **Recall:** è una misura di completezza, indicata il più delle volte con il termine di *Sensitivity*.

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

- **Specificity:** è un parametro che valuta il numero di negativi che sono correttamente classificati come tali; viene indicato anche come Vero Tasso Negativo.

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

Nel prossimo futuro, si auspica che la *routine* giornaliera sia la raccolta di dati dagli esami di *imaging* radiologico e la conversione in dati quantitativi per il supporto decisionale alla pratica clinica per migliorare l'accuratezza diagnostica e il potere prognostico.

La finalità di tale azione è la creazione di archivi digitali utili per valutazioni anche retrospettive mediante l'utilizzo di *software* di intelligenza artificiale.

1.2 L'INTELLIGENZA ARTIFICIALE

L'intelligenza artificiale (AI) può essere un potente strumento al servizio della medicina moderna, in parte già in uso. Le applicazioni, infatti, si moltiplicano in tutti gli ambiti, dalla diagnostica alla chirurgia, dallo sviluppo dei farmaci alla riabilitazione e sono destinate ulteriormente a crescere. In campo medico l'AI ha due finalità importanti:

- È in grado di aumentare le capacità di un medico o di un operatore sanitario e di alleggerirne il carico di lavoro;
- È utile nel supportare la gestione di un sistema sanitario al fine di migliorare le prestazioni e aumentare le capacità.

Da un punto di vista radiologico, l'AI è particolarmente importante perché è in grado di osservare dettagli che l'occhio umano talvolta non riesce a riconoscere o a cogliere immediatamente.

Per comprendere cos'è l'AI bisogna partire dal presupposto che le immagini radiologiche, così come le fotografie che noi osserviamo quotidianamente, non sono nient'altro che un enorme banca dati.

Per cui un'immagine radiologica e un'immagine digitale contengono al loro interno una serie di importanti informazioni.

Il medico radiologo quando osserva un'immagine radiologica, formula una diagnosi e compila un referto basandosi sulla propria esperienza e

sulle proprie conoscenze, confrontando le immagini con quello che ha appreso nel corso della sua carriera. L'intelligenza artificiale riesce a fare la stessa cosa ma con una capacità di confronto di immagini molto superiore, attraverso un processo di apprendimento dei dati che appartengono alle immagini stesse, chiamato Machine Learning, in cui si sottopongono all'elaboratore una serie di immagini, già classificate come sane o patologiche, e gli si insegna a distinguerle in base alle informazioni presenti nelle stesse. Successivamente il *software* immagazzinerà le informazioni e, una volta che le avrà apprese, potrà applicare la propria conoscenza a tutte le immagini che gli andremo a riproporre.

Questo è esattamente ciò che fa un essere umano e che in molti casi il computer riesce a fare con maggiore accuratezza. Questo perché la capacità di riconoscere dettagli anche minimi su un'immagine radiologica è sicuramente molto maggiore per una macchina addestrata rispetto a un medico radiologo talvolta anche esperto.

In futuro non si assisterà certamente alla scomparsa del medico radiologo, perché, come ben sappiamo, questa figura professionale non si occupa soltanto dell'osservazione di un'immagine e successiva formulazione di un referto scritto, ma ovviamente si occupa della conoscenza del paziente, sia da un punto di vista della storia clinica,

sia da un punto di vista diagnostico per comprendere meglio se una certa diagnosi possa essere compatibile con le sue patologie^[13].

Esistono ad oggi diversi dispositivi basati su tecniche di AI, approvati nell'Unione Europa e/o negli USA, nel campo del processamento e analisi delle immagini radiologiche, di supporto alla diagnosi, attraverso rivelazione di anomalie, con informazioni quantitative, mostrate generalmente nei sistemi CAD (*Computer-Aided Design*) per una più facile individuazione da parte del radiologo.

La costante ricerca di diagnosi sempre più raffinate e di minore invasività per il paziente ha ulteriormente stimolato, anche in modo caotico, sviluppi e proposte di utilizzo di nuove tecnologie nel campo radiologico. Tra queste, l'AI, da qualche anno in rapidissima espansione, rappresenta indubbiamente la tecnologia più esplorata [14,15].

La ragione di questo interesse è dovuta sia alle numerose implementazioni disponibili e consolidate e certificate, sia alla relativa facilità di approntare o riadattare strumenti *software* in tempi compatibili con l'emergenza. Gli strumenti basati sull'apprendimento automatico (*Data-driven*) dovrebbero essere efficaci quando non è ancora disponibile una modellizzazione quantitativa del problema, a

causa della sua complessità o novità, ma è già presente una quantità di dati statisticamente significativa.

Per quanto riguarda la pratica clinica, si trova immediatamente un vantaggio nella caratterizzazione della texture del tumore rilevato al fine di valutarne la malignità o la benignità.

Il processo pratico di radiomica coinvolge diverse fasi:

- Acquisizione delle immagini;
- Identificazione del volume di interesse, da cui estrarre i dati (ad esempio, il tumore);
- Segmentazione del volume di interesse mediante strumenti grafici;
- Estrazione dei dati dalle immagini digitali;
- Creazione di modelli utili per la classificazione.

A queste fasi possono essere aggiunte altre attività per affinare l'analisi. Ad esempio, dopo l'acquisizione dell'immagine, può essere eseguito un aggiustamento dell'immagine (in termini spaziali, con il Resampling); può anche essere eseguito il filtraggio, ovvero l'implementazione di filtri sull'immagine al fine di ridurre il rumore o evidenziare le caratteristiche anatomiche. Inoltre, può essere effettuata una selezione dei dati per creare modelli predittivi di successo al fine

di eliminare la ridondanza delle informazioni e accelerare le prestazioni.

I dati estratti in questo processo sono chiamati features radiomiche e ne esistono due tipi. Le features semantiche sono quelle comunemente utilizzate nel lessico radiologico (dimensione, forma, posizione, vascosità, spicolatura, necrosi) e sono quindi facilmente collegate alle evidenze radiologiche. Al contrario, le features agnostiche sono dati calcolati utilizzando formule matematiche basate sulle informazioni nell'immagine, come l'istogramma o le matrici costruite in base alla relazione spaziale dei livelli di grigio nei pixel. Queste sono caratteristiche di Haralick^[16], Laws^[17], Wavelets^[18], trasformate di Laplace, funzionali di Minkowski e frattali.

Nella categoria delle features agnostiche, si possono trovare caratteristiche di primo, secondo o ordine superiore. Le caratteristiche di primo ordine si basano sull'istogramma, ossia i livelli di grigio nell'immagine senza tener conto della loro relazione spaziale. Le caratteristiche di secondo ordine, d'altra parte, studiano la texture dell'immagine utilizzando i voxel, che sono la versione tridimensionale dei pixel. Pertanto, è possibile descrivere le relazioni di similarità tra voxel adiacenti. Le caratteristiche di ordine superiore,

d'altra parte, applicano filtri all'immagine, come i frattali, per riconoscere modelli nell'immagine.

Tendenzialmente le prestazioni dei modelli di AI sono simili a quelle offerte dai radiologi e spesso il supporto degli strumenti di AI ad una diagnosi semi-automatica sembra migliorare le prestazioni in termini delle sopra citate sensibilità e specificità^[19] e può ridurre il carico di lavoro del radiologo, evidenziando sulle immagini potenziali caratteristiche e/o estraendo informazioni quantitative delle stesse^[20].

L'analisi automatica delle immagini seriali con strumenti di AI può mettere in evidenza ed indirizzare l'attenzione del radiologo verso cambiamenti subdoli di difficile individuazione, soprattutto in contesti di forte carico lavorativo.

Tutti gli studi che si sono in qualche modo analizzati non offrono evidenza solida di generalizzabilità e corrispondente quantificazione della robustezza delle soluzioni proposte o valutate; queste limitazioni sono per lo più ricondotte alla carenza nei *dataset* utilizzati per lo sviluppo dei modelli (nel loro addestramento) e per i successivi test.

Molto spesso gli autori dei lavori mettono in chiara evidenza limitazioni che hanno molteplici sfaccettature: il numero di immagini (estensione, eterogeneità, completezza), loro qualità, varietà dei dati

clinici e di laboratorio a supporto, variabilità temporale rispetto all'insorgenza dei sintomi.

È ampiamente condivisa la necessità di migliorare qualità e quantità dei dati per addestramento, validazione e test in modo da rendere i risultati clinicamente valutabili. È stata ribadita da più fonti l'importanza di effettuare eventuale sperimentazione clinica per approfondire il potenziale utilizzo di strumenti AI, soprattutto nelle prime fasi della malattia dove le evidenze radiologiche sono spesso negative o marginali anche in pazienti positivi e pertanto inutilizzabili per la diagnosi.

È auspicabile tra l'altro una maggiore integrazione sinergica di immagini, dati clinici e di laboratorio, che sembra essere una delle potenzialità più promettenti per le tecnologie di AI.

1.3 SOFTWARE UTILIZZATI

La maggior parte delle immagini radiologiche viene memorizzata in un formato speciale, adatto per visualizzare, gestire e trasferire immagini mediche. Il formato principale nel campo dell'immagine medica è chiamato Digital Imaging and Communication in Medicine (DICOM).

Questo standard è estremamente accurato: ad esempio, garantisce la stessa variazione di luminosità su diversi dispositivi con gli stessi valori digitali, a 300 DPI e con 12 bit di livelli di grigio (8 per i normali dispositivi digitali)^[21].

Il formato DICOM si basa sulla programmazione orientata agli oggetti, con un'intestazione e diversi tag per pazienti, esame radiologico, scanner e stampante, che insieme contribuiscono alle informazioni radiologiche finali. Le immagini mediche sono rappresentate da una serie di tag, tra cui il tag del valore del pixel 'image pixel module'; possono essere compresse in modo sia lossless che lossy e possono essere visualizzate solo con un software visualizzatore DICOM in grado di leggere tutti i metadati dell'immagine. Per un singolo esame, avremo molte immagini DICOM (anche 500) che costituiscono l'esame radiologico, e ciascuna singola immagine è chiamata "slice".

Il formato NRRD (Nearly Raw Raster Data) è adatto per la visualizzazione e la memorizzazione di dati raster N-dimensionali. Si tratta di un file di testo con due intestazioni e poi diversi campi di informazioni su spessori, spazio generale e informazioni sull'orientamento.

Con l'avvento della Radiomica, sono stati sviluppate migliaia di features per calcolare i parametri radiomici. Nel 2018, una collaborazione internazionale ha introdotto l'Image Biomarkers Standardization Initiative (IBSI) per creare uno standard di estrazione delle caratteristiche radiomiche. In particolare, hanno descritto il biomarker come "una caratteristica che viene misurata e valutata oggettivamente come indicatore di normali processi biologici, processi patogeni o risposte farmacologiche a seguito di un intervento terapeutico" [22]. Hanno stabilito un flusso di lavoro radiomico completo, una definizione formale di ciascuna caratteristica radiomica e della sua estrazione. Lo standard IBSI è implementato in una libreria Python chiamata PyRadiomics^[23]. Come strumento grafico, PyRadiomics è incluso nel software open-source 3D Slicer, che consente la visualizzazione e la gestione di DICOM, la segmentazione ROI, l'estrazione delle caratteristiche, e così via, pertanto è stato utilizzato come software per i nostri studi^[24].

PYTHON

"PyRadiomics" è un termine che combina "Python" e "Radiomics": è una libreria che consente l'estrazione e l'analisi di varie caratteristiche da immagini mediche come tomografie computerizzate (TC) o risonanze

magnetiche (RM) in 2D (basate su pixel) o 3D (basate su voxel)^[23]. È conforme allo standard IBSI, il quale mira ad aumentare la consapevolezza delle capacità radiomiche e a espandere la comunità.

All'interno di PyRadiomics, ci sono caratteristiche di primo ordine, di secondo ordine e di ordine superiore. In particolare, ci sono 7 classi di caratteristiche, distinte come segue:

- **FIRST**, che sta per "Statistiche del primo ordine". Descrive la distribuzione delle intensità dei voxel all'interno della ROI (Region Of Interest) attraverso metriche comunemente usate e di base come Media, Mediana, Entropia;
- **SHAPE features**, che sono caratteristiche di ordine superiore, perché non dipendono dai livelli di grigio dell'immagine, ma solo dalla relazione spaziale dei pixel/voxel. Le caratteristiche sono calcolate su un ROI approssimativa definita da una rete triangolare generata attraverso l'applicazione di un algoritmo Marching Cubes^[25];
- **GLCM**, che sta per " Gray Level Co-occurrence Matrix ". È la matrice più conosciuta e utilizzata e calcola il numero di volte in cui la combinazione di due livelli i e j si verifica in due pixel adiacenti nell'immagine;

- **GLDM**, che sta per " Gray Level Dependence Matrix ". Quantifica le dipendenze dai livelli di grigio nell'immagine. Una dipendenza dai livelli di grigio è definita come il numero di voxel connessi entro una distanza δ che dipendono dal voxel centrale;
- **GLRLM**, che sta per " "Gray Level Run Length Matrix " e quantifica le *Runs* delle corse dei livelli di grigio. Le *Runs* sono definite come la lunghezza, in numero di pixel, di pixel consecutivi che hanno lo stesso valore di livello di grigio;
- **GLSZM**, che è una " Gray Level Size Zone " e quantifica le zone di livello di grigio in un'immagine. Una zona di livello di grigio è definita come il numero di voxel connessi che condividono la stessa intensità di livello di grigio. Un voxel è considerato connesso se la distanza è 1, secondo la regola dell'infinito;
- **NGTDM**, chiamata " Neighbouring Gray Tone Difference Matrix ", quantifica la differenza tra un valore di grigio e il valore di grigio medio dei suoi vicini entro una distanza δ . La somma delle differenze assolute per il livello di grigio i è memorizzata nella matrice.

Per ciascuna di queste matrici, la misura di distanza standard è l'Euclidea, ma ci sono anche altre scelte, come la distanza di Manhattan o la regola dell'infinito^[26]. Per default, la GLCM calcolata è simmetrica.

FIRST	10 Percentile; 90 Percentile; Energy; Entropy; Interquartile Range; Kurtosis; Maximum; Mean Absolute Deviation; Mean; Median; Minimum; Range; Robust Mean Absolute Deviation; Root Mean Squared; Skewness; Total Energy; Uniformity; Variance.
SHAPE 2D or 3D	Mesh surface; Pixel surface; Perimeter; Perimeter to Surface ratio; Sphericity; Spherical Disproportion; Maximum 2D diameter; Major Axis Length; Minor Axis Length; Elongation.
GLCM	Autocorrelation; Joint Average; Cluster Prominence; Cluster Shade; Cluster Tendency; Contrast; Correlation; Difference Average; Difference Entropy; Difference Variance; Joint Energy; Joint Entropy; Informational Measure of Correlation 1; Informational Measure of Correlation 2; Inverse Difference Moment; Maximal Correlation Coefficient; Inverse Difference Moment Normalized; Inverse Difference; Inverse Difference Normalized; Inverse Variance; Maximum Probability; Sum Average; Sum Entropy; Sum of squares.
GLDM	Small Dependence Emphasis; Large Dependence Emphasis; Gray Level Non-Uniformity; Dependence Non-Uniformity; Dependence Non-Uniformity Normalized; Gray Level Variance; Dependence Variance; Dependence Entropy; Low Gray Level Emphasis; High Gray Level Emphasis; Small Dependence Low Gray Level Emphasis; Small Dependence High Gray Level Emphasis; Large Dependence Low Gray Level Emphasis; Large Dependence High Gray Level Emphasis.
GLSZM	Small Area Emphasis; Large Area Emphasis; Gray Level Non-Uniformity; Gray Level Non-Uniformity Normalized; Size-Zone Non-Uniformity; Size-Zone Non-Uniformity Normalized; Zone Percentage; Gray Level Variance; Zone Variance; Zone Entropy; Low Gray Level Zone Emphasis; High Gray Level Zone Emphasis; Small Area Low Gray Level Emphasis; Small Area High Gray Level Emphasis; Large Area Low Gray Level Emphasis; Large Area High Gray Level Emphasis.
GLRLM	Short Run Emphasis; Long Run Emphasis; Gray Level Non-Uniformity; Gray Level Non-Uniformity Normalized; Run Length Non-Uniformity; Run Length Non-Uniformity Normalized; Run Percentage; Gray Level Variance; Run Variance; Run Entropy; Low Gray Level Run Emphasis; High Gray Level Run Emphasis; Short Run Low Gray Level Emphasis; Short Run High Gray Level Emphasis; Long Run Low Gray Level Emphasis; Long Run High Gray Level Emphasis.
NGTDM	Coarseness; Contrast; Busyness; Complexity; Strength.

Tabella riepilogative delle features radiomiche e relative classi usata in PyRadiomics

1.4 METODI FORMALI

Un modello è una versione semplificata della realtà. Utilizziamo modelli ogni giorno e in tutti i campi: basti pensare ad una semplice mappa geografica, a un grafico o a un progetto biomedico che riproduce un braccio artificiale. L'efficacia del modello non dipende dal suo valore, ma dalla selezione dei dettagli inclusi: meno dettagli superflui ci sono, maggiore è la qualità del modello.

Un modello può essere espresso in forme diverse, verbali, scritte o mediante diagrammi. I Metodi Formali consentono una rappresentazione semplificata delle connessioni di un sistema in una sintassi matematica^[27]. Sono spesso utilizzati nell'ambito dei sistemi critici, in cui c'è un numero relativo di agenti che hanno un alto interesse e conoscono tutti gli scenari d'uso del sistema. Infatti, questo tipo di metodi viene impiegato nell'analisi del software e dell'hardware per verificare se l'implementazione è allineata alle specifiche e viceversa^[28, 29].

La specifica è il processo di descrizione dei requisiti funzionali e non funzionali di un sistema che verrà implementato. La Specifica Formale richiede una sintassi e una semantica definite: una delle più utilizzate è il

Calculus of Communicating Systems (CCS), proposto da Milner e altri^[30].

La rappresentazione grafica del comportamento di un sistema è chiamata Sistema di Transizioni Etichettate (LTS, Labelled Transition System) e consiste in stati e transizioni. Il LTS parte da uno stato iniziale, che funge da punto di partenza predefinito; quindi, vengono illustrate le possibili transizioni da uno stato (denominato "P1") a un altro stato (designato come "P2") attraverso un'azione specifica (indicata come "a"). Nel contesto del Calcolo dei Sistemi Comunicanti (CCS), una serie di operatori e una semantica operativa vengono combinati per costruire automaticamente il modello LTS di un sistema. Questa sintassi operativa ha un grande potere modellistico per la descrizione di un sistema stabilendo assiomi e regole di transizione.

1.5 TECNICA DEL MODEL CHECKING

Il Model Checker è lo strumento atto a verificare se il sistema rispetta la specifica iniziale^[31]. Questo è possibile grazie alle proprietà logiche temporali che contengono i comportamenti che devono essere implementati per avere una specifica coerente. La sintassi deve riflettere gli operatori logici utilizzati nei modelli formali, a cui vengono poi

aggiunti altri operatori. Di seguito è riportato l'insieme di operazioni utilizzato per scrivere una proprietà formale.

$$\phi ::= tt \mid ff \mid Z \mid \phi \vee \phi \mid \phi \wedge \phi \mid [K]\phi \mid \langle K \rangle \phi \mid \nu Z.\phi \mid \mu Z.\phi$$

Il Mu-calculus è la tecnica formale utilizzata per verificare, in modo formale, se le proprietà sono soddisfatte in modo temporale e ricorsivo^[32].

Tuttavia, con questo metodo è impossibile discernere se una transizione include un loop interno o se due azioni distinte originano dallo stesso stato. Inoltre, manca la capacità di dedurre un conteggio delle azioni da uno stato specifico o di utilizzare quantificatori per valutare la validità di una formula su tutti gli stati. Fortunatamente, questa logica offre semplicità, robuste capacità espressive e proprietà algoritmiche, rendendola una proposta preziosa per la verifica dei programmi.

Questo processo di verifica è possibile con il Concurrency Workbench of the New Century (CWB-NC), uno strumento che può analizzare la proprietà su un automa a stati finiti^[33]. Nelle sezioni successive, subito dopo l'estrazione delle caratteristiche, i valori vengono sottoposti a un processo di discretizzazione per consentire la rappresentazione formale dell'esame radiologico. Successivamente, l'agente Model Checker valuta

se il modello del paziente corrisponde al modello che rappresenta la malattia.

Diverse metriche possono essere utilizzate per valutare le prestazioni di un classificatore. Un esempio è il conteggio dei falsi positivi (istanze previste come positive ma che sono effettivamente negative nel mondo reale) e dei falsi negativi (istanze previste come negative ma che sono effettivamente positive nel mondo reale). Queste istanze possono essere organizzate in una matrice, che illustra il numero di classificazioni corrette e può essere utilizzata per calcolare vari indici di prestazione, di cui i più importanti in questa fase di classificazione:

- **Accuracy**: rappresenta il numero di pattern correttamente classificati.

In genere si preferisce esprimerlo in percentuale, indicando così la percentuale di classificazioni corrette.

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

Tanto più tale valore è vicino ad 1, tanto più efficiente sarà il classificatore.

- **Precision:** è un indice di correttezza che valuta i pazienti correttamente classificati come positivi sul numero complessivo di pazienti positivi.

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

- **Recall:** è una misura di completezza, indicata il più delle volte con il termine di *Sensitivity*.

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

- **Specificity:** è un parametro che valuta il numero di negativi che sono correttamente classificati come tali; viene indicato anche come Vero Tasso Negativo.

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

Inoltre, sono stati calcolati gli Indici di Utilità Clinica (CUI) come indicato da Mitchell^[34]. Il valore per il CUI varia da 0 a 1:

- utilità eccellente (CUI ≥ 0.81);
- buona utilità (CUI ≥ 0.64);

- utilità soddisfacente/fair ($\text{CUI} \geq 0.49$);
- scarsa utilità ($\text{CUI} < 0.49$);
- utilità molto scarsa ($\text{CUI} \leq 0.36$).

2. AN INNOVATIVE PROSTATE CANCER METHOD TROUGHT FORMAL VERIFICATION

2.1 STATO DELL'ARTE

Il cancro alla prostata viene comunemente diagnosticato mediante biopsia prostatica o durante la resezione transuretrale per iperplasia prostatica. Il grading descrive l'aggressività ed è il parametro principale per determinare se il paziente deve essere sottoposto a un trattamento definitivo o a sorveglianza attiva^[29,35].

Quando il cancro viene individuato, il patologo assegna al cancro un grado, chiamato Gleason Score o gruppo di grado Gleason.

Per assegnare il grado al cancro, il patologo esamina i campioni di tessuto prostatico per valutare quanto il tessuto tumorale assomigli al normale tessuto prostatico e per individuare i due principali modelli cellulari^[36].

Il modello primario descrive il modello di tessuto più comune, mentre il modello secondario descrive il modello successivo più comune^[37]. A ciascun modello viene assegnato un grado, con il grado minimo si intende un tessuto prostatico prelevato molto simile a quello normale, mentre il grado massimo rappresenta quello più anormale. I due gradi vengono quindi sommati per ottenere un Gruppo di Grado Gleason.

In questo modo, un cancro alla prostata diagnosticato può essere contrassegnato con uno dei seguenti gruppi definiti dalla patologia: Gruppo di Grado Gleason $3 + 3 = 6$ (GG1), Gruppo di Grado Gleason $3 + 4 = 7$ (GG2), Gruppo di Grado Gleason $4 + 3 = 7$ (GG3), Gruppo di Grado Gleason $4 + 4 = 8$ (GG4) e Gruppo di Grado Gleason 9-10 (GG5).

Negli ultimi anni, il campo dell'analisi delle immagini mediche ha attirato l'interesse della comunità di ricerca^[38,39], con un numero crescente di strumenti di riconoscimento di pattern e set di dati liberamente disponibili per scopi di ricerca^[3,40].

Nella letteratura di ultima generazione, esistono diversi articoli di ricerca che discutono del potenziale della radiomica nel costruire modelli predittivi per la rilevazione del cancro, principalmente sfruttando l'intelligenza artificiale^[33,41,42,43]. Il principale svantaggio legato all'applicazione dell'intelligenza artificiale nel contesto medico è la mancanza di spiegabilità e affidabilità^[44,45,46]: infatti, la previsione di malattia o meno è prodotta automaticamente rendendo i radiologi inconsapevoli del processo che ha determinato un certo risultato. In contrasto con l'intelligenza artificiale, proponiamo l'uso di tecniche di verifica con metodi formali in cui gli esperti del settore (in questo caso i patologi e i radiologi) formulano una serie di proprietà da verificare,

incorporando quindi la propria conoscenza ed esperienza all'interno del sistema predittivo, rendendo pertanto la decisione del sistema non più inconsapevole e basata su un algoritmo, ma basata sulla conoscenza degli esperti del settore. Altro grande vantaggio dei metodi formali è che non richiedono una grande quantità di dati per l'impostazione delle proprietà, a differenza delle soluzioni basate sull'intelligenza artificiale^[35].

Per questi motivi, si è voluto costruire un software per la verifica con metodi formali da applicare alle immagini di risonanza magnetica (MRI) della prostata, e utilizzarlo per fare diagnosi di cancro alla prostata ed assegnare anche il Gleason score. Ciò si traduce in un approccio non invasivo per il paziente (ossia, non è richiesta una biopsia per la diagnosi di cancro alla prostata).

2.2 MATERIALI E METODI

Abbiamo utilizzato un dataset di 112 pazienti proveniente dall'Archivio di Imaging Oncologico^[47], un vasto archivio di immagini mediche di tumori disponibili a scopo di ricerca. Il dataset è accessibile all'indirizzo: <https://wiki.cancerimagingarchive.net/display/Public/SPIE-AAPM-NCI+PROSTATEx+Challenges>.

Il dataset contiene anche il referto del patologo con i dettagli del Gleason grade group.

Le immagini di risonanza magnetica della prostata sono state eseguite presso il Centro Medico dell'Università di Radboud (Radboudumc) nel Prostate MR Reference Center sotto la supervisione del prof. Dr. Barentsz, situato a Nijmegen, nei Paesi Bassi. Il dataset è stato raccolto per la ricerca nella diagnosi assistita da computer della risonanza magnetica della prostata sotto la supervisione del Dr. Huisman, Radboudumc. Abbiamo considerato le immagini pesate in T2 (T2W) e acquisite sul piano coronale. Le immagini pesate in T2 sono state acquisite utilizzando una sequenza turbo spin echo e avevano una risoluzione di circa 0,5 mm nel piano e uno spessore del taglio di 3,6 mm. La segmentazione è stata eseguita manualmente utilizzando il software 3D Slicer^[48].

Per decidere i pazienti da includere nello studio, il radiologo ha indicato aree di sospetta malattia assegnando a ciascuna di queste un punteggio. Quando un'area veniva considerata probabile per il cancro, veniva eseguita una biopsia. Tutte le biopsie sono state eseguite sotto guida RM. Il campione biptico è stato successivamente classificato da un patologo e i risultati sono stati utilizzati come verità fondamentale.

Di seguito, descriviamo le caratteristiche radiomiche. Per essere precisi, quattro caratteristiche radiomiche sono considerate per generare il

modello formale: consideriamo caratteristiche di forma, cioè caratteristiche indipendenti dalla distribuzione dell'intensità del livello di grigio nella regione di interesse del cancro (cioè ROI)^[49].

I dettagli sulle caratteristiche radiomiche sono mostrati nella Tabella 1.

Feature	Name	Description
RF1	MinorAxisLength	This feature yields the second-largest axis length of the ellipsoid
RF2	MajoraxisLength	This feature considers the largest axis length of the ROI-enclosing ellipsoid and is calculated using the largest principal component
RF3	Maximum2DDiameterColumn	It is defined as the largest pairwise Euclidean distance between tumour surface mesh vertices in the row-slice plane
RF4	Maximum2DDiameterSlice	It is defined as the largest pairwise Euclidean distance between tumour surface mesh vertices in the row-column plane

TABELLA 1

Per valutare il metodo nella rilevazione del Gleason grade group del cancro alla prostata, consideriamo le seguenti metriche: specificità, sensibilità, valore predittivo positivo e valore predittivo negativo.

Le caratteristiche radiomiche sono ottenute utilizzando uno script Python sviluppato dagli autori che invoca `pyradiomics`, che, come sopra anticipato, consiste in una libreria per il calcolo delle caratteristiche radiomiche dalle immagini mediche.

METODI

Il metodo proposto prende in input una risonanza magnetica (MRI) e genera un modello formale dalle sue sezioni. Successivamente, questo modello viene verificato con un insieme di proprietà (una proprietà per

ciascun Gleason grade group) utilizzando lo strumento di verifica formale. Se l'output dello strumento di verifica formale è vero, la MRI viene contrassegnata con il Gleason grade group indicato dalla proprietà.

La nostra analisi inizia direttamente dalle MRI. Una volta ottenuta la MRI del paziente, è possibile utilizzare il metodo proposto direttamente sulle MRI presenti nel Sistema di Archiviazione e Comunicazione delle Immagini (PACS in Figura 1).

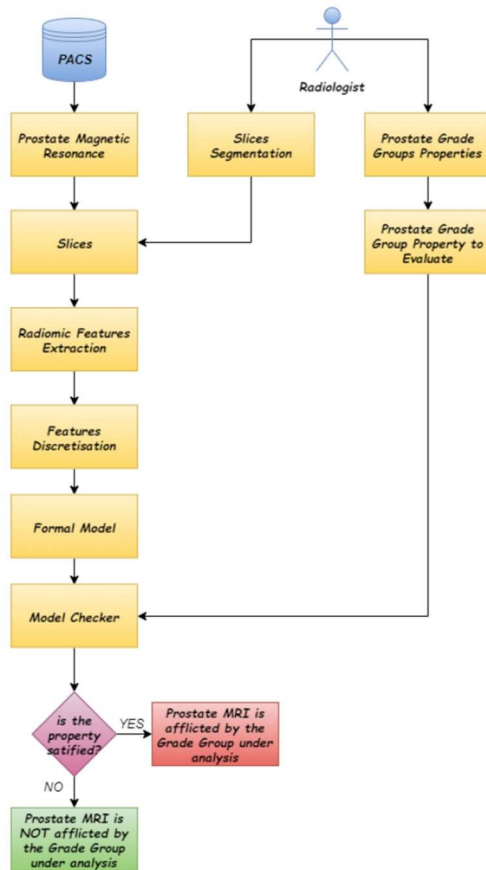


FIGURA 1

In particolare, in questo lavoro, sono state prese in considerazione le slices in sezioni coronali sulle quali è stata applicata una ROI. Una volta ottenute le sezioni MRI e la relativa segmentazione, si procede all'Estrazione delle Features Radiomiche (noi abbiamo scelto quelle relative alla forma), che vengono quantizzate con valori numerici (in mm) per le caratteristiche RF1, RF2, RF3 e RF4 di ciascuna slices appartenente alla MRI del paziente.

A questo punto, per poter generare il Modello Formale, si passa alla Discretizzazione di ciascuna caratteristica numerica, dividendole in tre intervalli: basso, medio e alto con partizione a larghezza uguale. La discretizzazione è necessaria per ottenere una soluzione valida. Trasforma il problema inizialmente continuo in un problema discreto. Questo è necessario a causa della natura finita del modello formale che dobbiamo generare e per limitare lo spazio dei possibili valori che le caratteristiche radiomiche possono assumere, minimizzando anche l'impatto degli outlier (valori anormali). Le caratteristiche discretizzate vengono convertite in un modello formale definito *Language Of Temporal Ordering Specifcation (LOTOS)*^[50].

In particolare, ogni riga rappresenta i valori discretizzati delle caratteristiche radiomiche per una singola sezione, che viene

formalizzata in processi LOTOS. Per meglio comprendere la sintassi LOTOS ne abbiamo utilizzato una versione semplificata.

Di seguito viene riportato un elenco degli operatori utilizzati:

- L'operatore ";" rappresenta la sequenzialità delle azioni. Ad esempio, a; b significa che l'evento b deve essere eseguito dopo l'evento a.
- L'operatore "[|]" rappresenta il parallelismo tra eventi, ovvero l'interleaving delle caratteristiche radiomiche. Ad esempio, a [|] b significa che sia l'evento b che l'evento a devono essere eseguiti, in qualsiasi ordine (vengono considerate sia la sequenza a b che la sequenza b a).

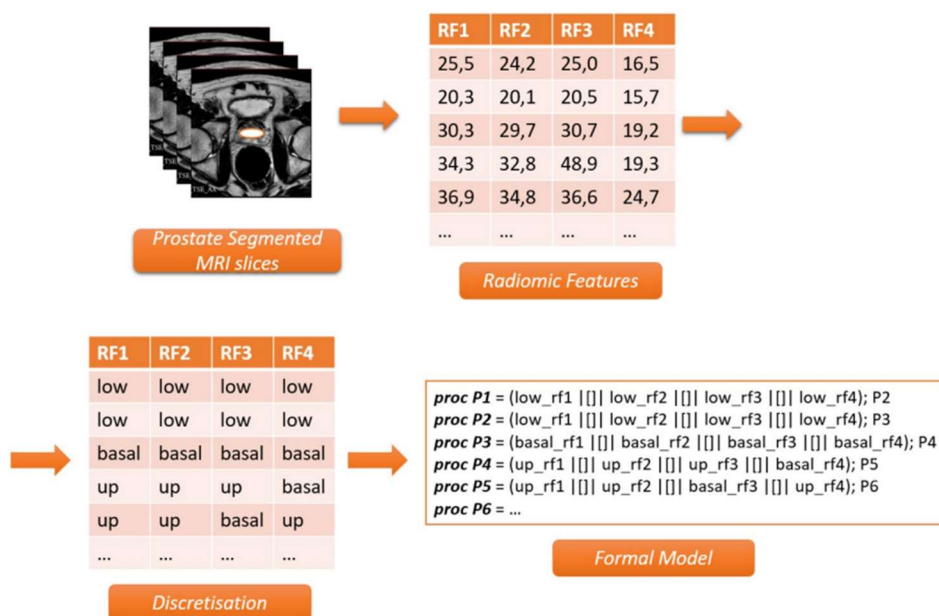


FIGURA 2

Nella prima riga del frammento di processo LOTOS raffigurato nella Figura 2, osserviamo che RF1, RF2, RF3 e RF4 mostrano un valore basso, coerentemente con i valori discretizzati precedentemente ottenuti. Questi eventi sono combinati utilizzando l'operatore "[]", mentre la seconda slices di risonanza, appartenente allo stesso paziente e modellata dal processo LOTOS P2, è composta utilizzando l'operatore ";" P2 rappresenta le caratteristiche radiomiche che mostrano gli stessi valori bassi.

Nella terza riga, il processo LOTOS P3 nella Figura 2 codifica le caratteristiche radiomiche che mostrano valori medi, mentre il processo LOTOS P4 codifica la quarta riga, dove RF1, RF2 e RF3 mostrano un valore elevato, mentre RF4 mostra un valore medio. E così via...

Il modello è stato generato per considerare tutti i punti dell'immagine ed è analizzato da un computer (non da un uomo) per non perdere dettagli. Quindi, il modello formale viene utilizzato per valutare le proprietà formulate dai patologi e radiologi. Infatti, con il supporto di patologi e radiologi, per ogni gruppo di Gleason della prostata, viene formulata una proprietà per rilevare il gruppo di grado specifico e, a seconda del gruppo

di Gleason da verificare, viene selezionata la relativa proprietà, come mostrato in Fig. 1.

Una volta che il radiologo ha selezionato la proprietà da verificare, viene applicato il model checker: se il modello formale soddisfa la proprietà, l'MRI è contrassegnato con il gruppo di Gleason correlato alla proprietà verificata; in caso contrario, l'MRI non è correlato al gruppo di Gleason verificato. Il processo di verifica è mostrato dettagliatamente nella Figura 3.

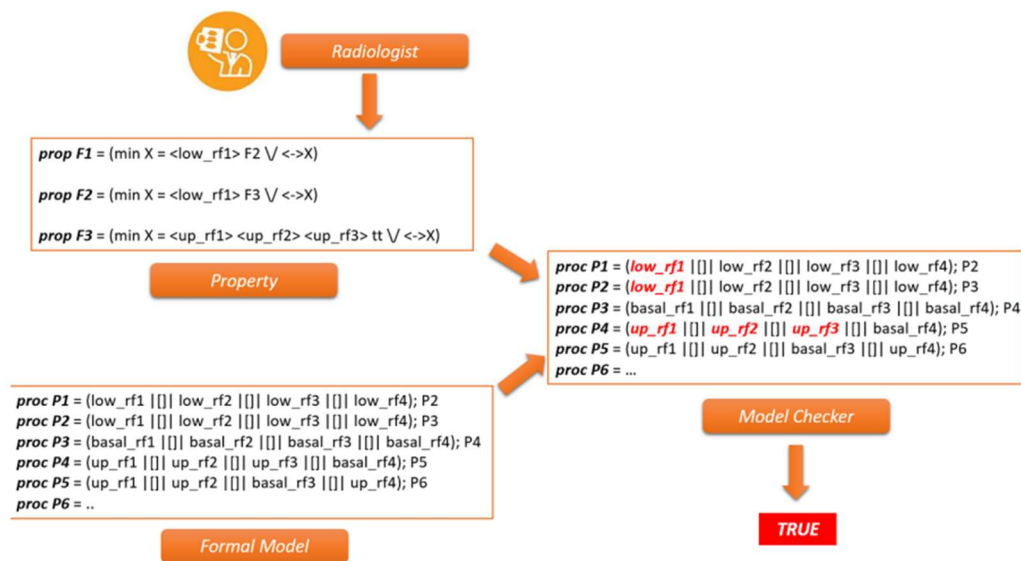


FIGURA 3

Il model checker, coerentemente con il flusso di lavoro mostrato in Fig. 1, accetta due input: il modello formale e la proprietà. Nella Figura 3, viene rappresentato un esempio di proprietà. Le proprietà sono espresse nella logica del calcolo di $\mu^{[51]}$, un'estensione della logica modale

proposizionale che aggiunge l'operatore di punto fisso minimo e l'operatore di punto fisso massimo. In dettaglio, la proprietà mira a verificare se esiste almeno una sequenza di un valore basso per RF1, un altro valore basso per RF1 e tre valori alti per RF1, RF2 e RF3. Chiaramente, tra questi valori potrebbero esserci altri valori: questo esempio di proprietà verifica se questa sequenza è presente nel modello formale. Per questo è cruciale che la proprietà rifletta effettivamente il gruppo di Gleason da rilevare. In pratica, nella proprietà i patologi e i radiologi formalizzano la propria conoscenza.

I patologi e i radiologi hanno formulato cinque diverse proprietà: la prima mira a rilevare il GG1, la seconda a rilevare il GG2, la terza a identificare il GG3, la quarta relativa alla rilevazione del GG4 e l'ultima mira a rilevare il GG5. Le proprietà sono espresse in una logica temporale per catturare la variazione delle sezioni dell'area cancerosa correlate alla risonanza magnetica del paziente. Ovviamente le proprietà sono state misurate e verificate per ogni slice e, anche in base al numero di slices che presentavano valori alti si è potuto classificare il cancro (ad esempio: la presenza ripetuta di RF2, RF3 e RF4 con un valore alto in diverse sezioni può suggerire il grado GG5).

La verifica delle proprietà viene fatta rispetto ai modelli dei pazienti ottenuti dall'insieme di caratteristiche radiomiche utilizzando la *Construction and Analysis of Distributed Processes* ^[52] (CADP), un ampio ambiente di verifica formale che fornisce diverse tecniche per specificare e verificare sistemi concorrenti finiti. Quando l'ambiente di verifica formale CADP restituisce true verificando una proprietà logica su un modello LOTOS, significa che il metodo proposto ha contrassegnato il modello formale come appartenente al gruppo di grado specificato dalla proprietà analizzata. In caso contrario, l'ambiente di verifica formale restituisce false, ovvero che il modello in analisi non appartiene al gruppo di grado descritto nella formula analizzata.

2.3 SCOPO DELLO STUDIO

Offrire un metodo per rilevare i diversi gradi del cancro alla prostata a partire dall'analisi delle immagini di risonanza magnetica. Piuttosto che, come fatto in precedenza, considerare un algoritmo per inferire automaticamente le proprietà direttamente da un ristretto insieme di modelli, in questo studio abbiamo voluto che le proprietà dell'algoritmo fossero formulate direttamente dagli specialisti (patologi e radiologi) con prestazioni, conseguentemente, più performanti. Questo conferma che l'efficacia delle proprietà proposte per la classificazione del gruppo

Gleason è dipendente dalla formalizzazione delle conoscenze di patologi e radiologi ed è per questo che si è reso necessario questo studio.

2.4 RISULTATI

Con il modello sviluppato si è ottenuta una sensibilità che varia tra 0,95 e 1 e una specificità pari a 1, superando le prestazioni raggiunte in altri lavori basati sui metodi formali.

Va ricordato che l'obiettivo del nostro studio è rilevare automaticamente il gruppo Gleason del cancro alla prostata attraverso la risonanza magnetica (MRI). Per farlo abbiamo generato un modello formale dalla MRI del paziente. Successivamente, un insieme di proprietà viene verificato, dove ciascuna proprietà è correlata a un diverso gruppo di grado del cancro alla prostata. A questo punto in un ambiente di verifica formale abbiamo controllato se le proprietà fossero verificate sul modello: se una proprietà è soddisfatta su un determinato modello, quel modello viene etichettato con il gruppo di grado indicato dalla proprietà.

Di seguito, l'esperimento che abbiamo condotto per dimostrare l'efficacia dell'approccio proposto per la rilevazione del Gleason.

	Total MRIs	GG1	GG2	GG3	GG4	GG5
Grade Group 1	36	35	0	<i>1</i>	0	0
Grade Group 2	41	<i>1</i>	39	<i>1</i>	0	0
Grade Group 3	20	0	0	19	<i>1</i>	0
Grade Group 4	8	0	0	0	8	0
Grade Group 5	7	0	0	0	0	7

TABELLA 2

Nella Tabella 2 sono classificate le risonanze prese in esame per ciascun gruppo di grado e il numero di MRI etichettate come vere dall'ambiente di verifica formale.

Ogni riga nella Tabella 2 è correlata al modello formale di MRI risultato vero per le proprietà definite da patologi e radiologi per ciascun gruppo di grado. La colonna "Total MRI" mostrata sempre nella Tabella 2 fornisce dettagli su MRI incluse nell'esperimento: un totale di 112 MRI distribuite in 36 MRI etichettate da patologi e radiologi con GG1, 41 MRI etichettate da patologi e radiologi con GG2, 20 MRI etichettate da patologi e radiologi con GG3, 8 MRI etichettate da patologi e radiologi con GG4 e 7 MRI etichettate da patologi e radiologi con GG5. In particolare, la proprietà GG1 è stata correttamente etichettata in 35 MRI su 36 appartenenti a GG1, la proprietà GG2 è stata correttamente etichettata per 39 MRI su 41, la proprietà GG3 è stata correttamente rilevata per 19 MRI su 20, la proprietà GG4 è stata correttamente

etichettata su tutte le 8 MRI etichettati da patologi e radiologi con la malattia GG4. Infine, la proprietà GG5 è stata correttamente rilevata in tutti e 7 le MRI con la malattia GG5.

A questo punto per dare una stima delle prestazioni dell'approccio proposto, consideriamo le seguenti metriche: Specificità, Sensibilità, Valore Predittivo Positivo e Valore Predittivo Negativo. La Tabella 3 mostra i risultati delle prestazioni.

	Sensitivity	Specificity	PPV	NPV
Grade Group 1	0.97	1	1	0.98
Grade Group 2	0.95	1	1	0.97
Grade Group 3	0.95	1	1	0.98
Grade Group 4	1	1	1	1
Grade Group 5	1	1	1	1

TABELLA 3

Come mostrato nella Tabella 3, il metodo proposto ottiene una sensibilità che varia tra 0.95 (per la rilevazione di GG1 e GG3) e 1 (per la rilevazione di GG4 e GG5). Per quanto riguarda la specificità, si raggiunge un valore pari a 1 per tutti i gruppi di grado.

Ora passiamo alla discussione ed esplicazione delle scelte effettuate nei materiali e nei metodi. Esaminiamo anche i limiti del nostro metodo e,

infine, confrontiamo i nostri risultati con quelli disponibili in letteratura per evidenziare i nostri progressi.

Nell'analisi delle risonanze magnetiche (MRI), possono essere considerati tre piani: sagittale, coronale ed assiale. Nella progettazione della nostra metodologia, abbiamo indagato sul piano che ha portato a risultati migliori. Nella letteratura attuale, diversi articoli hanno dimostrato che il piano coronale è la scelta migliore, non solo per la prostata ma anche per altri organi. In particolare, per le lesioni prostatiche, nell'articolo^[53], gli autori affermano che "I modelli basati su caratteristiche estratte dalla sequenza coronale T2 hanno ottenuto prestazioni complessive molto migliori rispetto alle sequenze sagittali e assiali". Inoltre, per la patologia del canale uditivo interno, nell'articolo^[54] gli autori affermano che "La sequenza coronale T2 mostra meglio la lesione ipointensa". E ancora nell'articolo^[55], gli autori scrivono: "In entrambe le applicazioni, oncologiche e reumatologiche, il piano coronale è spesso preferito perché consente una visione estesa e un'indagine diretta dello scheletro".

Come dimostrato in precedenza, la nostra metodologia fornisce risultati molto buoni, anche se presenta alcune debolezze. I principali limiti della nostra metodologia sono legati alle seguenti questioni.

- Definizione manuale delle formule. Il set di regole logiche che caratterizzano il grading del cancro alla prostata deve essere progettato e definito e scrivere regole corrette può essere un compito piuttosto complesso. Il lato positivo, tuttavia, è che una volta che le formule sono state definite, possono essere utilizzate senza alcuna modifica e la metodologia diventa completamente automatica, non richiedendo ulteriori input da parte dell'utente.
- Prestazioni temporali. La nostra metodologia è stata implementata in uno strumento prototipo di ricerca il cui obiettivo principale è dimostrare l'efficacia nell'identificazione del grado di cancro alla prostata; pertanto, le prestazioni temporali non sono il nucleo. Sebbene il tempo per ottenere i risultati sia ancora elevato, il lato positivo è l'efficacia dei risultati, questione di vitale importanza; tanto che riteniamo sia meglio attendere anche tempi leggermente più lunghi, ma ottenere una sensibilità media pari a 1.

Nel complesso, riteniamo che questi limiti non limitino gravemente l'applicabilità del nostro metodo. Il nostro metodo, come dimostrato dall'esperimento, dovrebbe essere considerato come un buon controllo nella diagnosi e classificazione del cancro alla prostata.

Il nostro metodo, in compenso, ha anche un altro grande vantaggio rispetto alle tecniche di apprendimento automatico perché non richiede coorti di addestramento. La fase di formazione è completamente trasferita all'esperienza dei patologi e dei radiologi che aiuta nella formulazione delle formule temporali. Pertanto, il nostro set di dati costituisce anche la coorte di convalida.

2.5 CONCLUSIONI

In questo articolo viene proposto un approccio basato sul model checking per rilevare il gruppo di grado del cancro. Modelliamo le risonanze magnetiche (MRI) dei pazienti attraverso un modello LOTOS e valutiamo i modelli ottenuti attraverso l'ambiente di verifica formale CADP. Un insieme di proprietà, legate a ciascun gruppo di grado, formulato con l'aiuto di patologi e radiologi. Come lavoro futuro, pianifichiamo di modellare pazienti affetti da altri tipi di cancro. Inoltre, indagheremo se il metodo proposto possa essere sfruttato nel contesto della medicina di precisione, un campo di ricerca promettente che consente ai medici di selezionare trattamenti che sono più probabili di

aiutare i pazienti sulla base di una comprensione genetica della loro malattia.

3 EXPLAINABILITY OF FORMAL METHODS

La recente letteratura ha stabilito l'efficacia dell'Intelligenza Artificiale (IA) nell'analisi quantitativa, nella ricerca, nella diagnosi e nella prognosi nel campo della Radiomica^[56,57,58,59]. Una delle sfide principali in questo contesto è capire la logica dietro le previsioni specifiche del modello: in questo senso, le soluzioni basate sull'IA non hanno avuto successo nel mondo medico reale. Ovviamente l'analisi di questi problemi di interpretazione porta vantaggi sia per i pazienti che per i medici, poiché i modelli potrebbero beneficiare della loro competenza e conoscenza.

Una soluzione alternativa agli approcci di IA, è rappresentata dai Metodi Formali^[43,60], che sfruttano il potere della logica matematica e offrono un

vantaggio unico. Con i Metodi Formali, diventa possibile amalgamare il campo della Scienza Informatica con l'esperienza dei radiologi per costruire sistemi esperti ed efficienti che aiutano nella diagnosi o prognosi di varie malattie. Rimane comunque inteso che il risultato automatizzato finale, ovvero la diagnosi, dovrebbe essere considerato come un 'parere virtuale secondario' che integra l'esperienza del radiologo. Ovviamente avere a disposizione questi strumenti rappresenta un vantaggio non trascurabile in condizioni particolari come emergenze epidemiologiche o nel caso di malattie rare e di difficile diagnosi. Il flusso di lavoro del metodo formale e della radiomica è estremamente semplice e in contrasto con le 'scatole nere' dell'IA, che invece sono più difficili da spiegare, come riportato in^[61].

L'obiettivo principale è condurre un'analisi orientata alla medicina di varie caratteristiche radiomiche seguendo le linee guida IBSI^[62], cercando di stabilire connessioni tra queste caratteristiche ed evidenze visive cliniche o radiologiche, costruendo efficacemente modelli basati su diagnosi precedentemente fatte. A tal fine alcuni aspetti chiave possono essere:

1. Comprendere l'intervallo di valori delle caratteristiche considerato dalla metodologia per la diagnosi automatica;

2. Investigare questi intervalli di valori per discernere pattern associati a diversi stati patologici;
3. Comprendere i meccanismi del riconoscimento automatico di pattern per confermare la diagnosi automatica;
4. Condurre la localizzazione delle fette di immagine più cruciali utilizzate nel processo diagnostico.

Il processo di localizzazione delle fette consente agli specialisti di ricevere queste immagini come feedback, consentendo loro di confermare o contestare la diagnosi qualitativa visiva. Questo approccio innovativo offre ai medici e ai loro colleghi l'opportunità di fornire spiegazioni per la diagnosi automatica. Secondo gli autori, la trasparenza è di primaria importanza in questo campo traslazionale, dove l'esperienza umana è indispensabile.

Le caratteristiche di primo ordine descrivono la distribuzione dei valori dei singoli voxel indipendentemente dalla loro disposizione spaziale, indicando la presenza di colori o valori specifici in un'immagine mediante istogrammi, ma non come sono disposti nell'immagine. Di conseguenza, stabilire una connessione diretta tra i valori numerici e ciò che un radiologo percepisce durante una scansione TC multi-slice può essere piuttosto impegnativo.

Ad esempio, non sono in grado di distinguere tra un motivo a scacchiera in bianco e nero e un'immagine divisa con un lato nero e l'altro bianco: nel caso della Figura 4, l'istogramma a destra mostra un gran numero di pixel in solo due colonne di intensità di livello di grigio, una per i pixel neri e una per i pixel bianchi.

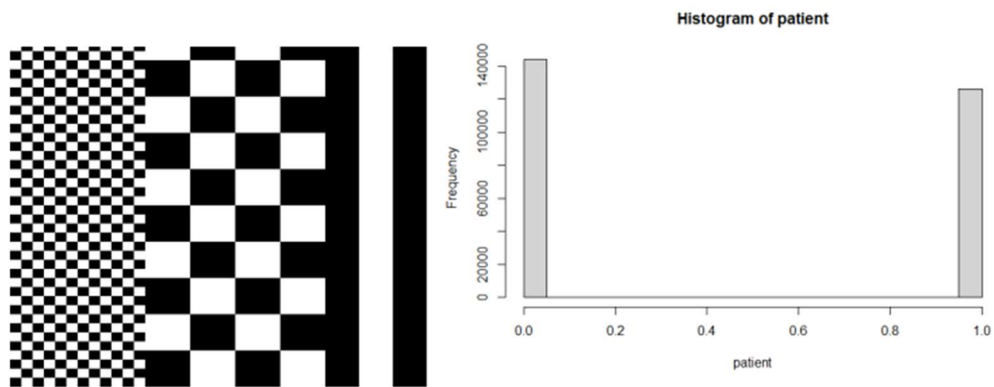


FIGURA 4

Questa limitata informazione può costituire un ostacolo per i ricercatori che analizzano i diversi schemi all'interno delle immagini. D'altro canto, le statistiche di primo ordine forniscono le prime indicazioni sulle caratteristiche generali di un'immagine e sono attualmente le metriche più ampiamente utilizzate per quantificare il fenotipo delle immagini, come indicato nell'articolo^[42].

Le caratteristiche di primo ordine derivano dall'istogramma dell'immagine e forniscono metriche come la media, la mediana, i valori massimi e minimi delle intensità dei voxel. Comprendono anche

caratteristiche come la skewness (indicante l'asimmetria), la kurtosis (esprimendo l'appiattimento), l'uniformità e l'entropia (indicante la casualità). Ci si aspetta che le caratteristiche di base come Massimo, Minimo e Media corrispondano ai valori di Hounsfield relativi^[63] per i livelli di intensità massima, minima e media, mentre le rimanenti caratteristiche sono essenzialmente misurazioni statistiche.

Il lavoro da cui si è partiti^[64] può essere considerato come la base per questi collegamenti tra valori numerici ed evidenze radiologiche. È essenziale notare che nella metodologia spiegata nel suddetto lavoro, gli autori hanno scelto di non ritagliare l'immagine e quindi la ROI (Region of Interest) include l'intera immagine TC, garantendo che nessuna parte dell'immagine venga trascurata nel processo di calcolo. Le caratteristiche radiomiche sono state calcolate per ciascun paziente e le variazioni associate a diversi stati di salute sono state evidenziate: sono stati presi in considerazione gli elementi chiave della letteratura esistente per gli studi polmonari e sono stati compiuti sforzi per fornire significato e interpretabilità all'intero processo di utilizzo dei metodi formali per la classificazione del COVID-19.

Un dataset di 35 pazienti potrebbe sembrare limitato per le tecniche di intelligenza artificiale, che di solito richiedono un volume sostanziale di

immagini per creare modelli robusti con prestazioni elevate. Tuttavia, questa regola non si applica ai metodi formali, che non imparano direttamente dalle immagini ma si basano sulla conoscenza e sull'esperienza dei radiologi, traducendo questa competenza in proprietà della logica temporale. Di conseguenza, anche un piccolo dataset consente prestazioni elevate. Tuttavia, è importante ricordare che questo metodo non è completamente separato dall'intelligenza artificiale. Il flusso di lavoro radiomico include passaggi in cui l'intelligenza artificiale è essenziale, come il processo di selezione della funzionalità.

Attualmente, i Metodi Formali si distinguono tra i pochi che possono affrontare alcune delle limitazioni dell'IA, in particolare la necessità di dati estesi per garantire la robustezza del modello. Utilizzando i metodi formali, la diagnosi è generata dal Model Checker, che restituisce un verdetto di "Vero" o "Falso" basato sulla soddisfazione di una proprietà della malattia. Quando l'output è "Falso", il Model Checker chiarisce anche la logica dietro la sua decisione identificando il punto specifico in cui si verifica la discrepanza tra il modello LTS e la proprietà. Come illustrato nella Figura 5, c'è un LTS di un paziente COVID-19 mentre si verifica una proprietà associata a un paziente sano. Il controesempio generato include il LTS di input, insieme a chiare annotazioni che spiegano le ragioni del suo fallimento nel soddisfare la proprietà. In

questa specifica istanza nella Figura 5, la proprietà sta cercando la varianza di b1of3 e l'energia di b1of3, mentre il LTS di input inizia con la varianza di b2of3 e l'energia di b3of3. Di conseguenza, lo stato iniziale è contrassegnato in rosso, e l'output finale è "Falso", indicando che il paziente non è affetto da COVID-19.



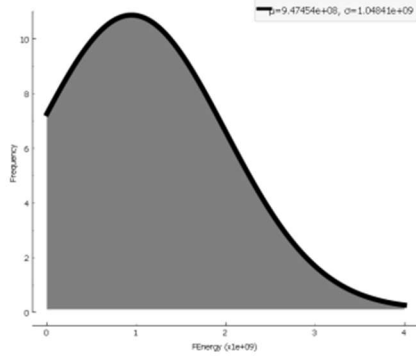
FIGURA 5

I Metodi Formali sono comprensibili sia per i professionisti medici che per gli scienziati informatici o gli esperti di radiomica, anche grazie all'incorporazione delle attività umane nel flusso di lavoro. Esaminando le informazioni diagnostiche derivate dall'applicazione dei metodi formali ai pazienti COVID-19 e l'analisi delle loro scansioni TC corrispondenti, l'obiettivo è stabilire collegamenti tra valori numerici ed evidenze radiologiche, con l'idea di creare le basi per l'implementazione di uno standard radiomico per migliorare la comprensione delle malattie polmonari.

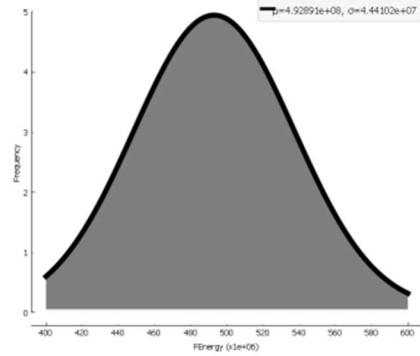
Partendo da alcune caratteristiche di Primo Ordine utilizzate per la diagnosi automatizzata dei pazienti COVID-19 (Energia, Intervallo interquartile, Asimmetria, Energia totale, Varianza, Percentile 90, Curtosi e Minimo), le distribuzioni dei valori sono state visualizzate e studiate utilizzando Orange^[65], e i risultati sono presentati come grafici con annotazioni riguardanti la media e la deviazione standard corrispondente. Nell'analisi dei valori tra pazienti sani e COVID-19, si è osservato che i valori di Energia, Minimo e Varianza nei pazienti COVID-19 erano inferiori rispetto ai pazienti sani. Al contrario, le restanti caratteristiche come Asimmetria e Curtosi erano superiori nei pazienti COVID-19 rispetto ai pazienti sani.

L'Energia^[62] è una misura delle dimensioni dei valori dei voxel ed è spesso associata alla misurazione del volume, ma in questo caso è importante notare che la formula non considera le dimensioni spaziali. L'Energia è anche un indicatore del tasso di cambiamento del colore, della luminosità e delle dimensioni dei valori di grigio in aree locali, che rappresenta essenzialmente l'uniformità dell'intensità in un'immagine. In questo studio, gli autori hanno osservato diverse distribuzioni di valori tra classi malate e sane, come illustrato nella Figura 6. La figura mostra le curve di distribuzione e include annotazioni per “mu”, che rappresenta il valore medio, e “sigma”, che sta per la deviazione standard. In questa

e nelle figure successive, il nome della funzione è rappresentato sull'asse x, mentre l'asse y rappresenta la frequenza dei valori.



(a) Energy value distribution for COVID-19 patients.



(b) Energy value distribution for healthy patients.

FIGURA 6

Le misurazioni di Curtosi e Asimmetria sono stime statistiche che valutano la forma di una distribuzione e quantificano quanto essa si discosta da una distribuzione gaussiana (normale), come dimostrato nella

Figura 7.

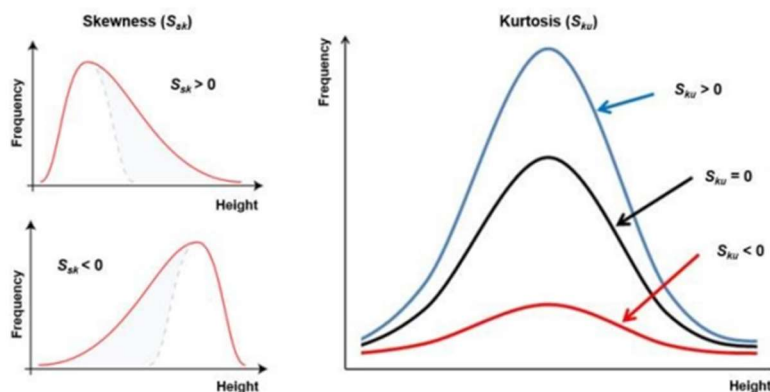


FIGURA 7

La Curtosi indica se la distribuzione è più alta o più piatta rispetto a una distribuzione normale. In particolare, valuta l'uniformità dell'immagine attraverso la forma della curva di distribuzione. Negli studi polmonari^[66], un valore di curtosi elevato suggerisce un grado maggiore di eterogeneità del tessuto e complessità microstrutturale, mentre un valore di curtosi basso riflette indirettamente un grado maggiore di connessione incrociata e uniformità del tessuto, come in un motivo a nido d'ape. Queste variazioni sono direttamente evidenti negli istogrammi, che mostrano differenze distinte tra diverse categorie di pazienti, come nella Figura 8.

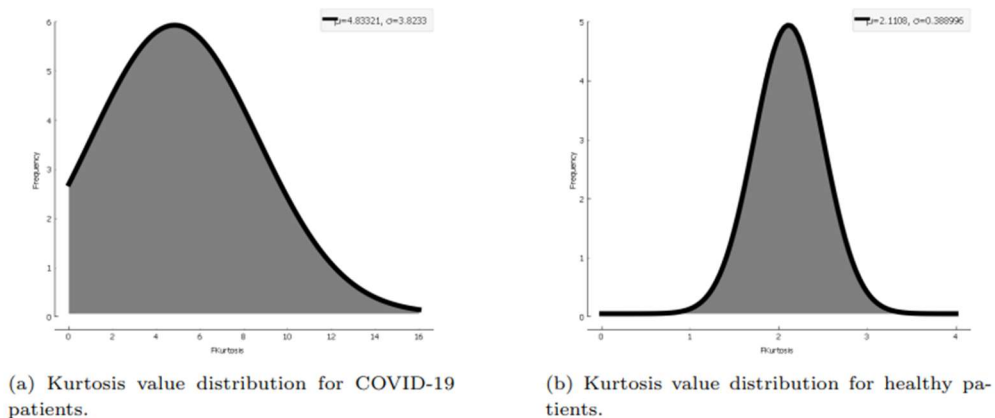


FIGURA 8

La Skeweness rappresenta il grado di asimmetria nella distribuzione dell'istogramma dell'immagine. Un valore di asimmetria positivo indica una distribuzione positivamente distorta, in cui molti valori sono concentrati nella parte inferiore, mentre un valore di asimmetria negativo rappresenta una distribuzione negativamente distorta con molti valori

concentrati nella parte superiore, come mostrato nella Figura 7. Quando la distribuzione è simmetrica e segue un modello gaussiano normale, l'asimmetria è zero. In questo studio, le curve di distribuzione dell'asimmetria sono rappresentate nella figura 9.

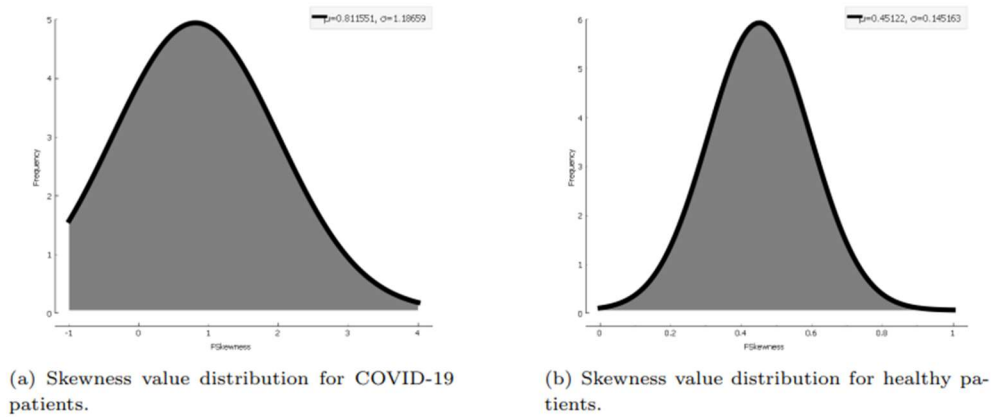
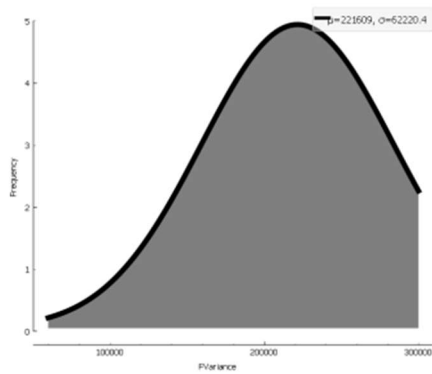
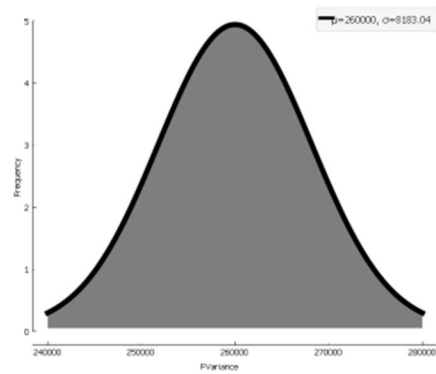


FIGURA 9

Due caratteristiche significative utilizzate nella diagnosi hanno costantemente mostrato valori più elevati nei pazienti affetti da COVID-19: Varianza e Minimo. L'indice di Varianza valuta la media delle distanze quadrate di ciascun valore di intensità dalla media, quantificando la diffusione della distribuzione attorno alla media, come mostrato nella Figura 10.



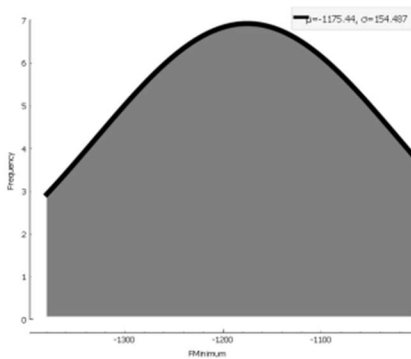
(a) Variance value distribution for COVID-19 patients.



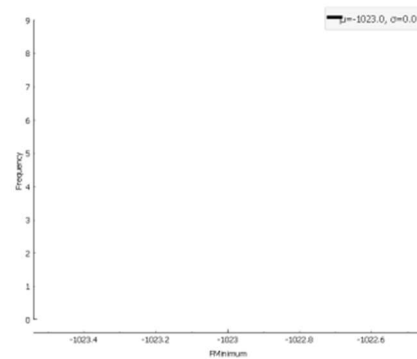
(b) Variance value distribution for healthy patients.

FIGURA 10

La misurazione del Minimo fornisce una stima dell'intensità minima, che corrisponde al livello di intensità più basso all'interno dell'immagine: intuitivamente, la caratteristica del Minimo rappresenta la densità dell'aria all'interno dei polmoni nella scansione TC (Figura 11).



(a) Minimum value distribution for COVID-19 patients.



(b) Minimum value distribution for healthy patients.

FIGURA 11

Queste ultime due caratteristiche sembrano cruciali in uno studio polmonare, date le variazioni osservate nella distribuzione dei valori. Questi dati oggettivi si allineano con alcuni risultati radiologici, come i

risultati dei test di funzionalità polmonare dati dalle misurazioni della densità polmonare: questi mostrano come il tessuto polmonare attenua i raggi X, riflettendo anomalie strutturali all'interno dei polmoni^[67].

Durante la fase di selezione delle caratteristiche, alcune delle caratteristiche menzionate precedentemente sono state scelte utilizzando Weka^[68]. I risultati sono rappresentati nella struttura ad albero mostrata nella Figura 12, con la caratteristica più importante alla radice dell'albero.

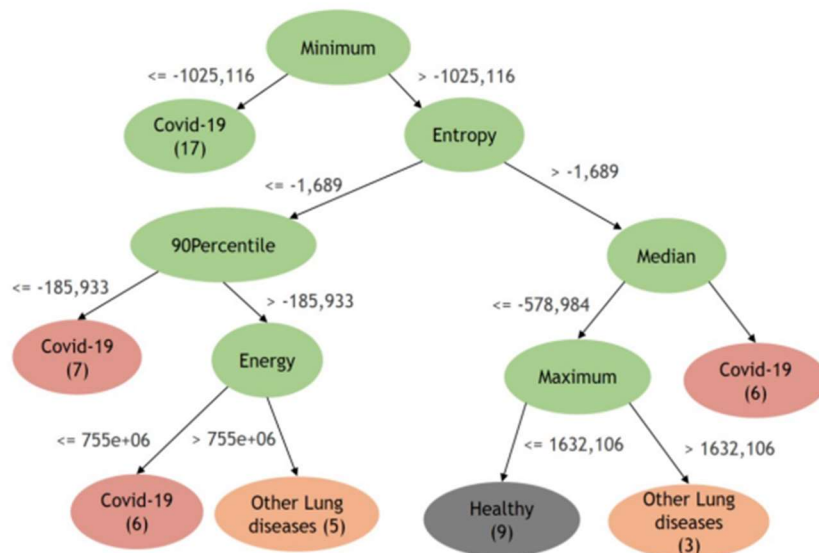


FIGURA 12

La caratteristica Minimo si trova alla radice dell'albero e, come definito da IBSI, rappresenta l'intensità minima nell'immagine. Per quanto concerne la radiologia e gli studi polmonari, il valore minimo dovrebbe corrispondere con precisione al valore dell'aria catturata dalla TC^[45,69].

Tuttavia, l'aria è presente sia nei pazienti sani che malati e non è influenzata dalla presenza della malattia, per cui ci si è domandati perché essa sia significativa per lo strumento di intelligenza artificiale.

Gli autori sono andati più a fondo per capire perché Weka avesse scelto la caratteristica "minimo" come la più discriminante: per i pazienti sani, i valori minimi erano costantemente -1023.0, risultando in un grafico di distribuzione vuoto come mostrato nella Figura 11b. Invece, i valori minimi per alcuni pazienti COVID-19 talvolta superavano -1023.0, come in Fig. 11a.

L'ipotesi che il valore fisso -1023.0 rappresenti correttamente l'aria nei polmoni ha portato gli autori a considerare questo come un potenziale bias legato alle impostazioni della macchina per i pazienti COVID-19. Questa supposizione, indipendentemente dal numero di pazienti, è stata successivamente confermata attraverso l'uso dei metodi formali.

FIRST 1st part		FIRST 2nd part
Energy		Skewness
Variance		
Precision	Specificity	Accuracy
1.00	1.00	0.79

TABELLA 4

Per la classificazione, come mostrato nella Tabella 4, gli autori hanno classificato le caratteristiche del Primo Ordine in base ai due livelli della metodologia. La prima parte della tabella include le caratteristiche utilizzate per la classificazione di pazienti sani e malati, mentre la seconda parte riguarda la classificazione di pazienti COVID-19 e pazienti con malattie polmonari. Inizialmente, la caratteristica Minimo è stata inclusa nella diagnosi automatizzata COVID-19. Di conseguenza, le formule matematiche utilizzate per la diagnosi di secondo livello incorporavano il Minimo in combinazione con la caratteristica Skeweness. Sulla base delle suddette ipotesi, le combinazioni sono state modificate eliminando la caratteristica "Minimo". Dopo una rivalutazione della diagnosi, i risultati sono rimasti consistenti: ciò conferma che la caratteristica Minimo non è un fattore casuale nella classificazione tra pazienti sani e COVID-19 e potrebbe, invece, non essere adatta per altri studi polmonari, contrariamente a quanto suggerito da altri autori in vari documenti relativi ai polmoni^[70,71].

In conclusione, l'esplicazione offerta dai metodi formali permette di convalidare o confutare i risultati dello strumento di intelligenza artificiale, come dimostrato in questo caso. Questa analisi approfondita fornisce i mezzi per interpretare e spiegare i risultati, rivolgendosi sia agli specialisti medici che ai loro collaboratori. Serve come illustrazione

pratica di un meccanismo di feedback per i medici, consentendo loro di contribuire allo sviluppo e all'uso di sistemi di supporto decisionale medica in collaborazione con informatici e scienziati dei dati.

La diagnosi automatica è agevolata da un modello matematico separato noto come "proprietà" o "formula" o "regola", progettato per rappresentare la malattia in modo indipendente, senza essere influenzato dalle caratteristiche del paziente. Oltre a consentire la diagnosi, queste proprietà possono anche fornire approfondimenti e informazioni sui modelli dei pazienti e quindi migliorare l'interpretazione dei risultati.

```
prop F1 =(min X = <b1of3_fvariance> <b1of3_fenergy> F2 \ / < - > X)
prop F2 =(min X = <b1of3_fvariance> <b1of3_fenergy> F3 \ / < - > X)
prop F3 =(min X = <b1of3_fvariance> <b1of3_fenergy> F4 \ / < - > X)
prop F4 =(min X = <b1of3_fvariance> <b1of3_fenergy> F5 \ / < - > X)
prop F5 =(min X = <b1of3_fvariance> <b1of3_fenergy> F6 \ / < - > X)
prop F6 =(min X = <b1of3_fvariance> <b1of3_fenergy> F7 \ / < - > X)
prop F7 =(min X = <b1of3_fvariance> <b1of3_fenergy> F8 \ / < - > X)
prop F8 =(min X = <b1of3_fvariance> <b1of3_fenergy> F9 \ / < - > X)
prop F9 =(min X = <b1of3_fvariance> <b1of3_fenergy> tt \ / < - > X)
```

FIGURA 13

In questo esempio (fig. 9.9), solo le caratteristiche di Varianza ed Energia sono state identificate come fattori decisivi nella determinazione della malattia e verranno utilizzate per classificare l'intero set di dati dei pazienti. Tuttavia, all'interno delle proprietà formali vi è la ripetizione di alcuni valori discretizzati, che aiuta nel processo di classificazione,

chiamato pattern. Il pattern suggerisce che la combinazione di determinati valori può identificare in modo affidabile il COVID-19 e distinguerlo dai pazienti sani.

Per indagare e supportare queste ipotesi, è stata approfondita la struttura dei modelli formali sia per i pazienti sani che per quelli con riconoscimento dei pattern COVID-19. Nel caso dei pazienti sani, il modello di un paziente sano tende a mostrare valori di varianza medio e energia medio-bassa. Questa tecnica ha fornito non solo informazioni sulla distribuzione dei valori, ma evidenzia anche il fatto che questa proprietà potrebbe non essere universalmente applicabile a tutti i pazienti.

Lo studio dei modelli consente ai ricercatori di ottenere un controllo migliore sugli stessi, il che può avere un impatto significativo sulle metriche di performance come accuratezza e specificità. Inoltre, il processo di estrazione del pattern non è né lungo né costoso, rendendolo un'applicazione economica nell'industria medica. Ciò, a sua volta, aumenta l'esplicabilità complessiva dei risultati, a vantaggio sia dei medici specialisti che dei pazienti. Per i medici specialisti, queste

informazioni permettono loro di comprendere la logica dietro la diagnosi automatica e di perfezionare le impostazioni se la diagnosi non è in linea con lo stato reale di salute del paziente, godendo di un sistema di supporto decisionale più robusto. La diagnosi automatica è agevolata da un modello matematico separato noto come "proprietà" o "formula" o "regola", progettato per rappresentare la malattia in modo indipendente, senza essere influenzato dalle caratteristiche del paziente. Oltre a consentire la diagnosi, queste proprietà possono anche fornire approfondimenti e informazioni sui modelli dei pazienti e quindi migliorare l'interpretazione dei risultati.

Come dimostrato, le caratteristiche del Primo Ordine mostrano variazioni tra diversi stati di salute, catturando efficacemente evidenze radiologiche e traducendole in una gamma di valori. Questo concetto evidenzia una grande sfida affrontata dai radiologi nell'analisi degli esami TC, che consiste nella gestione di centinaia di immagini: infatti, lo sforzo umano richiesto dal radiologo si intensifica e potrebbe aumentare gli errori, i casi di dettagli trascurati o i casi di diagnosi automatica divergenti, in cui il radiologo è incerto su quale immagine abbia portato alla discrepanza nella classificazione.

Con l'algoritmo sviluppato per l'uso dei Metodi Formale, si dispone di un metodo semplice ma potente per individuare le fette più cruciali all'interno di un esame TC multistrato. Questo algoritmo identifica le fette in cui vengono rilevati specifici pattern, rivelando quali fette vengono valutate dall'algoritmo diagnostico come patologiche. Nella figura 14, è possibile apprezzare la formula e alcune fette; nella fattispecie la "numero 246" che appunto si riferisce alla 246^a fetta nella serie di esami TC. In questa immagine, la malattia è chiaramente visibile, apparendo spesso come opacità a vetro smerigliato (GGO), segni distintivi del COVID-19.

```
prop F1 =(min X = <b1of3_fvariance> <b1of3_fenergy> F2 \ / < - > X)  
SLICE 246
```

```
prop F2 =(min X = <b1of3_fvariance> <b1of3_fenergy> F3 \ / < - > X)  
SLICE 251
```

```
prop F3 =(min X = <b1of3_fvariance> <b1of3_fenergy> F4 \ / < - > X)  
SLICE 253
```

```
prop F4 =(min X = <b1of3_fvariance> <b1of3_fenergy> F5 \ / < - > X)  
SLICE 255
```

```
prop F5 =(min X = <b1of3_fvariance> <b1of3_fenergy> F6 \ / < - > X)  
SLICE 261
```

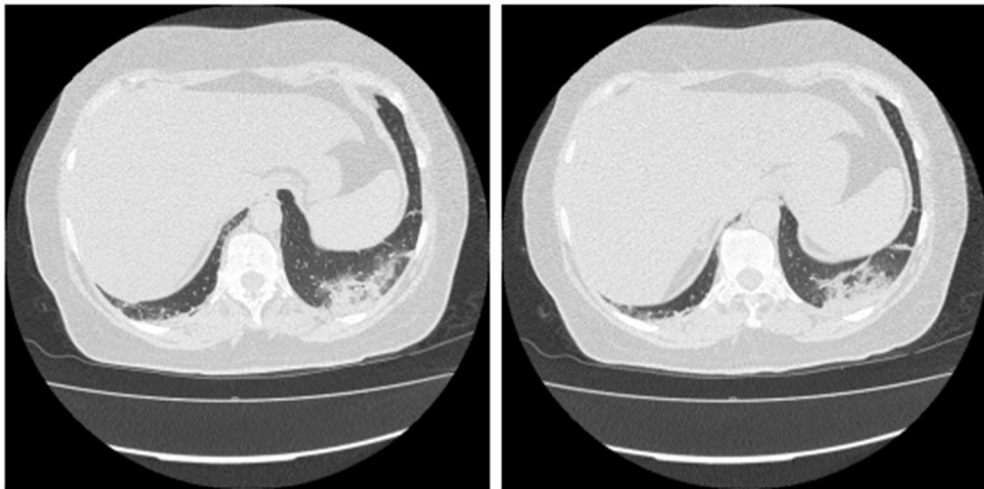


FIGURA 14

Questo risultato ha implicazioni significative per la spiegazione e l'uso pratico del metodo. Gli specialisti medici possono ora concentrare la loro attenzione su un sottoinsieme ridotto di fette (ad esempio, 30 fette) anziché navigare attraverso la totalità di un esame TC multistrato. Pur mantenendo l'essenza di una valutazione multistrato, questo approccio è semplificato per concentrarsi su un numero limitato di fette. Tuttavia, questo risultato può solo evidenziare gli elementi cruciali rilevati dalla

diagnosi automatica. Un'interpretazione approfondita e completa delle condizioni polmonari rimane compito del medico.

La radiomica ha guadagnato una significativa popolarità in vari settori della radiologia e della medicina nucleare negli ultimi anni. Consente l'estrazione di dati diagnostici, prognostici e completi da immagini radiologiche, spaziando dalle applicazioni in oncologia alla medicina forense^[3].

Sebbene la radiomica offra la promessa della medicina personalizzata, presenta anche sfide, principalmente dovute alla mancanza di standardizzazione e alla necessità di stabilire le connessioni tra le caratteristiche radiomiche e le conclusioni radiologiche visive^[72].

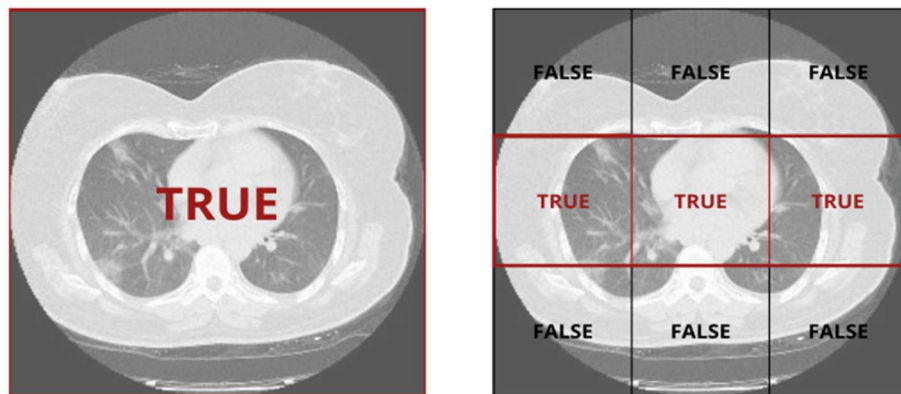
Sforzi per stabilire uno standard all'interno della comunità scientifica sono in corso, con la speranza di colmare il divario tra la radiologia di routine e la radiomica in un prossimo futuro. Una delle sfide principali negli studi radiomici è definire la ROI (Region of Interest). Quando si analizzano tumori o lesioni, la segmentazione è relativamente diretta. Tuttavia, per prevedere condizioni come le pneumopatie polmonari (es. polmoniti), può essere meno chiaro quali parti dei polmoni segmentare.

Questo processo di segmentazione è sia lungo che intensivo in termini di risorse^[73].

Per affrontare questi problemi, un approccio è utilizzare l'intera immagine come ROI per garantire che non vengano tralasciati dettagli critici dell'immagine; ovviamente questo rende di ancor più difficile comprensione i processi di intelligenza artificiale grazie ai quali si ottiene una determinata risposta al punto di portare i medici a definire l'AI come "scatole nere"^[61]. Questa mancanza di trasparenza potrebbe non essere accettabile in campo medico, dove capire la logica dietro una diagnosi o la prognosi è cruciale.

In risposta, è stato sviluppato un nuovo metodo automatizzato per la definizione della ROI su ogni immagine radiologica. Applicando una griglia all'immagine e utilizzando i valori radiomici di ogni cella della griglia è possibile creare un modello per fornire diagnosi in modo più robusto e interpretabile. Questa metodologia basata su matrice semplifica permette una transizione dall'esame radiologico al modello più chiara fornendo una classificazione e localizzazione immediata: un vantaggio prezioso per i radiologi e i loro colleghi.

Si è quindi provato a segmentare le immagini per valutare i risultati ottenuti e confrontarli sia con i risultati ad immagine intera che con le diagnosi fatte dai radiologi.



Il dataset è composto da 40 pazienti, di cui 31 affetti da COVID-19 e 9 in buona salute. Gli esami sono stati acquisiti con TC Canon Aquilion One utilizzando il protocollo HRCT.

Inizialmente, partendo dalle caratteristiche radiomiche selezionate della Tabella 5,

FIRST	Total Energy, 90 Percentile, Mean Absolute Deviation;
GLCM	Sum Average, Autocorrelation, Contrast, Joint Average;
GLDM	Large Dependence High Gray Level Emphasis, Dependence Non Uniformity, Gray Level Non Uniformity, High Gray Level Emphasis;
GLRLM	Run Length Non Uniformity, Gray Level Non Uniformity, High Gray Level Run Length Non Uniformity, High Gray Level Emphasis;
GLSZM	Small Area High Gray Level Emphasis, Gray Level Non Uniformity, High Gray Level Emphasis, Large Area Low Gray Level Emphasis;
NGTDM	Strength, Complexity.

TABELLA 5

abbiamo applicato la ROI all'intera immagine, ottenendo buoni risultati nella classificazione di un paziente affetto da COVID-19, come mostrato nella Tabella 6.

Radiomic class	Accuracy	Precision	Recall
GLDM	0.75	0.83	0.83
GLRLM	0.77	0.86	0.83
NGTDM	0.75	0.78	0.93

TABELLA 6

In questo caso, il Model Checker restituisce un singolo 'TRUE' o 'FALSE' per ciascun paziente.

Successivamente, abbiamo ripetuto il processo di verifica, mantenendo le stesse caratteristiche radiomiche selezionate dettagliate nella Tabella 6, ma costruendo questa volta una matrice ROI 3x3 sulle immagini DICOM.

Durante il processo di verifica, è stata creata una matrice, con ogni cella contenente il risultato di classificazione per la ROI corrispondente. Il Model Checker ha fornito un valore per ogni cella nella matrice, restituendo 'TRUE' se la proprietà era soddisfatta nella cella e 'FALSE' se non lo era. Alla fine, se la matrice per un singolo paziente conteneva almeno tre valori 'TRUE' consecutivi a una distanza di 1 (sia

orizzontalmente, verticalmente o diagonalmente), il paziente veniva classificato come 'TRUE' per la diagnosi finale.

Attraverso la nostra analisi, abbiamo determinato che tre quadrati all'interno della matrice coprivano sufficientemente la regione per valutare la presenza della malattia. La Tabella 7 presenta i risultati delle prestazioni per ciascuna classe radiomica.

Radiomic class	Accuracy	Precision	Recall
GLDM	0.62	0.75	0.77
GLRLM	0.90	0.96	0.90
NGTDM	0.77	0.77	1

TABELLA 7

Come osservato, in questo particolare caso della malattia da COVID-19, l'utilizzo di una griglia sull'immagine produce risultati positivi quando si utilizzano classi come NGTDM e GLRLM. Al contrario, si osserva una diminuzione delle prestazioni per la classe GLDM. Ciò suggerisce che le caratteristiche GLDM potrebbero funzionare meglio quando applicate a porzioni di immagine più grandi, probabilmente con ROI più grandi.

Questa metodologia innovativa non solo consente la classificazione, ma anche la localizzazione e la visualizzazione simultanea. La matrice di risultati appena generata può essere sovrapposta al DICOM di HRCT del

paziente. Ciò facilita la localizzazione della malattia all'interno delle immagini radiologiche dei pazienti.

L'algoritmo rivela che 2 delle 3 classi radiomiche hanno mostrato miglioramenti negli indici di prestazione, mentre solo la classe GLDM sembra aver subito un leggero peggioramento. Questo potrebbe essere attribuito alla dipendenza dell'insieme totale dei voxel rispetto a un voxel centrale. Con ROI più piccole, queste dipendenze potrebbero diventare più flebili, portando a migliori errori di classificazione. Un altro fattore che contribuisce a questo deterioramento potrebbe essere legato alle caratteristiche all'interno della classe GLDM. La selezione delle caratteristiche è stata basata sull'intera ROI, dove le variazioni di intensità sono più pronunciate.

Studi futuri avranno l'obiettivo di fare luce su questi risultati correlandoli con prove cliniche per migliorare l'interpretabilità.

L'obiettivo di questa ricerca è stato impiegare e validare un nuovo metodo per distinguere tra pazienti affetti da COVID-19 e individui sani

analizzando le caratteristiche radiomiche estratte da immagini TC.

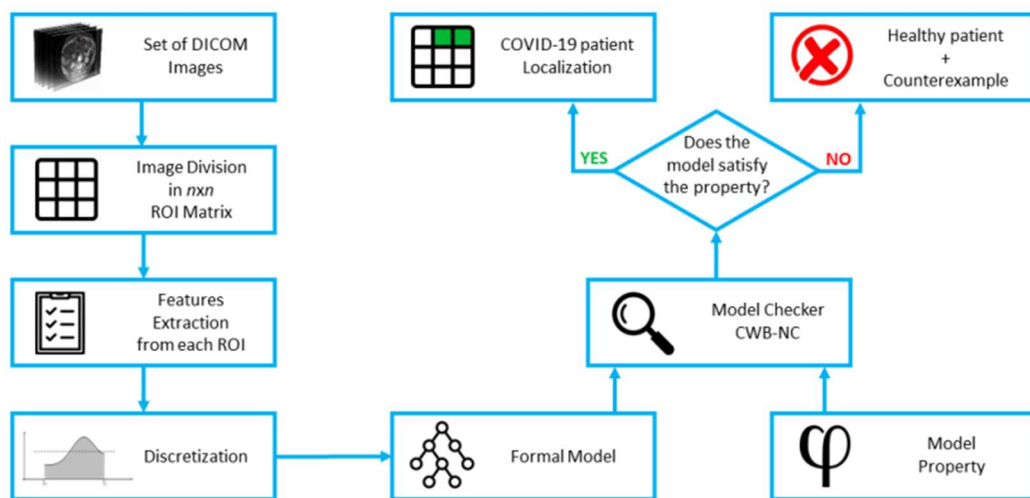


FIGURA 15

La Figura 15 illustra il percorso della metodologia proposta. Essa inizia con un database contenente immagini radiologiche di soggetti sani e pazienti affetti da COVID-19, principalmente costituite da immagini DICOM acquisite attraverso scansioni HRCT dei polmoni dei pazienti.

La fase iniziale di segmentazione prevede la divisione di ciascuna immagine DICOM in “n” sezioni equivalenti, risultando in una matrice delle dimensioni “ $n \times n$ ” che comprende regioni di interesse (ROIs) distinte. L'obiettivo principale di questa segmentazione è individuare la presenza o l'assenza della malattia in regioni specifiche. Questo approccio localizzato mira a migliorare l'accuratezza del metodo proposto e ad assistere i medici nella valutazione delle immagini. Una volta estratte le caratteristiche radiomiche, il passo successivo si discosta

dagli approcci convenzionali impiegando una traduzione automatica dei valori delle caratteristiche ottenute in un modello matematico durante la costruzione dei modelli formali dei pazienti. Collaborando con esperti di informatica, si effettua un confronto tra i modelli per identificare le correlazioni con la pratica clinica. Gli esperti valutano quali valori delle caratteristiche differenziano efficacemente tra soggetti sani e pazienti affetti da COVID-19, e queste informazioni sono riassunte in un modello formale denominato "proprietà". Infine un Model Checker viene impiegato per verificare le ipotesi e i criteri delineati nella proprietà, determinando se il paziente è effettivamente affetto dalla malattia o gode di buona salute. Nella valutazione dei risultati in questo studio, abbiamo condotto un'analisi comparativa tra i risultati ottenuti utilizzando una ROI che comprende l'intera immagine e una diagnosi ottenuta attraverso il nuovo algoritmo di segmentazione.

La definizione di una ROI implica l'identificazione e la delimitazione di una specifica regione all'interno di un'immagine. Nell'ambito delle immagini mediche, le ROI sono fondamentali per condurre analisi localizzate, fornendo valori delle caratteristiche più precisi e, in ultima analisi, migliorando l'accuratezza del modello. Il concetto di una ROI può essere paragonato al concetto geometrico di scatole di delimitazione. Dato un insieme di punti in uno spazio N-dimensionale, una scatola di

delimitazione è l'incapsulamento più piccolo (in termini di area o volume) che contiene tutti i punti. Per implementare la creazione di scatole di delimitazione, è stato utilizzato un metodo Python chiamato "Quokka". Quokka ha facilitato la generazione di rettangoli allineati agli assi sulle immagini, contribuendo alla comprensione dei sistemi di coordinate. Questa comprensione è stata cruciale per determinare le coordinate necessarie per la segmentazione delle immagini DICOM e ottenere le 9 ROI, risultanti in una matrice 3x3.

La segmentazione di una ROI all'interno di un'immagine DICOM basata sulle coordinate può essere un compito complesso a causa delle complicazioni della struttura del file DICOM. Le immagini DICOM sono orientate con l'asse Y rivolto verso il basso, come rappresentato nella Figura 16.

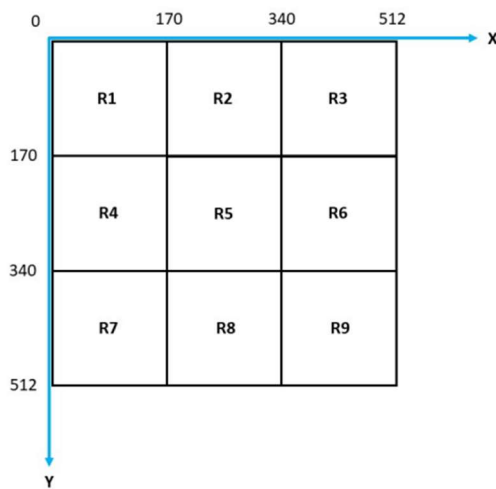


FIGURA 16

Di conseguenza, ogni quadrato viene creato considerando le coordinate minime (angolo superiore sinistro della ROI) e le coordinate massime (angolo inferiore destro della ROI).

Le immagini DICOM utilizzate nell'analisi hanno dimensioni di 512x512 pixel. Di conseguenza, la matrice, essendo 3x3, suddivide queste immagini in segmenti con valori di 0, 170, 340 e 512. Ad esempio, nel caso di R1, la prima ROI, le coordinate minime sarebbero $x = 0$ e $y = 0$, mentre le coordinate massime sarebbero $X = 170$ e $Y = 170$. Dopo aver definito la struttura della matrice sulle immagini DICOM, il passo successivo è stato estrarre le caratteristiche per ciascuna ROI all'interno della matrice. Questo è stato ottenuto utilizzando uno script Python che sfruttava le librerie Pyradiomics. L'output di questo processo era un file contenente tutte le caratteristiche estratte e i loro valori corrispondenti per la ROI selezionata (ad esempio, R1) in ciascuna immagine della serie DICOM.

Con questi dati a disposizione e associando tutte le ROI (ad esempio, R1) a ciascuna immagine, è stato possibile creare un modello formale e procedere a verificare la presenza della malattia in quell'area specifica dell'organo. Questo approccio ha permesso la localizzazione della malattia all'interno delle immagini radiologiche.

4 ALTRI LAVORI E SVILUPPI FUTURI

4.1 LAVORI PUBBLICATI

- Transcatheter Arterial Embolization for Bleeding Related to Pelvic Trauma: Comparison of Technical and Clinical Results between Hemodynamically Stable and Unstable Patients

Minici, R., Mercurio, M., Guzzardi, G., ...Galasso, O., Laganà, D.

Tomography, 2023, 9(5), pp. 1660–1682

- Endovascular Treatment of Visceral Artery Pseudoaneurysms with Ethylene-Vinyl Alcohol (EVOH) Copolymer-Based Non-Adhesive Liquid Embolic Agents (NALEAs)

Minici, R., Guerriero, P., Fontana, F., ...Brunese, L., Laganà, D.

Medicina (Lithuania), 2023, 59(9), 1606

- Transcatheter Arterial Embolization (TAE) of Cancer-Related Bleeding

Minici, R., Guzzardi, G., Venturini, M., ...Brunese, L., Laganà, D.

Medicina (Lithuania), 2023, 59(7), 1323

- A Multicenter Retrospective Cohort Study Evaluating the Clinical Outcomes of Patients with Coagulopathy Undergoing Transcatheter Arterial Embolization (TAE) for Acute Non-Neurovascular Bleeding

Minici, R., Fontana, F., Venturini, M., ...Brunese, L., Laganà, D.

Medicina (Lithuania), 2023, 59(7), 1333

- Transcatheter Arterial Embolization (TAE) in the Management of Bleeding in the COVID-19 Patient

Minici, R., Fontana, F., Venturini, M., ...Brunese, L., Guerriero, P.

Medicina (Lithuania), 2023, 59(6), 1062

4.2 POSSIBILITA' FUTURE

Come è intuibile dagli articoli pubblicati, la mia attività di ricerca si è concentrata anche sulla radiologia interventistica vascolare che sembrerebbe aver poco a che fare con la radiomica e la radiologia diagnostica.

Il mio obiettivo è invece trovare applicazioni di radiomica anche nella radiologia interventistica.

Se ci soffermiamo a pensare a quanto potrebbe essere utile avere il percorso vascolare già tracciato dall'AI per l'interventista che approccia un intervento, con delle "road map" possibilmente sviluppate su TC con mdc diagnostiche e non durante l'intervento.

Oppure ci si potrebbe avvalere dell'AI per il corretto sizing dei dispositivi da utilizzare in fase operatoria o da impiantare ai pazienti.

Non trascuriamo infine la possibilità di stabilire, prima dell'intervento, come va embolizzato un vaso che sanguina, uno pseudoaneurisma o semplicemente un varicocele. Ad oggi l'interventista in corso di intervento decide se utilizzare o meno "coil metalliche" e quante e di che dimensioni; se utilizzare solo colla chirurgica; quanto spongostan iniettare. Certo, esistono dei valori di riferimento, ma l'ambizione della mia ricerca sarà standardizzare la procedura e fornire gli strumenti per un software di AI che possa accompagnare l'interventista nelle scelte dei metodi e materiali da utilizzare, quantomeno per avere tutto il materiale necessario già pronto all'uso, evitando complicazioni intraoperatorie.

BIBLIOGRAFIA

1. *Philippe Lambin, Emmanuel Rios-Velazquez e Ralph Leijenaar, Radiomics: Extracting more information from medical images using advanced feature analysis, in European journal of cancer (Oxford, England : 1990), vol. 48, n. 4, 2012-3, pp. 441–446*
2. Virendra Kumar, Yuhua Gu e Satrajit Basu, QIN “Radiomics: The Process and the Challenges”, in *Magnetic resonance imaging*, vol. 30, n. 9, 2012-11, pp. 1234–1248.
3. Robert J. Gillies, Paul E. Kinahan e Hedvig Hricak, *Radiomics: Images Are More than Pictures, They Are Data*, in *Radiology*, vol. 278, n. 2, 2016-2, pp. 563–577.
4. M. Amadasun e R. King, *Textural features corresponding to textural properties*, in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, n. 5, 1989, pp. 1264–1274.
5. Depeursinge, Adrien., Al-kadi, Omar S. e Mitchell, J. Ross., *Biomedical texture analysis : fundamentals, tools*

and challenges, Academic Press, 2017, ISBN 9780128123218, OCLC 1002195690

6. Radhika Sivaramakrishna, Kimerly A. Powell e Michael L. Lieber, Texture analysis of lesions in breast ultrasound images, in *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society*, vol. 26, n. 5, 2002-9, pp. 303–307.
7. Sean D. McGarry, Sarah L. Hurrell e Amy L. Kaczmarowski, Magnetic Resonance Imaging-Based Radiomic Profiles Predict Patient Prognosis in Newly Diagnosed Glioblastoma Before Therapy, in *Tomography : a journal for imaging research*, vol. 2, n. 3, 2016-9, pp. 223–228.
8. Elizabeth J. Sutton, Erich P. Huang e Karen Drukker, Breast MRI radiomics: comparison of computer- and human-extracted imaging phenotypes, in *European Radiology Experimental*, vol. 1, n. 1, 2017.

9. Gary J. R. Cook, Musib Siddique e Benjamin P. Taylor, Radiomics in PET: principles and applications, in *Clinical and Translational Imaging*, vol. 2, n. 3, 2014-06, pp. 269–276.
10. Vishwa Parekh e Michael A. Jacobs, Radiomics: a new application from established techniques, in *Expert review of precision medicine and drug development*, vol. 1, n. 2, 2016, pp. 207–226.
11. <https://www.tecnicaospedaliera.it/radiomica-potente-strumento-a-disposizione-del-radiologo>
12. <https://www.fondazionesirm.org/radiomica>
13. Cisbani E, Dini V, Grande S, Palma A, Rosi A, Tabocchini MA, Gasparrini F, Orlacchio A. Stato dell'arte sull'impiego della diagnostica per immagini per

COVID-19. Versione del 7 luglio 2020. Roma: Istituto Superiore di Sanità; 2020. (Rapporto ISS COVID-19 n. 55/2020)

14.

Liu Y, Chen P-HC, Krause J et al., How to read articles that use machine learning users' guides to the medical literature. JAMA 2019;322(18):1806-16.

15.

Doshi A, and Shah AS. The role of artificial intelligence and technology in lung disease detection. EC Pulmonology and Respiratory Medicine 2019;8.11:24-8.
<https://www.econicon.com/ecprm/pdf/ECPRM-08-00489.pdf>

16.

Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, SMC-3(6):610–621, 1973.

17. Kenneth I Laws. Texture energy measures. In Proc. Image understanding workshop, pages 47–51, 1979.
18. P Grossmann, O Grove, N El-Hachem, et al. Identification of molecular phenotypes in lung cancer by integrating radiomics and genomics. *Sci Transl Med*, 2017.
19. Bai HX, Wang R, Xiong Z, et al. Augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other etiology on chest CT. *Radiology* 2020, Apr 27.
20. Zhang K, Liu X, Zhihuan L, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements and prognosis of COVID-19 pneumonia using Computed Tomography. *Cell* 2020; 181:1423-33.
21. Mario Coriasco, Nello Balossino, Osvaldo Rampado, and Sergio Rabellino. L'immagine digitale in

diagnostica per immagini. Springer, 2013.

22. Mathieu Hatt, Martin Vallieres, Dimitris Visvikis, and Alex Zwanenburg. Ibsi: an international community radiomics standardization initiative, 2018.
23. Joost JM Van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion Robin, Steve Pieper, and Hugo JWL Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21):e104–e107, 2017.
24. Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging

- network. *Magnetic resonance imaging*, 30(9):1323–1341, 2012.
25. Gregory M. Nielson. On marching cubes. *IEEE Transactions on visualization and computer graphics*, 9(3):283–297, 2003.
 26. Steven Salzberg. Distance metrics for instance-based learning. In *International Symposium on Methodologies for Intelligent Systems*, pages 399–408. Springer, 1991.
 27. Detlef F Sprinz and Yael Wolinsky-Nahmias. *Models, numbers, and cases: methods for studying international relations*. University of Michigan Pre
 28. Francesco Mercaldo, Rosangela Casolare, Giovanni Ciaramella, Giacomo Iadarola, Fabio Martinelli, Francesco Ranieri, and Antonella Santone. A realtime method for can bus intrusion detection by means of supervised machine learning. *Proceedings of the 19th*

International Conference on Security and Cryptography (SECRYPT 2022), 2022.

29. Luigi Coppolino, Roberto Nardone, Alfredo Petruolo, Luigi Romano, and Andrej Souvent. Exploiting digital twin technology for cybersecurity monitoring in smart grids. In Proceedings of the 18th International Conference on Availability, Reliability and Security, pages 1–10, 2023.
30. Robin Milner. A calculus of communicating systems. Springer, 1980.
31. Edmund M Clarke. Model checking. In Foundations of Software Technology and Theoretical Computer Science: 17th Conference Kharagpur, India, December 18–20, 1997 Proceedings 17, pages 54–56. Springer, 1997.
32. E Allen Emerson. Model checking and the mu-calculus. Descriptive Complexity and Finite Models, 31:185–214, 1996.

33. Rance Cleaveland, Joachim Parrow, and Bernhard Steffen. The concurrency workbench: A semantics-based tool for the verification of concurrent systems. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 15(1):36–72, 1993.
34. Alex J Mitchell. Sensitivity× ppv is a recognized test called the clinical utility index (cui+). *European journal of epidemiology*, 26:251–252, 2011.
35. Corinne Balleyguier, Salma Ayadi, Kim Van Nguyen, Daniel Vanel, Clarisse Dromain, and Robert Sigal. Birads™ classification in mammography. *European journal of radiology*, 61(2):192–194, 2007.
36. Pedicelli A, Lozupone E, Gatto A, Gulino P, D’Argento F, Capozzi A, et al. Vertebra plana: reappraisal of a contraindication to percutaneous vertebroplasty. *Eur J Radiol*. 2013;82:2303–8.
37. Kevin G Billingsley, Jonathan J Lewis, Denis HY Leung, Ephraim S Casper, James M Woodruff, and

- Murray F Brennan. Multifactorial analysis of the survival of patients with distant metastasis arising from primary extremity sarcoma. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 85(2):389–395, 1999.
38. Francesca Bagnato. Gadolinium should always be used to assess disease activity in ms—commentary. *Multiple Sclerosis Journal*, 26(7):769–770, 2020.
39. Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging*, 26(6):1045–1057, 2013
40. Barry J Bedell and Ponnada A Narayana. Automatic segmentation of gadolinium-enhanced multiple sclerosis lesions. *Magnetic resonance in medicine*, 39(6):935–940, 1998.

41. Edmund M Clarke. Model checking. In Foundations of Software Technology and Theoretical Computer Science: 17th Conference Kharagpur, India, December 18–20, 1997 Proceedings 17, pages 54–56. Springer, 1997.
42. Laurent Dercle, Samy Ammari, Mathilde Bateson, Paul Blanc Durand, Eva Haspinger, Christophe Massard, Cyril Jaudet, Andrea Varga, Eric Deutsch, Jean-Charles Soria, et al. Limits of radiomic-based entropy as a surrogate of tumor heterogeneity: Roi-area, acquisition protocol and tissue site exert substantial influence. Scientific reports, 7(1):1–10, 2017.
43. S. Gradara, A. Santone, and M.L. Villani. Delfin+: An efficient deadlock detection tool for ccs processes. Journal of Computer and System Sciences, 72(8):1397–1412, 2006.
44. Iman Beheshti, Daichi Sone, Norihide Maikusa, Yukio Kimura, Yoko Shigemoto, Noriko Sato, and

Hiroshi Matsuda. Flair-wise machine-learning classification and lateralization of mri-negative 18f-fdg pet-positive temporal lobe epilepsy. *Frontiers in neurology*, 11:580713, 2020.

45. PJ Brown, J Zhong, R Froot, S Currie, A Gilbert, AL Appelt, D Sebag Montefiore, and A Scarsbrook. Prediction of outcome in anal squamous cell carcinoma using radiomic feature analysis of pre-treatment fdg pet-ct. *European journal of nuclear medicine and molecular imaging*, 46(13):2790–2799, 2019.

46. Massimo Filippi, Paolo Preziosa, Brenda L Banwell, Frederik Barkhof, Olga Ciccarelli, Nicola De Stefano, Jeroen JG Geurts, Friedemann Paul, Daniel S Reich, Ahmed T Toosy, et al. Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. *Brain*, 142(7):1858–1875, 2019.

47. <https://wiki.cancerimagingarchive.net/>.

48. <https://www.slicer.org/>.

49. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RG, Fillion-Robin J-C, Pieper S, Aerts HJ (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77(21):e104–e107
50. Langerak R (1994) Transformations and semantics for LOTOS
51. Stirling C (1989) An introduction to modal and temporal logics for CCS. *Concurrency: theory language and architecture*. Springer, Berlin, pp 2–20
52. Fernandez J-C, Garavel H, Kerbrat A, Mounier L, Mateescu R, Sighireanu M (1996) CADP a protocol validation and verification toolbox. In: *International conference on computer aided verification*, Springer, pp 437–440
53. Sobecki P, Gora A, Zycka-Malesa D, Sklinda K, Mykhalevych I, Przelaskowski A (2017) Feature

extraction optimized for prostate lesion classification. vol
Part F128534, pp 22–27

54. Abele T, Besachio D, Quigley E, Gurgel R, Shelton C, Harnsberger H, Wiggins R (2014) Diagnostic accuracy of screening MR imaging using unenhanced axial ciss and coronal t2wi for detection of small internal auditory canal lesions. *Am J Neuroradiol* 35(12):2366–2370
55. Lecouvet F (2016) Whole-body MR imaging: musculoskeletal applications. *Radiology* 279(2):345–365
56. Luca Brunese, Fabio Martinelli, Francesco Mercaldo, and Antonella Santone. Machine learning for coronavirus covid-19 detection from chest x-rays. *Procedia Computer Science*, 176:2212–2221, 2020.
57. Javier Civit-Masot, Francisco Luna-Perejón, Manuel Domínguez Morales, and Anton Civit. Deep learning system for covid-19 diagnosis aid using x-ray pulmonary images. *Applied Sciences*, 10(13):4640,

58. Sivaramakrishnan Rajaraman and Sameer Antani. Weakly labeled data augmentation for deep learning: a study on covid-19 detection in chest x-rays. *Diagnostics*, 10(6):358, 2020.
59. Pengyi Zhang, Yunxin Zhong, Yulin Deng, Xiaoying Tang, and Xiaoqiong Li. Cosingan: learning covid-19 infection segmentation from a single radiological image. *Diagnostics*, 10(11):901, 2020.
60. Antonella Santone, Gigliola Vaglini, and Maria Luisa Villani. Incremental construction of systems: An efficient characterization of the lacking sub-system. *Science of Computer Programming*, 78(9):1346–1367, 2013.
61. Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):93, 2018.

62. Mathieu Hatt, Martin Vallieres, Dimitris Visvikis, and Alex Zwanenburg. Ibsi: an international community radiomics standardization initiative, 2018.
63. Susan Zannos. Godfrey Hounsfield and the Invention of Cat Scans (Unlocking the Secrets of Science). Mitchell Lane Pub Inc, August 1, 2002.
64. Santone, A., Belfiore, M. P., Mercaldo, F., Varriano, G., and Brunese, L. (2021). On the adoption of radiomics and formal methods for covid-19 coronavirus diagnosis. *Diagnostics*, 11(2), 293.
65. Janez Demšar, Tomaž Curk, Aleš Erjavec, Crt Gorup, Tomaž Hočcevar, Mitar Š Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Stajdohar, Lan Umek, Lan Š Zagar, Jure Š Zbontar, Marinka Š Zitnik, and Blaž Š Zupan. Orange: Data mining toolbox in python. *Journal of Machine Learning Research*, 14:2349–2353, 2013.

66. Alessandro Stefano, Mauro Gioè, Giorgio Russo, Stefano Palmucci, Sebastiano Emanuele Torrisi, Samuel Bignardi, Antonio Basile, Albert Comelli, Viviana Benfante, Gianluca Sambataro, et al. Performance of radiomics features in the quantification of idiopathic pulmonary fibrosis from hrct. *Diagnostics*, 10(5):306, 2020.
67. Mario Mascalchi, Gianna Camiciottoli, and Stefano Diciotti. Lung densitometry: why, how and when. *Journal of thora*
68. Ian H Witten, Eibe Frank, Mark A Hall, CJ Pal, and MINING DATA. Practical machine learning tools and techniques. In *DATA MINING*, volume 2, page 4, 2005.
69. Isabella Fornacon-Wood, Hitesh Mistry, Christoph J Ackermann, Fiona Blackhall, Andrew McPartlin, Corinne Faivre-Finn, Gareth J Price, and James PB O'Connor. Reliability and prognostic value of radiomic features are highly dependent on choice of feature

extraction platform. *European radiology*, 30:6241– 6250, 2020.

70. L Moura, C Dartora, C Mattjie, R Barros, and Ana Maria Marques da Silva. Texture-based feature extraction for covid-19 pneumonia classification using chest radiography. *EAI Endorsed Transactions on Bioengineering and Bioinformatics*, 2021.
71. Qingxia Wu, Shuo Wang, Liang Li, Qingxia Wu, Wei Qian, Yahua Hu, Li Li, Xuezhi Zhou, He Ma, Hongjun Li, et al. Radiomics analysis of computed tomography helps predict poor prognostic outcome in covid-19. *Theranostics*, 10(16):7231, 2020.
72. Van Timmeren, Davide Cester, Stephanie Tanadini-Lang, Hatem Alkadhi, and Bettina Baessler. Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights into imaging*, 11(1):1–16, 2020.
73. Francesco Rizzetto, Francesca Calderoni, Cristina De Mattia, Arianna Defeudis, Valentina Giannini,

Simone Mazzetti, Lorenzo Vassallo, Silvia Ghezzi, Andrea Sartore-Bianchi, Silvia Marsoni, et al. Impact of inter-reader contouring variability on textural radiomics of colorectal liver metastases. *European radiology experimental*, 4(1):1–12, 2020.