Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Gestionale

Tesi di Laurea Magistrale

Anno accademico 2024/2025
Sessione di Laurea 22 luglio 2025

Intelligenza artificiale in campo medicale

Relatori:
Prof.ssa Francesca Montagna

Candidato:
Riccardo Roncato

CAPITOLO 1: DALLE ORIGINI TEORICHE AI BISOGNI DELLA PRATICA CLINICA..... 3

Obiettivi della Tesi 3

I FONDAMENTI DELLE INNOVAZIONI E DELLE TECNOLOGIE AI IN MEDICINA 5

Il machine learning (ML) 5

L'apprendimento profondo (DL)..... 6

L'Elaborazione del Linguaggio Naturale (NLP) 6

I BISOGNI NEL SETTORE MEDICALE..... 7

LE APPLICAZIONI DI APPROCCI INTEGRATI IN MEDICINA 8

CAPITOLO 2: APPLICAZIONI DELL'IA NELLA PRATICA CLINICA: CASI STUDIO..... 10

2.1 Analisi Caso Studio: Diagnosi Basata su Intelligenza Artificiale per Il Melanoma 10

2.2 Analisi Caso Studio: screening del cancro al seno 19

2.3 intervista a Amie J. Miller, AI e prevenzione del tumore al polmone.....28

2.4 Analisi Caso Studio: large language model 31

2.5 Analisi Caso Studio: screening DNA.....45

CAPITOLO 3: ANALISI DEI COSTI E INDICI DI PERFORMANCE 51

3.1 Melanomi.....51

3.2 Radiologia 53

3.3 Screening.....55

3.4 NLP 57

3.5 indicatori di performance.....58

CONCLUSIONI E STRATEGIE:..... 66

Strategie di integrazione (previsione 2025-2030).....66

Il quadro regolatorio europeo per l'adozione dell'intelligenza artificiale in sanità.....68

Conclusioni finali.....69

CAPITOLO 1: DALLE ORIGINI TEORICHE AI BISOGNI DELLA PRATICA CLINICA

L’intelligenza artificiale (IA) è la disciplina che si occupa di progettare sistemi capaci di eseguire compiti che, se svolti da un essere umano, richiederebbero un certo livello di “intelligenza” o ragionamento. Questi compiti includono il riconoscimento di pattern, l’elaborazione del linguaggio naturale e la capacità decisionale autonoma, che simula per l’appunto, il processo pensante di un’intelligenza umana. Le origini teoriche dell’IA risalgono al 1943, quando Warren McCulloch e Walter Pitts proposero un modello matematico di neurone artificiale, gettando le basi per le reti neurali. Nel 1950, Alan Turing introdusse il celebre "Test di Turing", spostando l’attenzione dal “pensare” al “comportarsi”: una macchina è intelligente se il suo comportamento, in un’interazione testuale, è indistinguibile da quello umano. Il termine “intelligenza artificiale” fu coniato da John McCarthy nel 1956 durante il Dartmouth Summer Research Project on Artificial Intelligence, evento considerato la nascita ufficiale del campo. Oggi l’IA permea settori come la sanità (diagnosi assistita), la finanza (trading algoritmico), l’industria 4.0 (manifattura intelligente) e i servizi digitali (chatbot, assistenti vocali). Tuttavia, l’adozione su larga scala solleva sfide etiche come bias nei dati, trasparenza degli algoritmi e normative. L’Unione Europea sta definendo l’“AI Act” per garantire un equilibrio fra innovazione e tutela dei diritti fondamentali. Come ogni innovazione vi sono altre criticità riguardanti la sostenibilità energetica, dovuta all’enorme consumo dei data center, e la ricerca di modelli più efficienti. L’evoluzione dell’IA riflette un percorso di alternanza fra sperimentazione teorica, applicazioni pratiche e revisione critica degli obiettivi. Lo sviluppo futuro dipenderà non solo dalle innovazioni tecniche, ma dalla nostra capacità di integrare l’IA in modo etico, responsabile e sostenibile nella società

Obiettivi della Tesi

Questa tesi indaga l'impatto delle tecnologie di intelligenza artificiale (IA) dal punto di vista tecnologico e organizzativo nel campo medico, riguardo alle implicazioni su come innovarlo.

L'obiettivo fondamentale è valutare in che misura la tecnologia digitale dell'IA, potrebbe modificare notevolmente l'organizzazione dei processi clinici e la qualità delle cure fornite in termini di diagnosi, individualizzazione del trattamento e screening delle malattie. Allo stesso tempo, si vuole esplorare come la potenza dell'IA possa essere messa in pratica in tre domini chiave, tramite casi studio dettagliati:

• Diagnostica per Immagini: Nuove soluzioni con diagnosi delle patologie critiche (melanomi e radiografie) molto più rapide e precise basate su algoritmi di deep learning. Tali studi sottolineano come la tecnologia possa aiutare e talvolta superare il lavoro dei medici specializzati, portando una diagnosi più rapida e precisa.

• Personalizzazione del Trattamento: Analizzando dati clinici e genetici di un individuo, gli algoritmi di IA sono addestrati in modo tale da aiutare a sviluppare un piano di trattamento che funzioni per il paziente specifico. In questa sezione, verranno descritti esempi di come queste tecnologie consentono decisioni cliniche personalizzate che possono ottimizzare i trattamenti con maggiore efficacia e ridurre così i loro effetti collaterali.

• Screening Genomico: Infine, saranno esaminati strumenti diagnostici, che utilizzano algoritmi avanzati di IA e Machine Learning, per l'analisi del DNA libero circolante (cfDNA). In questa sezione, emergerà come queste tecnologie stiano rivoluzionando la diagnostica non invasiva, migliorando la rilevazione precoce del cancro, il monitoraggio fetale e la gestione post-trapianto.

In questo modo, la tesi mira a costruire una visione del prossimo futuro dell'intelligenza artificiale in campo medicale in modo tale da evidenziare, da una parte le ottime opportunità che si presentano con quest'innovazione e dall'altra i problemi relativi alla fattibilità dell'implementazione dell'IA nel sistema sanitario in un quadro integrato, ponendo la prospettiva trasversale sui diversi campi coinvolti, della sicurezza tecnologica, clinica e organizzativa. L'obiettivo sarebbe dunque, quello di porre maggiore attenzione ad argomenti che ritengo siano sensibili e di molta importanza sulla prevenzione con un l'ausilio di questa tecnologia emergente e dirompente, che migliora distintamente l'efficienza operativa e il risultato clinico, portando vantaggi significativi sia ai pazienti che ai medici.

I fondamenti delle Innovazioni e delle Tecnologie AI in Medicina

L'intelligenza artificiale (IA) è una disciplina che si occupa di creare sistemi capaci di compiere compiti che, tradizionalmente, richiedono l'intelligenza umana, come il riconoscimento di pattern, l'interpretazione di linguaggio e la presa di decisioni. In sostanza, l'IA si basa sull'uso di algoritmi che, attraverso l'elaborazione di enormi quantità di dati, imparano a identificare correlazioni e a fare previsioni. Negli anni '50 e '60 nacquero le prime teorie e i primi modelli, mentre il vero salto qualitativo si ebbe con l'avvento del Machine Learning, e successivamente del Deep Learning, che ha permesso di sviluppare reti neurali artificiali in grado di analizzare immagini, testi e suoni con una precisione sorprendente. Allo stesso tempo, il Natural Language Processing (NLP) ha reso possibile l'interpretazione automatica del linguaggio naturale, trasformando il modo in cui interagiamo con le macchine. ma quindi come funzionano queste tecnologie? Nella figura 1 (affianco) ci mostra un quadro generale di come si suddividono le varie tecnologie:

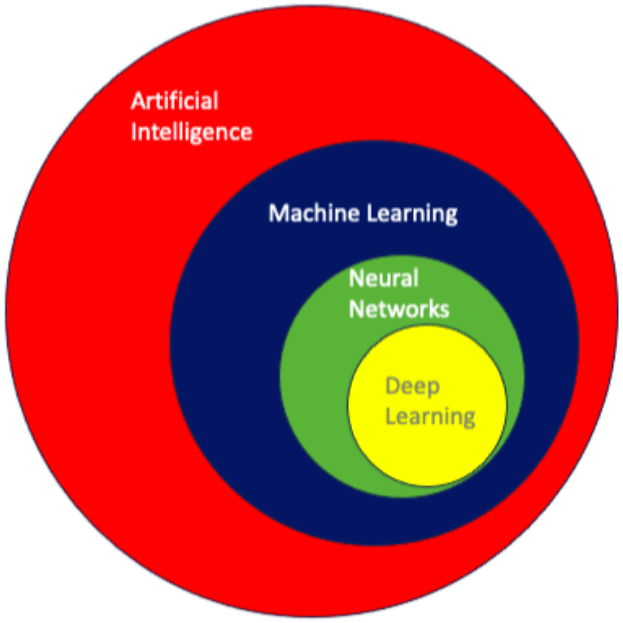


Figura 1 grafico delle tecnologie di intelligenza artificiale

Il machine learning (ML)

Il Machine Learning rappresenta il cuore dell'IA, basato sulla capacità dei sistemi di apprendere e migliorare automaticamente attraverso l'esperienza. Gli algoritmi di Machine Learning (ML), come regressioni, alberi decisionali e macchine a supporto vettoriale (SVM), vengono addestrati su dataset eterogenei con l'obiettivo di trovare correlazioni precedentemente sconosciute. Dividendo i dati in training e test, l'algoritmo apprende alcune caratteristiche dei dati e poi effettua le validazioni per misurare l'accuratezza dell'algoritmo. In medicina, il ML è in grado di analizzare i dati clinici e genomici: ad esempio, i profili genetici vengono analizzati per lo screening del

DNA (in questo caso, il ML viene utilizzato per identificare varianti associate a una specifica patologia, il che può poi portare a diagnosi precoci e misure preventive).

L'apprendimento profondo (DL)

Il Deep Learning è un tipo di machine learning più avanzato che utilizza ulteriori strati in una rete neurale[1] per gestire tipi di dati più complessi nel processo di machine learning. Tali reti sono conosciute come reti neurali convoluzionali (CNN) e sono progettate per riconoscere pattern di dati di immagini, e sono estremamente efficaci nel riconoscimento di schemi visivi. Le CNN apprendono le caratteristiche rilevanti dai dati passando attraverso i livelli: i primi strati rilevano bordi e forme, mentre quelli più profondi identificano strutture complesse e specifiche come anomalie nelle immagini radiologiche o nei campioni dermoscopici. In medicina, il DL ha rivoluzionato l'imaging, consentendo di analizzare radiografie, TAC e risonanze magnetiche con una precisione che spesso supera quella dell'occhio umano. Ad esempio, gli algoritmi basati su CNN vengono utilizzati per la mappatura dei nei e il riconoscimento precoce del melanoma, migliorando la diagnosi e riducendo il rischio di errori dovuti a variabilità inter-operatoria.

L'Elaborazione del Linguaggio Naturale (NLP)

L'Elaborazione del Linguaggio Naturale (NLP) è la tecnologia che consente alle macchine di comprendere il linguaggio umano. Originariamente il PNL si basava su regole linguistiche e schemi, ma successivamente ci sono stati significativi avanzamenti con l'uso di modelli statistici e reti neurali come GPT-3 e BERT hanno ora reso la NLP ampiamente praticabile. Questi modelli analizzano vasti quantitativi di testo non strutturato (cartelle cliniche, letteratura medica, ecc.), analizzando dettagli salienti ed evidenziando correlazioni rilevanti per abilitare il supporto

[1] Le reti neurali nel deep learning sono modelli matematici ispirati al funzionamento del cervello umano. Sono composte da strati di neuroni artificiali, che elaborano le informazioni in modo simile ai neuroni biologici. Le reti possono avere molti strati nascosti, ed è proprio questa profondità che le rende "deep" (profonde). Grazie a questo sistema, le reti neurali sono in grado di imparare da grandi quantità di dati ed eseguire compiti complessi, come riconoscere immagini, comprendere il linguaggio o fare previsioni.

decisionale clinico. In medicina, il PNL offre la capacità di accedere a informazioni utili registrate in documenti narrativi clinici non strutturati, consentendo un rapido recupero di dati rilevanti circa diagnosi, trattamenti precedenti, e dati di anamnesi che possono essere combinati con altri set di dati (cioè immagini, dati genomici) per una panoramica completa dello stato di salute del paziente.

I bisogni nel settore medicale

Negli ultimi anni, il settore sanitario si è trovato a dover far fronte a una serie di bisogni critici e sistematici, ovvero, in primo luogo, la necessità di diagnosi sempre più accurate e tempestive, che derivano dall'aumento costante dei casi di malattie cronico-degenerative e oncologiche. Il punto di interesse in questi casi è proprio quello di rilevare anomalie in fase iniziale per ridurre mortalità e disabilità, ma richiede capacità di analisi di enormi moli di dati clinici e immagini medicali ben oltre le possibilità del solo intervento umano. Parallelamente, cresce il bisogno di personalizzare le terapie, poichè c'è la questione della variabilità biologica fra pazienti e della composizione genomica alla risposta immunitaria che richiede modelli predittivi capaci di integrare dati molecolari, stilare protocolli su misura e adattarli in tempo reale in base agli esiti del trattamento. Senza strumenti di IA, la medicina rischia ancora di basarsi su approcci "one-size-fits-all", con conseguente aumento di effetti collaterali e sprechi di risorse. Il mondo della ricerca farmacologica, poi, reclama con urgenza accelerazione nei processi di scoperta e validazione dei nuovi composti. I lunghi cicli di screening e sperimentazione non solo gravano sui bilanci delle aziende, ma ritardano l'arrivo sul mercato di cure innovative; l'IA risponde con simulazioni virtuali e predizioni in silico, abbattendo tempi e costi e riducendo la necessità di test animali. Un'altra necessità stringente riguarda l'ottimizzazione dei flussi operativi in un ipotetico ospedale, poiché aumenti imprevisti di accessi al Pronto Soccorso, come si è visto nell'ultima pandemia, la gestione complessa delle sale operatorie e carichi amministrativi crescenti mettono a dura prova personale e infrastrutture. Qui i modelli predittivi e gli algoritmi di scheduling intelligente consentono di allocare in modo dinamico personale e risorse, riducendo code, tempi morti e inefficienze. Il bisogno di monitoraggio continuo dei pazienti, specialmente in contesti di cure domiciliari e telemedicina, impone di processare dati in tempo reale da dispositivi indossabili e

sensori ambientali, individuando subito segnali di allarme e garantendo interventi precoci fuori dalle mura ospedaliere. Infine, emerge la necessità di gestire e valorizzare grandi volumi di dati clinici e amministrativi, per esempio le cartelle elettroniche, i referti testuali, i dati di laboratorio e i registri di attività generano informazioni preziose solo se estratte, standardizzate e rese interoperabili. L'IA, attraverso tecniche di Natural Language Processing, risponde proprio a questa esigenza, liberando i professionisti sanitari dai compiti ripetitivi e consentendo loro di concentrarsi sulla cura diretta del paziente. I principali bisogni che spingono la ricerca in IA applicata alla medicina sono dunque la diagnostica precoce e precisa, la personalizzazione delle terapie, l'accelerazione dello sviluppo farmacologico, l'ottimizzazione operativa, il monitoraggio remoto e gestione avanzata dei dati. Solo affrontando queste necessità in modo integrato sarà possibile rendere il sistema sanitario più efficace, sostenibile e davvero centrato sulle esigenze del paziente.

Le applicazioni di approcci integrati in medicina

Queste tecnologie, adottate come parte di un ecosistema digitale, forniscono benefici cruciali in termini di processo decisionale ed efficienza del flusso di lavoro nell'ambiente clinico:

Screening del DNA: Dataset genomici possono essere analizzati utilizzando algoritmi di ML e DL in grado di rilevare mutazioni e variazioni che potrebbero aumentare la suscettibilità di un determinato paziente a una certa malattia. Questo approccio automatizzato consente uno screening più accurato e rapido, massimizzando la prevenzione e la gestione precoce.

Personalizzazione delle cure: La prescrizione di strategie terapeutiche appropriate include modelli di DL e PNL che integrano dati clinici, genomici e d'immagine. Consente di elaborare piani terapeutici su misura. Ad esempio, un sistema integrato può analizzare le immagini mediche di un paziente, interpretare i risultati delle analisi genetiche e, tramite NLP, estrarre informazioni dalle cartelle cliniche, fornendo un supporto decisionale che aiuta a definire la terapia più appropriata per le specifiche caratteristiche del paziente.

Imaging medico : grazie all’utilizzo di reti neurali convoluzionali si sono aperte le porte all'automatizzazione della diagnosi delle immagini. Non solo questi sistemi hanno la capacità di rilevare osservazioni anomale (come viene fatto per il mappaggio dei nei per il melanoma), ma hanno anche una quantificazione accurata, rendendo possibile il monitoraggio della progressione della malattia e l'impatto del trattamento nel tempo. Questo riduce il carico sul radiologo e aiuta nella standardizzazione e riproducibilità dei risultati diagnostici.

Per riassumere: queste versioni di intelligenza artificiale, ML, DL e PNL si evolvono in uno strumento rilevante per il supporto decisionale nel campo medico. Non solo queste tecnologie accelerano l'analisi dei dati; ma aumentano anche l'accuratezza diagnostica nei piani di trattamento personalizzati che consentono un monitoraggio più efficace e un preventivo della salute del singolo paziente. Questo approccio innovativo, supportato da infrastrutture tecnologiche sempre più sofisticate e da investimenti strategici, promette di trasformare radicalmente il sistema sanitario, rendendolo più reattivo, preciso e centrato sul paziente.

CAPITOLO 2: APPLICAZIONI DELL'IA NELLA PRATICA CLINICA: CASI STUDIO

Questo capitolo mira a fornire un'esplorazione approfondita delle applicazioni dell'IA nella pratica clinica, in particolare ci concentreremo su "Diagnosi e Imaging Medico", e metteremo in evidenza il melanoma, in cui la rilevazione è direttamente legata a un esito positivo. L'avvento delle tecnologie basate sull'IA negli ultimi anni è stato uno sviluppo fondamentale che ha dotato il mondo medico di strumenti innovativi per integrare, potenziare e migliorare i metodi diagnostici tradizionali. Usualmente, l'identificazione del melanoma si basava su una combinazione di valutazioni cliniche, dermatoscopie e, in molti casi, interventi chirurgici per conferma istologica. Tuttavia, questi metodi, che sono essenziali per minimizzare la mortalità, sono limitati dalla loro natura intrinseca, come la mancanza di variabilità interpretativa e la dipendenza dall'esperienza clinica, che possono ostacolare la diagnosi precoce. Con algoritmi di apprendimento automatico e reti neurali, l'IA porta innovazione per trovare soluzioni a questi problemi. L'analisi automatizzata delle immagini e il riconoscimento di pattern patologici portano a valutazioni più rapide e spesso più precise, aprendo la strada a opportunità innovative per una diagnosi precoce e mirata. Questo capitolo mostrerà come questa nuova classe di tecnologie possa essere applicata nel contesto clinico, specificamente per il melanoma, oltre ai potenziali benefici e alle sfide di implementazione. Utilizzando casi studi rappresentativi, l'analisi mette in evidenza esempi specifici di come l'IA possa trasformare l'erogazione delle cure sanitarie e includere una riflessione importante sulle traiettorie future e sulle sfide che rimangono nel diverso tra l'applicazione clinica e l'effettivo utilizzo della tecnologia. Lo scopo di questa revisione non è solo quello di chiarire il potenziale dell'IA per ottimizzare la precisione diagnostica, ma anche di promuovere un dialogo più ampio riguardo alla sinergia tra tecnologia, etica, medicina e prevenzione.

2.1 Analisi Caso Studio: Diagnosi Basata su Intelligenza Artificiale per Il Melanoma

Questo studio indaga un nuovo concetto per la diagnosi del melanoma, un cancro della pelle che, se individuato precocemente, può avere una prognosi sostanzialmente migliorata. Valuta e classifica specificamente la performance dell'algoritmo ADAE, sviluppato per imitare le decisioni dei professionisti in un ambiente dermatologico reale. Di seguito si affronta l'analisi del caso studio con una particolare spiegazione dei grafici chiave emersi dalla ricerca.

I dettagli dello Studio:

In questo caso studio, viene evidenziata l'incorporazione dell'Intelligenza Artificiale (IA) nella diagnosi precoce del melanoma e di come sia un importante passo avanti nella medicina moderna. In questo paragrafo, verrà descritto uno studio che confronta la performance dell'algoritmo ADAE, fondato su un insieme di Reti Neurali Convoluzionali (CNN), con l'esperienza e il giudizio dei dermatologi. L'obiettivo principale di questa ricerca era valutare in modo dettagliato e quantitativo la performance diagnostica dell'IA per il rilevamento del melanoma (e/o delle lesioni cutanee in generale), tenendo conto di caratteristiche chiave come sensibilità, specificità e accuratezza diagnostica. È stato utilizzato un dataset eterogeneo proveniente da otto ospedali universitari tedeschi, ottenendo risultati importabili e rilevanti.

La raccolta dati:

Il dataset esaminato includeva 1716 pazienti ed era composto da 1910 immagini di lesioni di melanoma clinicamente sospette. Otto centri ospedalieri hanno contribuito con i dati, e i centri presentavano un'alta diversità in vari aspetti, tra cui:

- Età dei pazienti (18-96 anni)
- Tipi di pelle, principalmente si riducono ai tipi II e III del sistema Fitzpatrick (Pelle chiara e pelle moderatamente chiara)
- Siti delle lesioni, con interessi che includono tronco, testa/collo, arti e acrali

Questa diversità è pubblicata per la generalizzazione dell'algoritmo ADAE e rappresenta la pratica clinica del mondo reale, contribuendo all'applicabilità dei risultati attraverso la popolazione di pazienti.

L'algoritmo ADAE:

L'algoritmo ADAE ha generato 90 modelli utilizzando una strategia di cross-validation, combinando 18 modelli CNN con codice aperto in un ensemble[2]. Tra questi, uno dei metodi più avanzati introdotti è l'augmentazione[3] al momento del test reale (R-TTA), che genera sei immagini per ciascuna lesione variando angoli di acquisizione e tipi (cioè polarizzate e non polarizzate). Così facendo, questo approccio sfrutta pienamente ciò che possiamo guadagnare all'avere accesso a viste diverse e poi aggregare le previsioni da ciascuna di esse, migliorando così la robustezza e l'accuratezza della diagnosi.

L'IA può essere confrontata con i dermatologi su tre importanti metriche diagnostiche:

- Sensibilità (Recall): misura la capacità del modello di rilevare correttamente i melanomi.

- Specificità: che mostra la capacità di rilevare e escludere lesioni non maligne, prevenendo falsi positivi.

- Accuratezza bilanciata: infine è una statistica che aggrega sensibilità e specificità; una misura della performance diagnostica a una certa soglia (il "quadro generale" della performance diagnostica).

Per la performance dell'IA, la sensibilità era circa del 92% mentre per i dermatologi era circa del 73%. Inoltre, l'accuratezza bilanciata eseguita dall'IA, sebbene leggermente

[2] combinare diversi modelli per migliorare le performance

[3] trasformazioni di immagini tramite ad esempio rotazioni, riflessioni, zoom, variazioni di luminosità e contrasti.

superiore a quella del giudizio clinico, era inferiore a quella della specificità portando a un numero maggiore di falsi positivi.

Risultati e Grafici

La curva ROC (Receiver Operating Characteristic) ci fornisce una lettura del grafico fondamentale per illustrare il compromesso tra sensibilità (tasso di veri positivi) e (specificità o tasso di falsi positivi) per diversi intervalli di soglia decisionale. L'area sotto la curva (AUROC) rappresenta una valutazione complessiva della capacità del modello di discriminare tra etichette vere e false nel contesto sopra descritto

L'algoritmo ADAE è riuscito a descrivere una area maggiore sotto la curva (AUROC), che si è dunque dimostrata superiore a quella raggiunta dai dermatologi, fino ad essere comparabile alla migliore performance standard dei dermatologi medi negli studi considerati. Questa evidenza indica che, in generale, l'intelligenza artificiale è migliore degli dottori nel riconoscere melanomi e lesioni non maligne

Un AUROC più elevato significa che il modello può discernere accuratamente anche le più lievi differenze e definire caratteristiche ad alta precisione nelle immagini dermatoscopiche. Soprattutto durante l'incertezza diagnostica, quando l'esperienza umana potrebbe essere soggettiva, mostrando potenzialmente come l'IA possa fornire supporto ai dermatologi durante il loro processo decisionale. Quella che segue è una tabella riassuntiva dei dati riportati a confronto tra le diagnosi mediche condotte dai dermatologi, con i risultati diagnostici ottenuti dall’algoritmo ADAE esprimendole in termini di balanced accuracy, sensibilità, specificità e AUROC.

Characteristic	Total lesions (n)	Total melanomas (n)	Dermatologists			ADAE			
			Bal. acc.	Sens.	Spec.	AUROC	Bal. acc.	Sens.	Spec.
Overall	1571	653	0.781	0.734	**0.828**	0.916	**0.798**	**0.922**	0.673
Age (in years)									
<35	177	14	0.767	0.571	**0.963**	0.974	**0.890**	**1.000**	0.779
35-54	372	109	**0.821**	0.780	**0.863**	0.931	0.818	**0.945**	0.692
55-74	571	287	0.776	0.739	**0.813**	0.913	**0.793**	**0.913**	0.673
>74	451	243	0.707	0.716	**0.697**	0.879	**0.743**	**0.918**	0.567
Sex									
Male	896	399	0.760	0.734	**0.785**	0.909	**0.783**	**0.920**	0.646
Female	675	254	0.806	0.732	**0.879**	0.923	**0.815**	**0.925**	0.705
Hospital									
Hospital 1	567	245	**0.783**	0.669	**0.898**	0.901	0.772	**0.914**	0.630
Hospital 2	265	123	0.764	0.709	**0.819**	0.922	**0.810**	**0.891**	0.729
Hospital 3	66	39	**0.758**	**0.923**	**0.593**	0.775	0.615	0.897	0.333
Hospital 4	215	93	0.765	0.748	**0.783**	0.937	**0.828**	**0.959**	0.696
Hospital 5	106	31	0.800	0.774	**0.827**	0.932	**0.852**	**0.903**	0.800
Hospital 6	176	67	0.739	0.716	**0.761**	0.890	**0.771**	**0.881**	0.661
Hospital 7	176	55	0.806	0.817	**0.795**	0.934	**0.822**	**0.957**	0.687
Location (grouped)									
Posterior torso	469	196	**0.791**	0.765	**0.817**	0.920	0.788	**0.923**	0.652
Anterior torso	294	80	0.804	0.725	**0.883**	0.951	**0.833**	**0.938**	0.729
Lower extremity	293	116	**0.856**	0.853	**0.859**	0.935	0.817	**0.957**	0.678
Head/neck	278	137	0.660	0.562	**0.759**	0.881	**0.775**	**0.883**	0.667
Upper extremity	191	111	0.778	0.757	**0.800**	0.893	**0.784**	**0.919**	0.650
Palms/soles	29	11	**0.798**	0.818	**0.778**	0.798	0.649	**0.909**	0.389
Oral/genital	12	1	**0.864**	1.000	**0.727**	0.818	0.818	1.000	0.636
Unknown	5	1	0.875	1.000	0.750	1.000	**1.000**	1.000	**1.000**
Diameter (in mm)									
<=3.00	120	14	**0.784**	0.643	**0.925**	0.806	0.772	**0.714**	0.830
3.01-6.00	387	62	0.660	0.419	**0.902**	0.797	**0.715**	**0.726**	0.705
6.01-9.00	274	89	0.695	0.584	**0.805**	0.895	**0.796**	**0.921**	0.670
9.01-12.00	313	155	0.782	0.735	**0.829**	0.924	**0.803**	**0.929**	0.677
12.01-15.00	151	94	0.713	0.777	**0.649**	0.918	**0.765**	**0.968**	0.561
>15.00	326	239	**0.728**	0.858	**0.598**	0.890	0.700	**0.962**	0.437
Fitzpatrick skin type									
I	149	77	**0.797**	0.857	**0.736**	0.865	0.760	**0.922**	0.597
II	797	390	0.778	0.708	**0.848**	0.912	**0.796**	**0.918**	0.673
III	531	160	0.774	0.731	**0.817**	0.938	**0.821**	**0.931**	0.712
IV	39	9	**0.839**	0.778	**0.900**	0.985	0.783	**1.000**	0.567
V	3	1	**1.000**	1.000	**1.000**	1.000	0.750	1.000	0.500
VI	2	0	**0.500**	-	**0.500**	-	0.000	-	0.000
Unknown	50	16	**0.801**	0.750	**0.853**	0.877	0.717	**0.875**	0.559
Technical domain									
Setup 1	898	339	0.779	0.705	**0.853**	0.903	**0.781**	**0.909**	0.653
Setup 2	321	154	0.778	0.753	**0.802**	0.943	**0.845**	**0.948**	0.743

Setup 3	352	160	0.776	0.775	**0.776**	0.912	**0.798**	**0.925**	0.672
Dermatologists' confidence rating									
1	36	13	0.508	0.538	0.478	0.913	**0.754**	**0.769**	**0.739**
2	111	34	0.588	0.500	**0.675**	0.851	**0.761**	**0.912**	0.610
3	424	141	0.662	0.546	**0.777**	0.860	**0.767**	**0.872**	0.661
4	575	206	0.806	0.733	**0.878**	0.934	**0.811**	**0.942**	0.680
5	425	259	**0.899**	0.876	**0.922**	0.939	0.820	**0.942**	0.699

Tabella 1 Questa tabella confronta le performance diagnostiche di ADAE con quelle dei dermatologi, esprimendole in termini di balanced accuracy, sensibilità, specificità e AUROC

Melanoma subtype	Dermatologist			ADAE			Total	p-Value
	n	Sens.	95% CI	n	Sens.	95% CI	n	
Superficial spreading	228	0.820	0.773-0.863	**266**	**0.957**	0.932-0.978	278	p<0.001
Other	127	0.583	-	192	0.881	-	218	-
Nodular	**56**	**0.982**	0.947-1.000	55	0.965	0.912-1.000	57	p<0.001
Lentigo maligna	27	0.600	0.466-0.733	**41**	**0.911**	0.822-0.978	45	p<0.001
Acral lentiginous	16	0.842	0.684-1.000	**17**	**0.895**	0.737-1.000	19	p<0.001
Combined forms of melanoma	12	0.667	0.444-0.889	**17**	**0.944**	0.833-1.000	18	p<0.001
Unknown	6	0.667	-	7	0.778	-	9	-
Nevoid	3	1.000	-	3	1.000	-	3	-
Desmoplastic	2	0.667	-	1	0.333	-	3	-
Spitzoid	2	0.667	-	3	1.000	-	3	-
Total	479	73.35		**602**	**92.19**		653	

Tabella 2 diagnosi per ogni tipo di melanoma

Da queste due tabelle emergono informazioni complementari che ci offrono un quadro completo delle performance diagnostiche dell'algoritmo ADAE rispetto ai dermatologi, nonché della sua capacità di riconoscere i diversi sottotipi di melanoma.

In particolare, nella prima, viene evidenziata la performance globale sul test set, misurata con indicatori quali balanced accuracy, sensibilità, specificità e AUROC. Da questa tabella si evince che ADAE raggiunge una balanced accuracy di circa 0.798, che è superiore a quella dei dermatologi (0.781). La sensibilità dell'algoritmo è particolarmente elevata infatti è di 0.922 contro 0.734 dei dermatologi, indicando che ADAE riesce a identificare correttamente un maggior numero di melanomi e riduce così il rischio di falsi negativi, elemento cruciale per la diagnosi precoce, aumentando così la fiducia nell'algoritmo. Tuttavia, questo vantaggio in termini di sensibilità si accompagna a una specificità inferiore che è 0.673 rispetto a 0.828, il che sta a significare un algoritmo che tende a generare un maggior numero di falsi positivi. Questi risultati, supportati da test statistici robusti (ad es., $p<0.001$ e intervalli di confidenza al 95%), sottolineano il potenziale dell'IA come strumento di supporto, che, integrato con il giudizio clinico, può migliorare l'accuratezza diagnostica complessiva.

Nella seconda tabella, invece, si approfondisce l'analisi dei vari sottotipi di melanoma. Qui si osserva come ADAE dimostri una sensibilità significativamente più alta rispetto ai dermatologi per la maggior parte dei sottotipi, evidenziando il suo potenziale nel rilevare precocemente il melanoma. Un'eccezione da notificare vi è per il melanoma nodulare, in cui la differenza in termini di sensibilità non è così marcata, in quanto dimostra che in questo sottotipo specifico l'algoritmo potrebbe necessitare di ulteriori ottimizzazioni o del contributo integrato del dermatologo. L'analisi statistica conferma la significatività delle differenze osservate ($p<0.001$ nella maggior parte dei confronti), rafforzando la validità dei risultati.

Complessivamente, combinando i dati delle due tabelle si può concludere che ADAE eccelle nella capacità di identificare i melanomi, in particolare grazie alla sua alta sensibilità e balanced accuracy, pur presentando alcune criticità in termini di specificità. L'analisi per sottotipi, inoltre, dimostra il potenziale che avrebbe l'algoritmo nel supportare la diagnosi precoce affiancata da dermatologi, pur avendo alcune lacune per alcuni sottotipi (come il melanoma nodulare) per il quale rimane importante l'integrazione

con l'esperienza clinica. Una sinergia tra intelligenza artificiale e valutazione umana può ottimizzare il percorso diagnostico e migliorare la gestione clinica dei pazienti con sospetto melanoma.

Impatto della Real Test-Time Augmentation (R-TTA)

La tecnica R-TTA prevede l'elaborazione di più immagini per ogni lesione e l'aggregazione delle predizioni, consentendo una valutazione più completa e robusta.
L'uso di R-TTA ha contribuito a migliorare la performance complessiva dell'algoritmo, riducendo l'effetto di variabilità dovuta a singole acquisizioni di immagini.
Ad esempio, si può pensare di dover giudicare una situazione osservandola da diverse angolazioni: le informazioni aggregate portano a una decisione più informata e accurata. Analogamente, R-TTA permette all'algoritmo di "vedere" la lesione da più prospettive, aumentando la fiducia nella diagnosi finale.Le figure seguenti mostrano le differenze dell'analisi diagnostica con e senza l'utilizzo di R-TTA

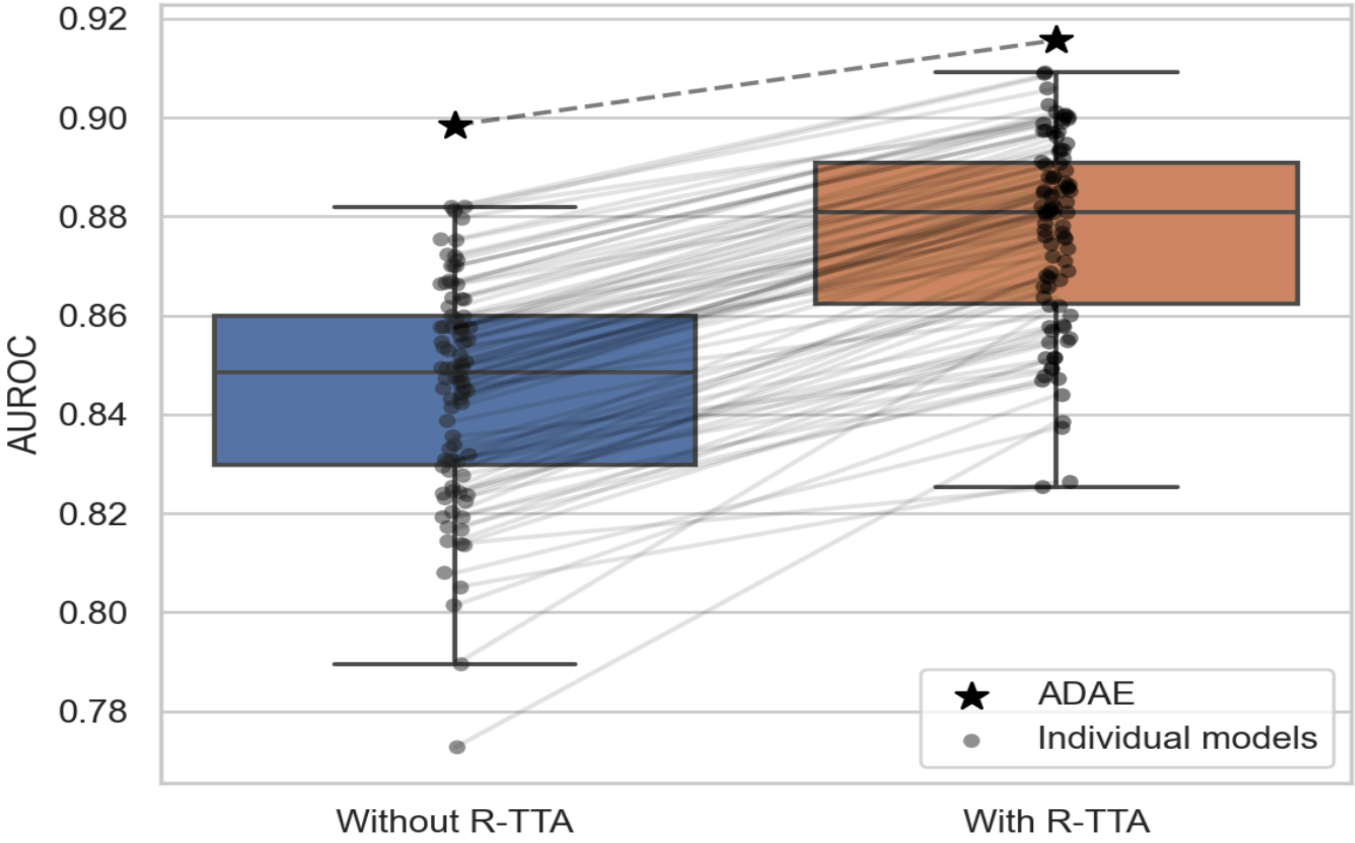


Figura 2 differenze di diagnosi utilizzando per l'algoritmo ADAE la tecnica R-TTA e non

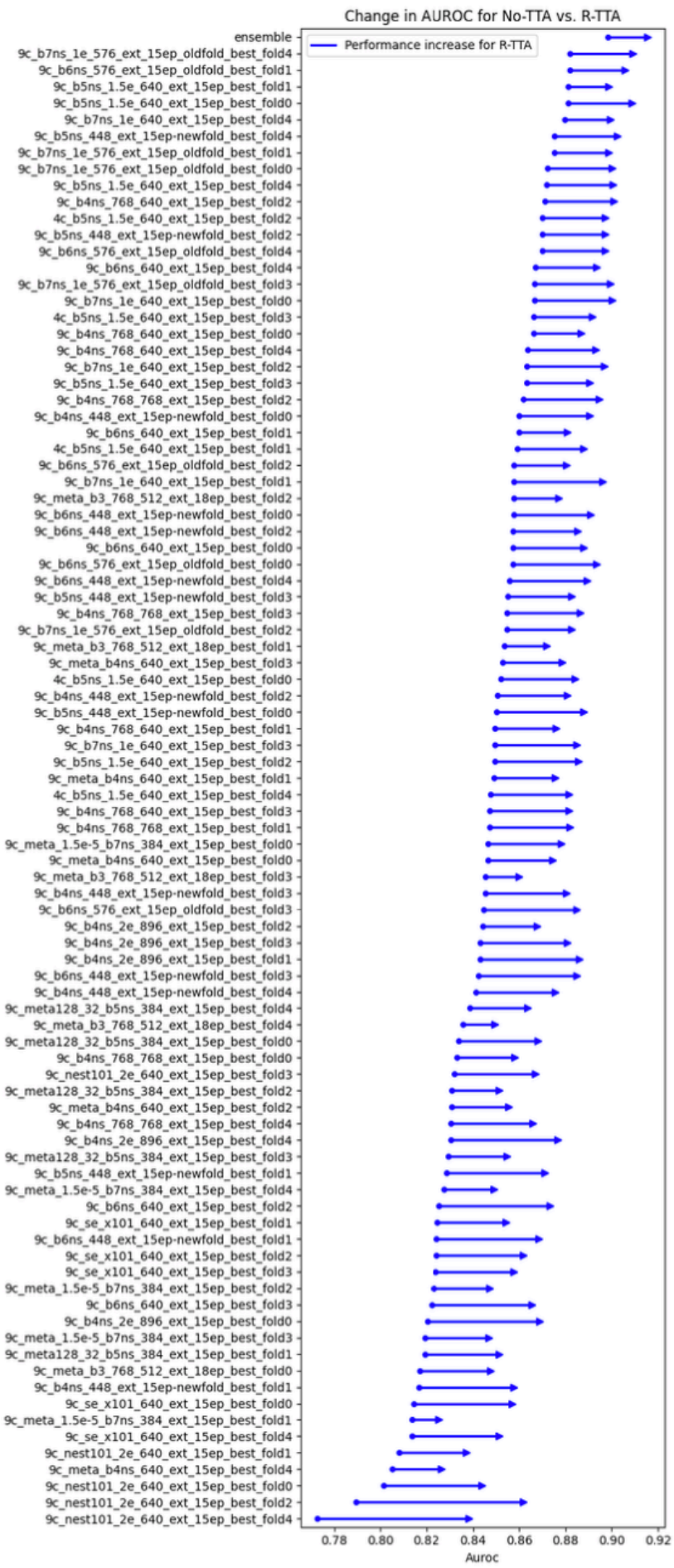


Figura 3 incremento delle performance dell'algoritmo ADEE con l'utilizzo della tecnica R-TTA

Saltano immediatamente all’occhio, i benefici portati all'algoritmo ADAE tramite l’utilizzo di questa tecnica R-TAA.

Considerazioni sui Falsi Positivi

Uno dei risultati più importanti emersi dallo studio è che l'AI produce più falsi positivi rispetto alla specificità inferiore. Questo può significare più test di follow-up o biopsie, ma lo svantaggio è generalmente considerato un piccolo prezzo da pagare nel mondo della medicina del cancro, dove l'obiettivo è non trascurare un caso di melanoma. Una strategia clinica che sintetizzi le valutazioni AI con quelle degli stessi clinici potrebbe aiutare ad alleviare del tutto la preoccupazione, come l'uso dell'AI come filtro per indirizzare più casi sospetti (dal pubblico preoccupato) verso ulteriori screening. I risultati indicano che, nonostante i modelli AI performino bene, c'è dello spazio per miglioramenti in termini di un livello più alto di specificità del modello. Studi futuri potrebbero coinvolgere l'ottimizzazione degli algoritmi, lo sviluppo di dati clinici più solidi e approcci ibridi che mescolino i punti di forza delle tecnologie AI, insieme all'esperienza clinica umana. Lo studio del caso qui esaminato dunque mostra chiaramente che l'Intelligenza Artificiale rappresentata dall'algoritmo ADAE può unirsi al dermatologo nella diagnosi precoce del melanoma. l'AI è ora capace di raggiungere un'alta sensibilità, e questo è un valore aggiunto nel sistema diagnostico. Si può dunque prevede che gli strumenti AI avranno grandi implicazioni per l'avanzamento della pratica clinica e della precisione diagnostica in presenza di incertezza, soprattutto con enfasi nel minimizzare i tassi di falsi positivi.

2.2 Analisi Caso Studio: screening del cancro al seno

Il caso studio in questione costituisce un importante passo avanti nella diagnostica oncologica, specialmente per quanto riguarda lo screening del cancro al seno. Lo scopo di questo studio è quello di concentrarsi sull'incorporazione di sistemi di intelligenza artificiale, ad esamine di immagini bidimensionali del seno da diverse prospettive, dalla localizzazione di una singola lesione all'intero seno, fino all'intero caso clinico, utilizzando tecnologie IA sviluppate tramite l'aiuto di modelli di apprendimento profondo (deep learning). Questo approccio prevede un

ensemble (combinazione) di tre modelli e restituisce un punteggio di rischio normalizzato (da 0 a 1) basato sulla media dei risultati dei singoli modelli.

È stato testato il sistema su due grandi set di dati clinici con caratteristiche metodologiche distinte: uno proveniente dal Regno Unito, comprendente 25.856 donne che hanno ricevuto una doppia lettura nell'ambito del programma NHSBSP, e uno dagli Stati Uniti, che ha coinvolto 3.097 mammografie con lettura singola, come avviene solitamente nella pratica clinica statunitense. Questa duplice strategia consente la rivalutazione sia confrontare le prestazioni tra diversi contesti, sia a generalizzare il contesto e i dati del modello ad altri ambienti clinici.

Metodologia e Struttura del Sistema

Il fulcro di questo studio è l'uso di un ensemble di tre modelli di apprendimento profondo. Ogni modello è stato addestrato a classificare immagini mammografiche a livelli diversi:

1. Dimensione relativa della lesione: Il primo modello è progettato per individuare anomalie benigne di piccole dimensioni che un radiologo esperto può avere difficoltà a riconoscere.
2. Modello di livello superiore: Un modello secondario analizza l'intero seno, combinando aspetti spaziali e morfologici per fornire una prospettiva globale.
3. Aggregato di punteggio di rischio globale: Il terzo modello raccoglie e media le informazioni derivate dalle analisi locali, che vengono utilizzate per generare una sorta di punteggio di rischio globale.

La combinazione di questi tre livelli di analisi è ciò che consente al sistema di essere così preciso nella valutazione del rischio di cancro di una persona, perché tutte le sfumature e le specificità che andrebbero perse in una lettura standard vengono raccolte.

Risultati e Confronto con la Lettura Umana

il primo risultato che emerge da questo studio è rappresentato dalla minimizzazione degli errori diagnostici: Nel Dataset USA: L'uso del sistema ha aumentato la specificità del (5,7%) e la sensibilità del (9,4%) rispetto alle letture umane. Tale miglioramento sta a significare che il

sistema di intelligenza artificiale è in grado di ridurre due problemi principali nello screening oncologico: i falsi positivi e i falsi negativi. Nel dataset del Regno Unito (programma NHSBSP) invece, è stata registrata una riduzione assoluta dell'1,2% nei falsi positivi e un miglioramento corrispondente della sensibilità che supera le misure minime accettate standard, il che indica che anche in ambienti con doppia lettura, il sistema fornisce contributi positivi. Successivamente si può evincere un confronto tra la diagnosi dell'IA e i radiologi: il sistema IA aveva un'area sotto la curva ROC (AUC) circa 0,115 maggiore della media dei radiologi (p=0,0002). Questi dati enfatizzano che il sistema è significativamente efficiente nel differenziare i casi positivi dai negativi, considerando il significato statistico e clinico del metodo basato sull'intelligenza artificiale. La generalizzazione del modello è un aspetto molto rilevante in questo studio; poichè, una parte interessante del modello è la sua capacità di generalizzazione su diversi casi studio. Il metodo è stato addestrato esclusivamente sui dati del Regno Unito, ma valutando le immagini dei pazienti statunitensi ha mostrato un miglioramento della specificità del 3,5% e un tasso di sensibilità dell'8,1%. Ciò indica che il modello è generalizzabile ad altri ambienti clinici, dove il particolarismo geografico o metodologico lo precedeva come limitazione. infine da questo studio è emerso il concetto di come può essere integrato nel workflow clinico, beneficio che potenzialmente consente di integrare efficacemente il sistema IA nel flusso di lavoro clinico. Allo stesso tempo, nel caso in cui il primo lettore e il sistema concordassero, la seconda lettura potrebbe saltare, con una riduzione del carico di lavoro fino all'88% senza deterioramento della qualità. Questo fattore rappresenta un notevole risparmio di risorse e può avere implicazioni positive per la gestione dell'aumento della domanda di screening oncologici

Grafici rilevanti:

In questo caso studio sono stati pubblicati grafici ed immagini a dimostrazione e supporto dei test effettuati che contribuiscono a rendere più comprensibili e immediatamente visibili i risultati e le metodologie adottate, e sono i seguenti:

- Schema di Sviluppo e Valutazione del Sistema AI

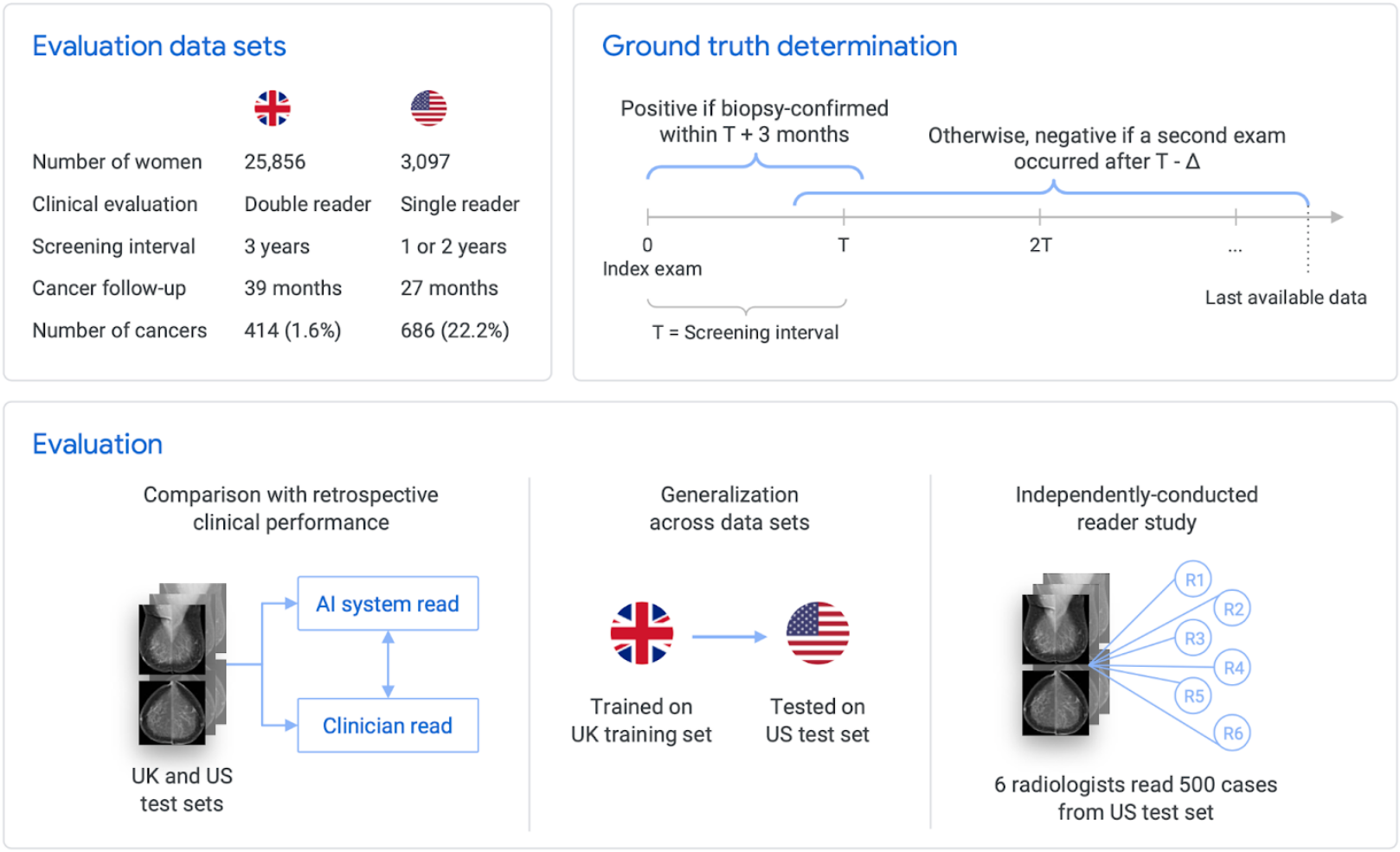


Figura 4 schema di sviluppo e valutazione del sistema AI

Nel contesto dello studio, la Figura 4 offre una panoramica sintetica e immediata dell'intero processo di sviluppo e valutazione del sistema di intelligenza artificiale. Innanzitutto a partire dal primo riquadro in alto a sinistra viene evidenziata la composizione dei dataset, suddividendo in due classi la provenienza dei dati, ovvero da popolazioni di screening sia nel Regno Unito che negli Stati Uniti. Questi dataset includono informazioni cruciali relative agli esiti, quali le biopsie e i follow-up, nonché diverse caratteristiche cliniche dei pazienti. Successivamente, il diagramma illustra il flusso di sviluppo del sistema: l'AI viene addestrata e validata su set di dati distinti, per poi essere sottoposta a test su dataset indipendenti. Questo approccio garantisce che il modello sia sottoposto a una robusta verifica della sua efficacia e generalizzabilità. Infine, la figura confronta

l'output del sistema AI con le decisioni storiche adottate dai radiologi, che possono essere basate su un singolo o doppio processo di lettura, e con i risultati di una valutazione condotta in uno studio indipendente che ha coinvolto lettori esterni.

- Prestazioni del sistema di previsione del cancro al seno

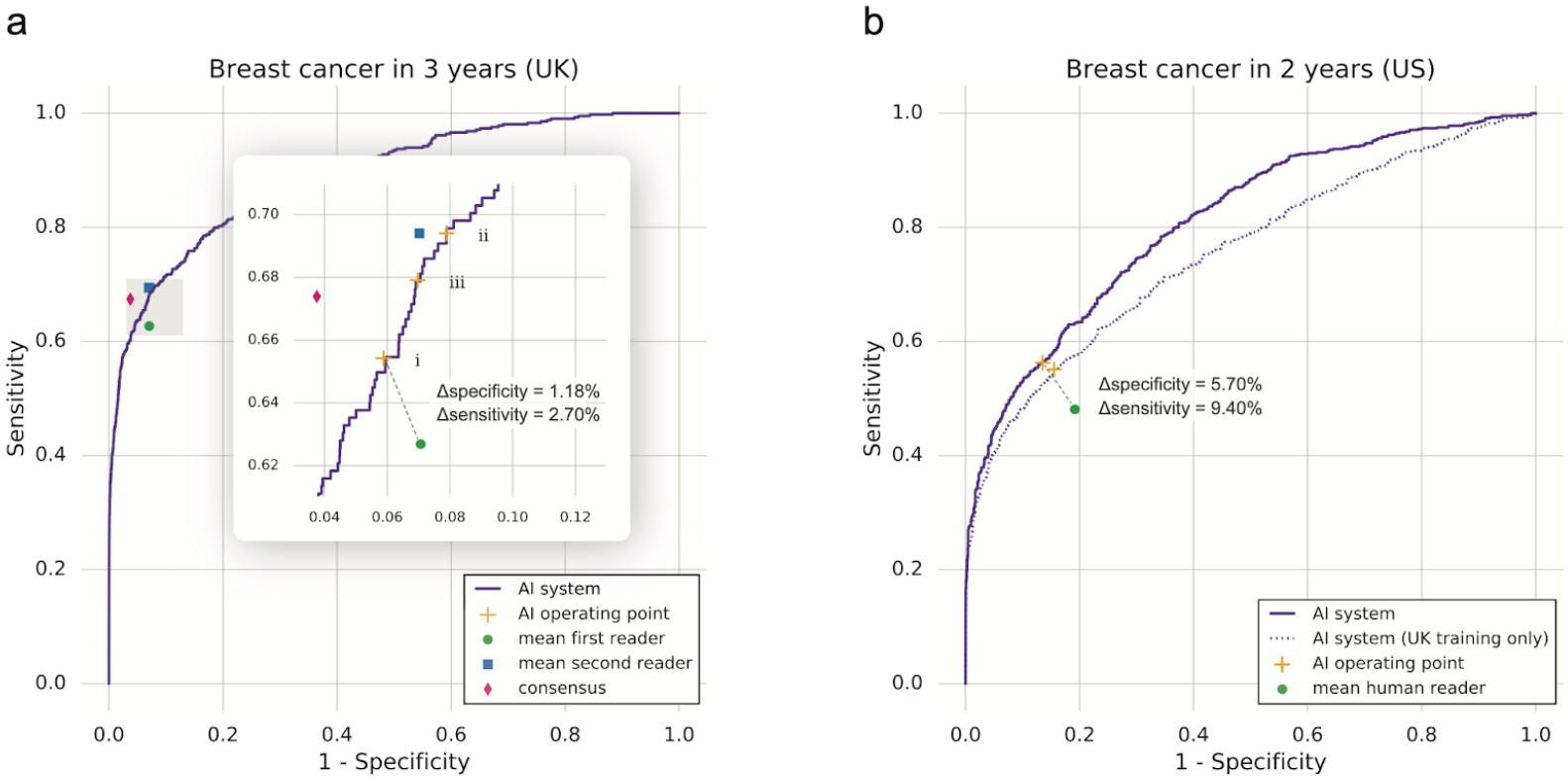
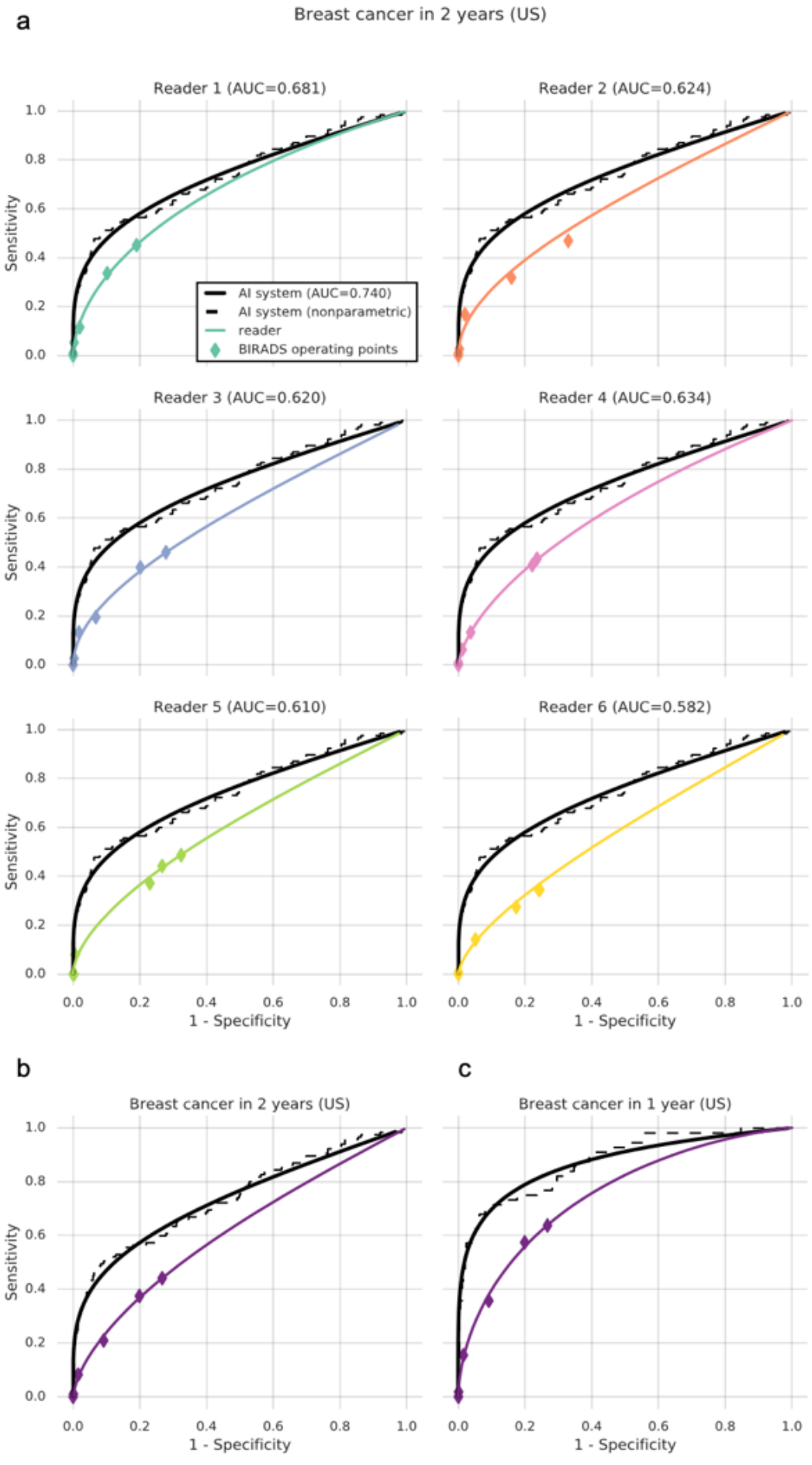


Figura 5 Prestazioni del sistema di previsione del cancro al seno

Nel pannello (a, fig. 5) – relativo ai dati del Regno Unito – viene mostrata la curva ROC (Receiver Operating Characteristic) del sistema AI. Questa curva rappresenta la capacità del modello di distinguere tra i casi positivi (cancro) e quelli negativi (assenza di cancro) variando la soglia decisionale. Il valore dell'Area Sotto la Curva (AUC) risulta pari a 0.889, il che indica un'ottima capacità discriminante: valori vicini a 1 suggeriscono performance quasi perfette, mentre un valore di 0.5 rappresenterebbe una previsione casuale. Vi sono riportati anche dei punti che indicano le coppie (sensibilità, specificità) ottenute dai radiologi durante la pratica clinica (nel riquadro grigio). Questi punti, posizionati lungo l'asse della curva, hanno un significato cruciale

per comprendere il confronto tra il modello AI e l'operato umano. In pratica, ogni punto rappresenta una decisione clinica: per esempio, un radiologo che decide se richiamare il paziente per ulteriori esami (quindi interpretando il caso come positivo) oppure meno (caso negativo). La sensibilità associata a ciascun punto esprime la capacità del radiologo di identificare correttamente i casi di cancro, mentre la specificità indica quanto bene viene riconosciuta l'assenza di cancro. Il posizionamento di questi punti rispetto alla curva ROC del sistema AI permette di valutare se il modello sia in grado di raggiungere o superare le performance medie dei radiologi. Se i punti relativi ai radiologi si trovano al di sotto della curva, significa che, per lo stesso livello di sensibilità, il sistema AI riesce a garantire una specificità superiore e viceversa, evidenziando quindi un potenziale vantaggio nell'adozione del modello. Nel pannello (b, fig. 5) invece, emergono risultati particolarmente interessanti relativi alla performance del sistema AI sui dati statunitensi. In primo luogo, il grafico evidenzia quando il modello viene addestrato utilizzando un dataset misto, comprensivo sia dei dati britannici che di quelli statunitensi, il quale raggiunge un'AUC (Area Under the Curve) di circa 0.81. Successivamente uno in cui il modello viene addestrato esclusivamente con dati del Regno Unito e l'AUC scende a circa 0.757. Nel primo scenario, l'AUC riporta un valore che indica una buona capacità discriminante nel distinguere correttamente tra casi positivi e negativi. Al contrario, il decremento nel secondo caso, evidenzia come l'addestramento su un singolo dataset possa limitare la capacità di generalizzazione del sistema, rendendolo meno performante quando applicato a dati provenienti da un ambiente diverso, come quello statunitense. I risultati indicano che, per contesti clinici eterogenei come quelli dei dati statunitensi, un addestramento che incorpora dati locali è essenziale per raggiungere prestazioni ottimali.

Figura 6 confronto Diretto tra il Sistema AI e i Radiologi



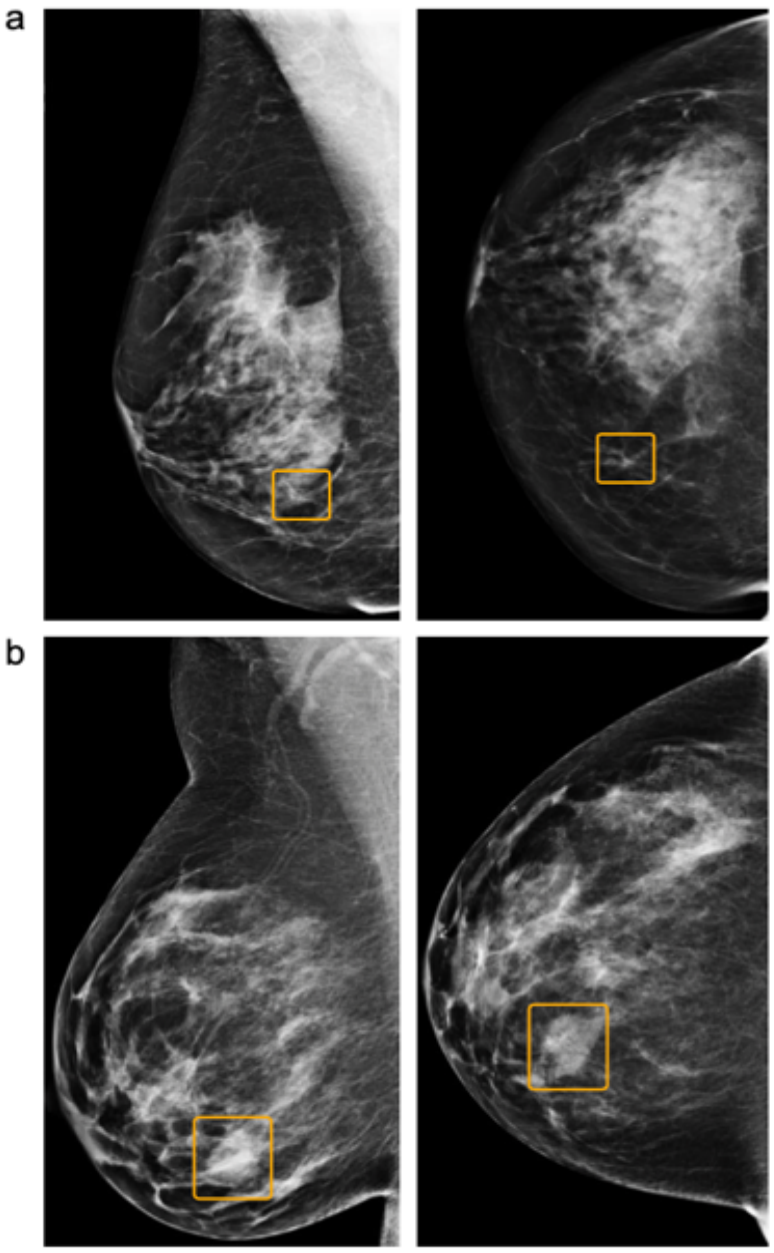
Confronto Diretto tra il Sistema AI e i Radiologi

a. Sei radiologi hanno valutato ciascun caso (n = 465) utilizzando la scala BI-RADS a 6 punti. Una curva ROC adattata per ciascun radiologo è stata confrontata con la curva ROC del sistema di intelligenza artificiale (IA). Per riferimento, è presentata anche una curva ROC non parametrica. I casi sono stati considerati positivi (n = 113) se hanno ricevuto una diagnosi di cancro confermata dalla patologia entro 27 mesi (2 anni e 3 mesi) dal momento dello screening. Tutti e sei i grafici evidenziano la variabilità intrinseca nelle valutazioni soggettive dei radiologi, che, pur essendo esperti, mostrano differenze nel bilanciare sensibilità e specificità. La media degli AUC dei radiologi (circa 0.625) è significativamente inferiore a quella del sistema AI (0.740), evidenziando un vantaggio del modello di intelligenza artificiale. Il gap, espresso come ΔAUC di +0.115, è statisticamente significativo (p = 0.0002) e suggerisce che l'AI riesce a identificare in maniera più accurata i casi positivi senza compromettere eccessivamente la capacità di escludere i falsi positivi.
b. Risultati aggregati di tutti e sei i radiologi dalla parte (a).
c. Risultati aggregati (n = 408) di tutti e sei i radiologi utilizzando un intervallo di 12 mesi per la definizione del cancro. I casi sono stati considerati positivi (n = 56) se hanno ricevuto una diagnosi di cancro confermata dalla patologia entro 1 anno (vedere la Tabella 3 dei Dati Estesi).

Questo confronto diretto, rappresentato dalle curve ROC individuali, sottolinea l'opportunità di integrare il sistema AI nel flusso diagnostico: l'AI, grazie a una capacità di modulazione della soglia decisionale, offre un compromesso ottimale tra sensibilità e specificità, potenzialmente migliorando la diagnosi precoce del cancro al seno e contribuendo a ridurre l'errore umano nelle interpretazioni.

- Esempi di Discrepanze tra AI e Lettori Umani

Nel pannello a, fig.7 è mostrato un caso di cancro al seno non rilevato da tutti e sei i radiologi nello studio statunitense, ma correttamente identificato dal sistema di intelligenza artificiale. Le immagini mostrano due viste di una piccola massa irregolare con microcalcificazioni associate



nel quadrante inferiore interno della mammella destra. Nel pannello b, fig. 7 invece vi è dimostrato l'opposto, ovvero, un caso di cancro al seno rilevato da tutti e sei i radiologi nello studio statunitense, ma non identificato dal sistema di intelligenza artificiale. Le immagini mostrano due viste di una massa densa nel quadrante inferiore interno della mammella destra (a sinistra, vista mediolaterale obliqua; a destra, vista craniocaudale).

Questi esempi illustrano come l'intelligenza artificiale possa migliorare la rilevazione del cancro al seno, ma anche come ci siano ancora sfide da affrontare per raggiungere una diagnosi ottimale.

Figura 7esempi di Discrepanze tra AI e Lettori Umani

Implicazioni Cliniche

L'integrazione dell'intelligenza artificiale nello screening mammografico rappresenta un progresso nella diagnosi e nella gestione clinica del cancro al seno. A questo proposito, l'implementazione dell'IA ha importanti vantaggi clinici e operativi, e i benefici sono molteplici. In primo luogo, l'integrazione dell'IA raggiunge un'accuratezza diagnostica significativamente maggiore. Riducendo sia i falsi positivi che i falsi negativi, questi sistemi aiutano a mitigare l'ansia dei pazienti e ad evitare procedure invasive non necessarie. Una diagnosi più accurata e tempestiva è fondamentale: la rilevazione precoce delle anomalie migliora le prognosi di trattamento

multifattoriali. Per la lettura automatizzata, l'aspetto dell'efficienza della tecnologia AI emerge come il più operativo. Le evidenze suggeriscono che il carico di lavoro del personale può essere ridotto fino all'88% con l'uso di tali sistemi nella simulazione di doppia esposizione. Ciò è importante in situazioni in cui la carenza di esperti è una realtà da affrontare. Le letture automatizzate non solo risparmiano risorse disponibili, ma consentono ai radiologi di concentrarsi su casi complessi che richiedono la loro esperienza. Un altro vantaggio dell'AI nello screening mammografico è la sua generalizzazione e adattabilità. La capacità di un modello allenato su dati britannici di continuare a funzionare eccezionalmente bene quando testato su set di dati americani illustra che questi sistemi sono capaci di essere utilizzati in vari contesti sanitari internazionali. Questa adattabilità va oltre i vincoli geografici o metodologici, aprendo la strada a un aumento nell'applicazione della tecnologia AI in oncologia. È stato dimostrato che l'integrazione della tecnologia AI nelle procedure cliniche della vita reale avviene senza compromettere la qualità diagnostica. Eseguire cliniche in simulazione dimostra che c'è un valore aggiunto nella fusione della capacità computazionale del sistema AI e del lavoro svolto dal radiologo. L'analisi di questo studio evidenzia dunque come la sinergia tra tecnologie avanzate e competenze cliniche possa rappresentare il futuro della diagnostica oncologica, favorendo una diagnosi precoce dei tumori e una migliore gestione delle risorse sanitarie. Tuttavia, l'adozione di tali sistemi richiede anche una riflessione attenta sulle implicazioni etiche e organizzative, soprattutto in termini di formazione, governance e integrazione nei flussi di lavoro esistenti.

2.3 intervista a Amie J. Miller, AI e prevenzione del tumore al polmone

La dottoressa e ricercatrice Miller è stata intervistata in occasione del 51° Meeting Annuale & Summit sul Business dei Centri Oncologici dell'ACCC (Association of Cancer Care Centers), tenutosi a novembre, che coincide con il Mese della Consapevolezza sul Cancro ai Polmoni. Come sottolinea Miller, il tumore polmonare causa ogni anno più decessi, negli Stati Uniti, di quelli provocati insieme dal cancro al seno, al colon e alla prostata. Durante l'intervista, Miller ha spiegato come l'intelligenza artificiale (IA) e le soluzioni di business intelligence possano migliorare la pratica clinica nei sistemi sanitari. Al Sarasota Memorial Health Care System, nel

sud-ovest della Florida, guida un'iniziativa basata su IA che non solo salva vite, ma genera oltre un milione di dollari l'anno. Presso il Brian D. Jellison Cancer Institute di Sarasota, coordina il Programma per la Prevenzione e la Rilevazione Precoce del Cancro ai Polmoni, che include anche un progetto dedicato ai noduli polmonari incidentali (IPN) scoperti tramite IA. Gli IPN sono lesioni asintomatiche intercettate per caso durante TAC eseguite in pronto soccorso o in altri contesti; possono indicare un tumore precoce, ma spesso restano "invisibili" ai pazienti e ai medici di base.

Grazie ai progressi terapeutici degli ultimi anni, la sopravvivenza al tumore polmonare è aumentata sensibilmente, a patto però di intercettare la malattia nelle fasi iniziali. Un'analisi 2024 dell'American Lung Association evidenzia però che le disuguaglianze nell'accesso a test genetici e screening sono in crescita, penalizzando la possibilità di cura tempestiva. A livello nazionale, appena il 18,1% dei pazienti rientra nei programmi di screening, pur con punte del 45% in realtà virtuose come St. Elizabeth's in Kentucky. Capire chi è candidato allo screening è fondamentale. Le linee guida della US Preventive Services Task Force (USPSTF) raccomandano screening per chi ha fra i 50 e gli 80 anni, è fumatore attuale o ha smesso da meno di 15 anni dopo almeno 20 anni di sigarette.

«Guardando i dati – osserva Miller – solo il 42% dei pazienti con tumore polmonare soddisfaceva quei criteri. È un problema reale.» Il 44% delle diagnosi arriva in stadio avanzato, quando le chance di guarigione sono molto ridotte. Un programma dedicato agli IPN può colmare questa lacuna: uno studio 2022 della Penn Medicine ha mostrato che circa il 10% dei pazienti con noduli al di sopra di 8 mm riceve una diagnosi di cancro polmonare.

Ma chi sono questi pazienti?

«Gli IPN non nascono da programmi di screening strutturati – spiega Miller –: spesso si tratta di persone senza medico di base, senza assicurazione e che si presentano in pronto soccorso per tutt'altro motivo. Mentre chi segue un programma di screening ha di norma una buona copertura assicurativa, è già in contatto con il medico e più informato.»

Ecco come il Sarasota Memorial si è mosso: L'IA esamina ogni anno circa 150.000 TAC nei quattro ospedali del sistema (1.100 posti letto). Il radiologo referta, ma il paziente spesso non viene mai informato della scoperta del nodulo. In precedenza, quelle segnalazioni si perdevano:

il paziente veniva dimesso e il medico di base non riceveva alcuna notifica. Anni dopo, il tumore ricompariva in stadio avanzato.

Miller ricorda:

«Succede oggi nei pronto soccorso di tutto il Paese, ve lo garantisco.»

Con il nuovo protocollo: L'IA individua gli IPN che richiedono follow-up e attiva un kit di comunicazione per contattare direttamente i pazienti. Tra il 2019 e il 2023 il numero di screening annuali è quasi raddoppiato, passando da 676 a 1.342 (+99%). Oggi il 70% dei tumori intercettati dallo screening e il 44% di quelli individuati tramite IPN vengono diagnosticati in stadio I o II, ben al di sopra della media nazionale.

Dal punto di vista economico, l'investimento in IA si è ripagato: il tasso di ritorno per i pazienti ad alto rischio è stato superiore all'85%, con una media del 91% dal 2019 al secondo trimestre 2024, contro il 22,3% nazionale.

Secondo il report della sessione, l'iniziativa ha generato: Screening polmonare: 8.321.128 $ di costi downstream e 803.106 $ di margine di contribuzione da 1.702 casi.

Noduli incidentali: 5.559.125 $ di costi e 349.121 $ di margine da 275 casi. Una chiave del successo è che l'IA non interrompe il flusso di lavoro dei clinici: ogni giorno analizza le TAC e compila una "work list" di pazienti da ricontattare.

«Ogni mattina – racconta Miller – chiamo qualcuno dicendo: "Buongiorno, lei era al pronto soccorso la scorsa settimana per un calcolo renale, ma abbiamo rilevato un nodulo di 10 mm al polmone." E quasi sempre il paziente non ne aveva la minima idea.»

Ora un gruppo di pneumologi e un chirurgo toracico si incontrano settimanalmente per discutere i casi ad alto rischio emersi dall'IA: riescono a valutare dieci situazioni in appena mezz'ora, e l'attenzione agli IPN sta crescendo.

«Coinvolgere i medici di base e le comunità locali – spiega – è essenziale per aumentare ulteriormente i nostri numeri di screening.»

Dal 2019 a oggi, lo screening per il cancro al polmone è cresciuto del 99%, con un netto aumento delle diagnosi precoci. Inoltre, l'IA ha permesso di intercettare i noduli incidentali durante normali esami TAC, garantendo un follow-up tempestivo. Sul piano finanziario, l'iniziativa ha ottenuto un ritorno dell'85% per i pazienti ad alto rischio, superando di gran lunga la media nazionale, e

sottolineando il potenziale dell'IA nel ridurre i costi a lungo termine. Fondamentale è stato anche il lavoro di sensibilizzazione e informazione nelle comunità, che ha incrementato l'adesione allo screening. Nonostante i risultati incoraggianti, l'adozione di programmi simili in Europa e nel resto del mondo resta limitata. Superare barriere infrastrutturali, normative e culturali richiederà la collaborazione di istituzioni, investitori e comunità scientifica, per estendere questi benefici e garantire un impatto sostenibile ed equo sulla salute pubblica.

2.4 Analisi Caso Studio: large language model

Il documento "Leveraging Large Language Models for Patient Engagement: The Power of Conversational AI in Digital Health" approfondisce come i modelli di linguaggio di grandi dimensioni (LLM) possano rivoluzionare il rapporto tra pazienti e sistema sanitario. Il testo prende avvio descrivendo il contesto di trasformazione della digital health, dove l'enorme mole di dati clinici, spesso non strutturati, richiede strumenti in grado di estrarne valore in modo rapido ed efficace. In questo panorama, i LLM si distinguono per la capacità di comprendere e generare linguaggio naturale, di ragionare su testi complessi e di dialogare in modo fluido. Per questo sono considerati alleati preziosi nei sistemi di Clinical Decision Support (CDS), perché possono tradurre note mediche e conversazioni con i pazienti in informazioni strutturate utili per il processo decisionale. Il documento mostra come, grazie ai LLM, sia possibile creare risposte contestuali e personalizzate che non solo assistono il medico nelle scelte terapeutiche, ma coinvolgono attivamente il paziente, migliorando la sua esperienza e il suo livello di engagement. Soprattutto, questi modelli superano i limiti delle tecniche tradizionali di Natural Language Processing, risultando efficaci anche in ambiti altamente specialistici e complessi come quello sanitario.
Il documento prosegue con l'analisi dei progressi tecnologici che hanno reso possibili applicazioni sempre più sofisticate dei LLM, con particolare riferimento a modelli come GPT-4 e PaLM, che grazie alle innovazioni continue hanno implementato le capacità multimodali, cioè sono ora in grado di elaborare dati non solo testuali, ma anche visivi, grazie a tecniche di multimodal learning. Tecniche di allineamento che consentono una maggiore affidabilità e coerenza nei modelli, garantendo che i risultati rispondano in modo più preciso alle esigenze specifiche degli utenti.

Questi aspetti offrono una rappresentazione grafica del flusso di informazioni e delle innovazioni architetturali dei modelli moderni, sottolineando come il meccanismo di self-attention contribuisca a una comprensione bidirezionale e contestuale del testo.
Il nucleo del documento è rappresentato da tre casi studio, ciascuno volto a dimostrare applicazioni concrete dei LLM in ambito sanitario:

2.4.1 Analisi delle Discussioni sulla Salute Mentale su Reddit:

Questo studio ha applicato modelli di linguaggio di grandi dimensioni (LLM) per analizzare i pattern linguistici all'interno delle discussioni sulla salute mentale su Reddit, focalizzandosi in particolare sul subreddit “Suicide Watch”. Questo spazio digitale consente alle persone che affrontano pensieri suicidi di condividere le proprie esperienze e di chiedere supporto o consigli. Analizzando 2,9 milioni di post pubblicati in 30 subreddit tra ottobre e dicembre del 2022, e confrontandoli con lo stesso periodo del 2010, l’obiettivo dello studio era identificare temi coerenti con i fattori di rischio già noti per l’ideazione suicidaria. La ricerca utilizza i LLM e le tecniche di Natural Language Processing (NLP) per interpretare le discussioni online sul suicidio, offrendo un’applicazione cruciale dell’intelligenza artificiale conversazionale nel settore sanitario. Grazie all’adozione di metodologie AI avanzate, capaci di comprendere le sfumature del linguaggio usato nei forum dedicati alla salute mentale, lo studio dimostra come l’intelligenza artificiale conversazionale possa rivelarsi uno strumento prezioso per identificare e comprendere problematiche psicologiche ed emotive partendo da contenuti generati dagli utenti stessi. Questa capacità di interpretazione offre non solo opportunità di monitoraggio in tempo reale e di supporto tempestivo, ma consente anche di migliorare l’efficacia degli interventi digitali nel campo della salute. I risultati evidenziano il potenziale della conversational AI nell’offrire insight profondi sugli stati emotivi e psicologici degli individui, un elemento chiave per lo sviluppo di interventi mirati e per l’incremento dell’engagement del paziente nelle piattaforme sanitarie digitali. Attraverso analisi avanzate come questa, l’intelligenza artificiale conversazionale può evolversi fino a diventare uno strumento più empatico ed efficace, rispecchiando in modo più aderente le esigenze e le condizioni degli utenti a cui si rivolge. Ciò rafforza ulteriormente il suo ruolo nella gestione proattiva della salute mentale e nel supporto continuo ai pazienti. Lo studio ha previsto

l'utilizzo di modelli linguistici di grandi dimensioni (LLM) per generare rappresentazioni numeriche (embedding [4]) dei post, allo scopo di catturare le somiglianze semantiche tra i contenuti testuali. Una volta creati questi embedding, sono state applicate tecniche di riduzione della dimensionalità per identificare le dimensioni linguistiche latenti e visualizzare le relazioni tra i vari subreddit analizzati. Per una panoramica della metodologia impiegata, si rimanda alla Figura 8.

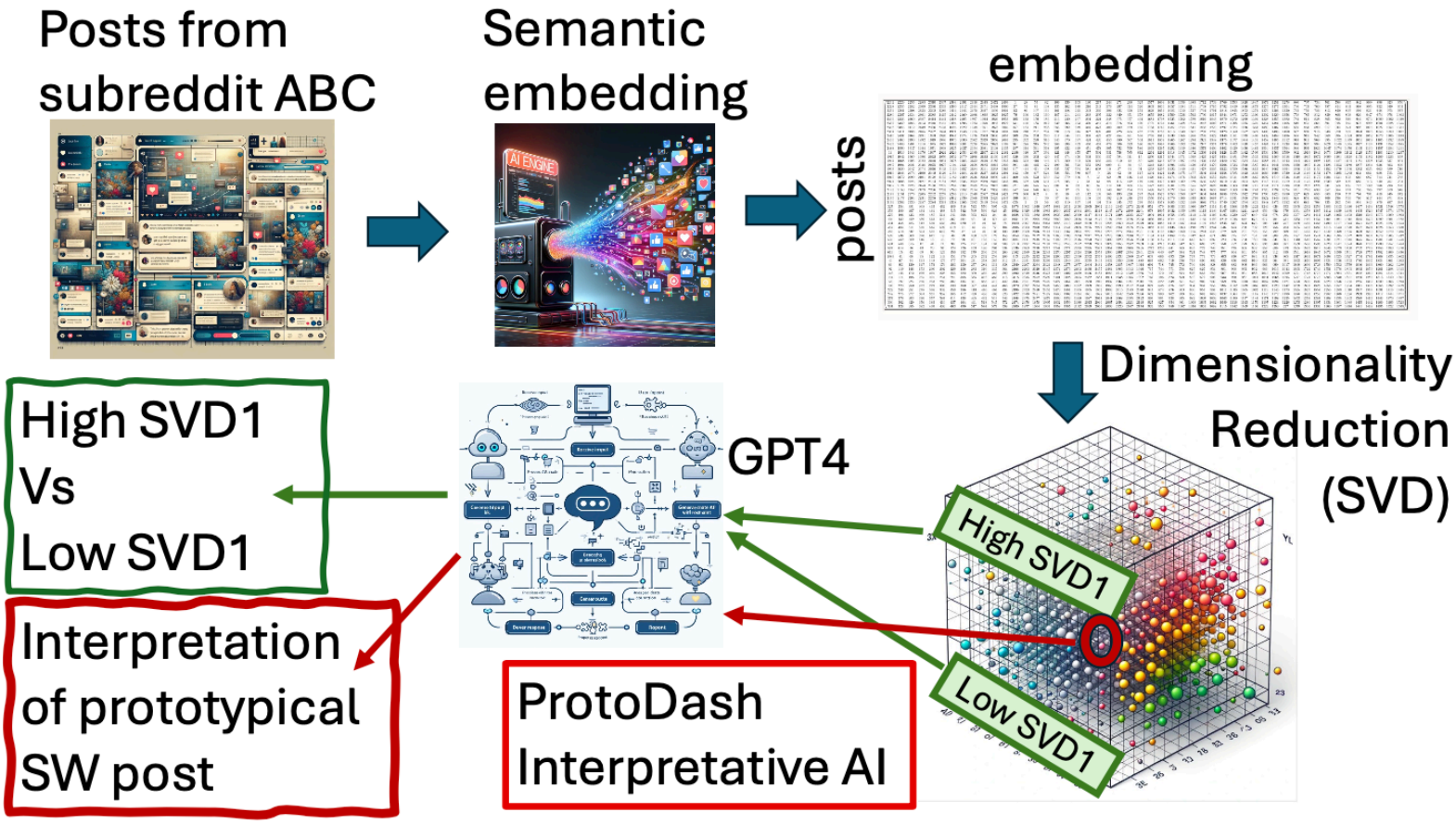


Figura 8 Overview of study methodology. The sub-images were created with the help of DALL·E 3

[4] Il word embedding (tradotto letteralmente: immersione di parole) anche conosciuto come rappresentazione distribuita delle parole permette di memorizzare le informazioni sia semantiche che sintattiche delle parole partendo da un corpus non annotato e costruendo uno spazio vettoriale in cui i vettori delle parole sono più vicini se le parole occorrono negli stessi contesti linguistici, cioè se sono riconosciute come semanticamente più simili (secondo l'ipotesi della semantica distribuzionale)

Questo approccio sottolinea il potenziale dell'intelligenza artificiale conversazionale nell'analisi dei dati non strutturati e nell'estrazione di insight clinicamente rilevanti, utili a orientare interventi terapeutici guidati da AI in ambito di salute mentale. Questa analisi è stata condotta utilizzando LLM per generare semantici dei post e, tramite tecniche di riduzione della dimensionalità (ad esempio, Singular Value Decomposition, SVD), vengono identificate dimensioni latenti come il benessere emotivo, il supporto richiesto e la gravità del disagio. L'analisi ha portato a diversi risultati significativi. In primo luogo, i post del subreddit Suicide Watch sono risultati caratterizzati da temi linguistici coerenti con i fattori di rischio noti per l'ideazione suicidaria, come la percezione di essere un peso per gli altri (perceived burdensomeness), la disperazione (hopelessness), la disconnessione sociale (disconnection) e i traumi vissuti (trauma). Questi temi, descritti nella Tabella 2

Values	Interpretation and axis label		
	SVD1, Emotional Well-Being: Despair to Resilience	SVD2, Seeking Understanding and Support: Closing In to Reaching Out	SVD3, Severity of Distress: Low to High
High	• Struggle and resilience • Experiencing significant anxiety, depression, and life changes • Learning to trust themselves and their abilities to handle their situations • Actively seeking help and trying to find ways to manage their mental health (eg, therapy, self-care, and relaxation techniques) • Making progress toward self improvement and happiness	• Seeking guidance or advice • Desire for information, advice, or validation • Reaching out for insights, recommendations, or shared experiences • This emotion is intertwined with feelings of uncertainty, curiosity, and a desire for understanding or improvement	• More intense and distressing content • Severe mental health struggles, including suicidal thoughts and feelings of extreme despair • Darker tone and more desperate
Low	• Narrative of struggle and despair • Feelings of sadness and emptiness • Express a sense of hopelessness about their situation • Previously sought help (eg, therapy and medication) but feel that these methods have been ineffective	• Feelings of internal conflict, frustration, and being overwhelmed by personal challenges • Deep sense of pain stemming from mental health struggles, physical health issues, or personal insecurities • There's a recurring theme of individuals grappling with their emotions and seeking understanding, validation, or support	• Moderate distress and struggles with mental health • Discusses personal experiences with mental health struggles, seeking help, and dealing with anxiety and social situations

Table 2. Interpretation based on GPT-4 for prompts for the main 3 singular value decomposition (SVD) axes. Text from GPT-4 has been edited for grammar and clarity.

Inoltre, lo studio ha individuato dimensioni linguistiche ben distinte che si correlano rispettivamente con il benessere, la ricerca di supporto e la gravità del disagio psicologico. Questo approccio innovativo non solo conferma l'efficacia dei modelli teorici esistenti sui disturbi mentali e sul comportamento suicidario, ma apre la strada a una comprensione più profonda e guidata dai dati dei complessi fattori che contribuiscono alla suicidarietà. In tal modo, lo studio dimostra concretamente il potenziale trasformativo dell'intelligenza artificiale spiegabile (explainable AI) nella ricerca sulla salute mentale.

Il secondo risultato chiave dello studio riguarda l'identificazione di tre dimensioni linguistiche latenti comuni ai subreddit dedicati alla salute mentale: il benessere emotivo (con un continuum che va dalla disperazione alla resilienza), la ricerca di supporto (che oscilla tra strategie di coping auto-dirette e richieste esplicite di aiuto), e la gravità dei sintomi (dal disagio lieve fino alla crisi acuta). I diversi subreddit, a seconda della condizione psicologica affrontata, si distribuiscono lungo queste tre dimensioni in cluster distinti, come mostrato nelle Figure 9 e 10. Questo suggerisce l'esistenza di una struttura linguistica condivisa che sottende le discussioni su vari disturbi mentali. In particolare, il subreddit Suicide Watch si posiziona in modo netto sui poli negativi delle tre dimensioni: bassa resilienza, alta richiesta di aiuto e elevata gravità del disagio, riflettendo la vulnerabilità

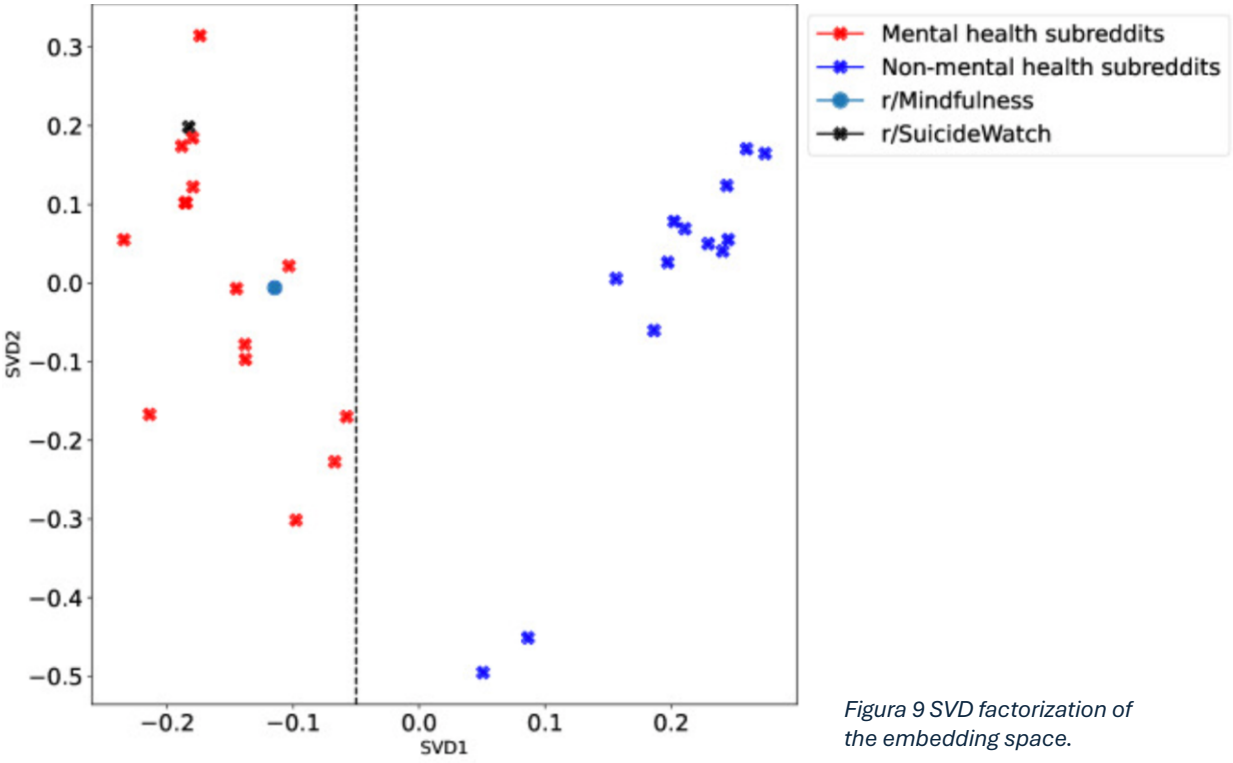


Figura 9 SVD factorization of the embedding space.

e il rischio elevato della sua utenza.

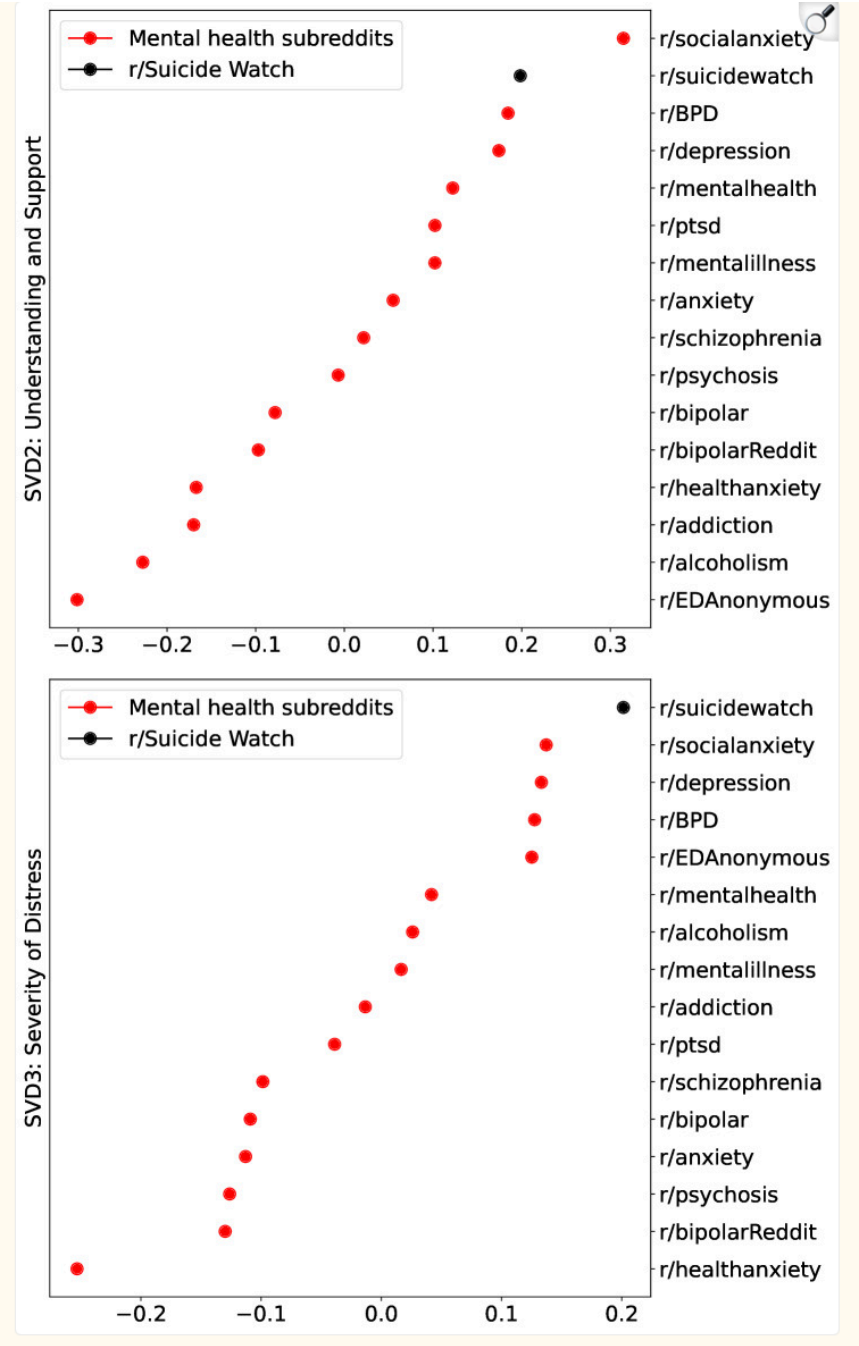


Figura 10 SVD factorization of the embedding space for mental health subreddits.

Un terzo risultato di rilievo è l'allineamento di questi cluster linguistici con la struttura gerarchica della psicopatologia proposta nel modello HiTOP (Hierarchical Taxonomy of Psychopathology). I subreddit relativi ai disturbi internalizzanti (come ansia, depressione e PTSD), ai disturbi del pensiero (schizofrenia, psicosi), e a quelli esternalizzanti (dipendenze, disturbo borderline di personalità) tendono a formare gruppi distinti. Ciò fornisce una validazione empirica data-driven del modello HiTOP, partendo dall'analisi di pattern linguistici spontanei sui social media.

Nel complesso, lo studio dimostra il potenziale dei LLM nell'estrarre insight clinicamente significativi da discussioni online non strutturate. Queste informazioni potrebbero contribuire a identificare individui a rischio, a supportare strategie di intervento e valutazione clinica, e a monitorare l'andamento della salute mentale a livello di popolazione. Tuttavia, va sottolineato che si tratta di uno studio proof-of-concept, e presenta diverse limitazioni:

Il dataset si riferisce a un periodo limitato di soli tre mesi, il che non consente di osservare tendenze stagionali o a lungo termine. L'analisi si basa esclusivamente su Reddit, il che ne limita la generalizzabilità ad altre piattaforme social o a contesti di conversazione offline.
I modelli LLM utilizzati non sono stati fine-tuned specificamente per la salute mentale, per cui la loro capacità di cogliere concetti clinici complessi o sfumati non è garantita.
Gli autori non hanno accesso alle storie cliniche reali o agli esiti sanitari degli utenti, per cui i risultati si basano su correlazioni e non dimostrano un'efficacia predittiva dei marker linguistici individuati.
Per affrontare queste criticità, le ricerche future dovrebbero concentrarsi su:

- Validare i pattern linguistici identificati confrontandoli con dati clinici reali.
- Verificare la stabilità dei risultati nel tempo e su diverse piattaforme social.
- Valutare l'utilità predittiva di questi pattern per l'identificazione di traiettorie cliniche individuali.

Integrare dati multimodali (come immagini e metadati) e sviluppare LLM specificamente addestrati sulla salute mentale, al fine di migliorare precisione, sensibilità e generalizzabilità degli strumenti.
In sintesi, questo studio pone le basi per un approccio innovativo e spiegabile all'analisi della salute mentale tramite AI, dimostrando come la linguistica computazionale possa contribuire alla comprensione e gestione delle malattie mentali su larga scala. Tuttavia, per passare dalla prova di concetto all'applicazione clinica concreta, sarà necessaria una robusta validazione empirica e un attento bilanciamento tra innovazione tecnologica e responsabilità etica.

2.4.2 Analisi della chatbot personalizzato per l'engagement cognitivo negli anziani

Questo studio in corso mira a indagare il potenziale delle chatbots basati su modelli linguistici di grandi dimensioni (LLM) per promuovere l'interesse alla lettura e prevenire il declino cognitivo negli adulti anziani. Il progetto prevede lo sviluppo di un agente conversazionale di intelligenza artificiale che coinvolga gli anziani in discussioni sui loro interessi di lettura e sui progressi fatti, partendo dall'ipotesi che una lettura regolare e stimolante possa aiutare a mantenere le funzioni

cognitive. Lo studio è condotto in collaborazione con il laboratorio per la tecnologia della salute cerebrale e cognitiva del McLean hospital e della Harvard Medical School, e si rivolge a partecipanti di età pari o superiore ai 70 anni che vivono in comunità assistite. I partecipanti sono suddivisi in un gruppo di controllo e in un gruppo di intervento. La chatbot funge da strumento di supporto per incoraggiare il gruppo di intervento a leggere in modo più regolare e frequente rispetto al gruppo di controllo, mantenendo un coinvolgimento attivo con i contenuti letti. Per garantire una differenza significativa nel comportamento di lettura tra i due gruppi – elemento cruciale per l'accuratezza dei risultati dello studio – la chatbot è progettato per rendere l'esperienza di lettura più coinvolgente, piacevole e gratificante per i partecipanti del gruppo di intervento. Offrendo una piattaforma personalizzata e interattiva per discutere di libri ed esplorare interessi di lettura, la chatbot mira a replicare l'esperienza stimolante di un club del libro, il cui svolgimento con moderatori umani risulterebbe troppo costoso e logisticamente complesso in uno studio su larga scala e di lungo periodo. La capacità della chatbot di adattare il proprio stile conversazionale, le raccomandazioni di contenuti e gli spunti di discussione alle preferenze e ai bisogni unici di ciascun utente dovrebbe contribuire a potenziare la motivazione, il coinvolgimento e la continuità dell'intervento durante tutto il periodo dello studio. Sebbene la tecnologia della chatbot abbia potenziali applicazioni per promuovere l'abitudine alla lettura in diverse fasce d'età, il suo scopo specifico in questo studio è quello di creare un ambiente controllato in cui poter indagare in modo rigoroso gli effetti dell'aumento della lettura sulla salute cognitiva nella popolazione anziana. L'utilizzo di un assistente basato sull'IA non solo rende

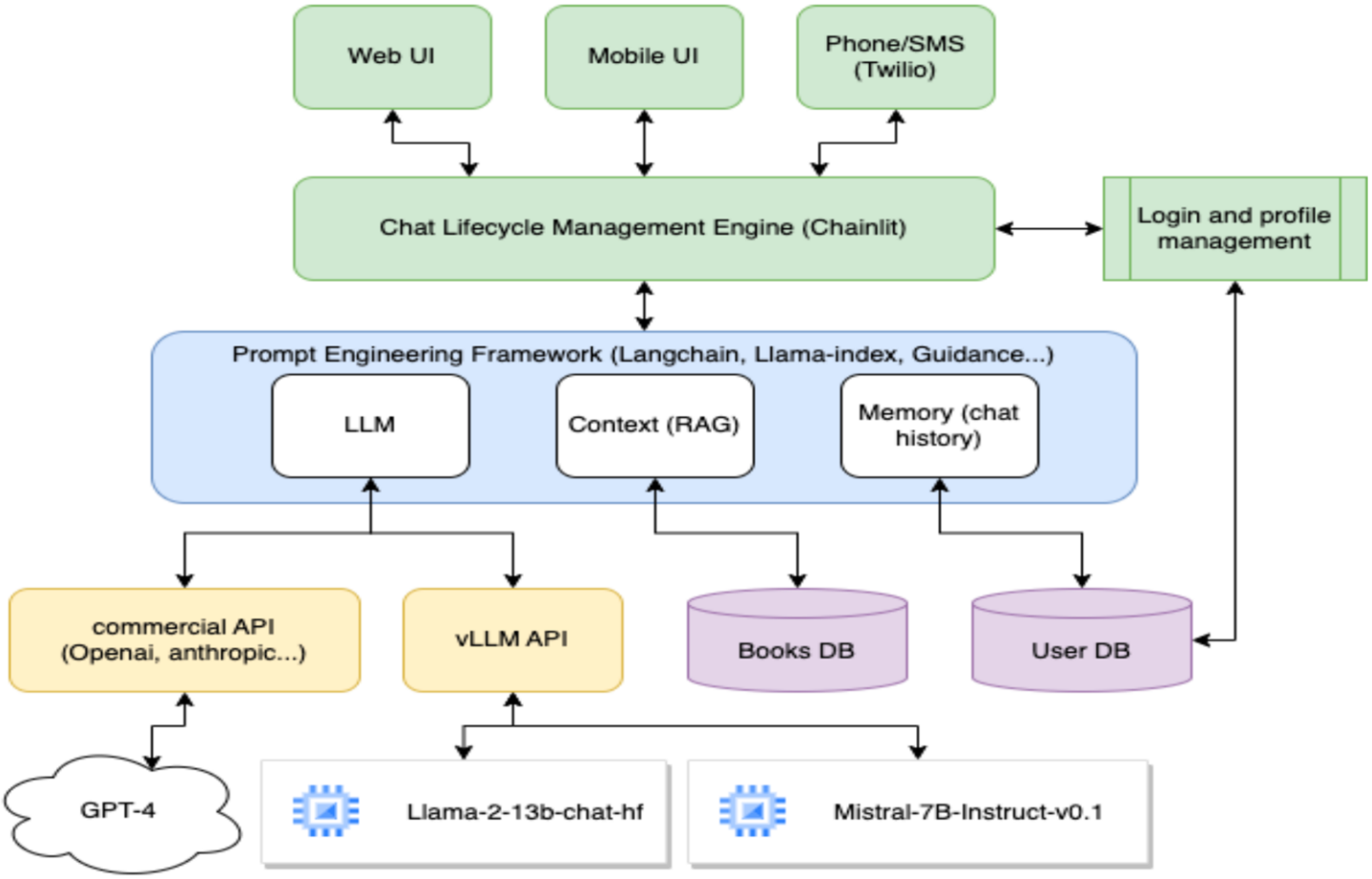


Figura 11 Architettura del chatbot ospitante del bookclub

fattibile un simile disegno di ricerca, ma apre anche nuove possibilità per interventi personalizzati e scalabili a supporto dell'invecchiamento sano.

La chatbot, mostrato nella figura 11, è costruito utilizzando modelli linguistici di grandi dimensioni (LLM) e tecniche di prompt engineering per generare domande aperte basate sul profilo di lettura e sulla cronologia delle conversazioni di ciascun utente. Il sistema mantiene profili utente che tengono traccia dei progressi nella lettura e dei principali punti discussi, permettendo così alla chatbot di adattare le sue risposte agli interessi individuali. Ad esempio, se un utente ha espresso interesse per un personaggio o un tema particolare in una conversazione precedente, la chatbot può porre domande di approfondimento o fare collegamenti a quegli elementi in discussioni successive.

Per garantire accessibilità e facilità d'uso per la popolazione anziana, la chatbot offre interfacce sia vocali (tramite telefonate) che testuali (tramite web app mobile o SMS). I test preliminari con gli utenti hanno prodotto risultati e spunti promettenti. I partecipanti hanno riferito che le conversazioni erano coinvolgenti e stimolanti dal punto di vista mentale, e alcuni hanno osservato che la chatbot li ha spinti a riflettere più a fondo sulla lettura e a fare collegamenti che altrimenti non avrebbero fatto. Diversi utenti hanno apprezzato la comodità e la flessibilità di poter interagire con il bot in qualsiasi momento, senza dover coordinarsi con un partner umano per la discussione. Tuttavia, sono emerse anche delle criticità, come interpretazioni errate delle risposte degli utenti, difficoltà nella gestione di osservazioni tangenziali o fuori tema, e la necessità di una maggiore varietà negli spunti di discussione per mantenere un coinvolgimento a lungo termine. Alcuni utenti hanno inoltre espresso una preferenza per l'interazione umana, sottolineando che la chatbot mancava del calore e dell'empatia di un vero interlocutore. Sulla base di questo primo feedback, il gruppo di ricerca sta iterando sul design della chatbot per migliorarne la comprensione del linguaggio naturale, la generazione delle risposte e la varietà dei prompt. Il progetto è open source e accoglie contributi dalla comunità per lo sviluppo di funzionalità avanzate, come la possibilità di regolare la velocità, il volume e il tono della voce, per adattarsi a utenti con problemi uditivi o cognitivi. La prossima fase dello studio prevede il lancio della chatbot migliorato sull'intero gruppo di intervento e il monitoraggio dell'uso e degli esiti cognitivi nell'arco di un anno. I partecipanti effettueranno valutazioni regolari delle funzioni cognitive, come la Cognitive Ecological Momentary Assessment (EMA), oltre a compilare un questionario sulle proprie abitudini di lettura, sull'interazione con la chatbot e sul livello di soddisfazione generale al termine dello studio. Il gruppo di controllo svolgerà le stesse valutazioni senza accedere alla chatbot, consentendo così un confronto dei percorsi cognitivi tra i due gruppi. Tuttavia, lo studio presenta alcune limitazioni e sfide da considerare. Le dimensioni ridotte del campione e la popolazione specifica (residenti in comunità assistite) potrebbero limitare la generalizzabilità dei risultati ad altri adulti anziani. L'efficacia a lungo termine e la fidelizzazione degli utenti rimangono da verificare, poiché l'effetto novità potrebbe svanire col tempo. Vi sono anche considerazioni etiche legate alla privacy dei dati, al consenso informato e al rischio che gli utenti sviluppino un attaccamento o una dipendenza dal sistema AI. La ricerca futura dovrebbe puntare a replicare i

risultati su campioni più ampi e diversificati, esplorare l'integrazione di misure oggettive di coinvolgimento e performance cognitiva, e indagare le caratteristiche ottimali di design e le strategie di prompting per mantenere benefici e adesione a lungo termine. Con l'evoluzione delle chatbots basati su LLM verso una sempre maggiore sofisticazione e umanizzazione, sarà necessario affrontare in modo continuo anche le implicazioni etiche e sociali del loro utilizzo in contesti sensibili come la cura degli anziani e il supporto alla salute mentale.

2.4.3: Progettazione di un Sistema di Coinvolgimento del Paziente Basato su Intelligenza Artificiale:

Questo caso studio presenta la progettazione e lo sviluppo di un sistema di coinvolgimento del paziente basato su intelligenza artificiale, finalizzato a migliorare la gestione delle malattie croniche attraverso un contatto automatico e regolare con i pazienti. Il sistema, attualmente in fase pilota per pazienti affetti da morbo di Crohn (CD), è sviluppato in collaborazione con la Cleveland Clinic Foundation (CCF) e la Morehouse School of Medicine (MSM). Esso sfrutta i modelli linguistici di grandi dimensioni (LLM) per generare e mantenere dialoghi longitudinali con i pazienti tramite SMS o chiamate vocali.

Il cuore del sistema è un motore di gestione dei dialoghi alimentato da LLM personalizzati tramite prompt engineering, in grado di comprendere le risposte dei pazienti, porre domande di follow-up, fornire informazioni e istruzioni sanitarie al personale sanitario umano se necessario.

I clinici possono monitorare il coinvolgimento dei pazienti tramite una dashboard web (Fig. 12) che mostra i pazienti attivi con le rispettive chiamate programmate, codificate a colori per indicare lo stato (completate, programmate o fallite). Accedendo al profilo del singolo paziente (Fig. 13), i medici possono aggiornare le informazioni, visualizzare le chiamate future e aggiungere nuove chiamate specificando gli strumenti da somministrare, come ad esempio questionari sulla qualità della vita specifici per la malattia.

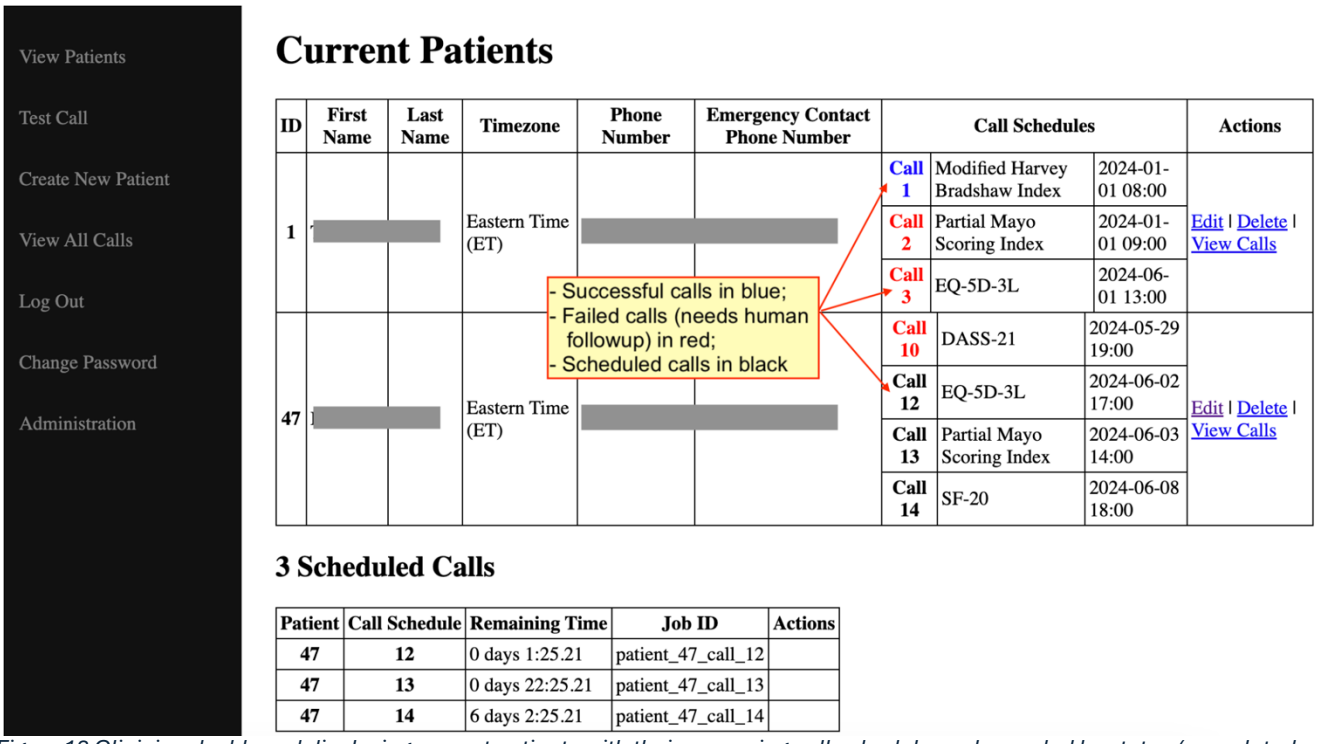


Figura 13 Clinician dashboard displaying current patients with their upcoming call schedules color-coded by status (completed, scheduled, failed), allowing easy monitoring of patient engagement.

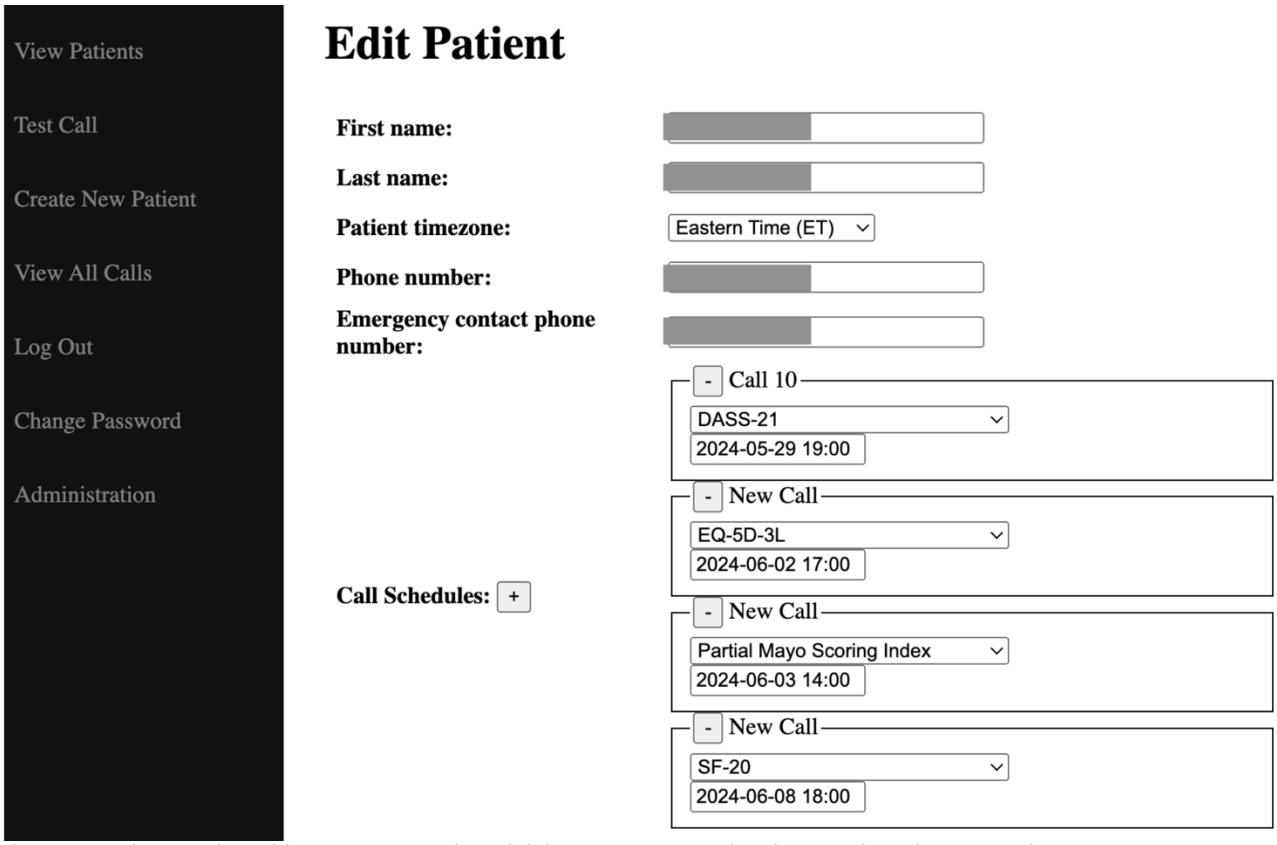


Figura 12 Patient profile editing screen enabling clinicians to update patient information, view upcoming scheduled calls, and add new calls with specific instruments like quality of life assessments.

All’orario previsto per la chiamata, il motore di backend utilizzerà fornitori VoIP come Twilio per generare una chiamata o inviare un SMS al paziente, collegando poi la sessione al chatbot basato su LLM. Le conversazioni guidate dall’IA hanno lo scopo di raccogliere informazioni sui sintomi e sul benessere del paziente tramite questionari medici standard, ma in forma di dialogo naturale e amichevole. Questo approccio punta a incentivare i pazienti a condividere il proprio stato di salute e fornire ai clinici una comprensione più dettagliata e tempestiva della situazione tra una visita e l’altra.

Come mostrato nella Fig. 14, l’assistente AI dialoga con i pazienti in modo naturale, somministrando valutazioni cliniche standard e ponendo domande su sintomi, qualità della vita e benessere generale. Le tecniche di prompt engineering guidano gli LLM nell'eseguire script specifici definiti dai medici, come ad esempio la somministrazione di un questionario o il follow-up post-trattamento.

ID	Patient	Call Schedule	Survey Completed	Canceled	Failed	No Answer	Busy
71	[redacted]	1	True	0	0	0	0

▼ 1. 5/28/2024, 8:26:03 AM (112 seconds) id:101

Patient Call (Download)

Role	Content
user	(patient pickup the phone)
assistant	Greetings! This is Health Guardian from [redacted]. May I talk to [redacted]?
user	Yes, I am.
assistant	Hi [redacted], thank you for answering. I'm calling to follow up on your recent visit and complete a quick survey to monitor your condition. Can you spare about 5 minutes for this?
user	Yes, please.
assistant	Thank you, [redacted]. Let's get started then. How would you describe your overall well-being yesterday?
user	All right. Very good.
assistant	That's great to hear! Did you experience any abdominal pain yesterday?
user	No.
assistant	That's good news! How many times did you have liquid or soft stools yesterday?
user	twice
assistant	Did you experience any joint pain yesterday?
user	No.
assistant	Did you have any eye inflammation?
user	No.
assistant	Did you notice any skin or mouth ulcers?
user	No.
assistant	Did you have any painful bumps on your skin or a new fistula?
user	No.

Figura 14 UI for reviewing of the patient chat history, showing a transcription of the chatbot conversing naturally with the patient to ask follow-up questions about their well-being, symptoms, and quality of life.

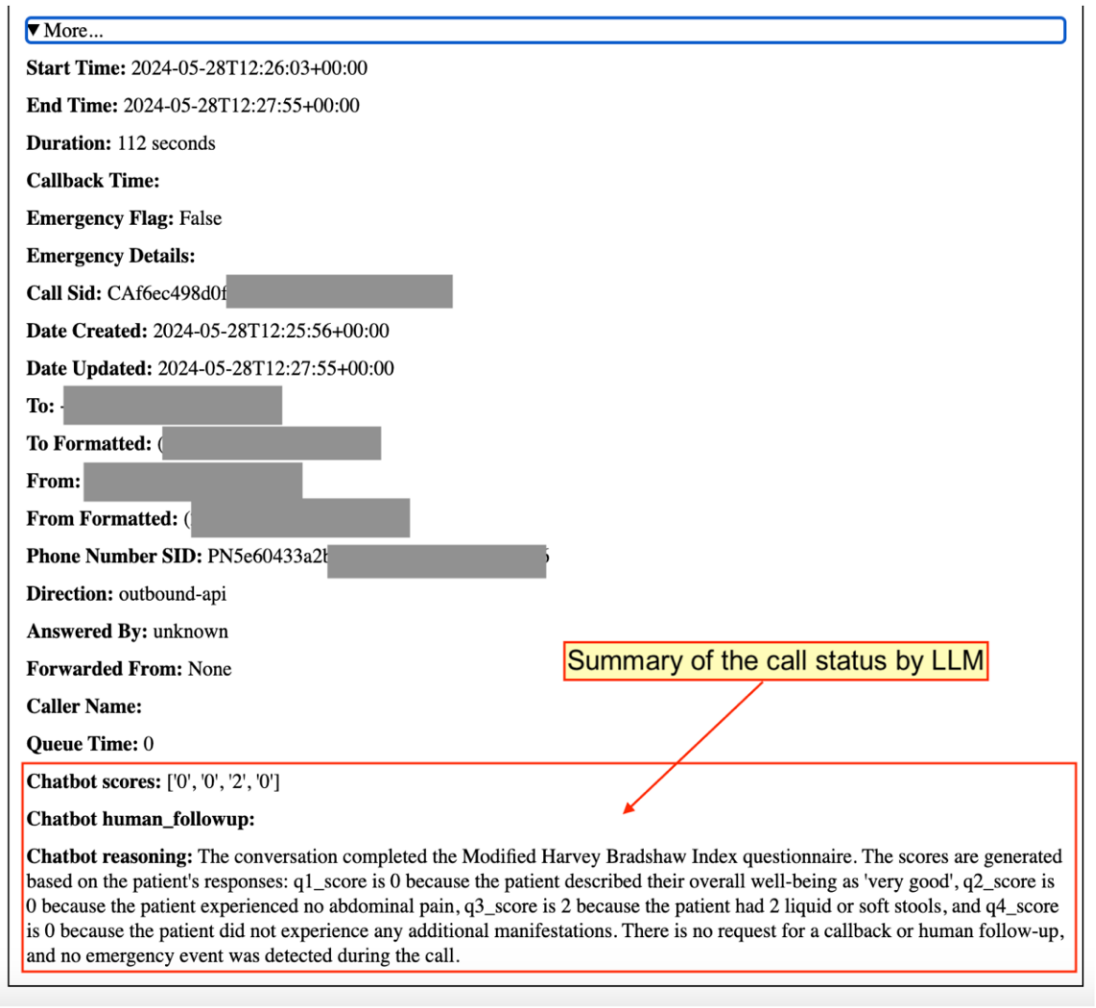


Figura 15 Call summary generated by the AI system, providing key details such as call duration, chatbot scores reflecting the patient's responses on standard assessments, chatbot reasoning explaining the scores, and noting no emergency events or callback requests

Alla fine della chiamata, la cronologia della conversazione viene salvata nel database e può essere esaminata dal team medico attraverso il portale web (Fig. 14). Il sistema genera automaticamente un riepilogo della chiamata (Fig. 15), che include:

- la durata della chiamata
- i punteggi ottenuti dal chatbot sulla base delle risposte del paziente ai questionari
- spiegazioni dei punteggi basate sulle risposte fornite.

Inoltre, il sistema segnala eventi di emergenza o richieste del paziente di essere ricontattato da un umano, permettendo ai medici di intervenire prontamente nei casi più urgenti. Attraverso queste immagini illustrative del sistema reale, l'obiettivo è fornire ai lettori una comprensione concreta di come un sistema di coinvolgimento del paziente alimentato da LLM possa essere realizzato nella pratica. Le immagini dimostrano la capacità del sistema di:

- sostenere dialoghi naturali e contestualizzati con i pazienti
- generare report automatici sullo stato del paziente

- integrarsi perfettamente nei flussi di lavoro clinici tramite un'interfaccia intuitiva

Il sistema può essere ulteriormente migliorato tramite il fine-tuning dei LLM su ampi dataset di conversazioni mediche e l'integrazione con basi di conoscenza specifiche del dominio, al fine di garantire una maggiore accuratezza e rilevanza clinica.
I tre casi studio selezionati in questo lavoro sono stati scelti per la loro capacità di rappresentare, in modo complementare, l'ampiezza e la versatilità delle applicazioni degli LLM nel contesto sanitario. Ogni esempio mette in luce un ambito chiave di trasformazione della sanità digitale: dal coinvolgimento attivo del paziente e il supporto alle decisioni cliniche, fino alla valutazione della salute mentale e alla sintesi della conoscenza medica. La loro eterogeneità riflette l'intento di offrire una panoramica concreta e multidimensionale del potenziale impatto degli LLM sulla medicina del futuro. Questi casi sono stati ritenuti particolarmente rilevanti non solo per il livello di innovazione tecnologica che rappresentano, ma anche per la loro valenza sistemica: si tratta infatti di applicazioni che, se implementate su larga scala, potrebbero migliorare sensibilmente l'efficienza dei servizi sanitari, la personalizzazione delle cure e l'accesso alle risorse mediche, soprattutto in contesti con risorse limitate o barriere geografiche. Guardando al futuro, è evidente che il pieno potenziale degli LLM in sanità potrà essere realizzato solo attraverso una collaborazione sistemica tra ricercatori, clinici, istituzioni regolatorie e comunità di pazienti, finalizzata a promuovere l'equità, la trasparenza e la sostenibilità dell'innovazione tecnologica. Sarà inoltre fondamentale investire nella personalizzazione dei modelli, nell'integrazione di basi di conoscenza medica specialistiche, e nello sviluppo di benchmark etici e scientifici condivisi, che garantiscano la sicurezza e l'efficacia degli strumenti proposti. Solo così sarà possibile realizzare una sanità digitale realmente centrata sulla persona, in cui l'intelligenza artificiale non sostituisce il medico, ma lo potenzia nella sua capacità di offrire cure tempestive, personalizzate ed eque.

2.5 Analisi Caso Studio: screening DNA

Nel campo della medicina moderna, la ricerca sui tumori sta assumendo sempre più importanza grazie all'utilizzo di tecnologie avanzate sia nella diagnosi che nel monitoraggio della malattia. Una delle innovazioni più promettenti è l'impiego del DNA libero da cellule (cfDNA), un frammento di materiale genetico che circola nel sangue e in altri fluidi corporei, e l'applicazione di tecniche di intelligenza artificiale (AI) e apprendimento automatico (ML) per analizzarne i dati. Questo capitolo intende spiegare in termini semplici come funziona il cfDNA e in che modo l'AI sta contribuendo a migliorare la ricerca e la diagnostica dei tumori, per poi esaminare criticamente i vantaggi e gli svantaggi di queste tecniche e le ragioni per cui non sono ancora state adottate su scala globale.

Ma cos'è e come funziona il cfDNA?

Il cfDNA rappresenta frammenti di DNA che vengono rilasciati nel sangue, nelle urine, nella saliva e in altri fluidi corporei da cellule morte o attivamente secreto da esse. Questi frammenti, non racchiusi in cellule, provengono da diversi tessuti del corpo. È possibile identificare la loro origine grazie ad alcune caratteristiche distintive, come la lunghezza dei frammenti e i pattern di metilazione (cioè, le modifiche chimiche che regolano l'espressione genica). Uno dei punti chiave emersi dalla letteratura è che il DNA presente nel plasma ha una lunghezza media di circa 166 coppie di basi (bp), corrispondente alla lunghezza del DNA avvolto attorno a un nucleosoma più una porzione di DNA "linker". Tuttavia, il cfDNA derivante da tessuti particolari, come quello fetale o tumorale, tende a essere più corto (circa 143 bp). Questa differenza nella dimensione dei frammenti consente di arricchire selettivamente il segnale target: per esempio, isolando i frammenti più piccoli si può ottenere un maggior livello di cfDNA tumorale, utile per il rilevamento precoce del cancro. Inoltre, il processo di frammentazione del cfDNA è influenzato dall'azione di enzimi chiamati nucleasi, che "tagliano" il DNA in punti specifici. Questi tagli non avvengono in modo casuale, ma presentano pattern ben definiti (ad esempio, periodi di 10 bp o specifiche sequenze terminali) che riflettono il tipo di nuclease coinvolta. Tale conoscenza permette ai ricercatori di risalire, almeno in parte, alla fonte tissutale di questi frammenti, fornendo informazioni utili per diagnosticare sia tumori che altre patologie.

utilizzo dell'intelligenza artificiale nella ricerca sui tumori

L'intelligenza artificiale e l'apprendimento automatico rappresentano strumenti fondamentali per analizzare i dati complessi derivanti dallo studio del cfDNA. Considerata l'elevata dimensionalità e la varietà dei segnali (dimensione dei frammenti, motivi terminali, pattern di metilazione, ecc.), l'AI è in grado di integrare queste informazioni per identificare pattern non evidenti con metodi tradizionali. Ad esempio, algoritmi come il support vector machine (SVM) o i modelli di rete neurale profonda (DNN e CNN) vengono utilizzati per classificare i segnali derivati dal cfDNA in modo da rilevare la presenza di cellule tumorali. Alcune immagini all'interno dello studio, come quella che illustra il processo di arricchimento del segnale fetale (SeqFF) e l'applicazione di reti neurali per rilevare specifici pattern di metilazione, mostrano come differenti metodologie possano essere combinate per ottenere una diagnosi più accurata e affidabile figura 16. In questa immagine possiamo notare in alto a sinistra il SeqFF, che utilizza una tecnica di elastic net per stimare la frazione fetale basandosi sulla copertura locale lungo il genoma.

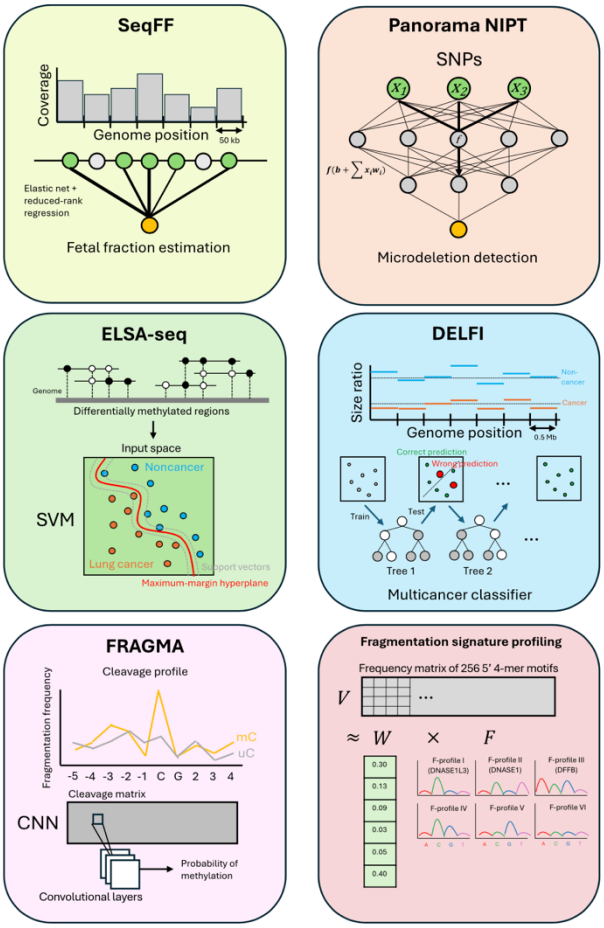


Figura 16 algoritmi ML applicati al cfDNA nei contesti del NIPT e della biopsia liquida per il cancro.

In alto a destra l'approccio di natera, che impiega reti neurali profonde (DNN) per analizzare informazioni di linkage tra SNP, finalizzato alla rilevazione di microdelezioni.

I due riquadri centrali sono rispettivamente SVM e Gradient Boosting e dimostrano l'uso di modelli classici e avanzati per identificare segnali specifici in cfDNA, ottenendo alta sensibilità e specificità nella rilevazione del cancro. I pannelli in basso infine rappresentano CNN e NMF e illustrano come le reti neurali convoluzionali possano analizzare pattern locali attorno a siti

metilati e come la fattorizzazione in matrici possa estrarre “profili fondatori” dalle distribuzioni dei motivi terminali Questa figura riassume graficamente la capacità degli algoritmi ML di integrare molteplici caratteristiche del cfDNA, evidenziando il valore aggiunto dell’approccio computazionale nella diagnosi e nella localizzazione del cancro e nel monitoraggio prenatale.

la figura 17 evidenzia il metodo “holistic kinetic” (HK) che, grazie al sequenziamento in tempo reale, analizza le cinetiche di incorporazione dei nucleotidi per determinare lo stato di metilazione dei CpG, il che si traduce in un significativo aumento della precisione nel rilevamento del cancro. In particolare, si può notare nella parte a sinistra come la polimerasi incorpori nucleotidi

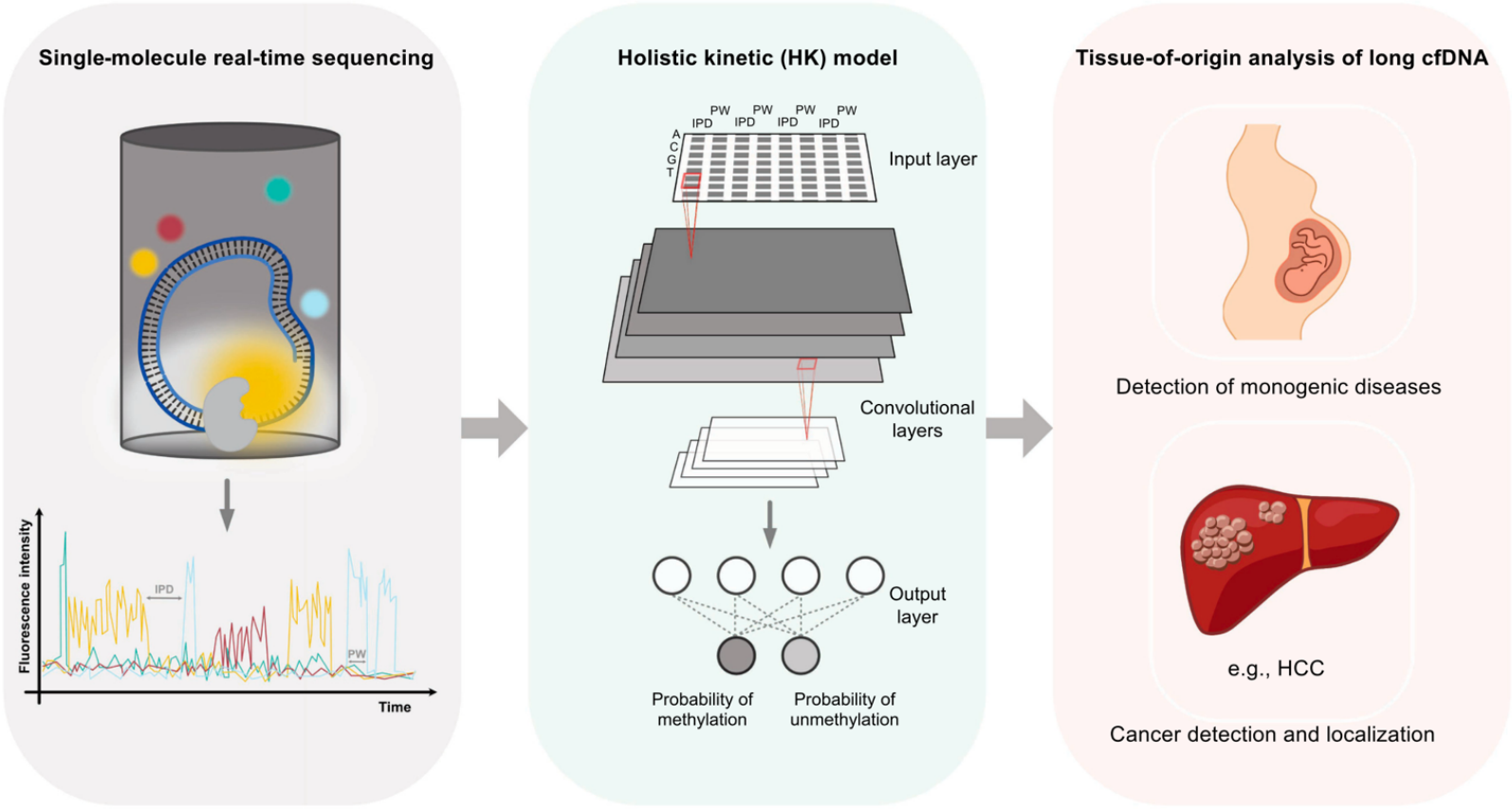


Figura 17 sequenziamento in tempo reale a molecola singola (SMRT) in combinazione con il modello "holistic kinetic" (HK).

fluorescenti in una finestra d’osservazione, dove le cinetiche (IPD e PW) sono influenzate dalle modifiche basiche.

Nella parte centrale il modello HK utilizza un CNN per analizzare la matrice delle kinètiche insieme al contesto sequenziale, traducendo questi dati in una probabilità metilica per siti CpG.

Ed infine nella parte di destra vengono evidenziate applicazioni cliniche, come la stima della percentuale di DNA fetale e la rilevazione del cancro (ad es. HCC). La figura, dunque, ci dimostra come l'integrazione di dati cinetici e strutturali possa migliorare in maniera sostanziale l'accuratezza nella chiamata della metilazione, superando i limiti delle tecniche tradizionali (come la bisolfito-sequenziamento) e offrendo strumenti diagnostici più robusti. Un ulteriore evidenza dell'interazione tra dati provenienti da diverse "dimensioni" (metilazione, dimensioni dei frammenti, motivi terminali) arricchisca l'analisi e la capacità diagnostica, portando a sistemi di screening non invasivi più affidabili e precisi è data dalla seguente figura 18, che illustra l'evoluzione storica e tecnologica delle metodologie impiegate.

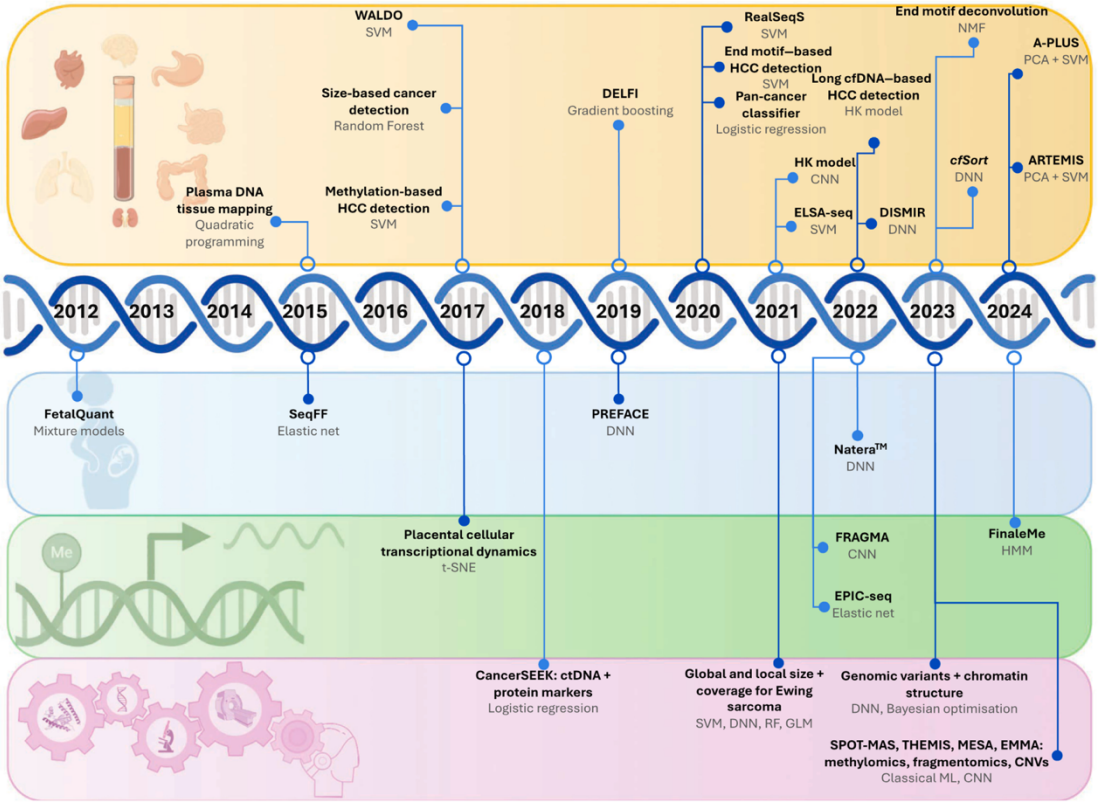


Figura 18 cronologia delle applicazioni di algoritmi di AI/ML alle analisi del cfDNA

In questa figura si può notare in arancione le analisi dell'origine tissutale e le biopsie liquide oncologiche; in blu i test prenatali non invasivi, inclusa la stima della frazione fetale; in verde le analisi frammentomiche basate su metilomica e trascrittomica; e in viola le analisi multimodali di cfDNA, che integrano diverse caratteristiche del cfDNA in un unico modello. La forza dell'AI risiede dunque nella sua capacità di gestire grandi quantità di dati, infatti grazie alla sua capacità di calcolo si riescono ad analizzare centinaia di migliaia di frammenti contemporaneamente. Inoltre è molto utile per rilevare pattern complessi che, altrimenti, non sarebbero identificabili con metodi statistici tradizionali e si riesce ad integrare diverse dimensioni

di informazione, ovvero si riesce a combinare dati relativi alla dimensione, alla metilazione e ai motivi terminali per una classificazione più robusta.

Pro e Contro dell'approccio cfDNA-AI

Di fronte all'opportunità di sfruttare il cfDNA potenziato dall'AI per la diagnosi e il monitoraggio dei tumori, sia le grandi aziende che le startup possono trarre benefici strategici, pur dovendo affrontare sfide tecnologiche, regolatorie e operative. Il mercato dei test cfDNA è in rapida espansione, passando da 8,74 miliardi USD nel 2024 a 10,86 miliardi nel 2025 (CAGR 24,2%) e si prevede arrivi a 27,3 miliardi USD entro il 2030 (CAGR 20,98%) , spinto dall'aumento dei tumori, dall'adozione di medicine personalizzate e dall'integrazione di AI nei flussi diagnostici. Le tecnologie di liquid biopsy basate sul cfDNA consentono di rilevare precocemente mutazioni tumorali e monitorare risposte terapeutiche con un semplice prelievo di sangue, riducendo l'invasività e permettendo un follow-up dinamico del cancro . L'AI entra in gioco per gestire la complessità dei dati frammentomici e metilomici: ad esempio, modelli di deep learning come EMIT analizzano i motivi terminali del cfDNA per distinguere pazienti con cancro da sani con elevata accuratezza. Le grandi aziende, dotate di solide piattaforme di NGS e reti commerciali globali, possono investire in partnership e acquisizioni per integrare rapidamente algoritmi AI nei propri prodotti. Roche, ad esempio, ha ampliato il suo ecosistema digitale di patologia con oltre 20 algoritmi AI per migliorare la diagnosi del cancro e supportare la medicina di precisione, e ha guidato il round da 254 M USD di Freenome per test multicancro basati su liquid biopsy. Questi investimenti rafforzano il vantaggio competitivo di aziende come Roche o Illumina, che possono sfruttare economie di scala, budget R&D multimilionari e un accesso privilegiato ai mercati regolamentati. Tuttavia, l'adozione su vasta scala richiede la validazione clinica e la standardizzazione dei protocolli, processi costosi e lunghi. Investimenti in R&D come i 15 M USD di Codetta Biosciences in una piattaforma dsPCR multi-omica illustrano quanto sia elevato il capitale necessario per il passaggio dal proof-of-concept alla commercializzazione. Allo stesso tempo, la variabilità regolatoria fra FDA, CE-IVD e altri enti crea barriere di ingresso in diversi Paesi, ritardando l'omologazione di test come Guardant360® CDx e FoundationOne Liquid CDx, ora FDA-cleared per supportare la terapia guidata dai biomarcatori. Le sfide tecniche

includono la gestione di flussi di dati ad alta dimensionalità e l'interoperabilità tra laboratori: sistemi cloud-based avanzati e standard comuni sono indispensabili per garantire qualità e replicabilità delle analisi, come evidenziato dalle previsioni di un mercato dei dati cfDNA che crescerà del 23,9% annuo fino al 2028, dando impulso anche ai servizi di data repository e analytics. Inoltre, la necessità di pipeline certificate per processare rapidamente campioni clinici rende critico l'investimento in infrastrutture IT specializzate. Investire nel cfDNA potenziato dall'AI significa in conclusione abbracciare con decisione quella convergenza tra biologia molecolare e calcolo avanzato che può davvero riscrivere le regole della diagnostica oncologica, trasformando un semplice prelievo in un sistema dinamico di monitoraggio e guida terapeutica personalizzata; le grandi aziende, con le loro risorse e la capacità di orchestrare pipeline integrate che vanno dalla raccolta del campione all'analisi dei dati, sono in grado di superare le sfide regolatorie e di standardizzazione, mentre le startup, più snelle e orientate all'innovazione di nicchia, possono sperimentare rapidamente nuovi algoritmi e partnership strategiche, pur dovendo fare i conti con la necessità di validazione clinica e infrastrutture informatiche robuste; è solo calibrando questi elementi – investimenti in R&D, governance dei dati, processi di certificazione e sinergie pubblico-private – che il potenziale rivoluzionario del cfDNA+AI potrà effettivamente tradursi in soluzioni efficaci, sostenibili e accessibili su scala globale.

CAPITOLO 3: ANALISI DEI COSTI E INDICI DI PERFORMANCE

3.1 Melanomi

Un'analisi di Gomez Rossi ha valutato l'implementazione di un sistema di intelligenza artificiale (IA) a supporto del processo decisionale per la rilevazione e la classificazione automatica dei melanomi a partire da immagini cutanee. Confrontando l'approccio IA con la pratica clinica tradizionale in popolazioni degli Stati Uniti e della Germania, mediante un modello di Markov sviluppato dall'United Arab Emirates University, gli autori hanno dimostrato che, nonostante un aumento marginale dei costi diretti per singolo esame, l'IA raggiunge un'accuratezza diagnostica pari o superiore a quella dei medici e rientra nella soglia di convenienza economica definita (ICER tra € 20 000 e € 30 000 per QALY).

Il modello per la diagnosi dermatologica è stato costruito partendo da semplici fotografie digitali delle lesioni cutanee, (il cuore del sistema è un insieme di reti neurali convoluzionali) dalle quali hanno imparato a riconoscere caratteristiche istologiche grazie all'analisi di decine di migliaia di immagini validate in laboratorio. A valle di questa elaborazione, il sistema restituisce una stima sulla probabilità che la lesione sia maligna e suggerisce se è il caso di procedere con una biopsia oppure di programmare controlli di follow-up. Pensato per l'uso quotidiano in centri dermatologici o in ambulatori di medicina di base, il flusso è semplicissimo: il medico cattura la foto della lesione con uno smartphone o un dermatoscopio digitale, la carica sulla piattaforma e, in pochi secondi, ottiene il referto. In più, il sistema fornisce delle mappe di calore (heatmap) che evidenziano le aree della lesione su cui il modello si è maggiormente concentrato per formulare la diagnosi.

i costi diretti:

Voce	IA-Assisted Care	Standard Care	Δ
Costo esame IA (software, licenze)	≈ US$ 50	US$ 40	+US$ 10
Costo di refertazione specialistica remota	≈ US$ 25	US$ 20	+US$ 5
Formazione e IT (ammortamento)	≈ US$ 10	US$ 0	+US$ 10
Totale per esame	US$ 85	US$ 60	+US$ 25

L'incremento medio di US $ 25 per esame viene compensato dalla riduzione dei casi mancati e dai minori costi associati a trattamenti avanzati.

I costi indiretti possono essere stimati come segue:

- Riduzione degli errori diagnostici tardivi: minore incidenza di melanomi identificati in stadio III/IV, il cui trattamento può arrivare a costare fino a US $ 50 000 per caso.
- Ottimizzazione del follow-up: elevata specificità del modello (≥ 90%) riduce le visite di controllo inutili.
- Migliore allocazione delle risorse: i casi non sospetti possono essere gestiti con monitoraggio remoto, liberando slot clinici per i pazienti ad alto rischio.

Si è ottenuto una sensibilità di circa 92% una specificità di circa 90% (AUC > 0,94) e un guadagno di QALY di + 0,03 QALY per paziente nel confronto su ciclo di vita, grazie a diagnosi più precoci. La riduzione della mortalità inoltre è diminuita del 15% dei decessi per melanoma a cinque anni L'analisi di costo-efficacia si riconduce ad un ICER (USA): US $ 24 000 per QALY guadagnato ed un ICER (Germania): € 22 500 per QALY guadagnato, entrambi i valori risultano inferiori alla soglia considerata costo-efficace (€ 20 000–30 000 per QALY). Le analisi di sensibilità indicano che l'ICER rimane sottosoglia anche aumentando del 20% i costi per esame IA o riducendo del 10% la sensibilità del sistema.

Si denotano dunque miglioramenti dell'accuratezza diagnostica, riducendo i ritardi nelle diagnosi più critiche; Rientri nelle soglie internazionali di costo-efficacia (ICER ≤ € 30 000/QALY); Generi risparmi significativi in termini di spese per trattamenti avanzati, visite e procedure non necessarie. Tale caso studio costituisce un valido esempio di come l'IA in dermatologia possa offrire elevato valore clinico-economico, favorendo una diffusione su larga scala all'interno dei sistemi sanitari pubblici e privati.

3.2 Radiologia

Il documento in questione analizza uno studio del ROI per misurare sia i benefici economici che quelli operativi derivanti dall'integrazione della piattaforma IA Calantic nel reparto di radiologia di un ospedale statunitense. Analizzando un periodo di cinque anni e vari scenari alternativi, gli autori mostrano come l'adozione dell'IA possa tradursi in margini di ritorno molto elevati (fino al 791 %) e in significativi miglioramenti di efficienza, senza perdere di vista la robustezza delle ipotesi alla base del modello L'analisi, condotta dalla prospettiva dell'ospedale, considera l'impatto su radiologi, medici referenti, IT, amministrazione, pazienti e pagatori. Sono stati definiti

indicatori chiave di performance relativi ai tempi operativi (attesa, triage, refertazione) che sono stati stimati confrontando il tempo medio per ciascuna attività nel processo tradizionale con quello misurato con l'IA, moltiplicati per il numero annuale di TAC (circa 150.000). La somma delle ore risparmiate, suddivise per fase, è stata convertita in giorni lavorativi (8 ore/giorno). Le diagnosi aggiuntive (es. emorragie intracraniche, occlusioni vascolari, noduli e carcinomi polmonari, lesioni demielinizzanti) che sono state calcolate applicando ai volumi totali di TAC i tassi di "miss rate" rilevati in letteratura e nelle evidenze cliniche interne, ottenendo incrementi numerici come +470 diagnosi di ICH o +678 noduli polmonari identificati. Le procedure chirurgiche e ricoveri evitati o anticipati, valutati moltiplicando il numero di diagnosi aggiuntive per la percentuale di casi che, grazie alla diagnosi precoce, hanno richiesto interventi meno estesi o ricoveri più brevi. I ricavi e costi totali che sono stati rispettivamente 3.560.959$ e 1.780.480$ nel caso base, e derivano dalla somma di esami aggiuntivi, follow up e procedure supplementari, ciascuno calcolato come numero di eventi indotti dall'IA x tariffa media DRG, mentre i costi includono l’installazione del software, la formazione del personale e il canone di manutenzione/supporto IT, stimati in base ai contratti e annualizzati sul quinquennio.

I risultati del caso base (orizzonte 5 anni) sono stati:

ROI complessivo: 451%, calcolato come (Ricavi totali - Costi totali)/Costi totali

Giorni lavorativi risparmiati: 145, derivanti da 15 giorni in attesa, 78 in triage, 10 in lettura e 41 in refertazione.

Benefici clinici: oltre alle 470 ICH e 678 noduli, si segnalano +196 LVO, +10 carcinomi polmonari e +99 referti MS aggiuntivi.

Impatto economico: sullo screening polmonare si sono generati 8.321.128$ di costi downstream e 803.106$ di margine su 1.702 casi; per i noduli incidentali 5.559.125$ di costi e 349.121$ di margine su 275 casi.

Per gli scenari alternativi e analisi di sensibilità, il modello è stato testato in quattro scenari per verificarne la robustezza:

Orizzonte 1 anno: ROI al 335% (ricavi e costi proporzionalmente ridotti).

Centro non ospedaliero: ROI al 41%, in assenza degli effetti downstream di ricoveri e terapie complesse.

Ospedale senza stroke center: ROI al 310%, escludendo i ricavi da terapie intra arteriose.
Valorizzazione del tempo risparmiato: calcolo delle ore risparmiate come entrate addizionali (moltiplicate per costo orario del radiologo), che porta il ROI al 791%.
Nell'analisi di sensibilità one way, una variazione del ±10% nel rapporto costo/ricavo si riflette in un'oscillazione ±10% del ROI; analogamente, ±10,% sui volumi di trombectomie (LVO) o interventi per ICH modifica il ROI rispettivamente di ±5,7% e ±3,1%.

L'analisi ci dimostra che l'integrazione della piattaforma IA Calantic può portare, in un ospedale accreditato per la gestione dello stroke, a un ROI del 451 % in cinque anni, con potenziali punte del 791 % includendo la valorizzazione del tempo dei radiologi. I benefici, sia economici sia clinici, derivano non solo dai risparmi diretti in termini di costi operativi, ma soprattutto dagli effetti downstream legati a diagnosi più tempestive e all'aumento di procedure e ricoveri. Tuttavia, la solidità dei risultati dipende fortemente dalla qualità degli input (rapporto ricavi/costi, volumi di esami, tassi di rilevazione IA), sottolineando l'importanza di adattare il modello alle specificità di ciascuna istituzione.

3.3 Screening

Nel panorama delle strategie di screening oncologico, l'applicazione dell'intelligenza artificiale allo studio del DNA libero circolante (cfDNA) ha aperto nuove prospettive in termini di diagnosi precoce e sostenibilità economica. Un esempio emblematico è rappresentato dal test multicancro "Galleri" di GRAIL, illustrato nel poster ISPOR EU 2023 da Kansal et al. Questo sistema si basa su un'analisi avanzata della metilazione del cfDNA, capace di individuare più di cinquanta diverse neoplasie tramite un semplice prelievo ematico, al costo di 949 US$ per test. Gli autori hanno costruito un modello Markov a vita che confronta lo screening tradizionale con l'approccio MCED (Multicancer Early Detection) integrato nella pratica clinica, adottando come soglia di willingness-to-pay statunitense i 100 000 US$ per QALY. Nel caso più favorevole (hazard ratio di efficacia pari a 1), l'Incremental Cost-Effectiveness Ratio scende a circa 66 044 US$ per QALY, ben al di sotto del limite prefissato, mentre anche nello scenario meno ottimistico (HR = 3) il costo rimane comunque vicino alla soglia, attestandosi intorno a 106 962 US$ per QALY.

I vantaggi economici derivano innanzitutto dal “multiplexing” diagnostico: grazie a un unico test si intercettano tumori che con gli standard attuali richiederebbero protocolli separati, con un notevole risparmio sui costi amministrativi e di laboratorio. Inoltre, la diagnosi precoce evita le spese elevate associate ai trattamenti in stadio avanzato (che possono oscillare tra i 50 000 e i 100 000 US$ per paziente) e contribuisce a ridurre la mortalità specifica del 20–30% per neoplasie non normalmente coperte dai programmi di screening tradizionali. Tuttavia, la sostenibilità di “Galleri” in contesti con soglie di costo-efficacia più stringenti (come quelle europee, spesso intorno ai 30 000 € per QALY) richiederebbe un abbassamento del prezzo di test intorno ai 600 € o una maggiore efficienza nei modelli di efficacia clinica.

A fianco di GRAIL, un’altra proposta promettente proviene dall’Università della California (UCLA): il test cfMethyl-Seq. Questo approccio focalizza l’analisi sulle regioni CpG ricche di informazioni diagnostiche, anziché sequenziare l’intero genoma, e utilizza algoritmi di intelligenza artificiale per evidenziare i pattern di iper- e ipo-metilazione specifici di tessuto e patologia. Con un costo stimato tra i 300 e i 400 US$ per test (circa il 40 % in meno rispetto ai metodi basati sul whole-genome bisulfite sequencing), cfMethyl-Seq raggiunge sensibilità intorno all’85% per stadi I–III e specificità superiore al 95%. I primi modelli economici preliminari suggeriscono un ICER di circa 45 000 US$ per QALY, rendendo questa soluzione particolarmente interessante per sistemi sanitari con soglie medie-alte di WTP. In conclusione, sia “Galleri” di GRAIL sia cfMethyl-Seq dimostrano come l’integrazione dell’intelligenza artificiale con tecnologie di analisi del cfDNA possa offrire strategie di screening oncologico non solo più precoci ed efficaci, ma anche potenzialmente costo-efficienti. La sfida futura consisterà nel calibrare prezzo del test, accuratezza diagnostica e requisiti di costo-efficacia locali, per garantire un’adozione su larga scala che massimizzi i benefici clinici e gestionali.

3.4 NLP

In un progetto pilota condotto da Axtria per un'azienda biotech, l'integrazione di modelli di Natural Language Processing (NLP) e Generative AI ha permesso di trasformare un flusso di lavoro tradizionale, basato su laboriose attività di data entry e pulizia manuale, in un processo completamente automatizzato. Il risultato è stata una riduzione dei costi operativi superiore al 75 %, accompagnata da drastici miglioramenti nei tempi di lavorazione, nell'accuratezza delle analisi e nella capacità di scalare l'attività anche in presenza di picchi di domanda. L'adozione di pipeline NLP e GenAI per la trascrizione, la strutturazione e la sintesi automatica dei contenuti ha abbattuto il costo per intervista di oltre tre quarti rispetto al metodo manuale, eliminando gran parte delle attività ripetitive e minimizzando l'intervento umano. L'incremento di efficienza nei singoli step è stato:

- 3× più veloce nella conversione audio→testo
- 5× più rapido nella strutturazione del testo
- 6× più veloce nell'analisi di parole chiave e sentiment
- 100× più efficiente nell'estrazione e generazione dei riassunti

Questi moltiplicatori si traducono in un risparmio di centinaia di migliaia di euro all'anno in ore-uomo, liberando risorse per attività a maggior valore aggiunto. I benefici operativi e qualitativi sono stati sicuramente: l'efficienza e la scalabilità che grazie all'automazione, il sistema può gestire variazioni improvvise nel volume delle interviste senza richiedere l'ampliamento del team, garantendo tempi di risposta rapidi anche in caso di picchi di attività. La maggiore frequenza di elaborazione consente un time-to-insight più breve, supportando decisioni cliniche e commerciali tempestive. L'accuratezza e l'oggettività che hanno portato l'eliminazione del bias umano nelle fasi di trascrizione e codifica migliora l'affidabilità complessiva dei dati qualitativi. Gli algoritmi di GenAI identificano con elevata precisione speaker, temi e sentiment, restituendo output uniformi e completi rispetto all'approccio manuale.

Infine gli Insight strategici che la piattaforma genera dashboard e report in near real-time, rendendo possibile il monitoraggio continuo delle tendenze nei feedback dei pazienti. I riassunti

automatici favoriscono la condivisione cross-funzionale dei risultati, riducendo drasticamente i tempi di preparazione dei meeting decisionali.

Come funzionava l'architettura del lavoro?

Vi è una prima fase di upload delle registrazioni di interviste sul cloud, seguita dalla trascrizione automatica dei modelli NLP speech-to-text (addestrati su dati clinici) vi è poi lo Structuring & Tagging, ovvero l'organizzazione di testo in entità (es. sintomi, sentiment, topic), l'Insight generation di algoritmi generativi che sintetizzano e segmentano per argomenti chiave, ed infine l'esportazione di un report output in formati standard (Excel, PDF, dashboard BI). Questo caso studio dimostra che l'integrazione di NLP e Generative AI nei processi di analisi qualitativa sarebbe un ottimo investimento in tecnologia, e si ripaga ampiamente grazie ai risparmi operativi e agli impatti qualitativi ottenuti, configurando un modello replicabile in altri contesti di analisi dei dati non strutturati.

3.5 indicatori di performance

Nel valutare l'efficacia dei sistemi di Intelligenza Artificiale in ambito medico, è fondamentale adottare un insieme organico di indicatori di performance che ne garantiscano l'affidabilità, la precisione diagnostica, la capacità discriminativa, l'equità e il rispetto della privacy. Di seguito, vi è una panoramica estesa dei principali indicatori, con metodologia di calcolo, trend evolutivi ed esempi di applicazione.

1. **Accuratezza diagnostica:**

Uno dei fattori principali è ovviamente l'accuratezza dei risultati prodotti dall'IA, come diagnosi basate su immagini o analisi predittive. I consumatori desiderano strumenti che riducano al minimo gli errori diagnostici, garantendo diagnosi precoci e precise. L'accuratezza diagnostica indica la capacità del modello di classificare correttamente i casi positivi e negativi, minimizzando sia i falsi negativi (mancate diagnosi) sia i falsi positivi (allarmi ingiustificati). In medicina, un miglioramento dell'accuratezza impatta direttamente su tempistiche di intervento e outcome clinici. La reliability in un contesto clinico misura la coerenza con cui un modello IA produce

risultati comparabili in condizioni operative variabili (diversi ospedali, apparecchiature, popolazioni).L'evoluzione storica vede dai primi strumenti IA, che risalgono agli anni '70–'90, l'accuratezza inferiore del 50 % sulle analisi condotte, ciò significava che era ancora a in una fase embrionale e a volte poteva dare risultati coerenti, tuttavia senza una rigore logico e costante; oggi, grazie al deep learning e a dataset ampi e diversificati, l'accuratezza nelle applicazioni di imaging oncologico supera spesso il 90 %. Un esempio concreto è dato da uno studio del 2020, che ha evidenziato un chatbot diagnostico con accuratezza del 90,3 %, contro l'80,6 % dei medici in compiti di diagnostica generale.

Di seguito vi è riportata un grafico di confronto tra le curve S di accuratezza nelle diagnosi da parte dei medici (in lilla) e da sistemi di algoritmi di intelligenza artificiale (in rosa) (figura 19).

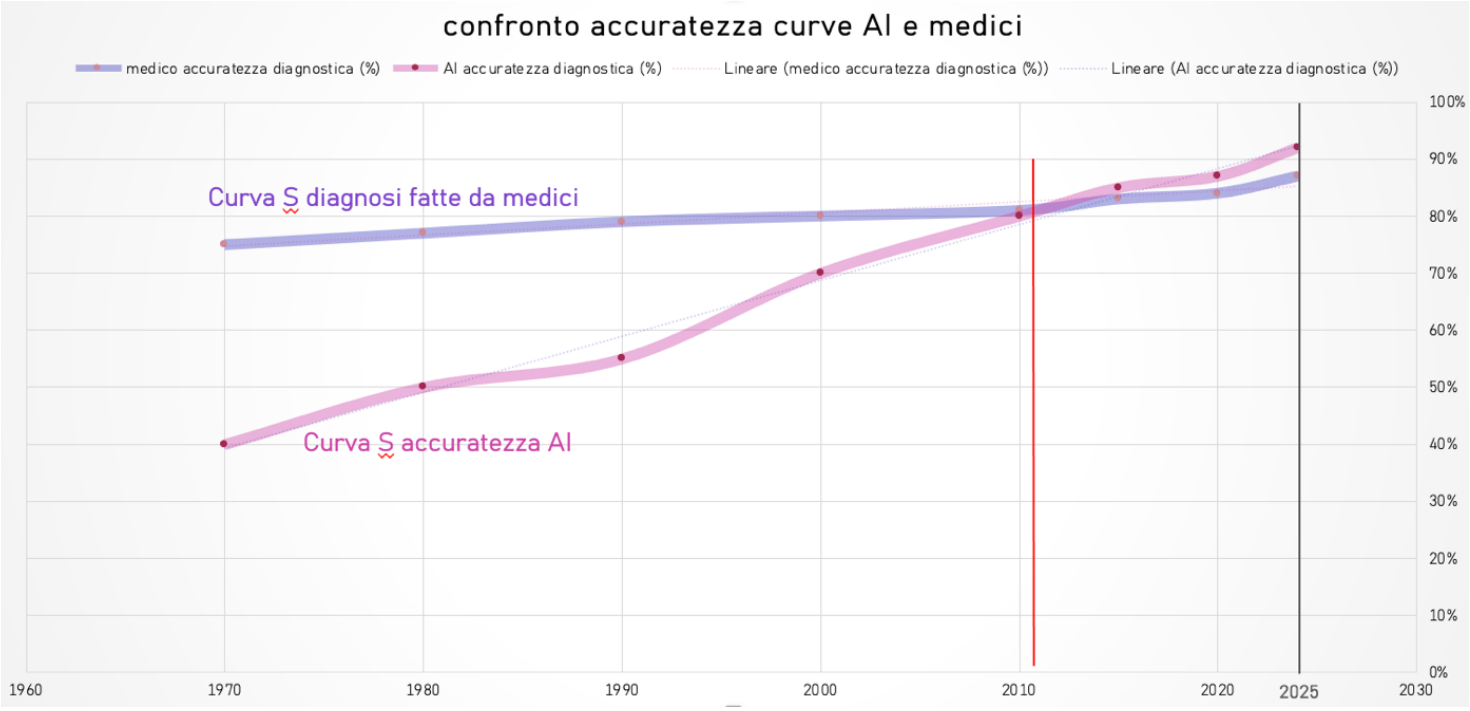


Figura 19 confronto accuratezza curve AI e medici

2. Sfide di bias nei dati e nei medici

I modelli di Intelligenza Artificiale in medicina, se addestrati su dati non rappresentativi, possono amplificare i pregiudizi già presenti nei record clinici e socio-demografici, generando diagnosi o raccomandazioni meno accurate per alcuni sottogruppi di pazienti. In un contesto in cui l'obiettivo

è migliorare la salute di tutti, è cruciale misurare e mitigare questi bias. i dataset storici riflettono spesso disuguaglianze di accesso alle cure e di raccolta dati. Per esempio, le popolazioni caucasiche sono sovrarappresentate in molti studi di imaging, mentre minoranze etniche e fasce di età estreme sono sottorappresentate; di conseguenza, un algoritmo può riconoscere meno efficacemente condizioni patologiche su pelli scure o negli anziani. un modello sbilanciato rischia di generare un maggior numero di falsi negativi (mancate diagnosi) o falsi positivi (allarmi ingiustificati) in sottogruppi svantaggiati, aggravando le diseguaglianze di salute e ostacolando l'equità nei trattamenti. Per quantificare l'entità del problema, molti progetti adottano un Bias Equity Index, cioè una scala da 1 (bias elevato) a 0 (bias minimo), calcolata confrontando metriche di performance (es. sensibilità, specificità) fra diverse categorie demografiche (etnia, genere, età).

Gli obiettivi del monitoraggio sono di identificare i gruppi per cui l'algoritmo performa peggio, grazie all'analisi comparativa di TPR/FPR, ribilanciando i dataset (oversampling delle minoranze, tecniche di re-weighting) o aggiornare l'algoritmo (fairness-aware learning). Inoltre, si deve verificare in fase di validazione clinica che il modello raggiunga soglie di equità paragonabili a quelle ottenute sui gruppi predominanti. Fino al 2015 la consapevolezza era limitata, vi erano pochi studi che riportavano Equity Index, spesso con valori medi > 0,8, a indicare un forte sbilanciamento verso la popolazione di training predominante. Dal 2015 ad oggi invece, c'è stata una crescente attenzione accademica e regolatoria a proposito di questi bias. In dermatologia e imaging radiologico, progetti di data augmentation e raccolta mirata di immagini su pelli di tonalità diverse hanno abbassato l'Equity Index sotto 0,5, dimostrando che con un bilanciamento adeguato è possibile ridurre sensibilmente i bias.

Di seguito vi è una rappresentazione grafica delle curve di come sono cambiate le attenzioni e il valore degli indici di bias negli ultimi anni ().

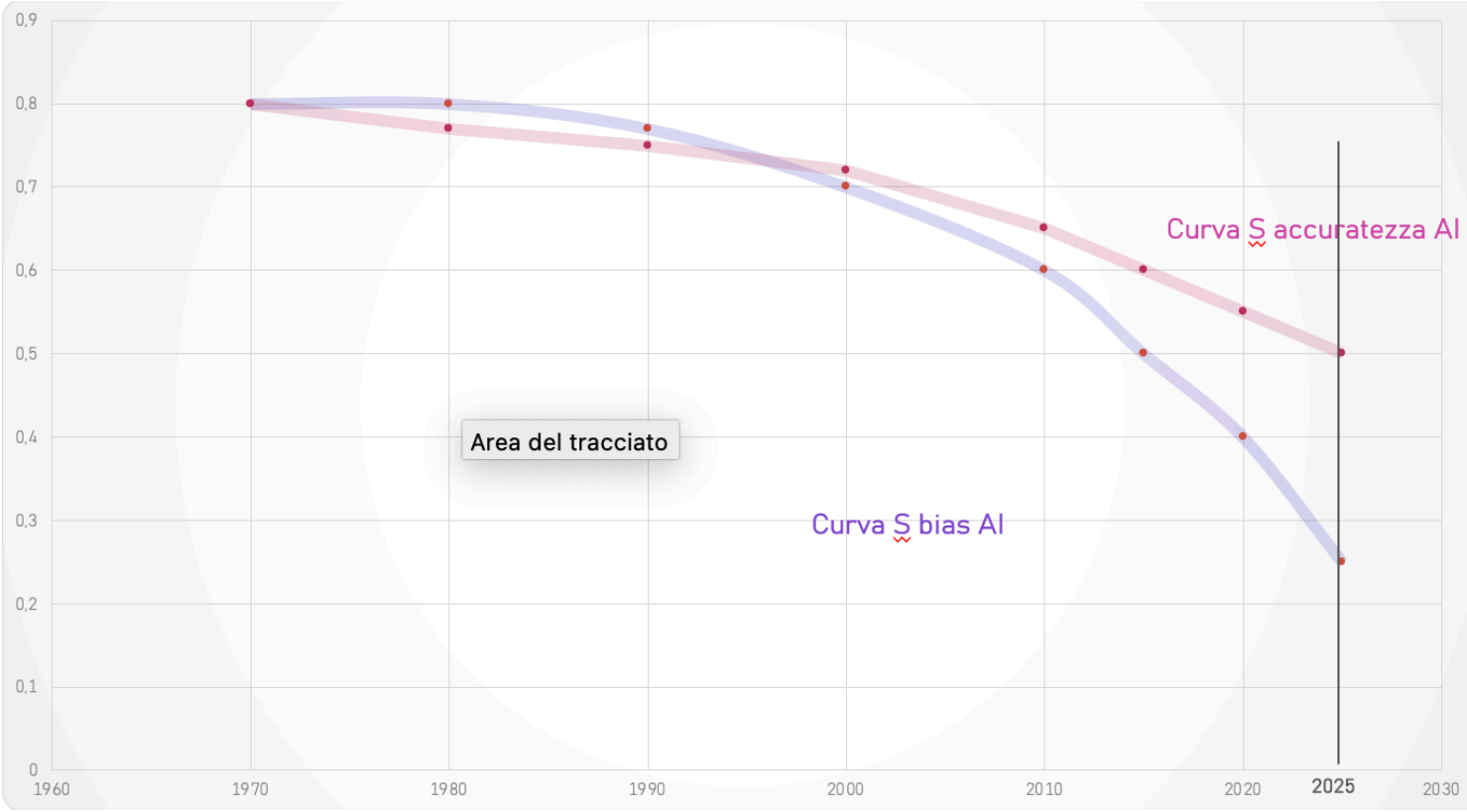


Figura 20 Sfide di bias nei dati e nei medici

3. privacy dei dati

La protezione della privacy dei pazienti è da sempre un pilastro imprescindibile di qualsiasi sistema sanitario. Il diritto alla riservatezza delle informazioni cliniche non è soltanto sancito da normative nazionali e internazionali (GDPR, HIPAA), ma costituisce la base etica su cui si fonda il rapporto di fiducia tra medico e paziente. Il crescente impiego di tecnologie digitali, e in particolare di soluzioni basate su Intelligenza Artificiale (IA), ha moltiplicato le fonti e i volumi di dati trattati, imponendo nuove sfide, come garantire che l'analisi avanzata e l'interoperabilità non compromettano la sicurezza dei dati né il diritto alla riservatezza.

La sicurezza dei dati in forma tradizionale (cartacea) è sempre stata la prassi, l'assenza di archiviazione digitale e di standard normativi rendeva i fascicoli clinici vulnerabili a furti,

smarrimenti o consultazioni non autorizzate. La protezione era affidata banalmente a lucchetti, armadi chiusi e procedure aziendali, facilmente aggirabili. La transizione alla digitalizzazione agli inizi degli anni 2000 portò ai primi Sistemi Informativi Ospedalieri (HIS) e alle cartelle cliniche elettroniche (EHR) che migliorarono decisamente la sicurezza, grazie alla possibilità di controllare gli accessi con user-ID e password. L'adozione capillare del GDPR in Europa e di versioni più mature dell'HIPAA negli USA spinsero verso una cifratura dei dati "at-rest" e "in-transit", oltre a rigorosi meccanismi di controllo degli accessi basati su ruoli (RBAC). Si diffusero così tecniche di pseudonimizzazione e di de-identificazione, fondamentali per l'analisi statistica e la ricerca clinica senza mettere a rischio l'identità dei pazienti. Ogni operazione dunque, su un record medico, veniva tracciata, rendendo possibile identificare eventuali accessi non autorizzati o anomalie operative. L'emergere di tecniche avanzate, come federated learning, differential privacy, homomorphic encryption, ha elevato la protezione dei dati a nuovi livelli, consentendo modelli di IA responsabili che apprendono senza mai trasferire dati sensibili all'esterno. Si è sviluppato un equilibrio più maturo tra utilità dei dati (per addestrare modelli predittivi di alta precisione) e privacy by design, integrando la protezione fin dalle prime fasi di sviluppo degli algoritmi. Le normative di ultima generazione e i framework etici internazionali richiedono alle soluzioni IA sanitarie non solo ottimi indicatori di privacy, ma anche garanzie di trasparenza, auditabilità e possibilità di intervento umano in caso di anomalie.

Il confronto con l'accesso esclusivo dei medici, vede un indicatore di privacy potenziale per i medici di 0,99 % (costante), questo perché? L'accesso alle cartelle cliniche rimane riservato ai soli operatori sanitari abilitati, con credenziali forti e sessioni temporizzate. Questo valore elevato riflette procedure consolidate di controllo identità, formazione obbligatoria sul trattamento dei dati e responsabilità legale. Tuttavia, l'apporto dell'IA modifica il perimetro di accesso: non è più solo il singolo medico a consultare i record, ma anche i moduli algoritmici che possono richiedere dati in forma aggregata o pseudonimizzata. Ciò rende necessaria una governance più articolata, capace di distinguere tra accessi "umani" e "automatici" e di applicare controlli e permessi differenziati.

Di seguito dunque la rappresentazione grafica di come sono distinte le due curve caratterische dei medici e delle intelligenze artificiali per confronto di dati di privacy dei pazienti (figura 21).

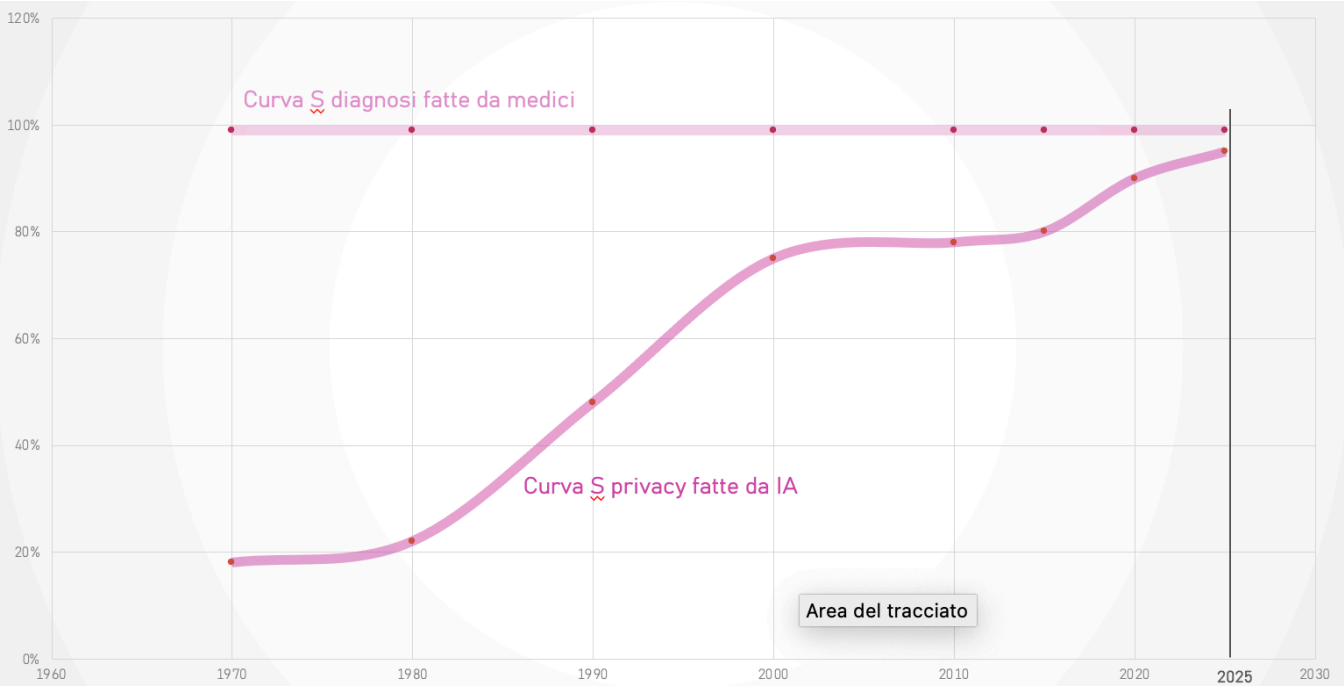


Figura 21privacy dei dati

4. L'AUC-ROC (Area Under the Receiver Operating Characteristic Curve)

L'AUC-ROC è un indicatore fondamentale nella valutazione dei modelli di classificazione, soprattutto in ambito medico, che sintetizza in un unico valore compreso tra 0 e 1 la capacità di un sistema di distinguere correttamente tra casi positivi e negativi. La curva ROC rappresenta graficamente il compromesso tra sensibilità (tasso di veri positivi) e specificità (tasso di falsi negativi) al variare della soglia di decisione; l'area sottesa a questa curva, l'AUC, indica la probabilità che un modello assegni un punteggio più alto a un campione positivo rispetto a uno negativo. Un valore di AUC-ROC pari a 0,5 equivale a una classificazione casuale, mentre valori

superiori a 0,8 indicano performance buone e superiori a 0,9 performance eccellenti; un punteggio prossimo a 1 sottolinea un'elevata capacità discriminativa, per esempio la distinzione tra DNA "normale" e DNA "alterato". Dall'inizio degli anni 2000 fino al 2015, l'introduzione dei primi metodi computazionali avanzati e di approcci di machine learning tradizionali ha favorito un miglioramento graduale delle prestazioni diagnostiche automatizzate, con AUC-ROC medie che si sono mosse da valori intorno a 0,7–0,8 fino a circa 0,85 verso la metà del periodo. Questo progresso è stato reso possibile grazie all'adozione di tecniche statistiche avanzate, all'incremento di potenza di calcolo e alla disponibilità di dataset clinici più ampi e diversificati. A partire dal 2015, l'integrazione completa dell'intelligenza artificiale, in particolare le reti neurali profonde e le metodologie di deep learning, ha determinato un salto qualitativo significativo. Le architetture neurali, supportate da data augmentation, algoritmi di ottimizzazione avanzati e tecniche di regularizzazione, hanno portato la sensibilità dei sistemi diagnostici a livelli prossimi al 90% e l'AUC-ROC media a circa 0,95, avvicinandosi quasi al limite teorico. Questi risultati si sono stabilizzati in virtù di continui affinamenti dei modelli e della crescente disponibilità di dataset di grandi dimensioni e di alta qualità. L'AUC-ROC è particolarmente apprezzata perché non dipende dalla distribuzione delle classi né dalla scelta di una soglia specifica, rendendo le comparazioni tra modelli immediate e affidabili. Inoltre, in ambito clinico, dove minimizzare i falsi negativi è spesso prioritario, questa metrica consente di visualizzare chiaramente il compromesso tra sensibilità e specificità, agevolando decisioni basate su evidenze quantitative. Guardando al futuro, la sfida principale riguarda l'ulteriore miglioramento dell'interpretabilità dei modelli, la validazione su popolazioni ancora più eterogenee e l'implementazione di procedure

standardizzate per garantire la riproducibilità dei risultati in diversi contesti operativi. Uno schema rappresentativo di tutto ciò che è stato detto si può riscontrare in questo grafico (figura 22).

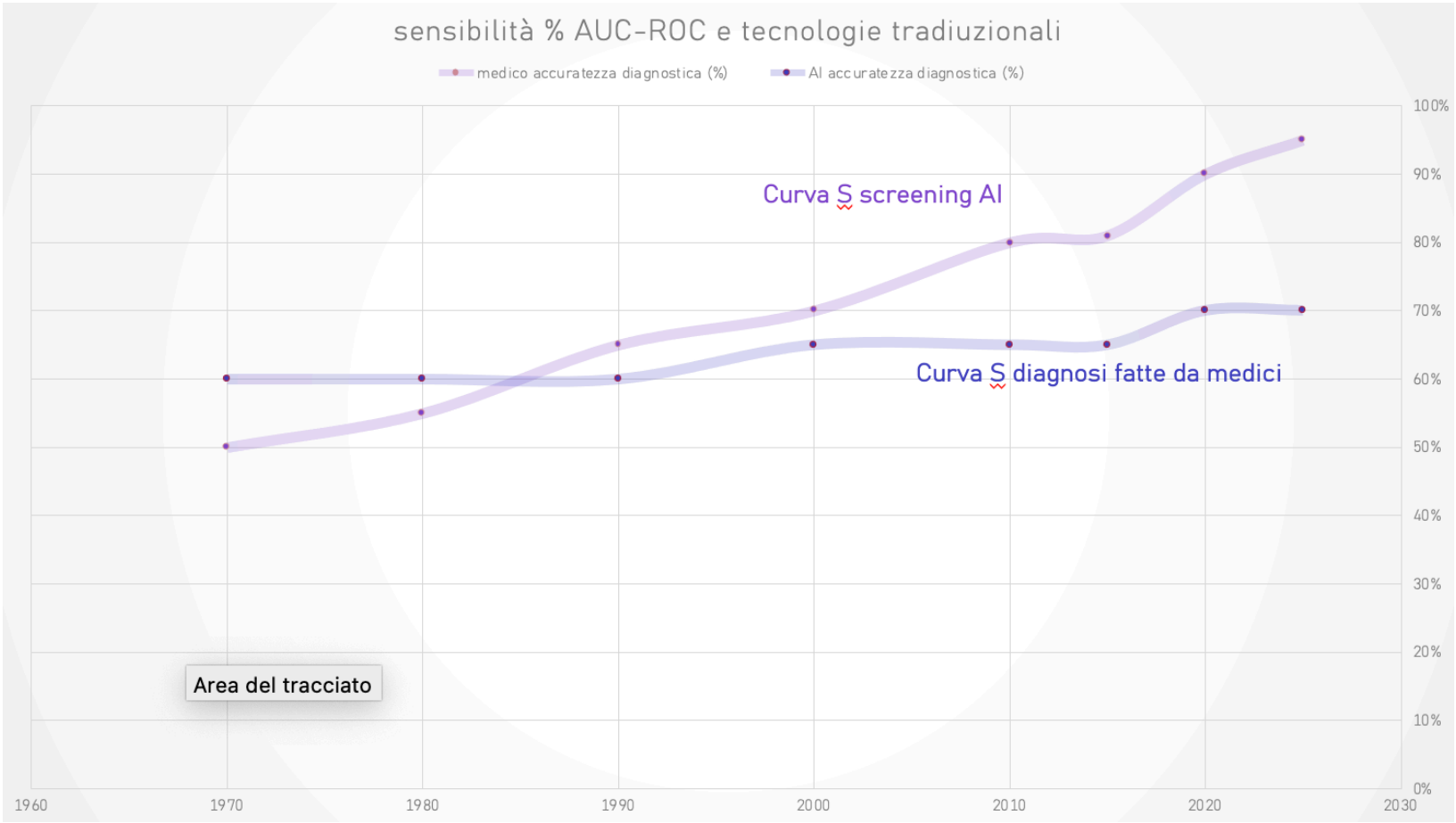


Figura 22 AUC-ROC

Conclusioni e strategie:

Strategie di integrazione (previsione 2025-2030)

La piena integrazione dell'IA in sanità richiede strategie articolate sia da parte delle grandi organizzazioni sanitarie e farmaceutiche, sia dalle startup. Le prime, includendo anche le strutture, cliniche dovrebbero avere innanzitutto l'idea di base di potenziare le infrastrutture digitali e formare gruppi dedicati di data scientist e clinici per supervisione dei progetti IA. Non dovranno sovvertire i flussi operativi quotidiani, bensì introdurre l'IA con un'integrazione senza attriti nei workflows già esistenti. In pratica, l'IA dovrebbe comportarsi come un "nuovo membro del gruppo di lavoro". Per dare un esempio di ciò, basti pensare all'AI come ad un agente che opera attraverso i sistemi ospedalieri esistenti (gestione appuntamenti, triage, fatturazione) senza richiedere complessi interventi. Così facendo si garantirebbe l'adozione graduale e la minima interruzione delle attività. Inoltre, un aspetto da non sottovalutare, consiste nella rapidità di raggiungimento dei risultati tangibili a cui le organizzazioni sanitarie puntano; vale a dire che i progetti di IA di successo sono quelli che portano valore in poche settimane, mostrando da subito riduzioni di tempi ed errori nelle aree ad alto impatto (accettazione, triage, fatturazione). Il concetto dunque sarebbe quello di lasciare al medico il pieno controllo, (un po' come il var nel calcio se vogliamo usare un parallelismo) uno strumento innovativo che potrà portare un secondo parere al medico di ruolo, l'IA non deve funzionare come scatola nera autonoma, ma includere meccanismi di oversight che permettano al personale di monitorare e correggere i processi in tempo reale. Per quanto riguarda le Startup e le PMI innovative, si dovrebbero adottare strategie complementari. Dovrebbero individuare casi d'uso ad alto impatto e ROI rapido come punto di ingresso; perciò, non basterebbe vendere una singola funzionalità, ma occorrerebbe una soluzione che risolva un problema critico e ampliabile ad altri processi. In questo modo una startup può consolidare la propria posizione coinvolgendo più reparti e stakeholder clinici nel tempo. Ad esempio, una piattaforma inizialmente per la diagnostica del tumore al seno (breast imaging) potrebbe evolvere in tool per follow-up e raccolta dati multidisciplinari. Le startup di successo progettano l'offerta come co-sviluppatori anziché come semplici fornitori, in modo tale che i buyer sanitari lavorino assieme agli innovatori (il 64% dichiara volontà di co-sviluppo). Ciò

significa inserire ingegneri e medici del gruppo nelle fasi di test clinico, e soprattutto rendere gli output dell'IA trasparenti e modificabili dai clinici. Ad esempio, una startup potrebbe integrare il proprio algoritmo non solo nella cartella clinica elettronica (EHR) del radiologo, ma anche nei moduli di archiviazione PACS e di refertazione, offrendo così un flusso coerente per immagini, referti e dati clinici collegati.

Di seguito ci sono alcuni esempi di come sono state adottate soluzioni con intelligenza artificiale in ambito clinico/ospedaliero già in utilizzo: restando in Italia, a Milano, l'Istituto Europeo di Oncologia è stato fra i primi centri in Europa ad applicare l'intelligenza artificiale alla Risonanza Magnetica Diffusion Whole Body (RM DWB), consiste in un esame d'avanguardia nella diagnostica per immagini, che, senza radiazioni e senza mezzo di contrasto, può studiare l'intero corpo in una sola seduta, riuscendo a visualizzare lesioni millimetriche. Un sistema di intelligenza artificiale per l'acquisizione delle immagini che non solo migliora ulteriormente la qualità delle immagini, ma permette anche di dimezzare il tempo dell'esame. Lo studio PRAIM condotto in Germania su oltre 460.000 donne ha dimostrato che l'uso della mammografia assistita da intelligenza artificiale, basata su algoritmi di deep learning, ha permesso di individuare il 18% in più di tumori al seno rispetto ai metodi tradizionali, senza aumentare i falsi positivi. Inquira Health, una startup olandese, ha sviluppato un assistente vocale sanitario basato su NLP conversazionale in grado di gestire in modo automatizzato attività cliniche e amministrative. Il voicebot multilingue supporta prenotazioni, accoglienza e follow-up dei pazienti, automatizzando chiamate, promemoria per farmaci e raccolta di dati anagrafici. Questa soluzione migliora l'efficienza operativa e libera tempo prezioso per il personale sanitario. Si dovrebbe dunque puntare sugli apparati clinici e ospedalieri, una volta in concordanza con la legge, per sperimentare questi algoritmi all'interno di workflow clinici già esistenti, portando migliorie dal punto di vista dell'efficienza e qualità di diagnosi, il che si tradurrebbe in miglioramento degli output delle visite di pazienti, che equivale a dire miglioramento generale dello stile di vita e di salute.

Il quadro regolatorio europeo per l'adozione dell'intelligenza artificiale in sanità

L'integrazione dell'intelligenza artificiale nel settore sanitario europeo non può prescindere da un solido e articolato contesto normativo. La delicatezza dei dati trattati, unita all'impatto diretto che le tecnologie IA possono avere sulla salute delle persone, rende necessario un approccio regolatorio chiaro, coerente e in continua evoluzione. In questo panorama, l'Unione Europea ha definito una serie di strumenti giuridici che tracciano con precisione i confini entro cui sviluppare e adottare soluzioni di IA in ambito clinico. Uno dei pilastri normativi fondamentali è il Regolamento Generale sulla Protezione dei Dati (GDPR - Reg. UE 2016/679), che classifica i dati sanitari come "categorie particolari" e ne impone un trattamento fortemente protetto. L'utilizzo di dati clinici per applicazioni basate su IA, come quelli provenienti da cartelle elettroniche o imaging diagnostico, richiede misure specifiche: consenso esplicito del paziente (oppure basi giuridiche equivalenti), pseudonimizzazione, minimizzazione dei dati e obbligatorie valutazioni d'impatto sulla privacy (DPIA). A rafforzare questo impianto, l'European Health Data Space (EHDS) introduce nuovi standard per la gestione e la condivisione dei dati sanitari a fini di ricerca, diagnosi e sanità pubblica, promuovendo al contempo architetture federate e interoperabili a livello europeo. Un secondo punto chiave è rappresentato dal nuovo Regolamento Europeo sull'Intelligenza Artificiale (AI Act - Reg. UE 2024/1689), entrato in vigore nell'agosto 2024. Questo regolamento classifica come "ad alto rischio" tutti i sistemi IA utilizzati in ambito medico, in particolare per attività diagnostiche, monitoraggio clinico o supporto alla prescrizione terapeutica. Tali sistemi devono soddisfare requisiti stringenti: gestione del rischio documentata, dati di addestramento affidabili, trasparenza tecnica, supervisione umana costante e robustezza algoritmica. In pratica, ogni soluzione IA clinica dovrà affrontare una procedura di conformità simile a quella prevista per i dispositivi medici marcati CE, integrando anche nuove responsabilità introdotte dall'AI Act, come i meccanismi di tracciabilità degli algoritmi e la segnalazione obbligatoria di malfunzionamenti. Questo regolamento si inserisce in un ecosistema giuridico già consolidato con l'MDR (Reg. 2017/745) e l'IVDR, che disciplinano rispettivamente i dispositivi medici e diagnostici in vitro. In molti casi, infatti, i software basati su IA per la diagnosi verranno

inquadrati come dispositivi medici a tutti gli effetti, soggetti a validazione clinica, controlli post-marketing e trasparenza nel ciclo di vita. A completare il quadro, il nuovo Product Liability Directive (PLD 2024) estende la responsabilità oggettiva anche ai fornitori di software basati su IA: ciò significa che, in caso di errore (ad esempio, una diagnosi errata generata dal sistema), il produttore dovrà dimostrare la piena conformità alle norme di sicurezza previste. Infine, sono da considerare anche altri atti normativi trasversali, come la Direttiva NIS2 sulla sicurezza informatica per le infrastrutture critiche (in cui rientrano gli ospedali), il Data Governance Act e il Data Act, che promuovono l'accesso e la condivisione dei dati anche in contesti sanitari regolamentati. Sul fronte nazionale, il Garante per la Protezione dei Dati Personali italiano ha già evidenziato la necessità di prevedere ulteriori tutele specifiche per i dati sanitari, in particolare rafforzando l'obbligo di DPIA in caso di utilizzo di IA nei percorsi clinici. In conclusione, per poter essere implementata in modo etico, sicuro ed efficace, ogni soluzione di intelligenza artificiale in ambito sanitario dovrà muoversi all'interno di un sistema regolatorio integrato, che comprenda il GDPR, l'AI Act, la normativa sui dispositivi medici e tutte le direttive correlate. Sarà fondamentale, sin dalle fasi iniziali di progettazione, integrare meccanismi di audit, sicurezza informatica e coinvolgimento attivo degli operatori sanitari, garantendo trasparenza e affidabilità lungo tutto il percorso di sviluppo tecnologico.

Conclusioni finali

Per concludere e tirare le fila di questo lavoro, emerge con forza, il perché l'intelligenza artificiale stia rivoluzionando ogni aspetto della pratica clinica, fondendo diagnostica per immagini, analisi genomica e interazione testuale in un unico flusso integrato. Dall'evoluzione delle reti neurali artificiali alle più recenti architetture multimodali, abbiamo visto come il corretto bilanciamento tra capacità computazionale, qualità dei dati e interpretabilità degli algoritmi sia la chiave per affiancare il medico senza mai sostituirlo. Nei casi studio analizzati, dal melanoma allo screening mammografico, dalla rilevazione dei noduli polmonari fino all'analisi dei discorsi online sulla salute mentale e all'impiego di chatbot per anziani, si è dimostrato che ensemble di CNN e LLM non solo migliorano sensibilità e accuratezza diagnostica, ma alleggeriscono in maniera

significativa il carico di lavoro sui professionisti, dedicando l'intervento umano ai casi più complessi. Il valore di questa trasformazione si è confermato anche sul piano economico: l'esperienza del Sarasota Memorial Health Care System, con un margine di contribuzione per i pazienti ad alto rischio quasi quattro volte superiore alla media nazionale, testimonia che un'adozione ben governata delle tecnologie IA produce benefici clinici e ritorni finanziari concreti. Allo stesso modo, l'uso di dashboard intelligenti per il patient engagement nelle cronicità, capaci di tradurre automaticamente conversazioni in report sintetici, ha reso evidente come l'IA possa diventare un vero braccio destro dei medici, migliorando l'aderenza terapeutica e la qualità della vita dei pazienti. Proprio a titolo di riassunto emblematico di quanto affrontato in questa tesi, non si può non citare l'innovazione MedGemma di Google. MedGemma rappresenta infatti la sintesi perfetta dei nostri tre domini di indagine: sfrutta modelli multimodali in grado di processare contemporaneamente immagini radiologiche ad alta risoluzione, sequenze genomiche e note cliniche testuali, restituendo un'unica interfaccia decision support in cui ogni risultato viene corredato di spiegazioni sul funzionamento interno del modello. Grazie al fine-tuning continuo infatti, si integrano i feedback in tempo reale dai medici e si aggiornano le reti con nuovi dataset anonimi. MedGemma, dunque, non solo aumenta ulteriormente la precisione diagnostica, ma introduce un livello di implementazione senza precedenti, poiché ogni suggerimento di terapia o di approfondimento è accompagnato da una rappresentazione visiva delle caratteristiche chiave che hanno influenzato la predizione, rafforzando la fiducia del medico. Questa innovazione esemplifica magistralmente l'orizzonte verso cui tende la sanità. Un ecosistema in cui prevenzione, diagnosi e cura si realizzano attraverso flussi ibridi, in cui algoritmi avanzati selezionano i casi, operatori sanitari intervengono negli step a più alto valore clinico e i pazienti diventano protagonisti attivi, grazie a interfacce conversazionali e sistemi di monitoraggio continuo basati su wearable. Resta però imprescindibile affrontare le sfide etiche e organizzative, perché si deve garantire la rappresentatività e trasparenza dei dataset per abbattere i bias, definire i protocolli di tutela della privacy conformi all'AI Act europeo, fare investimenti in formazione digitale per medici e pazienti, e promuovere infine collaborazioni tra il settore pubblico e quello privato per infrastrutture di data sharing aperte e interoperabili. Solo adottando un approccio che coniughi rigore scientifico, sostenibilità tecnologica e centralità della persona,

si potrà tradurre in pratica i promettenti risultati sperimentali, inaugurando davvero un'era di sanità intelligente, equa e sostenibile, in cui la tecnologia diventa motore di benessere diffuso.

BIBLIOGRAFIA E SITOGRAFIA

1. https://it.wikipedia.org/wiki/Intelligenza_artificiale#:~:text=L'intelligenza%20artificiale%20(in%20sigla,l'ottimizzazione%20di%20funzioni%20matematiche.
2. Teva ItalyIntelligenza artificiale in medicina: quali vantaggi per i pazienti?Questa tecnologia -che "legge" e analizza gli esiti di risonanze magnetiche, tac o ecografie- può supportare i medici nell'effettuare diagnosi …
3. Mondosanità su IA e medicina
4. Agenda Digitale: Impatti e regolamentazione
5. Previdir: Opportunità future dell'IA
6. Agenda Digitale — IA in medicina: come cambia la diagnosi e il trattamento delle malattie - Agenda Digitaleagendadigitale.eu
7. Agenda Digitale — Intelligenza artificiale in sanità: applicazioni, pro e contro | CGMcgm.com
8. CompuGroup Medical — L'OMS Rivela Il Futuro Dell'IA In Medicina | Mondosanitàmondosanita.it
9. Mondosanità — AI in medicina, tra vantaggi e rischi: la sfida dei Governi sarà tutelare i diritti di tutti - Agenda Digitaleagendadigitale.eu
10. Previdir — L'intelligenza artificiale in medicina: rivoluzione della salute e del benessere - Previdirprevidir.it
11. https://www.fondazioneveronesi.it/magazine/articoli/neuroscienze/lintelligenza-artificiale-in-medicina-tra-presente-e-futuro
12. https://www.cgm.com/ita_it/magazine/articles/intelligenza-artificiale-in-sanita.html
13. https://www.healthtech360.it/salute-digitale/intelligenza-artificiale/intelligenza-artificiale-medicina/
14. https://www.healthtech360.it/salute-digitale/intelligenza-artificiale/intelligenza-artificiale-generativa-in-sanita/
15. https://www.cambridge.org/core/books/cambridge-handbook-of-artificial-intelligence/connectionism-and-neural
16. networks/EB7A3A03C8A440689A6B4D94EA3727D7
17. https://academic.oup.com/edited-volume/34641/chapter/295172549
18. https://academic.oup.com/jamia/article/30/7/1323/7163188
19. Per il prodotto https://www.mdpi.com/2072-6694/16/21/3702
20. https://screenpoint-medical.com/
21. https://transparabreastcare.com/
22. https://screenpoint-medical.com/uncategorized/screenpoint-medical-showcases-transpara-breast-ai-designed-to-improve-cancer-detection-and-support-single-reader-screening-at-ecr/
23. •Imaging Technology News
24. •Transpara Breast Care AI
25. https://en.wikipedia.org/wiki/Mycin
26. https://www.britannica.com/technology/MYCIN
27. https://www.historytools.org/inventions/mycin-expert-system-the-first-ai-medical-diagnosis
28. https://expertsystem101.weebly.com/mycin.html
29. 1.https://www.itnonline.com/content/screenpoint-medical-transpara-breast-ai-demonstrates-value-real-world-clinical-usage
30. https://transparabreastcare.com/why-ai-for-breast-screening/
31. https://direct.mit.edu/books/edited-volume/4798/AI-in-the-1980s-and-BeyondAn-MIT-Survey
32. https://www.cedars-sinai.org/discoveries/ai-ascendance-in-medicine.html
33. https://news.mit.edu/2019/teaching-machines-to-reason-about-what-they-see-0402
34. https://www.mdpi.com/2072-6694/16/21/3702
35. The Official Microsoft Blog
36. Microsoft
37. https://news.nuance.com/2023-03-20-Nuance-and-Microsoft-Announce-the-First-Fully-AI-Automated-Clinical-Documentation-Application-for-Healthcare

38. https://blogs.microsoft.com/blog/2024/03/11/microsoft-makes-the-promise-of-ai-in-healthcare-real-through-new-collaborations-with-healthcare-organizations-and-partners/
39. https://news.microsoft.com/2023/04/17/microsoft-and-epic-expand-strategic-collaboration-with-integration-of-azure-openai-service/
40. https://blogs.microsoft.com/blog/2023/08/22/microsoft-and-epic-expand-ai-collaboration-to-accelerate-generative-ais-impact-in-healthcare-addressing-the-industrys-most-pressing-needs/
41. https://www.microsoft.com/en-us/industry/blog/healthcare/2024/10/10/unlocking-next-generation-ai-capabilities-with-healthcare-ai-models/
42. Recensione Completa sull'IA Generativa in Sanità (Shokrollahi et al., arXiv, 2023).
43. Generative AI in Medicine: Opportunities and Challenges (MDPI, 2023).
44. Pubblicazioni di MIT e Stanford AI Lab su applicazioni di IA medica.
45. Osservatorio Terapie Avanzate
46. https://medicalfuturist.com/5-insights-of-the-medical-futurists-100-digital-health-and-ai-companies-of-2024/
47. https://www.mckinsey.com/
48. https://www.compliancequest.com/
49. HubSpot Blog
50. MobiHealthNewsMedical Startups
51. Omdena
52. HubSpot Blog
53. pmc.ncbi.nlm.nih.govnature.comajmc.comnature.com
54. pmc.ncbi.nlm.nih.govagendadigitale.eusante.gouv.fr.
55. United Arab Emirates University Gomez Rossi et al., JAMA Netw Open 2022;5(3):e220269.
56. scirp.org Gomez Rossi et al., JAMA Netw Open 2022;5(3):e220269 e suppl.
57. refubium.fu-berlin.de Rossi et al., FU-Berlin repository (Markov model).
58. JAMA Network JAMA Dermatology 2025; analysis of progression costs.
59. Hospital Healthcare Europe Beer H., Hospital Healthcare Europe, EADV 2023 congress.
60. The Times, NHS AI instant checks, Apr 2025.
61. thelancet.com The Lancet eClinicalMedicine pilot SkinVision mHealth.
62. JAMA Network JAMA Dermatology 2025 prespecified CEA of 3D TBP.
63. NHS England - Cancro Survey NHS national tariff QALY thresholds.
64. United Arab Emirates University UAEU cost-effectiveness cross-disease study.
65. Insights - Axtria - Ingenious InsightsInsights - Axtria - Ingenious Insights
66. GRAIL
67. Unlocking the Value: Quantifying the Return on Investment of Hospital Artificial Intelligence
68. Clinical Melanoma Diagnosis with Artificial Intelligence: Insights from a Prospective Multicenter Study
69. international evaluation of an AI system for breast cancer screening
70. Leveraging Large Language Models for Patient
71. Engagement: The Power of Conversational AI in
72. Digital Health
73. Artificial intelligence and machine learning
74. in cell-free-DNA-based diagnostics
75. Unlocking the Value: Quantifying the
76. Return on Investment of Hospital
77. Arti cial Intelligence
78. 75% cost savings from Natural Language
79. Processing and Generative AI while
80. evaluating patient experiences