

Previsioni affidabili di apprendimento automatico per supportare la ricerca e le decisioni cliniche

Andrea Bianchi*, Antiniscia Di Marco*, Francesca Marzi*, Giovanni Stilo*, Cristina Pellegrini*, Stefano Masi†, Alessandro Mengozzi‡, Agostino Virdis†, Marco Salvatore Nobile‡, e Marta Simoni‡

*Università dell'Aquila, Italia, †Università di Pisa, Italia, ‡Università Ca' Foscari di Venezia, Italia

Abstract - Al giorno d'oggi, i medici hanno a disposizione un'enorme quantità di dati prodotti da un'ampia serie di test diagnostici e strumentali, integrati da dati ottenuti con tecnologie d'avanguardia. Se tali dati fossero opportunamente collegati e analizzati, potrebbero essere utilizzati per rafforzare le previsioni, in modo da migliorare la prevenzione e il tempo di diagnosi, ridurre i costi del sistema sanitario e far emergere le conoscenze nascoste. L'apprendimento automatico è la tecnica principale utilizzata oggi per sfruttare i dati e ottenere informazioni utili. Tuttavia, ha portato a diverse sfide, come il miglioramento dell'interpretabilità e della spiegabilità dei modelli predittivi impiegati e l'integrazione delle conoscenze degli esperti nel sistema finale. Risolvere queste sfide è di fondamentale importanza per aumentare la fiducia di medici e pazienti nelle previsioni del sistema. Per risolvere questi problemi, in questo lavoro proponiamo un software in grado di gestire gli aspetti di affidabilità dei modelli di apprendimento automatico e di considerare una moltitudine di dati e modelli eterogenei.

Termini dell'indice: dati eterogenei, apprendimento automatico, capacità di spiegazione, interpretabilità, previsione del rischio, sistema di supporto alle decisioni cliniche.

I. INTRODUZIONE

I sistemi sanitari del XXI secolo si trovano ad affrontare una dura sfida a causa dell'invecchiamento della popolazione. Un esempio è dato dalle malattie cardiovascolari (CVD), che contano 17,9 milioni di decessi all'anno. Secondo recenti previsioni, la situazione sta peggiorando: la prevalenza della malattia metabolica aumenterà e nel 2030 1 adulto statunitense su 2 soffrirà di obesità [25]. Con l'invecchiamento della popolazione [8], prevarranno i fenotipi cardiovascolari ad alto rischio, con un impatto negativo sulla mortalità e sulla morbilità cardiovascolare. Queste considerazioni possono essere estese a diverse malattie croniche (ad esempio cancro, malattie neurologiche e autoimmuni) che, se non prevenute, sono destinate a gravare sui sistemi sanitari. Affrontare questa sfida significa utilizzare adeguatamente le risorse e i dati medici per fornire modelli di previsione del rischio di malattia accurati, fattibili e facilmente implementabili. L'apprendimento automatico (ML) può essere utilizzato per estrarre i dati nascosti e migliorare la previsione

Questo lavoro è supportato dall'Unione Europea - NextGenerationEU - Piano Nazionale di Ripresa e Resilienza (PNRR) - Progetto: "SoBigData.it - Rafforzamento della RI italiana per il Social Mining e Big Data Analytics" - Prot. IR0000013 - Avviso n. 3264 del 28/12/2021, da "ICSC - Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing", finanziato dall'Unione Europea - NextGenerationEU, e da LIFEMAP-Dalla patologia pediatrica alle malattie cardiovascolari e neoplastiche nell'adulto: mappatura genomica per la medicina e prevenzione personalizzata Traiettorie 3 "Medicina rigenerativa, predittiva e personalizzata" - Linea di azione 3.1 "Creazione di un programma di prevenzione".1 "Creazione di un programma di medicina di precisione per la mappatura del genoma umano su scala nazionale" del Ministero della Salute.

delle malattie, come la CVD [12], [19], [22].

Tuttavia, c'è spesso una forte mancanza di fiducia quando i risultati sono ottenuti applicando metodi "black-box" che non possono spiegare, o motivare, la loro risposta. Inoltre, molti degli approcci proposti lavorano su una singola fonte di dati. In [17], invece, gli autori sottolineano l'importanza di integrare in modo efficiente ed efficace tutti i tipi di dati sanitari, dalle immagini ai dati clinici e omici, per ottenere una diagnosi accurata o suggerire un trattamento efficace. Oggi i medici dispongono di numerosi punteggi di rischio clinico. Tuttavia, questi soffrono di tre difetti principali: i) non riescono ancora a identificare i fenotipi ad alto rischio; ii) scambiano l'accuratezza con la fattibilità: l'avvento dei punteggi di rischio genetico e poligenico ha fornito buone stime del rischio nell'arco della vita che sono poco utili se il paziente è già affetto dalla malattia [3]; iii) sono spesso percepiti come oscuri, limitando la fiducia sia dei medici sia dei pazienti. A questo scenario si aggiunge il fatto che i dati dei pazienti sono eterogenei e spesso non disponibili nella loro totalità sia per il medico che per il ricercatore.

Il presente lavoro si propone di descrivere una metodologia per la definizione di sistemi di ML in ambito sanitario che sfruttino fonti di dati eterogenee e generino risultati di predizione accurati, interpretabili e spiegabili. La metodologia fa uso di conoscenze specifiche del dominio - in questo caso le conoscenze mediche - per garantire previsioni robuste. Ci aspettiamo che la nostra metodologia, alimentata dai dati della popolazione e dalle raccomandazioni delle linee guida, sia in grado di suggerire al medico i percorsi più efficaci, affidabili, diretti ed economici che portano a una diagnosi e a un trattamento finali accurati e su misura per il paziente. Sebbene la presentazione di risultati specifici o la condivisione di dati esuli dallo scopo di questo articolo, stiamo lavorando per validare il nostro approccio utilizzando dati del mondo reale nel prossimo futuro.

II. PROCESSO SOFTWARE PER SISTEMI DI PREVISIONE DELLA SALUTE AFFIDABILI

In questa sezione, descriviamo il flusso di lavoro per le previsioni ML affidabili mostrato nella Figura 1, che garantisce l'interpretabilità e la spiegabilità delle previsioni facendo leva sulle pipeline di apprendimento. Riassumiamo i requisiti e i vincoli da considerare nella progettazione del flusso di lavoro (Sezione II-A) e descriviamo il processo software che proponiamo per raggiungere la sfida di una predizione ML affidabile in caso di fonti di dati multiple (Sezione II-B).

A. Sfide

Obiettivo #1: i recenti progressi nell'apprendimento multimodale hanno dimostrato che l'integrazione di diversi tipi di dati (ad esempio dati clinici, diagnostica molecolare, immagini radiologiche e istologiche) porta a risultati promettenti. In [5] e [13] gli autori dimostrano che i modelli di previsione che combinano più tipi di dati ottengono risultati migliori rispetto ai modelli che considerano solo un singolo tipo di dati. Tuttavia, negli scenari reali, l'accessibilità a diverse fonti di dati è spesso limitata, il che rende difficile eseguire un'integrazione e un'analisi significative dei dati. Quando è necessario analizzare dati eterogenei provenienti da fonti diverse attraverso tecniche di ML, un'unica pipeline non è adeguata a generare i risultati, soprattutto se si vogliono imporre spiegabilità e interpretabilità. Infatti, i sistemi che applicano queste tecniche a contesti reali devono utilizzare modelli complessi che possono combinare più sotto-modelli più semplici.

Obiettivo #2: Immaginate uno scenario in cui sono disponibili varie fonti di dati, ma i dati sono archiviati in luoghi fisici diversi e, per qualche motivo, non possono essere distribuiti liberamente. Questa è una situazione tipica del settore sanitario, in cui i pazienti vengono sottoposti a vari esami diagnostici in diversi ospedali o cliniche che non condividono lo stesso archivio per i risultati. In questo scenario, è necessario imparare da ogni singola fonte separatamente e, successivamente, mettere insieme tutte le previsioni ottenute per addestrare un modello olistico che chiamiamo *iper-modello*. Maggiori dettagli sul processo completo sono forniti nella Sezione II-B.

Obiettivo #3: Data la complessità dei modelli matematici utilizzati nelle pipeline, è importante garantire che i loro risultati possano essere interpretati e spiegati correttamente. Nel settore sanitario, questo requisito è obbligatorio poiché l'uso di modelli predittivi black-box compromette la fiducia dei medici e ostacola il "diritto alla spiegazione" per i pazienti.

Obiettivo #4: Un'altra questione è come migliorare i risultati della predizione introducendo conoscenze aggiuntive. Documenti di ricerca come [9], [24] descrivono l'importanza di integrare la conoscenza all'interno delle pipeline di apprendimento per: i) ridurre la quantità di dati necessari, ii) rendere gli approcci più robusti e iii) creare sistemi di apprendimento interpretabili e spiegabili. Come affermato in [11], una delle opportunità nell'applicazione di approcci intelligenti e tecniche di integrazione dei dati è quella di coinvolgere gli esseri umani nel ciclo per condurre l'etichettatura e la verifica dei dati e dei risultati. In linea con questo suggerimento, ci proponiamo di aumentare l'affidabilità delle fasi della pipeline all'interno del processo introducendo le conoscenze degli esperti di dominio.

La nostra soluzione per affrontare queste sfide prevede l'implementazione di un flusso di lavoro software gerarchico per il sistema di predizione. Questo processo inizia con pipeline di apprendimento individuali che lavorano su fonti di dati distinte. Successivamente, l'ipermodello raccoglie e unisce le previsioni di più modelli utilizzando una funzione di aggregazione dedicata. L'obiettivo principale dell'ipermodello è quello di utilizzare le previsioni di più modelli per prendere una decisione finale informata che produca un risultato complessivo di previsione.

L'approccio proposto non deve essere confuso con l'apprendimento federato.

Il semplice apprendimento federato prevede la condivisione dello stesso modello di dati tra i partecipanti, con il risultato di ottenere previsioni identiche [28]. L'approccio proposto, invece, utilizza una pipeline di apprendimento predefinita per ciascuna fonte di dati, in cui i set di dati di addestramento divergono a causa delle differenze tra luoghi fisici e tipi e volumi di dati. A differenza del puro ensemble learning, che applica diversi modelli allo stesso sottoinsieme di dati [10], l'approccio proposto combina le previsioni ottenute da diversi modelli addestrati su diversi formati e volumi di dati. L'obiettivo è quello di migliorare l'accuratezza delle previsioni, fornendo al contempo informazioni sul processo decisionale dell'ensemble. Inoltre, l'approccio proposto incorpora i concetti di interpretabilità e spiegabilità per migliorare la comprensione del processo decisionale dell'ensemble.

B. Flusso di lavoro dei sistemi di ML affidabili

In questa sottosezione, presentiamo una spiegazione dettagliata delle fasi coinvolte nel nostro approccio alla progettazione di sistemi ML affidabili, tenendo conto delle sfide e dei vincoli illustrati nella sottosezione precedente. Il processo proposto è generale e particolarmente rilevante nel settore medico. Può essere applicato a qualsiasi ambito in cui i dati provenienti da fonti diverse non possono essere condivisi e in cui si presentano le sfide citate.

Il flusso di lavoro che proponiamo è mostrato nella Figura 1 ed è progettato per gestire dati provenienti da fonti eterogenee che differiscono per volume e tipologia. Per affrontare questa sfida, applichiamo una pipeline di apprendimento a ciascuna fonte di dati. Ogni pipeline comprende diverse fasi: pre-elaborazione dei dati, ingegnerizzazione delle caratteristiche, formazione del modello e valutazione. Questo approccio viene ripetuto per tutte le fonti di dati e porta a una serie di modelli ottimizzati per la fonte di dati specifica. Per combinare le previsioni dei diversi modelli, utilizziamo una funzione di aggregazione (cioè il riquadro di aggregazione nella Figura 1) che tiene conto delle singole previsioni (a sinistra della Figura 1). La fase di aggregazione può andare dalla semplice fusione dei dati all'unificazione dei formati e all'ensemble. Questo approccio mira a garantire una maggiore robustezza e accuratezza delle valutazioni nelle fasi successive del flusso di lavoro. Per garantire l'attendibilità delle previsioni, utilizziamo una strategia impilata per addestrare un nuovo modello - il riquadro dell'iper-modello nella Figura 1 - che incorpora una nuova entità che rappresenta il concetto di spiegabilità, a partire dalle singole previsioni aggregate. In letteratura, diversi articoli descrivono possibili flussi di lavoro di pipeline di ML [2], [16]. L'ipermodello qui proposto si basa sul flusso di lavoro di ML definito in [2]. Tuttavia, la pipeline di ML viene estesa con ulteriori passaggi necessari per progettare spiegabilità e interoperabilità. Inoltre, introduciamo esplicitamente il concetto di conoscenza del dominio come entità dinamica che alimenta le fasi della pipeline e che si evolve di conseguenza quando nuove conoscenze vengono messe a disposizione dal sistema, producendo alla fine previsioni future più accurate e avanzate. La pipeline di apprendimento avanzato che implementa l'iper-modello - si veda il riquadro dell'iper-modello nella Figura 1 - è composta da nove passi (i nove riquadri gialli enumerati nella figura), uno dei quali è nuovo rispetto a [2] (il riquadro verde) e quattro passi sono stati ampliati con la conoscenza del dominio.

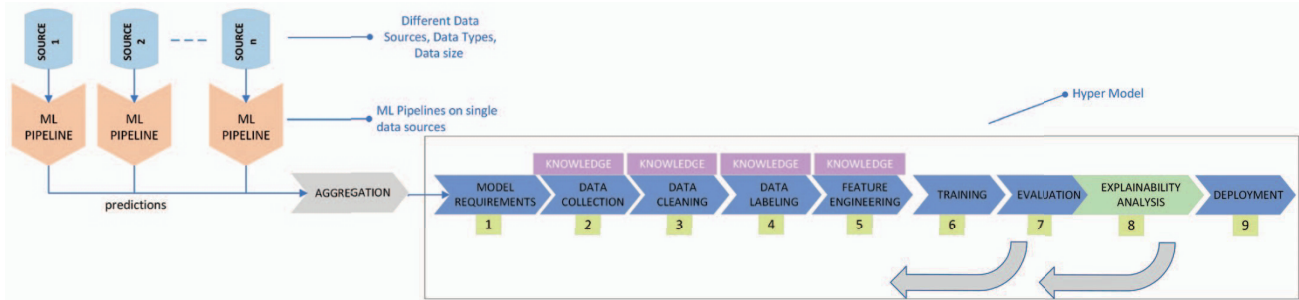


Fig. 1. Flusso di lavoro per previsioni ML affidabili.

L'idea è quella di integrare la conoscenza del dominio utilizzando regole e vincoli di esperti, modellati con sistemi basati su regole e alberi decisionali. In medicina, le regole degli esperti possono essere utilizzate per limitare le previsioni del modello in base a specifiche linee guida mediche. Questa conoscenza si evolve nel tempo, incorporando nuove scoperte ottenute dall'analisi dei dati o ottenendo informazioni da nuovi studi clinici. Per completezza e seguendo l'ordine della Figura 1, descriviamo brevemente ogni fase della pipeline ML avanzata. *I requisiti del modello* sono guidati dalle richieste degli stakeholder del sistema. Alla fine di questa fase, tutti gli stakeholder dovrebbero avere una visione chiara del problema da risolvere, quali parti potrebbero essere implementate sfruttando le tecniche di apprendimento automatico esistenti e un elenco di possibili modelli da utilizzare per risolvere il problema.

Nella *raccolta dei dati*, gli stakeholder controllano i dati ed eventualmente cercano altri dati da utilizzare nell'analisi.

La *pulizia dei dati* è il processo di rimozione o correzione di informazioni incorrette, imprecise, non ben formattate, duplicate o incomplete all'interno del set di dati.

Nell'*etichettatura dei dati*, le etichette di verità vengono assegnate a ciascun elemento del set di dati. Si tratta di una fase molto importante, perché consente ai metodi di apprendimento supervisionati di funzionare al meglio.

L'*ingegneria delle caratteristiche* si riferisce a tutte le attività necessarie per estrarre, selezionare e trasformare i dati iniziali in caratteristiche informative, con l'obiettivo di semplificare e velocizzare la trasformazione dei dati e far funzionare bene le tecniche di apprendimento nei vari compiti.

Durante la *formazione*, i modelli vengono addestrati e messi a punto iterativamente sui dati pre-elaborati.

Al termine dell'addestramento, viene eseguita la fase di *valutazione*. Tutte le parti interessate devono valutare i risultati del modello su diversi set di dati di prova in base a metriche predefinite. Questa fase è molto importante perché consente di dimostrare (in un ambiente di test) che il modello scelto soddisfa le esigenze degli stakeholder e altri requisiti normativi o etici del sistema.

La fase successiva, denominata *Analisi della spiegabilità*, è necessaria per soddisfare i requisiti di spiegabilità. In questa fase si possono scegliere due alternative di alto livello in base al precedente metodo di apprendimento automatico utilizzato nella pipeline. Se la tecnica è *interpretabile in base alla progettazione*, questa fase si concentra sulla misurazione dell'efficacia dell'interpretazione fornita.

Se il metodo di predizione utilizzato è di tipo black-box¹, in questa fase deve essere utilizzato un metodo di spiegabilità. Anche se il tipo di explainer dipende (si veda [6], [7], [15], [18] per una discussione più ampia, dal tipo di dati (immagini, grafici, tabelle, sequenze e tempo), dal livello di spiegazione desiderato (locale o globale) e dal tipo di spiegazione desiderata (ad es, locale o globale) e dal tipo di spiegazione desiderata (fattuale, controfattuale, prototipica, ecc.), va notato che la spiegabilità può essere ottenuta attraverso un modello post-hoc in modo diverso dall'interpretabilità che il modello di previsione stesso fornisce. Ad esempio, nel caso di studi biomedici traslazionali (come quelli relativi alla CVD), è importante considerare le reti omiche. In questo caso, il predittore di riferimento è basato su reti neurali grafiche (GNN), come discusso in [20], e il tipo di spiegabilità è di tipo controfattuale (gli spiegatori controfattuali per le GNN si basano principalmente su approcci euristici di perturbazione, come in [1], [4], [26]). Inoltre, in letteratura sono stati presentati diversi framework di valutazione per alcuni domini specifici [21], [27], ma siamo lontani dall'aver un "coltellino svizzero" per ogni circostanza. Nonostante il dibattito in corso e gli sforzi compiuti per chiarire l'area (si veda [23]), non esiste un processo consolidato per integrare e valutare i metodi di spiegabilità in un processo software più ampio. Una volta terminata con successo l'*analisi di spiegabilità*, verrà eseguito il *deployment*.

Per spiegare meglio come si può ottenere l'interpretabilità, facciamo riferimento al caso della CVD. Supponiamo che il risultato di tutti i (sotto)modelli possa essere unito in un ulteriore iper-modello fuzzy interpretabile, che fornisca la previsione finale del rischio di CVD. Le regole fuzzy dell'ipermodello saranno basate su variabili antecedenti corrispondenti alle previsioni dei sottomodelli. La generazione delle variabili linguistiche e la calibrazione del modello (ad esempio, i parametri degli insiemi fuzzy, i coefficienti conseguenti) saranno sia guidate dai dati [14] sia raffinate sulla base di conoscenza estesa. Quindi, l'iper-modello sarà il risultato di un approccio olistico alla CVD, in grado di sfruttare le conoscenze degli esperti del dominio - conoscenze cliniche - per collegare semanticamente i risultati di previsione ottenuti da diversi modelli.

¹Ad esempio, le Reti Neurali Artificiali e le loro evoluzioni.

III. CONCLUSIONI E LAVORO FUTURO

Diversi studi nel campo delle malattie cardiovascolari hanno sottolineato l'importanza di prevedere il rischio cardiovascolare utilizzando una varietà di fonti di dati, il che richiede l'applicazione di varie tecniche di ML. Questa è in realtà una caratteristica generale del dominio sanitario che porta a una nuova sfida in termini di organizzazione del sistema e di combinazione dei risultati per ottenere un risultato finale di previsione. Inoltre, nel settore sanitario è fondamentale garantire l'interpretabilità o la spiegabilità di tutte le previsioni, data l'eterogeneità delle fonti di dati e la loro limitata accessibilità a causa di problemi etici e organizzativi. In questo studio abbiamo presentato un flusso di lavoro completo che prevede l'uso di più pipeline, il cui numero corrisponde al numero di fonti di dati eterogenee. Ogni pipeline segue un processo di apprendimento che prende in ingresso una specifica fonte di dati e fornisce in uscita una previsione. I risultati generati da tutte le pipeline vengono poi aggregati e utilizzati per addestrare un iper-modello, progettato per fornire previsioni altamente accurate e robuste, implementando spiegabilità e interoperabilità. Abbiamo anche integrato le conoscenze mediche nelle pipeline di apprendimento per garantire l'affidabilità delle previsioni. Tale integrazione mira a migliorare l'accuratezza delle previsioni, vincolando le previsioni sulla base di conoscenze mediche e linee guida specifiche. L'approccio qui presentato per il settore sanitario può essere applicato anche in altri contesti, a condizione che siano disponibili le conoscenze degli esperti del settore.

Il lavoro futuro prevede l'applicazione del design da noi suggerito a un set di dati reali, nello specifico il database della UK Biobank, che contiene informazioni a livello di paziente su oltre 500.000 persone. Il database ha un design longitudinale con dati prognostici sostanziali sugli esiti cardiovascolari, osservazioni in numerosi periodi di tempo e dati multilivello che comprendono dati antropometrici, biochimici, di imaging e genetici. Il periodo di follow-up è lungo (15 anni). Utilizzando questo set di dati, speriamo di illustrare l'utilità della strategia da noi proposta in un contesto pratico e di offrire spunti per l'integrazione di diverse fonti di dati per la ricerca sulle malattie cardiovascolari.

RIFERIMENTI

- [1] C. Abrate e F. Bonchi. Grafi controfattuali per la classificazione spiegabile delle reti cerebrali. In *Proc. of the 27th ACM SIGKDD Conf. on Knowl. Disc. & Data Mining*, pagine 2495-2504, 2021.
- [2] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi e Thomas Zimmermann. Ingegneria del software per l'apprendimento automatico: Un caso di studio. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice*, pagine 291-300, 2019.
- [3] Natalie Arnold e Wolfgang Koenig. Punteggio di rischio poligenico: Strumento clinicamente utile per la previsione della malattia cardiovascolare e del beneficio della terapia lipidica? *Cardiovasc. Drugs Ther.*, 35(3):627-635, 2021.
- [4] M. Bajaj, L. Chu, Z.Y. Xue, J. Pei, L. Wang, P.C.H Lam e Y. Zhang. Spiegazioni controfattuali robuste su reti neurali a grafo. *Advances in Neural Inf. Proc. Sys.*, 34, 2021.
- [5] Kevin M Boehm, Pegah Khosravi, Rami Vanguri, Jianjiong Gao e Sohrab P Shah. Sfruttare l'integrazione dei dati multimodali per far progredire l'oncologia di precisione. *Nature Reviews Cancer*, 22(2):114-126, 2022.
- [6] David L Buckeridge. *L'IA spiegabile nell'assistenza sanitaria e nella medicina: Building a Culture of Transparency and Accountability*. Springer, 2021.
- [7] Nadia Burkart e Marco F. Huber. Un'indagine sulla spiegabilità dell'apprendimento automatico supervisionato. *J. Artif. Int. Res.*, 70:245-317, 2021.
- [8] Sarah Costantino, Francesco Paneni e Francesco Cosentino. Invecchiamento, metabolismo e malattie cardiovascolari. *J. Physiol.*, 594(8):2061-2073, 2016.
- [9] Changyu Deng, Xunbi Ji, Colton Rainey, Jianyu Zhang e Wei Lu. Integrare l'apprendimento automatico con la conoscenza umana. *iScience*, 23(11):101656, 2020.
- [10] Xibin Dong, Zhiwen Yu, Wenming Cao, Yifan Shi e Qianli Ma. Un'indagine sull'apprendimento d'insieme. *Front. Comput. Sci.*, 14(2):241-258, aprile 2020.
- [11] Xin Luna Dong e Theodoros Rekatsinas. Integrazione dei dati e apprendimento automatico: Una sinergia naturale. *Proc. VLDB Endow.*, 11(12):2094-2097, agosto 2018.
- [12] Ben Ali et al. Implementazione dell'apprendimento automatico nella cardiologia interventistica: I vantaggi valgono la pena. *Frontiers in Cardiovascular Medicine*, 8, 2021.
- [13] Jabbour et al. Combinazione di radiografie del torace e dati di cartelle cliniche elettroniche (ehr) mediante apprendimento automatico per la diagnosi di insufficienza respiratoria acuta. *Journal of the American Medical Informatics Association*, 29(6):1060-1068, 2022.
- [14] Caro Fuchs, Simone Spolaor, Marco S. Nobile e Uzay Kaymak. pyfume: un pacchetto python per la stima di modelli fuzzy. In *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pagine 1-8, 2020.
- [15] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti e Dino Pedreschi. Un'indagine sui metodi di spiegazione dei modelli a scatola nera. *ACM computing surveys (CSUR)*, 51(5):1-42, 2018.
- [16] Mark Haakman, Lu's Cruz, Hennie Huijgens e Arie van Deursen. I modelli del ciclo di vita delle imprese devono essere rivisti: Uno studio esplorativo nel settore fintech. *Empirical Softw. Engg.*, 26(5), 2021.
- [17] Andreas Holzinger, Benjamin Haibe-Kains e Igor Jurisica. Perché i dati di imaging da soli non bastano: Integrazione basata sull'AI di dati di imaging, omici e clinici. *European Journal of Nuclear Medicine and Molecular Imaging*, 46, 12 2019.
- [18] Mir Riyanul Islam, Mobyen Uddin Ahmed, Shaibal Barua e Shahina Begum. Una revisione sistematica dell'intelligenza artificiale spiegabile in termini di diversi domini applicativi e compiti. *Scienze applicate*, 12(3):1353, 2022.
- [19] Chayakrit Krittanawong, Hafeez Hassan Virk, Sripal Bangalore, Zhen Wang, Kipp Johnson, Rachel Pinotti, Hongju Zhang, Scott Kaplin, Bharat Narasimhan, Takeshi Kitai, Usman Baber, Jonathan Halperin e W.H. Tang. Previsione dell'apprendimento automatico nelle malattie cardiovascolari: una meta-analisi. *Scientific Reports*, 10, 09 2020.
- [20] L. Madeddu e G. Stilo. *Metodi di Deep Learning nella biologia delle reti*. 2022.
- [21] Mario Alfonso Prado-Romero e Giovanni Stilo. Gretel: framework per la valutazione delle spiegazioni controfattuali dei grafici. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pagg. 4389-4393, 2022.
- [22] Giorgio Quer, Ramy Arnaout, Michael Henne e Rima Arnaout. L'apprendimento automatico e il futuro delle cure cardiovascolari: Revisione dello stato dell'arte di Jacc. *Journal of the American College of Cardiology*, 77(3):300-313, 2021.
- [23] Chakkrit Kla Tantithamthavorn e Jirayus Jirapakdee. Explainable ai per l'ingegneria del software. In *2021 36a Conferenza internazionale IEEE/ACM sull'ingegneria del software automatizzato (ASE)*, pagine 1-2, 2021.
- [24] Laura von Rueden et al. Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering*, pagine 1-1, 2021.
- [25] Zachary J Ward, Sara N Bleich, Angie L Craddock, Jessica L Barrett, Catherine M Giles, Chasmine Flax, Michael W Long e Steven L Gortmaker. Previsione della prevalenza a livello statale dell'obesità adulta e dell'obesità grave negli Stati Uniti. *N. Engl. J. Med.*, 381(25):2440-2450, dicembre 2019.
- [26] G. P. Wellawatte, A. Seshadri e A. D. White. Generazione agnostica di spiegazioni controfattuali per le molecole. *Chemical Science*, 13(13):3697-3705, 2022.
- [27] H. Yuan, H. Yu, S. Gui e S. Ji. Spiegabilità nelle reti neurali a grafo: Un'indagine tassonomica. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2022.
- [28] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li e Yuan Gao. Un'indagine sull'apprendimento federato. *Knowledge-Based Systems*, 216:106775, 2021.