Big Data e Intelligenza Artificiale in Medicina di Laboratorio

Roberto Guerranti[1,2], Tommaso Fasano[3], Elvira Inglese[4,21], Claudia Bellini[5], Luisa Lanzilao[6], Alessio Mancini[7], Giuseppe Banfi[8,9], Mario Plebani[10,11], Mario Ciampi[12], Stefano Dalmiani[13], Maria Teresa Chiaravalloti[14], Nicoletta Musacchio[15], Elena Stanghellini[16], Davide Giavarina[17], Giovanni Riva[18], Andrea Padoan[19,20]

[1] Dipartimento Innovazione, Sperimentazione e Ricerca Clinica e Traslazionale, Laboratorio Patologia Clinica, Azienda Ospedaliera Universitaria Senese, Siena
[2] Dipartimento Biotecnologie Mediche, Università degli Studi di Siena
[3] AUSL Romagna – U.O. Patologia Clinica
[4] Laboratorio Analisi Chimico-Cliniche, ASST Grande Ospedale Metropolitano Niguarda - Milano
[5] UOC Laboratorio Analisi Chimico-Cliniche, PO Misericordia Grosseto, AUSL Toscana Sud Est
[6] Laboratorio Generale, AOU-Careggi, Firenze
[7] ASUR Marche AV2, U.O.S.D. Proteine specifiche, Senigallia (AN)
[8] IRCCS Galeazzi, Milano
[9] Università Vita e Salute San Raffaele, Milano
[10] Professore Onorario di Biochimica Clinica e Biologia Molecolare Clinica, Dipartimento di Medicina, Università degli Studi di Padova
[11] Professore Aggiunto, Dipartimento di patologia, Università del Texas, Medical Branch, Glaveston, USA
[12] Consiglio Nazionale delle Ricerche, Istituto di Calcolo e Reti ad Alte Prestazioni, Napoli
[13] Area ICT Fondazione Monasterio, Pisa;
[14] Consiglio Nazionale delle Ricerche, Istituto di Informatica e Telematica, Cosenza
[15] Associazione Medici Diabetologi, Italia
[16] Dipartimento di Economia, Università degli Studi di Perugia
[17] UOC Medicina di Laboratorio, AULSS 8 Berica, Vicenza
[18] SSD Ematologia Diagnostica e Genomica Clinica, Dip. Interaziendale ad Attività Integrata Medicina di Laboratorio e Anatomia Patologica, AOU/AUSL di Modena
[19] Dipartimento di Medicina (DIMED), Università degli Studi di Padova
[20] UOC Medicina di Laboratorio, Azienda-Ospedale Università di Padova
[21] Dipartimento di Neuroscienze, Università degli Studi di Pavia - Pavia

ABSTRACT

Artificial intelligence and big data in laboratory medicine

In the last few years, artificial intelligence (AI) is gaining attention in several medical disciplines, including laboratory medicine (LM). The raised interest on AI has been fueled not only by the huge amounts of information daily generated, but also by the special natural context offered by laboratories, where digitalization have already occupied an important part of the routine workflow of patients' data. Motivated by these topics and under the auspices of SIBioC, a conference on AI and big data was organized in May 2022 in Bologna, Italy. This conference covered several topics of AI and big data, including but not limited to the current and future perspectives, comprising ethical challenges and the role of laboratory specialists, including young professionals, the productive integration of AI with information technologies and with other digital infrastructure, such as the LOINC and the block chain. Furthermore, some examples of real application of AI in LM were reported, including diagnosis and monitoring of familiar hypercholesterolemia, management of insulin treatments for diabetes, reference intervals identification and verification by indirect methods, COVID-19 diagnosis and the monitoring of outpatients monoclonal gammopathy treatment by digital healthcare.

Parole chiave: *etica; sistemi esperti; sistemi informatici di laboratorio.*

Corrispondenza a: Andrea Padoan, Dipartimento di Medicina (DIMED), Università degli Studi di Padova e UOC Medicina di Laboratorio, Azienda-Ospedale Università di Padova, via Giustiniani 2, 35128 Padova, e-mail: andrea.padoan@unipd.it

Ricevuto: 28.10.2022 Revisionato: 01.11.2022 Accettato: 16.11.2022

Pubblicato online: 30.11.2022 DOI: 10.19186/BC_2022.080

INTRODUZIONE

L'Intelligenza Artificiale (IA) da alcuni anni è ormai entrata nella quotidianità di tutti noi e sta interessando anche la Medicina di Laboratorio (MdL), indubbiamente una fra le prime discipline mediche a produrre e gestire enormi quantità di dati in formato digitale.

La rapida evoluzione delle tecnologie digitali sta generando così una progressiva trasformazione che potrebbe portare nei prossimi anni a nuovi scenari e ad un ruolo dei laboratori più integrato nel processo clinico-diagnostico. Le tecnologie digitali per la gestione, integrazione ed analisi dei dati stanno subendo un repentino e profondo cambiamento e attualmente permettono di generare e analizzare Big Data anche attraverso algoritmi di IA. La prospettiva, più o meno a breve termine, sarà di doversi confrontare con sistemi esperti in grado di modificare significativamente i percorsi diagnostici e terapeutici con forti ripercussioni sulle attività e le organizzazioni dei Laboratori Clinici.

A fronte di queste trasformazioni che garantiranno vantaggi per i pazienti bisogna tuttavia riconoscere che uno sviluppo incontrollato e non governato dell'IA non è scevro da potenziali rischi. Conseguentemente, per poter introdurre in modo sicuro nella pratica clinica e di laboratorio i sistemi di IA, è necessario promuovere e sviluppare anche un nuovo substrato culturale fra i professionisti sanitari. E' quindi fondamentale agire su più fronti, a partire dalla realizzazione di infrastrutture organizzative, strutture di governance da parte delle agenzie regolatorie, predisposizione di moduli formativi universitari e post-universitari per migliorare le conoscenze e competenze in materia di IA del personale medico e delle professioni sanitarie, linee guida riguardanti le modalità di integrazione e il corretto utilizzo dei sistemi di IA nella diagnostica. In tutto ciò le Società Scientifiche possono dare un contributo considerevole nell'organizzazione e nella promozione e nello sviluppo della IA in MdL, organizzando eventi specifici su questa materia, evidenziando le tecnologie informatiche di supporto disponibili fornendo esempi applicativi, sia già utilizzati che in fase di sviluppo.

SIBioC crede molto in tutto ciò e lo ha dimostrato creando un apposito gruppo di studio 'Big Data ed Intelligenza Artificiale in Medicina di Laboratorio', pubblicando diversi contributi sulla rivista ufficiale (1-4)[1], inserendo nei propri convegni sessioni dedicate all'IA, promuovendo anche un evento specifico dal titolo "Big Data ed Intelligenza Artificiale in Medicina di Laboratorio" tenutosi a Bologna il 18 maggio 2022. Partendo dalla definizione di Big Data in MdL, discutendone le potenzialità, responsabilità e sfide future, con lo scopo di creare un terreno per garantire una semantica comune, anche in relazione alle tecnologie esistenti, il convegno è stata l'occasione per affrontare la tematica da vari punti di vista, come riportato nelle sessioni di seguito descritte.

Scopo di questo documento è la divulgazione dei temi e dei contenuti del convegno per dare modo anche a coloro che non hanno potuto partecipare, di venire a conoscenza sulle recenti attualità su queste tematiche.

Di seguito si riportano i contributi organizzati nelle sessioni del convegno.

IL CONTESTO PER LO SVILUPPO DEI BIG DATA E IA IN MEDICINA DI LABORATORIO

Nella prima sessione introduttiva sono stati analizzati il tema dell'etica e delle nuove opportunità offerte dai Big Data e l'IA.

L'etica dell'intelligenza artificiale in Medicina di Laboratorio

Il Parlamento Europeo ha concluso nel mese di maggio la discussione della proposta della Commissione Europea di un regolamento per l'IA. Uno degli argomenti principali è stato di considerare fondante e basilare l'utilizzo dei principi etici per regolamentare l'utilizzo dell'apprendimento automatico e delle applicazioni di informatica per la gestione di realtà complesse (5). Quindi, l'etica è imprescindibile per una verifica e un controllo dell'uso dell'IA , soprattutto in ambiti come quello sanitario, dove la relazione medico-paziente e struttura-cittadino è fondamentale. Le tecnologie di IA sono capaci di processare e individuare correlazioni fra una quantità di dati ben superiore alla capacità della mente umana migliorando l'accuratezza diagnostica, individuando malattie latenti in tempi utili al loro trattamento (prevenzione secondaria e terziaria), personalizzando la medicina sulle esigenze multidimensionali del paziente (cure orientate al paziente), rinforzando epidemiologia clinica e la gestione della salute della popolazione, incrociando dati clinici, amministrativi e sociosanitari. L'IA è particolarmente promettente nella MdL, essendo questa una delle principali fonti di dati per l'attività clinica, di ricerca e di policy, seppur con il limite di una disponibilità diretta del dato da parte del paziente. I principi che deve seguire l'etica dell'IA sono beneficenza (il sistema deve apportare benefici per il paziente), non maleficenza (il sistema non deve creare disparità o danni tra i pazienti), autonomia (il sistema garantisce a medico e paziente un proprio ruolo e dignità), giustizia (il sistema non deve creare ingiustizie per il paziente), come del resto per le altre procedure sanitarie, cui si aggiunge l'esplicabilità, che rende possibile il raggiungimento degli altri principi.

Il futuro dell'Intelligenza Artificiale nella Medicina di Laboratorio

Lo sviluppo della MdL nelle ultime decadi, grazie all'innovazione tecnologica, alla miglior comprensione dei meccanismi biologici, molecolari e fisiopatologici che sono alla base delle condizioni cliniche e delle patologie che affliggono l'umanità, è alla base del ruolo centrale dell'informazione di laboratorio nella medicina moderna. L'enorme mole di dati di laboratorio che sono parte essenziale dei cosiddetti Big Data e l'esigenza di strumenti che possano rendere possibile la loro integrazione, manipolazione e traduzione in informazioni

1 Il riferimento 1, contiene un utile glossario, per i lettori non particolarmente esperti nel campo specifico

cliniche più agevolmente "catturabili" dai clinici per essere interpretati e utilizzati in modo ottimale per la diagnosi e cura dei pazienti, hanno determinato un interesse sempre maggiore per l'IA e strumenti quali il machine learning (ML) (6). Non vi è dubbio alcuno che il miglioramento della fase post-analitica sia elemento fondamentale per assicurare la qualità totale nell'attività dei laboratori clinici (7). Per tutti questi motivi, sia le Società Scientifiche nazionali, come SIBioC, che le Federazioni della medicina di Laboratorio (IFCC e EFLM) hanno organizzato sessioni scientifiche e sponsorizzato ricerca e pubblicazione di studi sull'utilizzo di IA in MdL. IA e sistemi di ML applicati alle varie fasi del processo dell'analisi di laboratorio possono sicuramente migliorare la qualità dell'informazione di laboratorio grazie allo sviluppo e la validazione di algoritmi e strumenti di analisi ed integrazione dei dati. Uno dei problemi che, vanno sottolineati è che la condizione irrinunciabile per il razionale utilizzo dell'IA è l'accuratezza, affidabilità e riproducibilità dei dati di laboratorio. Il mantra che va sempre ricordato "*garbage in, garbage out*", infatti, sintetizza l'evidenza che se i sistemi di IA e ML non sono supportati da dati accurati, affidabili e riproducibili, anche i più sofisticati strumenti, algoritmi e software genereranno informazioni inaccurate e non riproducibili, creando talora problemi per la qualità della diagnosi e della cura dei pazienti. Pertanto, i professionisti di laboratorio devono approcciare l'IA collaborando strettamente con filosofi, ingegneri, matematici con formazione e competenze in questo ambito specifico, e lavorare al miglioramento della standardizzazione, armonizzazione e qualità dell'informazione di laboratorio in un contesto multidisciplinare realmente centrato sui bisogni di salute dei pazienti e dell'intera comunità.

LE BARRIERE DA SUPERARE, LA GESTIONE E L'ACCESSO AI DATI

Nella seconda sessione sono stati presentati in anteprima i risultati del questionario SIBioC e discusse le barriere che possono ostacolare lo sviluppo e la diffusione di nuove tecnologie basate sull'IA. Gli interventi si sono concentrati anche sulle soluzioni per alcune di queste barriere quali la necessità di standardizzazione e armonizzazione, l'adozione di standard fra cui il Logical Observation Identifers Names and Codes (LOINC).

Stato dell'arte dei laboratori clinici italiani e l'IA: siamo pronti alla rivoluzione digitale?

Negli ultimi anni sono stati sviluppati modelli molto promettenti in merito all'applicazione dell'IA in MdL, ma solo pochissimi hanno raggiunto la pratica clinica (8,9). Questa considerazione è stata il motivo della indagine conoscitiva proposta ai soci SIBioC al fine di inquadrare la reale situazione dei laboratori clinici italiani rispetto all'adeguatezza degli strumenti digitali a disposizione, all'integrazione dei dati, alla conoscenza e competenza sul tema dell'IA, sull'esistenza di progetti attivati e in definitiva per conoscere le reali barriere da superare per favorire l'utilizzo delle tecniche di IA nella pratica quotidiana (10). Da tale indagine è emerso che l'aspettativa che i Big Data possano cambiare la MdL è inconsistente con la realtà corrente. E' necessario quindi mettere in pratica una serie di interventi a partire da regolamenti in grado di rispettare l'etica e garantire la protezione dei dati e la privacy, l'adeguamento delle architetture infrastrutturali, l'armonizzazione dei dati basata su standard per scambio dati come HL7/FHIR, passare da dati non strutturati a strutturati, e utilizzare codifiche internazionali come la "International classification of diseases" (ICD), il LOINC, e la "Systematized Nomenclature of Medicine" (SNOMED). In merito all'impiego dell'IA in MdL, il 90% degli intervistati crede che questa non sostituirà il professionista ma che invece sarà utilissima nel supportarlo sulle decisioni da intraprendere.

Quando però l'indagine si è focalizzata sul livello di conoscenza e competenza rispetto a questi argomenti, questo è risultato essere mediamente insufficiente ed è emersa chiaramente la necessità di introdurre negli organici di laboratorio la figura del data scientist. Sicuramente il laboratorio saprà affrontare anche questa nuova sfida ma molto probabilmente assumerà una nuova forma, diversa da quella che conosciamo oggi.

Le barriere nella condivisione di dati di laboratorio e la problematica della standardizzazione

Il tema IA è sempre più di interesse in MdL. Nonostante gli innumerevoli ambiti applicativi, è la grande quantità di informazioni generate in MdL a sostenere lo sviluppo di applicazioni di IA (6). Le necessità tecnologiche, unite all'evoluzione dei sistemi informatici, hanno reso possibile l'integrazione di grandi quantità di dati provenienti da strumentazioni, da controlli di qualità, ed altri. L'utilizzo di IA e dei Big-Data può generare valore per il paziente e per la MdL, soprattutto se i dati vengono condivisi tra diverse strutture. La condivisione può essere ostacolata da diverse limitazioni. Le barriere possono essere interne ai laboratori (armonizzazione/standardizzazione, competenze, risorse umane, collaborazioni) o esterne (risorse tecnologiche e infrastrutturali, sicurezza, legislazione e/o privacy). Per standardizzazione e/o armonizzazione dei risultati si utilizzano principalmente materiali o procedure di riferimento; altre aree possono tuttavia essere standardizzate/armonizzate, come ad esempio la nomenclatura, le unità di misura, gli intervalli di riferimento. Nell'ambito delle competenze dei nuovi professionisti di MdL, è importante includere l'IA, grazie anche al supporto delle Società Scientifiche e delle collaborazioni multidisciplinari. Le limitazioni esterne al laboratorio possono essere infrastrutturali (come ad esempio lo stoccaggio di dati sicuri); sforzi futuri dovranno garantire la gestione di dati sanitari frammentati, e l'integrazione con dispositivi tecnologici "indossabili" o con i POCT. Infine, altre limitazioni riguardano gli aspetti legali e della privacy. Sforzi congiunti sono indispensabili per generare competitività globale con i Paesi in forte espansione, garantendo al contempo una IA basata sui valori promossi dall'Unione Europea, tra cui il rispetto dei dati personali, senza tralasciare la costruzione di una collaborazione di fiducia tra pazienti ed istituzioni.

Luci ed ombre dell'Intelligenza Artificiale: riflessioni etiche tra rischi e opportunità

Nell'ultimo secolo il laboratorio analisi si è rinnovato attraverso l'informatizzazione e l'automazione. L'ultima rivoluzione è l'implementazione dell'IA iniziata con l'utilizzo di Sistemi Esperti (ES), programmati formalizzando la conoscenza del professionista. Questo percorso si è evoluto con il ML e l'apprendimento non supervisionato che promettono, in cambio di enormi moli di dati, di poter elaborare soluzioni plasmate sul proprio scenario: gestione dei flussi di lavoro, degli approvvigionamenti e gestione del dato clinico come supporto diagnostico (11).

Tutto questo è possibile grazie alla crescente granulosità delle informazioni prodotte e ai sistemi di codifica standardizzati (come LOINC ed Ontology): abbiamo avuto un assaggio della miniera ancora da scalfire grazie alla pandemia COVID-19. L'uso di termini e codici standardizzati ha permesso di poter catalogare mediante "natural language processing" (NPL) e comparare risultati ottenendo risultati mai visti.

Trattandosi di azioni con impatti sulla vita del paziente, oltre che su processi e costi delle strutture sanitarie, risulta evidente come ogni scelta, anche se fortemente raccomandata da una IA, debba essere opportunamente valutata (11). Tuttavia, l'intrinseca "non spiegabilità" dei risultati ottenuti mediante IA, definita come problema delle "scatole nere" o "black box" (12), mette il professionista nelle condizioni di dover comunicare un risultato al paziente con una "responsabilità condivisa" con l'IA. Inoltre, risultati di medicina predittiva quali score e identificazioni di pattern prognostici, ci pongono il dilemma su chi sia il proprietario del dato appena elaborato dall'IA e su possibili effetti nocivi. Inoltre, è sempre più evidente quanto sia asimmetrico il rapporto sanitario-paziente rispetto alle informazioni (13). Come professionisti siamo chiamati ad uno sforzo di formazione per diventare competenti nel governo dell'IA e nella gestione dei set di training e di controllo che ci permetteranno di diventare familiari con i risultati ottenuti. Solo così potremo fare un passo avanti nella trasformazione della medicina da assistenziale a predittiva.

I giovani specialisti in Medicina di Laboratorio e le competenze del Data Scientist

Nella Medicina centrata sul paziente sono richieste conoscenze e competenze specialistiche contestualmente a integrazione e multidisciplinarietà. La MdL, trasversale rispetto alla clinica, è centrale nell'integrazione diagnostica, pertanto anche le competenze degli Specialisti in MdL sono in continua evoluzione (14). Inoltre, digitalizzazione e sviluppo tecnologico hanno generato un sovraccarico informativo aumentando la complessità del processo decisionale (15,16). Ne è derivata una spinta al passaggio dalla pratica clinica basata sull'esperto, ad un modello supportato da un sistema di algoritmi ed applicazioni di IA, basato sulle evidenze e su dati sanitari e non sanitari, fornendo un supporto decisionale al professionista. Dieci anni fa uno scenario simile sembrava vicino. Invece permangono degli ostacoli, tra cui l'impreparazione al corretto utilizzo dei sistemi di IA (17). Emerge la necessità di competenze digitali di base per i professionisti sanitari per poter collaborare con il mondo dell'Information Technology (IT) e utilizzare consapevolmente i modelli predittivi, interpretandone le prestazioni diagnostiche (18,19). Alcuni giovani Specialisti in MdL, invece, diventeranno effettivamente "Data Scientist" (DS), figure esperte nella disciplina (Data Science) che estrae valore dai dati utilizzando IA. Schematicamente il processo consiste di sei fasi:

- identificazione degli obiettivi;
- raccolta dei dati;
- preparazione dei dati: eliminare il "rumore", organizzare i dati, attuare strategie per i dati mancanti o che variano nel tempo (arricchimento dei dati);
- trovare correlazioni, formulare ipotesi; sperimentare modelli analitici, comprendere le loro prestazioni; rilevare eventuali bias;
- uso di algoritmi; elaborazione, modellizzazione, valutazione dei modelli;
- comunicazione.

Il DS sanitario deve avere almeno una conoscenza di base nella materia di studio per formulare ipotesi corrette e validare le analisi effettuate. Inoltre necessita di competenze digitali:

- nella gestione delle fonti dei dati, estrazione, preparazione e archiviazione;
- in biostatistica per identificare correlazioni, stabilire la significatività;
- in ML e IA;
- in data visualization: convertire i risultati in grafiche attrattive comprensibili.

Importanti sono le "soft skills", ovvero le competenze trasversali, tra cui il lavoro in gruppo, quando al DS, che elaborerà il miglior modello per analizzare i dati, si affianca un Data Manager, che gestisce e automatizza estrazione ed elaborazione. Entrambi dialogheranno con i responsabili IT e con i ricercatori per le competenze scientifiche e cliniche (20).

I LIS E MIDDLEWARE NELL'ERA DEI BIG DATA

In questa sessione sono stati inseriti dei contributi inerenti ai sistemi informatici di laboratorio (LIS) e sui sistemi utilizzati nei contesti sanitari, con lo scopo di illustrare eventuali limitazioni infrastrutturali comuni presenti nei laboratori.

Fascicolo Sanitario Elettronico: infrastruttura e servizi di interoperabilità nell'era dei Big Data

L'applicazione dell' "Information and Communication Technologies" (ICT) nel settore della salute sta comportando una straordinaria esplosione di dati sanitari eterogenei, come documenti (quali ad esempio referti di laboratorio, prescrizioni), immagini radiologiche, parametri vitali, annotazioni del paziente e così via. Tutti questi dati devono essere raccolti sistematicamente per poter essere analizzati. In tal senso, il Fascicolo Sanitario Elettronico (FSE) rappresenta in Italia un investimento

particolarmente importante: l'architettura federata della piattaforma nazionale definita in attuazione del D.L. 179/2012 offre servizi che favoriscono l'interoperabilità tra i sistemi di FSE regionali (21). Tuttavia, le moderne tecnologie possono favorire progressi significativi per un uso intelligente dei dati. Le tecniche di IA (come il "deep learning") e di Big Data Analytics consentono l'analisi di grossi volumi di dati per estrarre valore dalle informazioni (22). Le tecniche di elaborazione del linguaggio naturale stanno fornendo risultati rilevanti per la classificazione di informazioni da documenti sanitari narrativi (23). La tecnologia blockchain e gli "smart contract" permettono di garantire l'integrità di dati e processi (24). L'internet delle cose (IoT) e l'Edge Computing permettono la raccolta e l'analisi decentralizzata di dati di telemonitoraggio (25). Diverse sperimentazioni hanno dimostrato l'applicabilità di queste tecnologie che, pertanto, assicureranno un importante valore aggiunto e la corretta evoluzione dei servizi del FSE nell'era dei Big Data.

Blockchain in Medicina di Laboratorio

Negli ultimi anni abbiamo assistito a una vera e propria esplosione nella produzione di dati, che dovrà prima o poi richiedere un un'aumentata sicurezza ed efficienza della loro gestione.

Utile a questo scopo è la cosiddetta Blockchain. La Blockchain (letteralmente "catena di blocchi") è una struttura dati condivisa e "immutabile". È definita come un registro digitale le cui voci sono raggruppate in "blocchi", concatenati in ordine cronologico, e la cui integrità è garantita dall'uso della crittografia. Gli elementi costitutivi principali sono il Distributed Ledger (libro mastro distribuito), la vera e propria catena di blocchi e i nodi della rete o Miners (i minatori) con la loro attività di mining (26) .

Questa tecnologia promette di consentire un accesso sicuro ai dati dovunque, in qualunque dispositivo e in qualsiasi momento. Applicata in sanità risolverebbe diversi problemi come la portabilità, la sicurezza e l'estrazione del vero valore dei dati (27).

Le applicazioni basate su tecnologia Blockchain sollevano però alcuni dubbi circa la conformità alla normativa sulla protezione dei dati (Regolamento europeo 2016/679). La discussione ruota principalmente su tre punti: su quali dati tutelare, sul titolare del trattamento dei dati personali e sul diritto all'oblio (28).

La tecnologia Blockchain però è ancora ai suoi inizi, e molti problemi tecnici devono essere risolti prima di poterne sfruttarne appieno il potenziale. Fra questi ricordiamo: la scalabilità, il consumo di risorse computazionali ed energetiche, l'incertezza normativa, i timori per la sicurezza e la privacy, la riluttanza a condividere le proprie informazioni, la resistenza culturale e le sfide di standardizzazione.

Big Data, Dark Data, Data Lakes e Open Data: come i dati di laboratorio possono essere sfruttati nella ricerca

I dati di MdL trattati nei processi di diagnosi e cura rappresentano la più ampia base di dati dati strutturati sfruttabile per la ricerca. Le diverse modalità di raccolta ed aggregazione permettono di utilizzare queste miniere di informazioni secondo i paradigmi di elaborazione più avanzati e moderni (29). Le attuali tecniche di collazione e registrazione si basano su alcuni modelli che si riferiscono a impostazioni concettuali e implementazioni tecniche, come i Big Data ed i Data Lakes, e altri modelli che rispecchiano una politica di diffusione e condivisione del dato, nel rispetto delle normative vigenti, come gli Open Data e i Data Spaces. In particolare per i Data Spaces in ambito europeo e mondiale sono in corso diverse iniziative, tese a definire degli ambiti regolamentati in cui si possano sfruttare i microdati e le informazioni generate in diversi settori, tra cui il settore salute, al fine di supportare le correnti linee di ricerca, consumando una quantità inferiore di risorse finanziarie ed umane, e stimolare il settore della ricerca a nuove linee o approcci abilitati dalla disponibilità di informazioni che travalicano i consueti domini di ricerca (ecologia, salute, energia, trasporti) (30). Oltre ad uno sfruttamento per ricerca dei dati primari, sono in realtà disponibili dati secondari, utilizzati essenzialmente per supportare i processi di erogazione dei servizi sanitari e di raccolta dei dati primari, su cui la definizione di "Dark data" colloca il proprio ambito (31) .

Lo standard LOINC: vantaggi e modalità di utilizzo per l'interoperabilità semantica di dati e sistemi

L'interoperabilità semantica garantisce che l'informazione scambiata possa essere compresa dal destinatario allo stesso modo in cui il mittente l'ha trasmessa. Al raggiungimento di tale fine contribuiscono i sistemi di codifica standard. Specificatamente in ambito clinico, l'uso di standard facilita la comunicazione tra esseri umani e consente l'interoperabilità tra sistemi, applicazioni e istituzioni, permettendo il confronto e l'integrazione di dati per supportare studi epidemiologici, nonché la loro condivisione e portabilità nel FSE. LOINC è uno standard internazionale per l'identificazione univoca di osservazioni cliniche e di laboratorio, conta più di 98 000 codici (ultima versione Febbraio 2022), è tradotto in 13 lingue e usato in 193 Paesi (32). Ciascun record LOINC è identificato da un codice univoco e un nome derivante dall'unione di sei assi: oggetto, unità, campione, tempo, metodo e scala di misurazione. LOINC ha un elevato livello di dettaglio nell'identificazione degli esami di laboratorio, in modo da disporre di risultati comprensibili e riutilizzabili per migliorare l'analisi e l'interoperabilità dei dati. Nonostante sia fra i sistemi di codifica previsti dal DM 178/2015 per il FSE, in Italia è utilizzato a macchia di leopardo, poiché vige un'idiosincrasia data dalla coesistenza di cataloghi di esami locali e di nomenclatori tariffari regionali. Questi, tuttavia, non consentono di raggiungere il livello di dettaglio clinico di LOINC, perché organizzati sulle specificità locali e nati per finalità economiche (33). Per utilizzare LOINC è necesario mappare i codici locali verso quelli dello standard. Questo permette di non cambiare le proprie abitudini lavorative ma di produrre dati semanticamente interoperabili.

PROGETTI E APPLICAZIONI BIG DATA E INTELLIGENZA ARTIFICIALE IN MEDICINA DI LABORATORIO

L'ultima parte del congresso è stata dedicata alle applicazioni e progetti in corso da parte di colleghi al fine di comprendere come stiamo gestendo nella pratica quotidiana l'innovazione per governare il cambiamento.

Analisi dei Big Data per la diagnosi di Ipercolesterolemia Familiare

L'ipercolesterolemia familiare eterozigote è una patologia genetica frequente con una prevalenza stimata di circa 1 caso ogni 250-300 soggetti nella popolazione generale. Questa patologia conferisce un significativo rischio cardiovascolare in ragione degli elevati livelli circolanti di colesterolo LDL nei soggetti affetti. Si tratta di una patologia sotto-diagnosticata e sotto-trattata a causa della scarsa consapevolezza della sua prevalenza e delle morbidità e mortalità cardiovascolari associate. D'altro canto, l'ipercolesterolemia familiare è una patologia che, per le sue caratteristiche, si presta bene all'approccio Big Data per l'identificazione dei soggetti, per il successivo monitoraggio e per le eventuali indagini da intraprendere nel gruppo familiare (34) . Con il ML e con l'analisi dei Big Data si possono infatti interrogare grandi banche dati in maniera efficiente, incrociando le informazioni allo scopo di suggerire una diagnosi. Nel caso dell'ipercolesterolemia familiare, l'incrocio dei dati provenienti dal database del laboratorio clinico (dati relativi al profilo lipidico e valori di colesterolo LDL), da quello dei codici di diagnosi (riferibili a patologie cardiovascolari) e da quello delle prescrizioni farmaceutiche (farmaci ipolipemizzanti) può fornire informazioni utili per sospettare la condizione e identificare i soggetti a cui eventualmente proporre la diagnosi genetica (35). Questo metodo può avere un impatto ancora più rilevante se si pensa alla programmazione del cosiddetto "screening a cascata" nei familiari dei soggetti affetti, individui che potrebbero trarre giovamento da interventi di prevenzione primaria cardiovascolare. Dati provenienti da esperienze italiane ed internazionali evidenziano quindi come l'estrazione di dati da database clinici rappresenti un approccio affidabile per stimare i soggetti con ipercolesterolemia familiare e programmare ulteriori azioni di prevenzione.

Inerzia terapeutica e terapia insulinica: nuovi ed innovativi scenari guidati dall'Intelligenza Artificiale

L'obiettivo di questo studio è stato quello di individuare gli elementi chiave che caratterizzano le situazioni d'inerzia nell'inizio della terapia insulinica.

Dal database degli Annali dell'Associazione Medici Diabetologi, contenente le visite di 1,5 milioni di pazienti della rete dei centri di diabetologia italiani per il periodo 2005-2019, sono state analizzate tutte le situazioni in cui sarebbe stato appropriato utilizzare la terapia insulinica. E' stato utilizzato una Logic Learning Machine (LLM), una tecnica in clear box, a regole esplicite. I dati sono stati sottoposti a una prima fase di modellazione per consentire al ML di selezionare automaticamente i fattori più rilevanti, seguite da quattro ulteriori fasi di modellazione che hanno individuato le variabili chiave in grado di discriminare la presenza o l'assenza di inerzia.

Il modello predittivo presenta una precisione molto buona (77-79%) e le funzioni di "explainable artificial intelligence" della LLM hanno evidenziato che, oltre ai valori di emoglobina glicata (HbA_{1c}) in base ai quali viene presa la decisione di procedere con la terapia insulinica (>72 mmol/mol; 8,7%,), un altro elemento molto importante è rappresentato dalla differenza di HbA_{1c} tra due visite consecutive, il quale, se è <6,6 mmol/mol (0,6%), è più probabile che il medico mostri inerzia, mentre, se è >11 mmol/mol (1,0%) è più probabile che al paziente venga prescritta la terapia insulinica.

I risultati rivelano, per la prima volta, il ruolo dominante rappresentato dalla variazione di HbA_{1c} rispetto alla visita precedente e l'importanza delle variabili dinamiche, che riflettono l'andamento glicemico del paziente, più che i valori riscontrati alla singola visita. I risultati dimostrano inoltre che LLM può fornire informazioni a supporto della medicina basata sull'evidenza (EBM) utilizzando dati del mondo reale.

Approccio computazionale nel processo diagnostico di COVID-19: sinergia fra laboratorio e pronto soccorso

SARS-CoV-2 è l'agente patogeno responsabile della pandemia COVID-19 (36). Il lavoro valuta l'accuratezza di classificatori basati su IA e su modelli statistici adattati agli esami del sangue e ad altre informazioni comunemente raccolte presso il Pronto Soccorso. Attualmente, nonostante le limitazioni pratiche di procedure che richiedono tempo e un alto tasso di risultati falsi negativi (37,38), l'indagine molecolare (RT-PCR) su tamponi respiratori è, secondo il Centre of Disease Control statunitense, il gold-standard di classificazione. Le procedure automatizzate possono essere tempestivamente affiancate alla RT-PCR, fornendo un ulteriore strumento di valutazione.

Nello studio sono stati arruolati 971 pazienti al primo accesso al Pronto Soccorso dell'Azienda Ospedaliero Universitaria di Careggi (Firenze) dal 7 al 30 aprile 2020 con caratteristiche predeterminate di sospetto COVID-19. I medici hanno inizialmente dicotomizzato i pazienti in caso probabile/improbabile in base alle caratteristiche cliniche. Considerando i limiti di ciascun metodo per identificare un caso di COVID-19, è stata eseguita un'ulteriore valutazione dopo una revisione clinica indipendente dei dati di monitoraggio a 30 giorni. La "sensazione clinica" (39) è stata quindi utilizzata per implementare i seguenti modelli di classificazione sia di statistica classica che di ML

Molti metodi di classificazione mostrano una area sotto la curva (AUC) >0,80 sia sui campioni utilizzati per la convalida interna che esterna (40), alcuni dei quali si sono dimostrati essere migliori (Regressione Logistica, Random Forest e Neural Network). Questo supporta l'idea di poter effettuare una prima identificazione dei casi,

individuando anche pazienti che avrebbero sviluppato l'infezione (diagnosticata con RT-PCR positiva) entro 7 giorni (41). Data la rapida evoluzione del virus, riteniamo che le procedure automatizzate di elaborazione dei dati possano fornire supporto ai clinici che devono affrontare una iniziale classificazione del paziente.

Definizione di intervalli di riferimento e limiti decisionali con metodi indiretti

Affinché una misura su un liquido biologico diventi un'informazione clinica servono tre cose: il quesito clinico, una misura affidabile e robusta e un sistema di valutazione, che attribuisca alla misura un significato o una qualità. I livelli decisionali, derivati da esperienza ed evidenze cliniche, sono decisamente i sistemi di comparazione preferibili, perché strettamente legati alla risposta attesa. Tuttavia, i limiti decisionali non sono spesso disponibili per carenza di studi, sovrapposizione di più condizioni alternative, incertezza o contraddittorietà tra le raccomandazioni. Più utilizzati sono i confronti che si basano sulla distribuzione statistica della misura, frequentemente derivati da osservazioni su campioni di popolazione teoricamente sana. La loro diffusione è tale da aver assunto per antonomasia il nome di intervalli di riferimento (IR). La definizione degli IR tramite il metodo diretto (42) non è tuttavia semplice né agevole, quando si debbano operare selezioni, ripartizioni, campionamenti su popolazioni difficili, ecc.

Se si considera che nei database dei laboratori clinici gran parte dei dati disponibili deriva da accertamenti di screening o senza patologia correlata, si può ipotizzare di poter stimare un intervallo di riferimento, attraverso il riconoscimento della sottopopolazione maggioritaria normale. È questa l'idea dei metodi indiretti, per il calcolo degli IR. Grandi basi di dati e metodi robusti sono oggi entrambi disponibili e facilmente utilizzabili (43).

Monitoraggio clinico-laboratoristico integrato con Telemedicina in pazienti ambulatoriali con gammopatie monoclonali: il progetto "Team MGUS"

Come progetto innovativo dell'AUSL-Modena per la gestione dei pazienti con gammopatia monoclonale di significato indeterminato (MGUS), in linea con le recenti indicazioni PNRR-2021 per rafforzare il servizio di assistenza territoriale, presso l'AOU/AUSL di Modena, stiamo sviluppando un modello di integrazione funzionale clinico-laboratoristica, con aspetti operativi di telemedicina. Il nostro "Team MGUS" nasce come gruppo multidisciplinare, con specifiche competenze di laboratorio ed ematologia clinica, per promuovere una gestione coordinata ed ottimizzata della vasta popolazione MGUS provinciale (>10000 casi), principalmente mediante:

- esecuzione completa ed appropriata degli esami di laboratorio (profilo MGUS reflex);
- presa in carico del paziente, con eventuale contatto periodico mediante telefonate/videochiamate (televisita);
- accesso razionale e programmato alla visita clinica tradizionale presso l'Ambulatorio MGUS dedicato;
- sviluppo ad hoc del database MGUS, che permette una classificazione dinamica dei pazienti per rischio di progressione a mieloma, evidenziando quelli con rischio aumentato, per cui è consigliabile rivalutazione mirata e/o approfondimento diagnostico;
- implementazione di progetti di ricerca clinica, basati su analisi immunologiche avanzate (supporto Fondazione AIRC) e su studi di popolazione con l'utilizzo dei Big Data.

Pertanto, l'approccio del "Team MGUS" offre importanti vantaggi, sia in termini di appropriatezza e razionalizzazione delle risorse sanitarie, sia per il paziente stesso, che può beneficiare di un moderno servizio assistenziale dedicato, per il monitoraggio di una condizione pre-neoplastica che inevitabilmente genera apprensione.

CONCLUSIONI

Dalle relazioni emerge chiaramente come diverse tecnologie siano già da ora disponibili e pronte per essere impiegate nella creazione di strumenti di supporto basati su IA, per migliorare il processo diagnostico e di cura del paziente. Inoltre, queste applicazioni potrebbero avere un ruolo importante nel monitoraggio da remoto, anche attraverso la telemedicina, integrando dati preziosi della salute del paziente, permettendo di scegliere prontamente terapie più mirate e personalizzate. Sebbene questi vantaggi siano stati evidenziati da molti relatori, rimane tuttavia incerto se gli strumenti attualmente in possesso dei professionisti di medicina di laboratorio possano essere sufficienti a favorire lo sviluppo pro-attivo di queste tecnologie, senza trascurare l'importanza dell'aspetto etico e delle competenze necessarie per l'integrazione dell'IA nei laboratori clinici.

CONFLITTO DI INTERESSI

Nessuno.

Bibliografia

1. Guerranti R, Padoan A, Angeletti D, et al. Introduction to Big Data and Artificial Intelligence in laboratory medicine. Biochim Clin 2021;45:57-67.
2. Vidali M. I Big Data e la medicina di laboratorio. Biochim Clin 2021;45:13-4.
3. Pecoraro V, Pirotti T, Trenti T, et al. Big Data analysis to evaluate the clinical utility of IgM anti SARS-CoV-2 determination: the Modena experience. Biochim Clin 2022;46:154-9.
4. Carobene A, Sabetta E, Monteverde E, et al. Machine Learning based on laboratory medicine test results in diagnosis and prognosis for COVID-19 patients: A systematic review. Biochim Clin 2021;45:348-64.
5. Pennestrì F, Banfi G. Artificial intelligence in laboratory medicine: fundamental ethical issues and normative key-points. Clin Chem Lab Med 2022 aop doi 10.1515/cclm-2022-0096.
6. Padoan A, Plebani M. Flowing through laboratory clinical data: the role of artificial intelligence and big data. Clin Chem Lab Med 2022 aop doi: 10.1515/cclm-2022-0653.
7. Paranjape K, Schinkel M, Hammer RD, et al. The Value of Artificial Intelligence in Laboratory Medicine. Am J Clin

Pathol 2021;155:823-31.
8. Cadamuro J. Rise of the machines: the inevitable evolution of medicine and medical laboratories intertwining with artificial intelligence - a narrative review. Diagnostics (Basel) 2021;11:1399.
9. Kim HS, Kim DJ, Yoon KH. Medical big data is not yet available: why we need realism rather than exaggeration. Endocrinol Metab (Seoul). 2019;34:349-54-
10. Bellini C, Padoan A, Carobene A, et al. A survey on Artificial Intelligence and Big Data utilisation in Italian clinical laboratories. Clin Chem Lab Med 2022. aop doi: 10.1515/cclm-2022-0680.
11. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health 2019;9:010318.
12. Wadden JJ. Defining the undefinable: the black box problem in healthcare artificial intelligence. J Med Ethics 2022;48:764-8.
13. Bompelli A, Wang Y, Wan R, et al. social and behavioral determinants of health in the era of artificial intelligence with electronic health records: a scoping review. Health Data Science 2021;2021:1-19.
14. Wilson ML, Fleming KA, Kuti MA, et al. Access to pathology and laboratory medicine services: a crucial gap. Lancet 2018;391:1927-38.
15. Densen P. Challenges and opportunities facing medical education. Trans Am Clin Climatol Assoc 2011;122:48-58.
16. Stead WW, Searle JR, Fessler HE, et al.. Biomedical Informatics: Changing What Physicians Need to Know and How They Learn. Academic Medicine 2011;86:429-34.
17. Ministero della Salute. I sistemi di intelligenza artificiale come strumento di supporto alla diagnostica . disponibile da: https://www.salute.gov.it/portale/documentazione/p6_2_2_1.jsp?lingua=italiano&id=3218 (ultimo accesso Ottobre 2022).
18. Lomis K, Jeffries P, Palatta A, et al. Artificial intelligence for health professions educators. NAM Perspect 2021;2021:10.31478/202109a.
19. Matheny ME, Whicher D, Thadaney Israni S. Artificial Intelligence in Health Care. JAMA 2020;323:509-10.
20. Baig MA, Alzahrani SJ. Revisiting the Skills of a Healthcare Data Scientist as a Field Expert. Stud Health Technol Inform 2019;262:43-6.
21. Ciampi M, Sicuranza M, Esposito A, et al. A technological framework for ehr interoperability: experiences from italy. Communications in Computer and Information Science 2017;736:80-99.
22. Gargiulo F, Silvestri S, Ciampi M, et al. Deep neural network for hierarchical extreme multi-label text classification. Appl Soft Comput 2019;79:125-38.
23. Silvestri S, Gargiulo F, Ciampi M. Iterative Annotation of Biomedical NER Corpora with Deep Neural Networks and Knowledge Bases. Applied Sciences 2022;12:5775.
24. Ciampi M, Esposito A, Marangio F, et al. A blockchain architecture for the Italian EHR system Health information systems. Roceedings of the Fourth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing-HEALTHINFO. 2019;11-17.
25. Prabhu M, Hanumanthaiah A. Edge computing-enabled healthcare framework to provide telehealth services. In International Conference on Wireless Communications Signal Processing and Networking IEEE. 2022:349-53 doi: 10.1109/WiSPNET54241.2022.9767142
26. Xie Y, Zhang J, Wang H, et al. Applications of Blockchain in the Medical Field: Narrative Review. J Med Internet Res 2021;23:e28613.
27. Kumar T, Ramani V, Ahmad I, et al. Blockchain utilization in healthcare: key requirements and challenges. In 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services, Healthcom 2018. IEEE Institute of Electrical and Electronic Engineers. 2018:8531136 doi: 10.1109/HealthCom.2018.8531136.
28. Network Digital 360, GDPR e blockchain, tutte le sfide di un rapporto complesso - Agenda Digitale. Available from: https://www.agendadigitale.eu/documenti/gdpr-e-blockchain-tutte-le-sfide-di-un-rapporto-complesso/ (ultimo accesso 24 Ottobre 2022).
29. Blatter TU, Witte H, Nakas CT, et al. Big Data in Laboratory Medicine - FAIR Quality for AI? Diagnostics 2022;12:1923.
30. Eder J, Shekhovtsov VA. Data quality for federated medical data lakes. International Journal of Web Information Systems 2021;17:407-26.
31. Schembera B, Durán JM. Dark data as the new challenge for big data science and the introduction of the scientific data officer. Philos Technol 2020;33:93-115.
32. Vreeman DJ, Chiaravalloti MT, Hook J, et al. Enabling international adoption of LOINC through translation. J Biomed Inform 2012;45:667-73.
33. Florio I, Guaglianone MT. Il fascicolo sanitario elettronico: infrastruttura tecnologica e codifica dei dati. CNR-SeGID 2012;161-85.
34. Myers KD, Knowles JW, Staszak D, et al. Precision screening for familial hypercholesterolaemia: a machine learning study applied to electronic health encounter data. Lancet Digit Health 2019;1:e393.
35. Fasano T, Trenti C, Negri EA, et al. Search for familial hypercholesterolemia patients in an Italian community: a real-life retrospective study. Nutr Metab Cardiovasc Dis 2022;32:577-85.
36. Chen J, Lu H, Melino G, et al. COVID-19 infection: the China and Italy perspectives. Cell Death Dis 2020;11:438.
37. Ferrari D, Sabetta E, Ceriotti D, et al. Routine blood analysis greatly reduces the false-negative rate of RT-PCR testing for COVID-19. Acta Biomed 2020;91:e2020003.
38. Ferrari D, Motta A, Strollo M, et al. Routine blood tests as a potential diagnostic tool for COVID-19. Clin Chem Lab Med 2020;58:1095-9.
39. Nazerian P, Morello F, Prota A, et al. Diagnostic accuracy of physician's gestalt in suspected COVID-19: Prospective bicentric study. Acad Emerg Med 2021;28:404-11.
40. Cabitza F, Campagner A, Soares F, et al. The importance of being external. methodological insights for the external validation of machine learning models in medicine. Comput Methods Programs Biomed 2021;208:106288.
41. Long DR, Gombar S, Hogan CA, et al. Occurrence and timing of subsequent severe acute respiratory syndrome coronavirus 2 reverse-transcription polymerase chain reaction positivity among initially negative patients. Clin Infect Dis 2021;72:323.
42. Clinical Laboratory Standards Institute (CLSI). Defining, establishing, and verifying reference intervals. In the Clinical Laboratory; Approved Guideline -Third Edition. CLSI document EP28-A3c,
43. Yang D, Su Z, Zhao M. Big data and reference intervals. Clin Chim Acta 2022;527:23-32.