

Big data e intelligenza artificiale in medicina di laboratorio: una indagine SiBioC nei i laboratori clinici italiani

Claudia Bellini¹, Andrea Padoan², Anna Carobene³, Roberto Guerranti⁴ on behalf of the Italian Society of Clinical Biochemistry and Clinical Molecular Biology Big Data and Artificial Intelligence Working Group

¹Clinical Chemistry Laboratory Analysis Unit, Specialised in Clinical Biochemistry, Misericordia Hospital, Grosseto, Italy

²Department of Medicine-DIMED, University of Padova, Padova, Italy

³Laboratory Medicine, IRCCS San Raffaele Scientific Institute, Milan, Italy

⁴Department of Medical Biotechnologies, University of Siena, Siena, Italy; and Clinical Pathology Unit, Innovation, Experimentation and Clinical and Translational Research Department, University Hospital of Siena, Siena, Italy

Traduzione a cura di

Luisa Lanzilao, Azienda Ospedaliero Universitaria Careggi, Firenze

Claudia Bellini, Ospedale della Misericordia, Grosseto

ABSTRACT

A survey on Artificial Intelligence and Big Data utilisation in Italian clinical laboratories

Objectives: the Italian Society of Clinical Biochemistry and Clinical Molecular Biology (SiBioC) Big Data and Artificial Intelligence (BAI) Working Group promoted a survey to frame the knowledge, skills, and technological predisposition in Clinical Laboratories.

Methods: questionnaire, focusing on digitization, IT infrastructures, data accessibility, and BAI projects underway was sent to 1351 SiBioC participants. The responses were evaluated using SurveyMonkey software and Google Sheets.

Results: the 227 respondents (17%) from all over Italy (47% of 484 labs), mainly biologists, laboratory physicians and managers, mostly from laboratories of public hospitals, revealed lack of hardware, software and corporate Wi-Fi, and dearth of PCs. Only 25% work daily on clouds, while 65%-including Laboratory Directors-cannot acquire health data from sources other than laboratories. Only 50% of those with access can review a clinical patient's health record, while the other access only to laboratory information. The integration of laboratory data with other health data is mostly incomplete, which limits BAI-type analysis. Many are unaware of integration platforms. Over 90% report pulling data from the LIS, with varying degrees of autonomy. Very few have already undertaken BAI projects, frequently relying on IT partnerships. The majority consider BAI as crucial in helping professional judgments, indicating a growing interest.

Conclusions: the questionnaire received relevant feedback from SiBioC participants. It highlighted the level of expertise and interest in BAI applications. None of the obstacles stands out more than the others, emphasising the need to all-around work: IT infrastructures, data warehouses, BAI analysis software acquisition, data accessibility and training.

Parole chiave: *intelligenza artificiale, Big Data, digitalizzazione*

INTRODUZIONE

Gli strumenti di Intelligenza Artificiale (IA) rappresentano tecnologie all'avanguardia che trovano ampia applicazione in diversi ambiti tecnologici e scientifici. È interessante notare come l'IA sia stata ampiamente utilizzata in ambito sanitario, soprattutto grazie alla

disponibilità di innumerevoli fonti di dati sui pazienti (Big Data); di conseguenza, negli ultimi anni sono fioriti molti studi in diverse discipline mediche (1,2). Tuttavia, stranamente, solo un piccolo numero di questi studi si occupa specificamente di medicina di laboratorio (MdL), disciplina che può essere considerata la principale fonte di dati quantitativi, strutturati e codificati per la clinica (3-5).

Questo articolo è la traduzione di A survey on Artificial Intelligence and Big Data utilisation in Italian clinical laboratories. Clin Chem Lab Med 2022;60:2017-26, con l'autorizzazione dell'Editore. L'articolo originale è disponibile come materiale supplementare (1S). Clinical Chemistry and Laboratory Medicine non è responsabile della accuratezza della traduzione. In caso di citazione, riferirsi alla pubblicazione originale.

Corrispondenza a: Claudia Bellini, Clinical Chemistry Laboratory Analysis Unit, Specialised in Clinical Biochemistry Misericordia Hospital, Grosseto, Email: cbellini.md@gmail.com

Ricevuto: 10.11.2022

Rivisto: 11.11.2022

Accettato: 11.11.2022

Pubblicato online: 13.12.2022

DOI: 10.19186/BC_2022.079

Nonostante in letteratura siano riportate un gran numero di applicazioni, la maggior parte di esse è stata impiegata in contesti specializzati o per scopi di ricerca e solo un numero esiguo è attualmente utilizzato nella pratica quotidiana (6). In altre parole, sembra esserci un divario significativo tra un piccolo numero di laboratori in grado di creare e utilizzare applicazioni di IA e la maggior parte dei laboratori che, invece, hanno solo una comprensione generale di questi temi e non dispongono degli strumenti e delle competenze necessari per sviluppi autonomi (7).

Recentemente, la Società Italiana di Biochimica Clinica e Biologia Molecolare Clinica (SIBioC) ha creato uno specifico gruppo di studio (GdS) sul tema dei Big Data e dell'IA (BIA). Una volta costituito, il GdS ha iniziato a condurre una serie di iniziative educative e culturali per promuovere lo sviluppo di una rete multidisciplinare e integrata tra i professionisti della MdL e quelli del mondo dell'informatica (IT) applicata alla medicina, prerequisito fondamentale per l'analisi BIA (8). Sono stati elaborati una serie di documenti per permettere ai professionisti della MdL di familiarizzare con i temi dirompenti del mondo BIA (9-11) e alcune ricerche sono state condotte da alcuni membri del GdS (5,12-19).

Il GdS ha promosso un'indagine sull' argomento BIA per meglio inquadrare la situazione reale dei laboratori clinici italiani, con un duplice scopo: da un lato verificare lo stato tecnologico dei laboratori (adeguatezza delle attrezzature digitali/informatiche) e dall'altro indagare sulla visione dei professionisti (le conoscenze, le competenze, i progetti BIA in corso e la predisposizione tecnologica al loro sviluppo).

Lo scopo di questo lavoro, attraverso la valutazione delle risposte ricevute, è quello di valutare le opinioni attuali sul valore dell'IA nei laboratori clinici italiani e di identificare le sfide che probabilmente la società scientifica italiana dovrà affrontare per promuovere l'introduzione dell'IA in questo campo.

METODI

Il questionario è stato progettato dai membri del GdS e distribuito tramite newsletter dedicate (inviate tra aprile e luglio 2021) a 1351 soci SIBioC attraverso la piattaforma SurveyMonkey (SurveyMonkey Inc.). Ogni socio poteva partecipare all'indagine una sola volta ed è stata raccomandata vivamente la partecipazione dei Direttori del Laboratorio, soprattutto per le domande di carattere organizzativo. Era possibile tralasciare alcune domande, pertanto il numero totale di risposte non corrisponde sempre al numero totale di partecipanti.

Le domande dell'indagine (Tabella 1) sono state suddivise in cinque categorie:

- caratteristiche generali (sei domande). Questa voce comprendeva i profili professionali dei partecipanti, la loro distribuzione demografica e geografica e la tipologia di struttura in cui prestano servizio;
- adeguatezza della dotazione di strumenti digitali (nove domande). Questa sezione ha indagato le infrastrutture dal punto di vista informatico, in termini di attrezzature hardware e software, connettività e utilizzo di piattaforme cloud;

- accesso ai dati sanitari (sette domande). Si è concentrata sul grado di fruibilità e integrazione dei dati clinici e di laboratorio, componente cruciale per le applicazioni BIA;
- gestione e analisi dei dati di laboratorio (tre domande). Sono stati indagati i sistemi informatici di laboratorio (LIS) dei laboratori clinici;
- Big Data e Intelligenza Artificiale (nove domande). L'ultima sezione intendeva sondare i livelli di conoscenza e competenza dei partecipanti, le percezioni e opinioni sul ruolo delle analisi BIA nei laboratori e i principali ostacoli alla loro applicazione, le eventuali esigenze di formazione, lasciando inoltre domande a risposta aperta per presentare i progetti BIA in corso nei laboratori italiani.

I risultati sono stati infine valutati utilizzando il software SurveyMonkey e Google Sheets.

RISULTATI

Sezione 1: Caratteristiche generali (domande 1-6)

Hanno risposto all'indagine 227 (17%) tra 1351 soci SIBioC, impiegati in 484 laboratori clinici italiani, con una partecipazione del 30% fra i Direttori dei Laboratori.

La figura professionale più rappresentata è stata quella del biologo, seguita dai medici di laboratorio e dai tecnici sanitari di laboratorio biomedico (TSLB) (Figura 1). All'interno di una fascia d'età che va dai 20 agli oltre 70 anni, dividendo la popolazione in classi di 10 anni, la maggior parte degli intervistati ha un'età compresa tra i 51 e i 70 anni (n=89, 39,2%; 63 femmine, 26 maschi). Il sesso femminile (n=143, 62,9%) prevale in tutte le fasce d'età (Fisher's exact p=0,021), ad eccezione della fascia

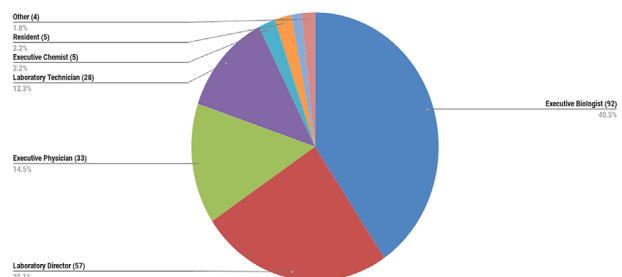


Figura 1
Profili professionali dei partecipanti al questionario.

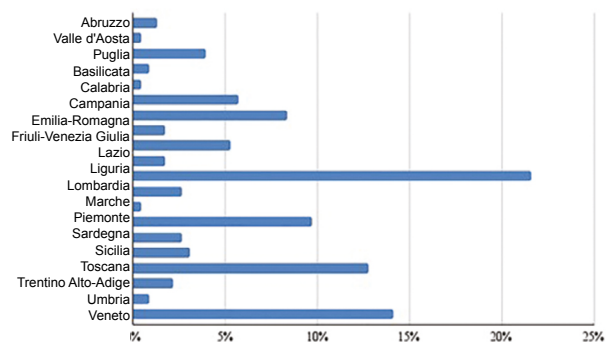


Figura 2
Distribuzione geografica per Regione dei partecipanti

Tabella 1

Caratteristiche dell'indagine: 34 domande in 5 diverse sezioni. Per le singole domande sono indicate le note aggiuntive, quando presenti nel questionario.

Sezione	Domanda	Note aggiuntive
SEZIONE 1: Caratteristiche generali (domande 1-6)	1. Profilo professionale di chi risponde al questionario	
	2. Sesso	
	3. Età	
	4. Tipo di Struttura in cui lavori	
	5. Regione in cui è collocata la struttura	
	6. N° esami/anno effettuati in media nel tuo laboratorio	
SEZIONE 2: Adeguatezza della dotazione di strumenti digitali (domande 7-15)	7. Quante postazioni con collegamento ad Internet ci sono in laboratorio rispetto al numero di operatori?	Indicare il rapporto operatori/ postazione
	8. È disponibile una rete Wi-Fi aziendale?	
	9. Il laboratorio è dotato di postazioni per lavorare e/o fare riunioni online?	Dotazione Webcam, microfono, possibilità di installare il software adeguato
	10. Come valuti la qualità delle connessioni in termini di velocità e stabilità?	1: per niente adeguata 4: del tutto adeguata
	11. Come valuti la dotazione software delle postazioni?	1: per niente adeguata 4: del tutto adeguata
	12. Come valuti la dotazione hardware delle postazioni?	1: per niente adeguata 4: del tutto adeguata
	13. Con che frequenza utilizzi il cloud nel tuo laboratorio?	1: mai; 4: sempre Con risposta MAI le domande 14 e 15 vengono automaticamente saltate
	14. Quale/i piattaforma cloud utilizzi nel tuo laboratorio?	È possibile selezionare più voci
	15. A quali scopi utilizzi il cloud?	È possibile selezionare più voci
SEZIONE 3: Accesso ai dati sanitari (domande 16-22)	16. Oltre ai dati di laboratorio, hai accesso agli altri dati dei pazienti?	Se la risposta è NO, la domanda successiva viene automaticamente saltata
	17. A quali dati hai accesso?	È possibile selezionare più voci
	18. Con che modalità puoi accedere agli altri dati del paziente?	È possibile selezionare più voci
	19. Quale è il livello di integrazione dei dati di laboratorio raccolti nel LIS ^a con gli altri dati sanitari del paziente raccolti in altri database nella tua Azienda?	Per integrazione si intende il collegamento dei dati che provengono dalle diverse fonti interne dell'azienda. Integrazione parziale: dati delle varie branche della Medicina di laboratorio (biochimica clinica, microbiologia, genetica, ematologia, immunoematologia, etc...). Integrazione totale: include dati di diagnostica di laboratorio, di diagnostica per immagini e dati clinici
	20. Se alla precedente domanda, hai risposto che i dati sono integrati, specifica in che modo lo sono	
	21. Nella tua Azienda esiste un sistema di analisi dei dati integrati totali?	Per sistema di analisi di dati integrati si intende per esempio un software che, per mezzo di pagina web o applicativi installati nei computer permetta di effettuare analisi semplici o complesse dei dati che provengono dalle diverse fonti interne dell'azienda.
	22. Invece, nella tua Azienda esiste un sistema di analisi dei dati parziali o del laboratorio?	

Tabella 1 (continua)

SEZIONE 4: Gestione e analisi dei dati di laboratorio (domande 23-25)	23. In che modo il LIS permette di estrarre i dati?	
	24. Come valuti l'adeguatezza dell'estrazione dati ai fini di un'analisi con metodologie di Intelligenza Artificiale in termini di velocità e volume di dati estratti?	1: totalmente inadeguata 4: completamente adeguata
	25. Quali strumenti disponi per l'analisi dei dati di laboratorio?	È possibile selezionare più voci
SEZIONE 5: Big Data e Intelligenza Artificiale (domande 26-34)	26. Quale è il tuo livello di conoscenza sull'argomento BIA ^b ?	1: conoscenza insufficiente 4: conoscenza ottima
	27. Quale è il tuo livello di competenza con le metodologie BIA?	1: competenza insufficiente 4: competenza ottima
	28. Nel tuo laboratorio sono in corso progetti sui BIA?	
	29. Nel laboratorio vi è una figura con competenze nel campo delle soluzioni statistiche applicabili ai big data?	
	30. Il laboratorio è dotato di sistemi per l'analisi dei Big Data?	Per sistemi si intendono software specifici, piattaforme/PC con potenze di calcolo adeguate
	31. Per l'utilizzo dei dati di laboratorio negli studi in cui non è necessario mantenere l'identità del paziente:	Specificare se si sta utilizzando una procedura di anonimizzazione standard, homemade o di altro tipo o se è ancora necessario il consenso informato
	32. Quale credi possa essere il ruolo dell'intelligenza artificiale in laboratorio?	
	33. Nella tua realtà quali sono le principali barriere all'utilizzo dei Big Data e Intelligenza artificiale?	Dai un punteggio per ogni opzione, da 1 a 4, dove 1=barriera minima; 4=barriera massima
34. Saresti interessato ad un corso formativo sui BIA?	È possibile selezionare più voci	

^a LIS; Sistema Informatico di Laboratorio

^b BIA, Big Data e Intelligenza Artificiale

31-40 anni (43,7%).

La maggior parte degli intervistati lavora nei laboratori del Servizio Sanitario Nazionale degli ospedali pubblici (42,7%) o in ospedali universitari (33,9%). Altri lavorano in laboratori privati (10,6%), in ospedali o case di cura private o nelle case di riposo (9,3%), in ospedali di ricerca (IRCCS, 1,3%), in aziende di diagnostica (0,9%) o in altri laboratori non clinici (1,3%).

Le risposte sono pervenute da tutte le 20 regioni italiane, come illustrato nella Figura 2.

La partecipazione al questionario ha permesso di raccogliere dati da strutture laboratoristiche di varie dimensioni rispetto al numero di esami eseguiti annualmente. I partecipanti erano equamente distribuiti tra laboratori che eseguono meno di 1 milione di esami all'anno (29%), tra 1 e 3 milioni (26%) e tra 3 e 6 milioni (31%), mentre una piccola parte era in strutture che eseguono più di 6 milioni di esami all'anno (14%).

Sezione 2: Adeguatezza della dotazione di strumenti digitali (domande: 7-15)

Questa sezione ha esaminato l'adeguatezza delle risorse tecnologiche, requisito fondamentale per l'analisi

dei Big Data. Meno della metà dei partecipanti (41%) ha dichiarato un rapporto di 1:1 tra personal computer (PC) e operatori, mentre in oltre il 30% delle risposte è stato segnalato solo 1 PC ogni 3 o più operatori. Una rete Wi-Fi aziendale è disponibile in quasi due terzi dei casi (66%). La disponibilità di webcam e microfoni con la possibilità di installare i software necessari per le videoconferenze

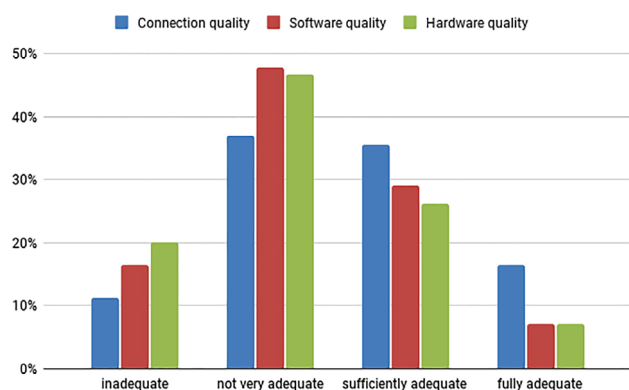


Figura 3
Giudizio sulla qualità della connessione e sull'adeguatezza delle infrastrutture software e hardware

è limitata nella maggior parte dei casi a un numero esiguo di postazioni (62%) e addirittura del tutto assente nel 20% dei casi. Il giudizio sull'adeguatezza della connessione in termini di velocità e stabilità è eterogeneo, mentre sia sulla qualità della dotazione software che hardware prevale l'insoddisfazione (Figura 3). In questa prima serie di domande relative all'adeguatezza delle dotazioni digitali, non sono state riscontrate differenze tra le risposte ricevute da istituzioni pubbliche e private; inoltre, non risultavano dipendenti dal numero di esami effettuati dal laboratorio. Solo circa un quarto (26%) dei partecipanti utilizza regolarmente in laboratorio soluzioni informatiche basate su sistemi Cloud. Tra le risposte ricevute, non ci sono differenze tra quelle provenienti da diverse categorie professionali (biologi, medici e TSLB) e tra Direttori di Laboratorio e altri professionisti. Tra coloro che hanno indicato l'utilizzo di piattaforme Cloud (188/227), i Cloud aziendali e Google Drive sono le piattaforme più diffuse (rispettivamente 41 e 38%), seguite da Dropbox (9%), Microsoft One Drive (8%) e altre (4%). Solo nel 25% delle risposte, il Cloud viene utilizzato non solo come archivio di conservazione dei documenti ma anche come ambiente di lavoro integrato, utilizzando le versioni online dei software e dei servizi offerti.

Sezione 3: Accesso ai dati sanitari (domande 16-22)

Riguardo alla possibilità di accedere ai dati sanitari dei pazienti, è interessante notare che quasi due terzi degli 208 intervistati (60%, 126 risposte) non hanno accesso ad altri dati oltre a quelli di laboratorio (Figura 4). Emerge una impossibilità generalizzata ad accedere ai dati sanitari non di laboratorio dei pazienti, che non dipende da restrizioni come credenziali riservate a particolari ruoli professionali. Questo problema è stato infatti riscontrato anche filtrando il dataset sulla popolazione specifica dei Direttori di Laboratorio: il 58% dichiara di non avere accesso. Tra coloro che invece hanno accesso ad altri dati sanitari non di laboratorio, vengono menzionati come tipologie, in ordine di frequenza, i dati relativi alla richiesta di esami di laboratorio (quesito diagnostico/ esenzione, 52%), alla cartella clinica ospedaliera (34%)

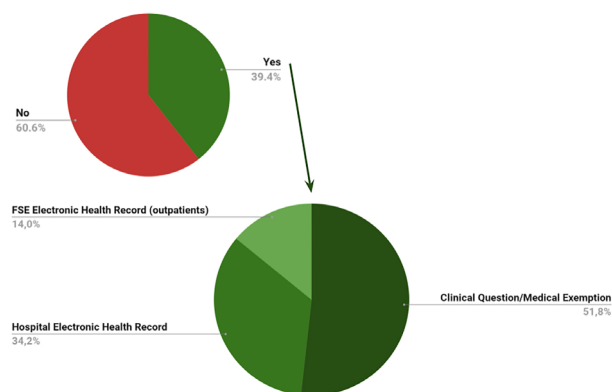


Figura 4
Limitazioni all'accesso ai dati sanitari non di laboratorio e tipo di dati recuperabili nei casi in cui questi siano accessibili.

o al fascicolo sanitario elettronico (14%). Per quanto riguarda le modalità di accesso ai dati, nella metà dei casi (53%) l'accesso è sempre possibile con le proprie credenziali, nell'altra metà (46%) l'accesso è consentito solo al personale di laboratorio autorizzato, mentre in pochi casi i dati vengono forniti solo su richiesta a speciali strutture esterne al laboratorio (1%).

Inoltre, è stato esaminato il grado di integrazione dei dati di laboratorio raccolti nel LIS con altri dati sanitari dei pazienti raccolti nei database aziendali. Per "integrazione" si intende il collegamento tra i diversi dati che l'azienda ottiene dalle varie fonti interne. L'integrazione è stata definita come totale (comprendente i dati di diagnostica di laboratorio, di imaging e clinici) o parziale (comprendente solo le branche della MdL, come la chimica clinica, la microbiologia, la genetica, l'ematologia e l'immunoematologia). La maggior parte degli intervistati (63%) ha indicato l'esistenza di una forma di integrazione dei dati sanitari; tuttavia si tratta soprattutto di un'integrazione parziale (105 risposte, 80%) invece che totale (49 risposte, 20%). Una percentuale considerevole tra gli intervistati non è a conoscenza del livello di integrazione nella propria azienda (30 risposte, 14%), mentre nel 23% si riferisce la mancanza totale di integrazione (48 risposte). Tra coloro che hanno riferito l'esistenza di integrazione alla domanda precedente quasi la metà degli intervistati (59 delle 125 risposte ricevute a questa domanda, 47%) non è stata in grado di offrire alcuna informazione sul tipo di piattaforme per l'archiviazione dei dati integrati, mentre i dati sono integrati in data warehouse aziendale nella maggior parte dei casi (57 risposte, 46%), e solo in pochi casi (9 risposte, 7%) su piattaforme Cloud.

È stata esaminata anche la presenza nelle aziende di sistemi di analisi dei dati totalmente integrati, come ad esempio software che, attraverso pagine web o applicazioni installate sui PC, consentono analisi semplici o complesse di dati provenienti da diverse fonti all'interno dell'azienda. Quasi la metà delle 197 risposte in questo campo è ambigua ("non so", 95/197, 48%), seguita dalla disconferma dell'esistenza di tali software ("no", 71/197, 36%), mentre solo in una piccola percentuale viene segnalata la presenza di software per la sola analisi descrittiva (17/197, 9%) o per l'analisi sia descrittiva che predittiva (14/197, 7%). Restringendo il campo ai software per l'analisi dei dati di laboratorio, si registra un aumento delle risposte affermative, con la presenza di software per l'analisi descrittiva nel 30% dei casi (58 su 196 risposte) e per l'analisi sia descrittiva che predittiva nel 16% dei casi (31 risposte). Nel 22% delle risposte (n=44) è stata segnalata l'assenza di tali software, mentre il 32% degli intervistati (n=62) non era a conoscenza della situazione nella propria azienda.

Sezione 4: Gestione e analisi dei dati di laboratorio (domande 23-25)

La gestione dei dati del LIS è stata esaminata nella quarta sezione. Più della metà degli utenti (63%, 123 su 195 risposte) è in grado di estrarre i dati dal LIS in autonomia rispetto ai servizi IT, per lo più tramite

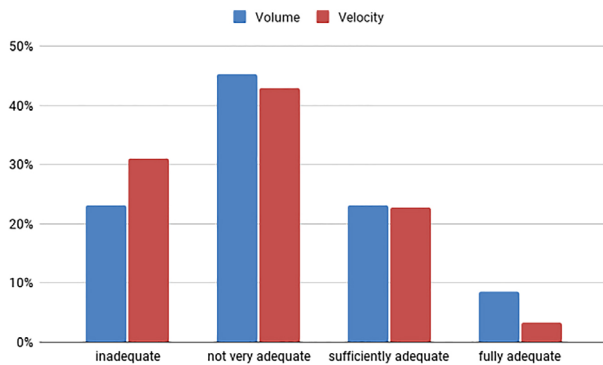


Figura 5

Valutazione sul volume dei dati e sulla velocità di estrazione dal Sistema Informatico di Laboratorio.

modalità preimpostate (42%) e solo in un quinto dei casi in modo libero e personalizzabile (21%). Negli altri casi per estrarre dati è necessaria una richiesta al supporto informatico (28%) oppure raramente (2%) non è possibile estrarre dati o gli intervistati non erano a conoscenza delle modalità di estrazione dati (7%). A coloro che sono in grado di estrarre autonomamente i dati (63%) è stato chiesto, nella domanda successiva, di valutare l' idoneità dell' estrazione dei dati, in termini di velocità e volume dei dati estratti, allo scopo di "analisi con metodologie di Intelligenza Artificiale": la velocità di estrazione e il volume dei dati non soddisfano rispettivamente il 73% e il 68% degli intervistati (punteggio 1-2). I punteggi assegnati alla velocità e al volume di estrazione dal LIS sono mostrati nella Figura 5. Per quanto riguarda gli strumenti disponibili per l'analisi dei dati, era possibile selezionare diverse risposte. La maggior parte ha una funzionalità integrata nel LIS (61%) e utilizza fogli di calcolo (49%). Solo il 16% fa riferimento all'uso di software più specifici. I programmi più utilizzati tra quelli citati dagli intervistati sono MedCalc, R, Analyze-it, SPSS, GraphPad e Phytion. Poiché R e Phytion sono i software più adatti per l'analisi dei Big Data (BD), è importante sottolineare che solo un numero molto ridotto di intervistati li utilizza effettivamente (solo due risposte per Phytion e quattro per R).

Sezione 5: Big Data e Intelligenza Artificiale (domande 26-34)

Solo il 20% dei 183 partecipanti ha dichiarato di avere una conoscenza buona (15%) o eccellente (5%) dell'argomento BIA, e una percentuale ancora minore (12%) ritiene di avere un livello di competenza buono (9%) o eccellente (3%) nelle metodologie BIA.

Tra le possibili applicazioni dell'IA, gli intervistati hanno dato maggiore considerazione ai sistemi esperti per il rilascio dei risultati di laboratorio e la stima degli intervalli di riferimento, mentre minore importanza è stata data agli algoritmi predittivi, al riconoscimento delle immagini, e all'elaborazione del linguaggio. In effetti, questi ultimi due (linguaggio e comprensione delle immagini) sono attualmente meno percepiti rispetto ad altri campi di applicazione (quali la radiologia), anche se già esistono applicazioni in ematologia e MdL urinaria

(20, 21).

Nella maggior parte dei laboratori (91%, 166 risposte su 182) non ci sono progetti in corso che prevedano l'uso di BIA. Inoltre, nella maggior parte dei laboratori (62%, 110 risposte su 178) non esiste una posizione professionale con competenze nel campo delle soluzioni statistiche applicabili ai Big Data. Tuttavia, quasi un terzo (28%, 49 su 178) ha dichiarato di avere collaborazioni esterne attive con personale informatico. Solo nel 10% dei casi (19 su 178) è presente personale di laboratorio con competenze specifiche sulle analisi BIA.

Allo stesso modo, software, piattaforme e PC specifici dedicati all'analisi BIA sono quasi assenti nei laboratori (145 risposte su 180, 80%), in una percentuale di casi si ricorre a collaborazioni esterne (29 risposte, 16%), mentre solo sette hanno dichiarato di avere a disposizione software dedicati. Il questionario ha approfondito la gestione degli studi in cui non è necessario mantenere l'identità del paziente e le eventuali procedure di anonimizzazione, con riferimento al problema della gestione della privacy nell'utilizzo di grandi quantità di dati. La maggior parte dei 170 intervistati (60%) ha riferito di utilizzare una procedura di anonimizzazione homemade, mentre nel 40% dei casi è ancora richiesto il consenso informato. Nessuno ha segnalato l'uso di una procedura standardizzata.

La maggior parte dei 170 partecipanti (91%) prevede che l'IA sarà utilizzata in MdL per supportare i professionisti nel loro processo decisionale, mentre solo una piccolissima percentuale (2%) prevede che il professionista di laboratorio possa essere sostituito in alcuni compiti, come la lettura degli strisci di sangue periferico o dei sedimenti urinari. Solo una minima parte degli intervistati prevede un ruolo marginale (2%) o non sa quale potrebbe essere il ruolo ipotizzabile dell'IA nella MdL (4%). Inoltre alcuni intervistati ipotizzano anche una facilitazione alla ricerca in MdL con l'aiuto della IA (risposta libera non prevista tra le opzioni).

Le risposte più comuni, tra le 168 ricevute, alla domanda su quale sia il principale impedimento all'applicazione dei BIA, hanno individuato come barriere più importanti l'inadeguatezza delle infrastrutture informatiche e la mancanza di software specifici, seguiti dalla scarsa integrazione tra le diverse fonti di dati, dalla mancanza di competenze e dalle difficoltà di accesso ai

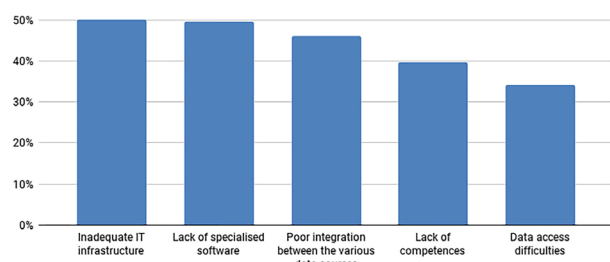


Figura 6

Opinione relativa alle maggiori barriere alle implementazioni Big data e intelligenza artificiale (BIA).

dati. Tuttavia gli intervistati sostanzialmente concordano sul fatto che non esista un singolo ostacolo su cui concentrarsi, ma piuttosto che tutte le problematiche menzionate siano importanti barriere allo sviluppo dei BIA nei laboratori clinici italiani (Figura 6).

Infine, la stragrande maggioranza dei partecipanti (95%, 162 su 170 intervistati) ha espresso interesse per la formazione sui BIA in MdL. Tra questi la maggioranza vorrebbe acquisire competenze per la gestione dei dati (35%) o l'analisi dei dati (37%), mentre un terzo degli intervistati necessiterebbe di un corso finalizzato alla conoscenza dell'argomento (28%).

DISCUSSIONE

Il questionario ha cercato di evidenziare il livello di competenza, conoscenza e interesse dei professionisti italiani del laboratorio clinico nel campo BIA e delle sue applicazioni nella MdL. L'indagine ha ricevuto un rilevante riscontro dai partecipanti SIBioC (227 risposte), consentendo di ottenere un margine di errore limitato sulle risposte complessive (circa il 6% con un livello di confidenza del 95%). I vantaggi di un sondaggio online sono già stati dimostrati rispetto ai metodi tradizionali (22) e, in particolare, per i ricercatori di scienze sociali della salute (23). Inoltre, è stato possibile escludere il bias di rappresentatività (che potrebbe essere presente in un sondaggio basato sul web), dal momento che tutte le persone intervistate utilizzano quotidianamente la posta elettronica per le attività lavorative.

Un'altra indagine sul valore dell'IA nella MdL è stata recentemente condotta da Paranjape et al. (24) negli Stati Uniti su una piccola coorte di soggetti (n=128), e comprendeva personale coinvolto nella MdL, con un target diverso (le principali professioni intervistate erano medici, Direttori di Laboratorio e patologi), ma con una stessa distribuzione per fascia d'età. Inoltre, Paranjape et al. hanno utilizzato risposte "aperte" alle domande, che hanno permesso di raccogliere opinioni, ma non dati numerici. Tuttavia, alcuni risultati sono simili a quelli del nostro studio. Anche Paranjape et al. hanno sottolineato la mancanza di conoscenze specifiche sul tema delle BIA nella comunità medica. In effetti riferiscono che il valore percepito dell'IA osservato non differisce da quello della popolazione generale e sembra riflettere solo la popolarità di queste nuove tecnologie. Molti intervistati non erano sicuri del "perché l'IA sarebbe o non sarebbe utile, cosa sia necessario per adottare comodamente l'IA o come essere istruiti sull'IA". Come risulta evidente non solo dal nostro questionario, ma anche dall'indagine americana, è necessario introdurre l'IA nella formazione medica. Paranjape et al. chiedono anche la partecipazione responsabile delle aziende del diagnostico a fornire attività di formazione e istruzione.

Altri risultati analoghi riguardano i prerequisiti relativi alle infrastrutture tecnologiche, che sono stati segnalati come carenti da entrambi i questionari. In particolare, i nostri risultati rivelano una generale mancanza di risorse hardware e software, la scarsità di PC, la carenza di reti Wi-Fi aziendali e un basso livello di soddisfazione soggettiva per quanto riguarda le apparecchiature software e

hardware. Le interfacce software e hardware sono necessarie non solo per gestire i Big Data, ma anche per utilizzare gli strumenti sviluppati dall'IA o per condividere le opinioni in diversi ospedali o persino all'interno di altri reparti dello stesso ospedale. Inoltre, le risorse Cloud non sono ben implementate nei laboratori italiani e questo potrebbe rendere difficile l'implementazione di strategie di IA basate sul Cloud.

La nostra indagine evidenzia che circa due terzi dei partecipanti, compresi quelli della categoria dei Direttori di Laboratorio, non possono accedere ai dati sanitari che derivano da fonti diverse dal laboratorio. Secondo gli intervistati, i LIS non garantiscono la conformità a tre delle cinque "V" che definiscono i Big Data perché si applicano solo ai dati di laboratorio e il volume e la velocità di estrazione non sono adeguati (25). Quest'ultima informazione rappresenta una vera e propria barriera per gli studi di IA. Infine, solo un piccolo numero di partecipanti ha dichiarato che il proprio laboratorio dispone di una figura professionale specifica con competenze nella scienza dei dati (data scientist).

In sintesi, i risultati nel loro complesso offrono l'opportunità di evidenziare importanti limiti per l'implementazione dell'IA nei laboratori clinici; la maggior parte dei laboratori, essendo dotati di infrastrutture non adeguate all'implementazione di compiti ad alto contenuto tecnologico, come quelli richiesti dall'IA, è in grado di condurre solo parzialmente una ricerca basata sull'IA. Sebbene la carenza di infrastrutture tecnologiche possa essere dovuta ad una carenza di risorse economiche, purtroppo anche l'aggiornamento dei software (come i LIS) potrebbe comunque non essere sufficiente a colmare le lacune che impediscono il successo nei laboratori di una IA ben gestita. Per essere pronti alle future applicazioni di IA, i LIS dovrebbero essere liberamente interfacciabili con applicazioni di terze parti e i produttori di software dovrebbero includere questa possibilità nelle loro versioni aggiornate. Gli aspetti educativi e gli sforzi di collaborazione non sono secondari rispetto alla considerazione precedente. Il ruolo degli Specialisti in MdL non sarà quello di sviluppare algoritmi di IA, ma dovranno piuttosto assumersi la responsabilità di aiutare i data scientist e gli ingegneri a selezionare l'algoritmo giusto sulla base delle informazioni biologiche e mediche sui misurandi di laboratorio (26). Le questioni etiche sono un ostacolo allo sviluppo di applicazioni di IA. Trovare un equilibrio tra gli aspetti etici e le regole del settore potrebbe non essere facile, in quanto norme troppo complicate potrebbero essere più facilmente trascurate o violate, e tuttavia la tutela della privacy è oggi giorno della massima importanza, considerando anche la necessità di ridurre al minimo il rischio di fughe di dati. In un articolo, Pennestri et al. sottolineano queste esigenze e affermano che *"poche regole chiare, una sostanziale chiarezza di intenti, la semplificazione, la flessibilità e la formazione professionale sull'IA rappresentano probabilmente un approccio migliore che sia il Framework europeo che le linee guida britanniche sembrano aver colto"* (27). Infine, la società scientifica e i produttori dei diagnostici *in vitro* dovrebbero stabilire una forte collaborazione non solo per migliorare reciprocamente le conoscenze in questi

campi, ma anche per sviluppare strumenti funzionali e utili basati sull'IA.

Questo studio presenta diversi limiti. In primo luogo, nonostante la forte adesione al questionario documentata dall'elevato numero di risposte, si tratta comunque di una percentuale trascurabile rispetto al totale dei laboratori italiani (stimati in circa 4 000), quindi i risultati sono solo una rappresentazione dello stato dell'arte italiano. Inoltre, il numero di risposte ricevute relative alla Sezione 5 è inferiore a quello delle altre sezioni. Abbiamo ipotizzato che questo divario possa essere dovuto alla specificità delle domande, che abbia scoraggiato i partecipanti o che non siano stati in grado di rispondere, avvalorando così ulteriormente i risultati dell'indagine.

In conclusione, le opinioni raccolte mostrano che nessuno degli ostacoli allo sviluppo di BIA in MdL spicca più degli altri, sottolineando la necessità di migliorare molti aspetti che impediscono l'utilizzo di queste nuove metodologie: dall'adeguamento dei sistemi informatici (data warehouse che combinino le varie fonti di dati, acquisizione di software specializzati per l'analisi BIA, risoluzione delle limitazioni all'accessibilità e all'utilizzo dei dati nel rispetto della privacy) (27), alla gestione della formazione e all'acquisizione di nuove competenze. Il ruolo delle società scientifiche è di indubbio valore per quanto riguarda la necessità di formazione, esplicitata dalla maggior parte degli intervistati. I clinici dovrebbero dedicare del tempo per apprendere i fondamenti di queste nuove tecnologie per poter valutare le opportunità di sperimentazione clinica (28). In effetti, il coinvolgimento degli specialisti in MdL è fondamentale per garantire che i dati di laboratorio siano sufficientemente disponibili e consapevolmente incorporati in progetti scientifici clinici di successo (5).

CONFLITTO DI INTERESSI

Nessuno.

BIBLIOGRAFIA

1. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng* 2018;2:719–31.
2. Austin C, Kusumoto F. The application of Big Data in medicine: current implications and future directions. *J Intervent Card Electrophysiol* 2016;47:51–9.
3. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017;2:230–43.
4. Cabitza F, Banfi G. Machine learning in laboratory medicine: waiting for the flood? *Clin Chem Lab Med* 2018;56:516–24.
5. Carobene A, Milella F, Famigliani L, et al. How is test laboratory data used and characterised by machine learning models? A systematic review of diagnostic and prognostic models developed for COVID-19 patients using only laboratory data. *Clin Chem Lab Med* 2022. <https://doi.org/10.1515/cclm2022-0182>. (Epub ahead of print).
6. Ronzio L, Cabitza F, Barbaro A, et al. Has the flood entered the basement? A systematic literature review about machine learning in laboratory medicine. *Diagnostics* 2021;11:372.
7. Wieringa G. Teaching the pony new tricks: competences for specialists in laboratory medicine to meet the challenges of disruptive innovation. *Clin Chem Lab Med* 2019;57:398–402.
8. Guerranti R. What is meant by Big Data in Laboratory Medicine? 52th National Congress of the Italian Society of Clinical Biochemistry and Clinical Molecular Biology (SIBioC – Laboratory Medicine). *Clin Chem Lab Med* 2020;58:eA1–15.
9. Guerranti R, Padoan A, Angeletti D, et al. Introduction to big data and artificial intelligence in laboratory medicine. *Biochim Clin* 2021;45:57–67.
10. 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.
11. Vidali M. I big data e la medicina di laboratorio. *Biochim Clin* 2021;45:13–4.
12. Carobene A, Campagner A, Sulejmani A, et al. Identification of SARS-CoV-2 positivity using machine learning methods on blood count data: external validation of state-of-the-art models. *Biochim Clin* 2021;45:281–9.
13. Campagner A, Carobene A, Cabitza F. External validation of machine learning models for COVID-19 detection based on Complete blood count. *Health Inf Sci Syst* 2021;9:37.
14. Famigliani L, Campagner A, Carobene A, et al. A robust and parsimonious machine learning method to predict ICU admission of COVID-19 patients. *Med Biol Eng Comput* 2022;1–13. <https://doi.org/10.1007/s11517-022-02543-x> (Epub ahead of print).
15. Aita A, Padoan A, Guerranti R, et al. Wrong blood in tube: a SIBioC project for a persistent problem. *Biochim Clin* 2022;46:51–7.
16. 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.
17. Trenti T, Pecoraro V, Pirotti T, et al. IgM anti-SARS-CoV2-specific determination: useful or confusing? Big Data analysis of a real-life scenario. *Intern Emerg Med* 2021;16: 2327–30.
18. Carobene A, Aarsand AK, Bartlett WA, et al. The European biological Variation Study (EuBIVAS): a summary report. *Clin Chem Lab Med* 2021;60: 505–17.
19. Carobene A, Campagner A, Uccheddu C, et al. The multi-center European Biological Variation Study (EuBIVAS): a new glance provided by the Principal Component Analysis (PCA), a machine learning unsupervised algorithms, based on the basic metabolic panel linked measurands. *Clin Chem Lab Med* 2022;60:556–68.
20. Herman DS, Rhoads DD, Schulz WL, et al. Artificial intelligence and mapping a new direction in laboratory medicine: a review. *Clin Chem* 2021;67:1466–82.
21. Badrick T, Banfi G, Bietenbeck A, et al. Machine learning for clinical chemists. *Clin Chem* 2019;65: 1350–6.
22. Fleming CM, Bowden M. Web-based surveys as an alternative to traditional mail methods. *J Environ Manag* 2009;90:284–92.
23. Wright KB. Web-based survey methodology. In: Liamput-tong P, editor. *Handbook of research methods in health social sciences*. Singapore: Springer; 2019:1339–52 pp.
24. Paranjape K, Schinkel M, Hammer RD, et al. The value of artificial intelligence in laboratory medicine. *Am J Clin Pathol* 2021;155: 823–31.
25. Gupta S, Kar AK, Baabdullah A, et al. Big Data with cognitive computing: a review for the future. *Int J Inf Manag* 2018; 42:78–89.
26. Padoan A, Plebani M. Flowing through laboratory clinical data: the role of artificial intelligence and big data. *Clin Chem Lab Med* 2022. <https://doi.org/10.1515/cclm-2022-0653> (Epub ahead of print).

27. Pennestrì F, Banfi G. Artificial intelligence in laboratory medicine: fundamental ethical issues and normative key-points. *Clin Chem Lab Med* 2022. <https://doi.org/10.1515/cclm-2022-0096> (Epub ahead of print).
28. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digit Med* 2020;3:126.