

REVIEW

Intelligenza Artificiale e Big Data in ambito medico: prospettive, opportunità, criticità

Artificial Intelligence and Big Data in Medicine: scenarios, opportunities and critical issues

N. Musacchio¹, G. Guaita², A. Ozzello³, M.A. Pellegrini⁴, P. Ponzani⁵, R. Zilich⁶, A. De Micheli⁷

¹Presidente Fondazione AMD, Past President AMD. ²Responsabile Servizio Diabetologia, Endocrinologia e Malattie Metaboliche, ATS Sardegna-ASSL Carbonia. ³Responsabile Struttura Semplice Dipartimentale di Malattie Endocrine e Diabetologia, ASL TO3, Pinerolo (TO). ⁴CDN Fondazione AMD, FriulCoram, Udine. ⁵Dirigente Medico SSD Endocrinologia, Diabetologia e Malattie Metaboliche, ASL3 Genovese. ⁶Partner Mix-x - ⁷ACISMOM, Genova.

Corresponding author: nicoletta.musacchio@gmail.com



OPEN
ACCESS



PEER-
REVIEWED

Citation N. Musacchio, G. Guaita, A. Ozzello, M.A. Pellegrini, P. Ponzani, R. Zilich, A. De Micheli (2018) Intelligenza Artificiale e Big Data in ambito medico: prospettive, opportunità, criticità. JAMD Vol. 21-3

Editor Luca Monge, Associazione Medici Diabetologi, Italy

Received October, 2018

Accepted October, 2018

Published October, 2018

Copyright © 2018 Musacchio et al. This is an open access article edited by [AMD](#), published by [Idelson Gnocchi](#), distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement All relevant data are within the paper and its Supporting Information files.

Funding The Authors received no specific funding for this work.

Competing interest The Authors declare: ¹Novo, Lilly, Sigma Tau, Doc Generici; ²Astra Zeneca, Takeda, Lilly, Boehringer Ingelheim, Novartis; ³no competing interests; ⁴Lilly; ⁵Lilly, Novo Nordisk, Novartis, GSK, Bayer; ⁶no competing interests; ⁷no competing interests.

Fondazione AMD

Nicoletta Musacchio (Presidente), Salvatore De Cosmo, Alberto De Micheli, Annalisa Giancaterini, Carlo Giorda, Giacomo Guaita, Valeria Manicardi, Alessandro Ozzello, Maria Antonietta Pellegrini, Paola Ponzani, Giuseppina Russo

Consiglio Direttivo AMD

Domenico Mannino (Presidente), Paolo Di Bartolo (vice Presidente), Alberto Aglialoro, Amodio Botta, Riccardo Candido, Riccardo Fornengo, Alfonso Gigante, Antonino Lo Presti, Ernesto Rossi, Giovanni Sartore, Franco Tuccinardi (Consiglieri), Agata Chiavetta (Coordinatore della Consulta), Giovanni Perrone (Segretario), Gaudenzio Stagno (tesoriere)

ABSTRACT

In the last two years digital information around the world has more than doubled and this trend is going to increase exponentially by generating huge amounts of digital data: the Big Data. Medicine is a key player of this growth because of four important phenomena: the digitization of diagnostic imaging, digital reporting replacing of paper folders, the development of the biotechnology in the field of so-called “omics” sciences and the explosion of the IoMT (Internet of Medical Things). Through the techniques of Machine learning, computers are able to learn from the data without being explicitly programmed for this, creating predictive models.

Artificial intelligence includes information systems that can perform tasks normally requiring human intelligence, such as visual perception, language recognition, decision making, translating from one language to another. It was born a cultural, technological and scientific phenomenon centered on the application of machine learning to very large databases. Artificial intelligence, computer vision and machine learning systems are proving that today the machines analyze large amounts of data faster and better than human beings.

The new paradigm of a science that “simulate the human reasoning” instilled doubts whether this phenomenon is under human control. This review provides a summary and a glossary on the artificial intelligence technologies in medicine and on the main issues that their use can generate, in the view that machine learning is not a wand that can turn data into gold, but an invaluable tool, more and more necessary for medicine and modern health care system, whose complexity today exceeds the human mind ability. The human intelligence combined with the artificial intelligence – i.e. a knowledgeable and empathic clinician, empowered by reliable predictive tools and clinical decision support systems and unburdened by repetitive work – can bring us closer to the true “patient centered care.”

Key words artificial intelligence; big data analytics; clinical decision making; human behaviour; healthcare.

RIASSUNTO

Negli ultimi due anni le informazioni digitali di tutto il mondo sono più che raddoppiate e questa tendenza è destinata ad aumentare in modo esponenziale generando enormi moli di dati elettronici: i Big Data. La medicina è uno dei principali protagonisti di questa crescita a motivo di quattro importanti fenomeni: la digitalizzazione della diagnostica per immagini, la reportistica digitale in sostituzione delle cartelle cartacee, lo sviluppo di biotecnologie impiegate nel campo delle cosiddette scienze “omiche”, l’esplosione dell’IoMT (internet of medical things). Attraverso le tecniche di Machine learning i computer sono in grado di imparare dai dati, senza essere stati esplicitamente programmati per questo, generando modelli predittivi.

L’intelligenza artificiale comprende sistemi informatici che possono eseguire attività che normalmente richiedono intelligenza umana, come la percezione visiva, il riconoscimento del linguaggio, prendere delle decisioni, tradurre da una lingua all’altra. È in atto una rivoluzione culturale, tecnologica e scientifica centrata sull’applicazione dell’apprendimento automatico a database di grandissime dimensioni. Intelligenza Artificiale, visione artificiale e sistemi di apprendimento automatico stanno dimostrando che oggi le macchine analizzano grandi quantità di dati più velocemente e meglio degli esseri umani.

Il nuovo paradigma di una scienza che “simula il ragionamento umano” ha instillato dubbi sul fatto che questo fenomeno sia sotto il controllo umano.

La presente revisione fornisce una sintesi ed un glossario sulle tecnologie di intelligenza artificiale

in medicina e sulle principali criticità che il loro utilizzo può generare, nella prospettiva che l’apprendimento automatico non sia una bacchetta magica che può trasformare i dati in oro, ma uno strumento prezioso, sempre più necessario per la medicina ed il sistema sanitario moderno, la cui complessità oggi supera la capacità della mente umana. L’intelligenza umana che lavora con l’Intelligenza Artificiale – cioè un clinico ben informato, empatico, dotato di rigorosi strumenti predittivi e di guida nelle scelte e alleggerito dal lavoro ingrato e ripetitivo – può portare ad avvicinarsi maggiormente alla realizzazione della vera “cura per il paziente”.

Parole chiave intelligenza artificiale; analisi dei big data; scelte cliniche guidate; comportamento umano; assistenza medica.

DI CHE COSA STIAMO PARLANDO:

DEFINIZIONI E CONTESTO

Big Data e Medicina

Negli ultimi due anni le informazioni digitali di tutto il mondo sono più che raddoppiate e questa tendenza è destinata ad aumentare, con un fenomeno esponenziale che genera enormi moli di dati elettronici: i Big Data.

Il termine Big Data è stato originariamente coniato dagli scienziati della NASA nel 1997 in seguito alla difficoltà di visualizzare e memorizzare un set dati troppo grande, limitandone di conseguenza l’analisi⁽¹⁾. I Big Data sono caratterizzati dalle 4 ‘V’: Volume (grandi quantità), Velocità (vengono continuamente riversati nella rete), Varietà (sono file di tutti i tipi: testi, immagini, video, audio, etc.), Veridicità (le fonti non sono sempre verificabili⁽²⁾, quindi prendere decisioni basandosi sui Big Data in modo acritico può essere pericoloso).

Il termine Big Data è stato originariamente coniato dagli scienziati della NASA nel 1997 in seguito alla difficoltà di visualizzare e memorizzare un set dati troppo grande. I Big Data sono caratterizzati dalle 4 ‘V’: Volume, Velocità, Varietà, Veridicità).

Si prevede che nel 2020 l’universo digitale conterrà tanti bit quante sono le stelle dell’universo⁽³⁾.

La medicina è uno dei principali protagonisti di questa crescita: i Big Data della salute stanno aumentando in percentuale maggiore rispetto ad altri settori, in virtù di quattro importanti fenomeni.

Il primo è quello dello sviluppo digitale della diagnostica per immagini, che ha cancellato la vecchia generazione di macchine analogiche e, nei tentativi di prevenzione sempre più spinta e d'identificazione precoce dei segnali di malattia, sta vedendo l'affermarsi di nuove potenti tecnologie diagnostiche digitali. La proliferazione di questi strumenti ha determinato da tempo la creazione dello standard DICOM (Digital Imaging and Communication in Medicine) che definisce le regole per l'archiviazione e la condivisione delle immagini.

Il secondo fenomeno è legato alle tecniche di reportistica digitale, ovvero le cartelle e i fascicoli elettronici dei pazienti, che in pochi anni sostituiranno i supporti cartacei. Negli USA, per accelerare questa transizione, nel 2011 è stato approvato il Health Information Technology for Clinical Health (HITECH) Act, nell'ambito del quale sono stati stanziati più di 35 miliardi di dollari in soli incentivi ai medici per adottare cartelle elettroniche che abbiano interoperabilità in tutto il territorio nazionale, con l'obiettivo di non limitarsi a una mera "adozione" delle cartelle elettroniche, ma con l'intento di ottenere "significativi miglioramenti nelle cure" nell'ambito del programma MU (Meaningful Use)⁽⁴⁾.

Il terzo fenomeno che sta facendo avanzare la frontiera dei Big Data della salute è lo sviluppo delle biotecnologie impiegate nel campo delle cosiddette scienze "omiche" (genomica, trascrittomica, proteomica...), discipline che hanno per oggetto uno studio molto approfondito della cellula, attraverso un'analisi dettagliata dei processi biologici osservati a diversi livelli, nello stesso intervallo di tempo⁽⁵⁾. Le sofisticate tecnologie di biologia molecolare utilizzate, portano a una produzione di dati estremamente elevata. L'obiettivo ultimo è quello di fornire terapie personalizzate, mirate alle caratteristiche del singolo individuo, a cui si ispira la medicina di precisione. Per ora l'evoluzione più completa di questo approccio è la medicina dei sistemi, che analizza nel loro insieme e contemporaneamente i geni del DNA (genomica) e le relative funzioni (genomica funzionale), i trascritti del DNA, ovvero l'RNA (trascrittomica), le proteine (proteomica) e i metaboliti che si trovano nell'organismo (metabolomica). Vengono studiate le modalità con cui interagiscono le molecole (interattomica), le interazioni con la flora intestinale (microbiomica) e come le stesse possano modificare il DNA (epigenomica). È un metodo olistico che opera con modalità integrative, per rispondere a quesiti biologici complessi, come la patogenesi, la storia naturale e l'evoluzione delle malattie.

Il quarto fenomeno, ma non ultimo in termini di dimensioni, è rappresentato dall'esplosione dell'IOT

(Internet Of Things o internet delle cose) che, in generale, comprende tutti gli oggetti di uso comune che con l'evoluzione tecnologica sono diventati smart, incorporando sensori intelligenti in grado di raccogliere una grande varietà di informazioni e trasmetterle alla rete. In medicina questa tendenza si riferisce alla sensoristica che rileva in tempo reale informazioni dal corpo umano, ambito che sta assumendo importanza e specificità tali da avere un nome tutto suo: la IoMT (Internet of Medical Things). Siamo già abituati agli smartwatch e alle bande da polso che rilevano dati come il battito cardiaco, la temperatura e i movimenti, ma adesso è in atto una rivoluzione che produrrà pervasivamente strumenti indossabili, impiantabili e ingoiabili in grado di captare informazioni su parametri fisici, meccanici, chimici ed elettromagnetici. Non solo, sono già disponibili strumenti che, rilevando determinati valori, possono agire di conseguenza: i sensori del CGM systems – continuous glucose monitoring – impiantabili, dotati di allarmi anche predittivi per i livelli di glucosio al di fuori dei range desiderati e capaci di inviare i dati ad uno smartphone e il pancreas artificiale ne sono un esempio.

Uno dei vantaggi degli IoMT è la possibilità di favorire l'aderenza alle terapie, ambito che in diabetologia è già familiare, visto che gli strumenti di automonitoraggio della glicemia e la condivisione dei dati fra medico e paziente sono una realtà già da qualche anno, ma non è così in altre aree terapeutiche. Un altro esempio di IoMT orientato all'aderenza è il sensore della Proteus, un device ingeribile che rileva la presenza del medicinale nello stomaco del paziente e invia i dati a un'applicazione smartphone.

I BIG DATA DELLA SALUTE STANNO CRESCENDO IN PERCENTUALE MAGGIORE RISPETTO AD ALTRI SETTORI, IN VIRTÙ DI 4 IMPORTANTI FENOMENI.

- 1 Digitalizzazione della diagnostica per immagini
- 2 Tecniche di reportistica digitale in sostituzione delle cartelle cartacee
- 3 Sviluppo di biotecnologie impiegate nel campo delle cosiddette scienze "omiche"
- 4 Esplosione dell'IOT (internet delle cose, o internet of things)

I SOFTWARE D'INTELLIGENZA ARTIFICIALE

Oggi non si parla più di Big Data tout court. È ormai un fatto acquisito che il valore dei dati è rappresentato dalla conoscenza che essi contengono, estratta con

L'utilizzo di tecnologie informatiche. Inizialmente, con i software di descriptive analytics, era possibile visualizzare le informazioni in modo da descrivere con grafici e tabelle i fenomeni rappresentati dai dati. Successivamente sono arrivati i software di predictive analytics, con capacità predittive basate su algoritmi che individuano le correlazioni fra i dati e consentono di associare a determinate situazioni, come per esempio esiti medici di un certo tipo, valori ricorrenti di specifici parametri, con una logica di tipo 'if... then'. In generale, i software in grado di generare nuova conoscenza dai dati ricadono in un contenitore piuttosto ampio che è quello dell'Intelligenza Artificiale: essa comprende i sistemi informatici che possono eseguire attività che normalmente richiedono intelligenza umana, come la percezione visiva, il riconoscimento del linguaggio, prendere delle decisioni, tradurre da una lingua all'altra. Vi sono diversi modi di simulare l'intelligenza umana, alcuni sono più intelligenti di altri. Se le regole che determinano il comportamento di un software sono state esplicitamente programmate da un essere umano, si parla di Good Old-Fashioned Artificial Intelligence (GOFAI): il focus risiede nella capacità del programmatore di scomporre il problema in tante parti e programmare i componenti che le devono eseguire, non vi è 'intelligenza autonoma' nel sistema. Poi è arrivato il *machine learning*, che ha rivoluzionato il mondo dell'Intelligenza Artificiale.

Intelligenza Artificiale: comprende i sistemi informatici in grado di eseguire attività che normalmente richiedono intelligenza umana, come la percezione visiva, il riconoscimento del linguaggio, prendere delle decisioni, tradurre da una lingua all'altra.

IL MACHINE LEARNING

Nel 1959, Arthur Samuel, considerato un pioniere del machine learning, lo definì come "la scienza che mette i computer in grado di imparare, senza essere stati esplicitamente programmati per questo". Infatti, i software di machine learning non sono appositamente programmati per far comportare il sistema in un certo modo: essi si auto-modificano in base ai dati con cui sono alimentati. In altre parole, ridefiniscono il loro comportamento, che non è preordinato, a seconda dei dati a cui sono esposti. Si tratta di software basati su algoritmi matematici che simulano ragionamenti di tipo induttivo, *imparando* dalle informazioni ("ciò che

dovrà accadere, spesso è già scritto nei dati"): questo avviene attraverso il riconoscimento di *pattern*, ovvero delle regolarità nei dati che permettono di classificare determinate situazioni e di ricondurle a specifici esiti (per esempio, outcome clinici, come il peggioramento di una patologia, l'emergere di complicanze o altro). In questo modo vengono generati dei modelli predittivi che consentono di capire che determinate combinazioni di valori di alcuni parametri sono spesso associate a specifiche condizioni cliniche.

Machine learning: "la scienza che mette i computer in grado di imparare, senza essere stati esplicitamente programmati per questo". Si tratta di software basati su algoritmi matematici che simulano ragionamenti di tipo induttivo, imparando dai dati e generando modelli predittivi.

McKinsey stima che il machine learning in medicina genererà un valore pari a 100 miliardi di dollari all'anno⁽⁶⁾. Gli ambiti maggiormente interessati saranno: la medicina di precisione, la scoperta di nuovi farmaci, la progettazione ed analisi dei trial clinici, l'interpretazione più precisa dei referti radiologici, la capacità di prevedere le epidemie.

Il Massachusetts Institute of Technology (MIT) ha creato un gruppo specifico che si occupa di machine learning applicato ed un esempio è il Clinical Machine Learning Group che ha fra le proprie priorità la comprensione della patogenesi del diabete di tipo 2 e l'identificazione dei trattamenti più efficaci⁽⁷⁾.

I MODELLI PREDITTIVI "BLACK BOX" E L'INTELLIGENZA ARTIFICIALE "TRASPARENTE"

I modelli predittivi possono essere generati con tecniche diverse. Un'importante differenza riguarda gli algoritmi definiti "black box", o scatola nera, e gli algoritmi di intelligenza artificiale "trasparente". C'è una prima parte di analisi che è identica per entrambe le tipologie: viene generato un modello predittivo che consente d'inserire le informazioni relative a un nuovo individuo per capire, in base a ciò che ha imparato il modello dai dati pregressi, se l'individuo esaminato ricada con più probabilità in un certo tipo di casistica, che potrebbe essere quella di sviluppare una certa patologia. Il modello fornisce un responso SI/NO (es. soggetto a rischio oppure no). Il modello predittivo

di tipo “black box” ci dice se il soggetto è a rischio, ma non ci spiega il perché (ovvero, non ci dice se quel soggetto è a rischio perché ha dei particolari valori di determinati parametri). Questo è il tipico funzionamento dei modelli generati con tecniche quali le reti neurali o le support vector machine (SVM): sono “scatole chiuse” che, in virtù di complesse funzioni matematiche, ci consegnano un responso del tipo SI/NO, senza aggiungere altre informazioni. Invece, le tecniche di intelligenza artificiale trasparente, oltre al responso predittivo SI/NO, rendono anche noti i parametri su cui si basa la previsione. Ovvero, spiegano a chi utilizza il modello quali siano le logiche interne utilizzate dal modello stesso per poter fornire la sua valutazione. Questo è il tipico funzionamento degli alberi decisionali e di un particolare algoritmo detto “Logic Learning Machine” che, in più, descrive anche in linguaggio naturale le regole con cui ‘ragiona’ il modello predittivo⁽⁸⁾.

Il funzionamento “black box” può essere ricondotto al concetto di “decisioni automatiche”: il software decide autonomamente, senza fornire spiegazioni. Invece, il funzionamento trasparente, abilita le cosiddette “decisioni aumentate” o “augmented decision”: il software non decide autonomamente, ma consente ad un essere umano esperto in un certo campo, per esempio un medico, di prendere una decisione basandosi sull’evidenza delle correlazioni scoperte dal modello e di usare queste informazioni aggiuntive integrandole con il bagaglio delle proprie conoscenze pregresse.

Le combinazioni di valori critici dei diversi parametri possono anche essere rappresentate con dei nomogrammi, tipici strumenti che consentono di assegnare dei punteggi e visualizzare con scale grafiche le variabili che hanno una rilevanza prognostica.

Il fatto di lavorare con strumenti trasparenti e di conoscere i parametri su cui si basa un modello predittivo permette di estrarre il massimo valore dall’Intelligenza Artificiale, perché abilita un’analisi di tipo prescrittivo (o prescriptive analytics): ovvero, la conoscenza generata col modello può consentire una modifica della situazione, per evitare che essa degeneri verso un outcome indesiderato. Per esempio, se i fattori di rischio individuati dal modello sono elementi su cui si può intervenire, con farmaci o con un diverso stile di vita, è possibile cambiare l’evoluzione della patologia per un certo individuo e, magari, riportare quel soggetto ad una situazione dagli esiti meno infausti.

Modelli predittivi – Predictive Analytics: i modelli predittivi utilizzano dati storici per prevedere eventi futuri. Forniscono un responso SI/NO (per es. soggetti a rischio o non a rischio di sviluppare complicanze). Possono essere di tipo “black box”, come le reti neurali e le SVM – Support Vector Machines – o di tipo “trasparente”, come gli alberi decisionali e la “Logic Learning Machine”. I modelli black box ci dicono se il soggetto può essere a rischio, ma non ci spiegano il perché.

Invece, i modelli trasparenti spiegano a chi utilizza il modello quali siano le logiche interne utilizzate dal modello stesso, evidenziando i parametri su cui basano la previsione. Inoltre, l’algoritmo di “Logic Learning Machine” esplicita anche in linguaggio naturale le regole di funzionamento del modello.

Modelli prescrittivi – Prescriptive analytics: i modelli predittivi trasparenti consentono di estrarre il massimo valore dall’intelligenza artificiale perché abilitano un’analisi di tipo prescrittivo (o prescriptive analytics). Siccome la conoscenza generata dal modello viene esplicitata, essa può consentire di intervenire in una specifica situazione, per evitare che essa degeneri verso un outcome indesiderato.

IL NUOVO GENERAL DATA PROTECTION REGULATION (GDPR) EUROPEO E L’INTELLIGENZA ARTIFICIALE “BLACK BOX”

Il GDPR richiede che qualsiasi decisione frutto di modelli predittivi che possa avere delle implicazioni “legali” debba avere anche una spiegazione esplicita. C’è ancora un acceso dibattito fra gli esperti su che cosa abbia implicazioni legali, ma è chiaro che qualsiasi decisione derivante da un algoritmo di machine learning che riguardi un individuo implichi anche un “diritto di spiegazione”⁽⁹⁾. Il GDPR si esprime nel seguente modo: “le persone hanno il diritto di non essere soggette a decisioni basate esclusivamente su processi automatizzati” e ancora: “I processi decisionali completamente automatizzati dovrebbero essere soggetti a salvaguardie, che dovrebbero includere informazioni specifiche da fornire alla persona interessata e al suo diritto di ottenere un intervento umano, di esprimere il suo punto di vista e di ottenere una spiegazione della decisione presa – in base a quali valutazioni – e di poter quindi contestare tale decisione”.

Il DGPR è solo un inizio, con molti aspetti ancora poco chiari che verranno dipanati negli anni a venire, ma di sicuro uno dei suoi fondamenti è il fatto che processare dati personali deve essere: legale, corretto e trasparente.

INTELLIGENZA ARTIFICIALE E BIG

DATA: LA VERA CHIAVE

RISOLUTIVA PER AUMENTARE

LA CONOSCENZA?

Le opinioni entusiastiche e/o allarmistiche sulle potenzialità e/o le conseguenze dell'uso dell'Intelligenza Artificiale in medicina si moltiplicano. Al di là della sensibilità di ognuno, che può vedere nella tecnologia un'opportunità o una minaccia, quello che è certo è l'enorme cambiamento, che è solo agli inizi. Da un certo punto di vista, si può dire che stiamo attraversando una delle fasi della naturale evoluzione del pensiero scientifico e della professione medica, siamo passati dal medico generalista, ai diversi specialisti, per arrivare al super-specialista dei giorni nostri, che per alcune specialità è estremamente "tecnologizzato". Nessuno obiettava sulla pervasività della tecnologia in medicina, i nuovi strumenti erano considerati supporti al servizio del medico. È il nuovo paradigma di una scienza che "simula il ragionamento umano" ad aver instillato dubbi sul fatto che anche questo fenomeno sia sotto il controllo umano. Non esiste una risposta sulla pericolosità reale o presunta di un'Intelligenza Artificiale che possa sostituirsi alle decisioni umane. Stephen Hawking, collegato al web summit di Lisbona del 2017, ha sollevato la questione dicendo che "Le nostre Intelligenze Artificiali devono fare quel che vogliamo che facciano", sostenendo che non possiamo ancora prevedere che cosa davvero sarà possibile quando la mente umana sarà amplificata dall'Intelligenza Artificiale, ma che non possiamo ignorare che vi siano anche dei pericoli e il modo migliore di fronteggiarli è quello di identificarli e non ignorare il fatto che le nostre vite verranno trasformate⁽¹⁰⁾. Seguendo il ragionamento di Hawking, il vero "potere" sarà la conoscenza approfondita di questo fenomeno. È importante ricordare che «Ci si preoccupa delle macchine che si umanizzano, ma il vero problema oggi sono i medici che sono diventati delle macchine»⁽¹¹⁾. Se le macchine possono contribuire a liberare per il medico del tempo che possa essere dedicato alla relazione col paziente, questo sarà sicuramente un vantaggio.

L'UNIONE DI DIVERSI DATABASE

DI BIG DATA: VANTAGGI,

PROBLEMATICHE TECNICHE E

GIURIDICHE

Un altro aspetto che non va sottovalutato quando si parla di Big Data è che non sempre i dati sono immagazzinati nello stesso database. Anzi, un grosso valore delle analisi che si possono fare sui dati è proprio quello di mettere assieme archivi diversi, che offrono punti di vista diversi, ma che per poter essere visti come un tutt'uno devono avere dei "punti di connessione". Per esempio, un nuovo interessante ambito di analisi in medicina è quello di verificare, a fronte di determinate attività gestionali, come si ottengano risultati clinici migliori. I processi organizzativi, o percorsi di gestione o cura del paziente, hanno sicuramente un impatto sulla qualità dell'assistenza e sugli esiti assistenziali. Questo impatto può essere misurato mettendo assieme dati clinici e dati gestionali, che normalmente non sono archiviati negli stessi database. Pertanto, l'analisi di questi ambiti potrà avere luogo solo integrando le diverse basi dati, ed è qui che possono nascere problemi che riguardano la privacy: perché, se è tecnicamente e organizzativamente facile anonimizzare un singolo database contenente dati di pazienti, quando devono essere uniti archivi diversi è innanzitutto necessario che i codici dei pazienti, pur anonimi, non perdano il riferimento allo specifico individuo a cui fanno riferimento. Questo è solo un esempio di come i problemi di privacy, già rilevanti in altri ambiti, possano rendere tutto molto più complesso quando si parla di informazioni che riguardano la salute delle persone. Queste difficoltà verranno superate quando esisterà un'unica entità con la responsabilità di raccogliere e integrare i dati sanitari di tutte le regioni, unificati con codici-paziente coerenti ma anonimi.

L'unione di diversi database di Big Data permetterà di mettere assieme archivi diversi in particolare dati clinici, gestionali / amministrativi, biologici ed epidemiologici.

AMBITI DI UTILIZZO DELLA

INTELLIGENZA ARTIFICIALE IN

MEDICINA E VANTAGGI ATTESI⁽¹²⁾

I Big Data sono un "fenomeno culturale, tecnologico e scientifico" centrato sull'applicazione dell'apprendimento automatico a database di grandissime dimensioni⁽¹³⁾.

Le organizzazioni sanitarie hanno oggi database di ingenti dimensioni, contenenti dati clinici, biologici, epidemiologici ed amministrativi; altri dati di possibile interesse medico possono provenire da fonti diverse, quali ad esempio i dispositivi indossabili che tracciano la mobilità, misurano variabili biologiche, monitorizzano gli stili di vita oppure le informazioni sui comportamenti individuali o richieste sulla salute dei social networks o dei motori di ricerca⁽¹⁴⁾. Intelligenza Artificiale, visione artificiale e sistemi di apprendimento automatico stanno dimostrando che oggi le macchine analizzano grandi quantità di dati più velocemente e meglio degli esseri umani⁽¹⁵⁾.

Attraverso l'Intelligenza Artificiale si possono quindi analizzare ed interpretare i dati disponibili per formulare ipotesi ed ottenere risposte utili per una diagnosi precoce ed un trattamento migliore, in più sartorializzati sulla peculiarità del singolo paziente^(16,17); in altri termini è possibile perseguire gli obiettivi di efficacia, efficienza ed appropriatezza tipici della medicina attuale, attraverso l'utilizzo di tecnologie capaci di affinare ed innovare le classiche metodologie epidemiologiche su quantità di dati molto vaste. L'analisi può riguardare sia i processi collegati ad esiti solidi, quali mortalità e morbilità, sia la soddisfazione dell'utente.

Altre applicazioni riguardano l'automatizzazione della diagnostica⁽¹⁸⁾, la gestione economica e la ricerca.

Questa rivoluzione tecnologica è alla base della realizzazione della medicina delle 4P volta a superare il concetto della "cura delle malattie" per rivolgersi alla "gestione complessiva del benessere degli individui".

Il termine 4P Medicine è stato coniato dallo scienziato Leroy Hood. Le 4P stanno per prevenzione, predizione, personalizzazione e partecipazione, una rivoluzionaria architettura clinica basata su una visione generale e d'insieme del paziente. Le attese sono il miglioramento delle cure, la loro estensione ad un numero più grande di malati, la riduzione dei costi, lo stimolo all'innovazione e quindi alla creazione di nuove attività⁽¹⁹⁾.

4P Medicine esprime il passaggio culturale dalla "cura delle malattie" alla "gestione complessiva del benessere degli individui". Le 4P stanno per prevenzione, predizione, personalizzazione e partecipazione.

Schematicamente si possono definire 4 ambiti di applicazione dell'Intelligenza Artificiale (Figura 1 e Tabella 1):

- Cura del paziente
- Diagnostica generale e per immagini
- Gestione
- Ricerca e sviluppo

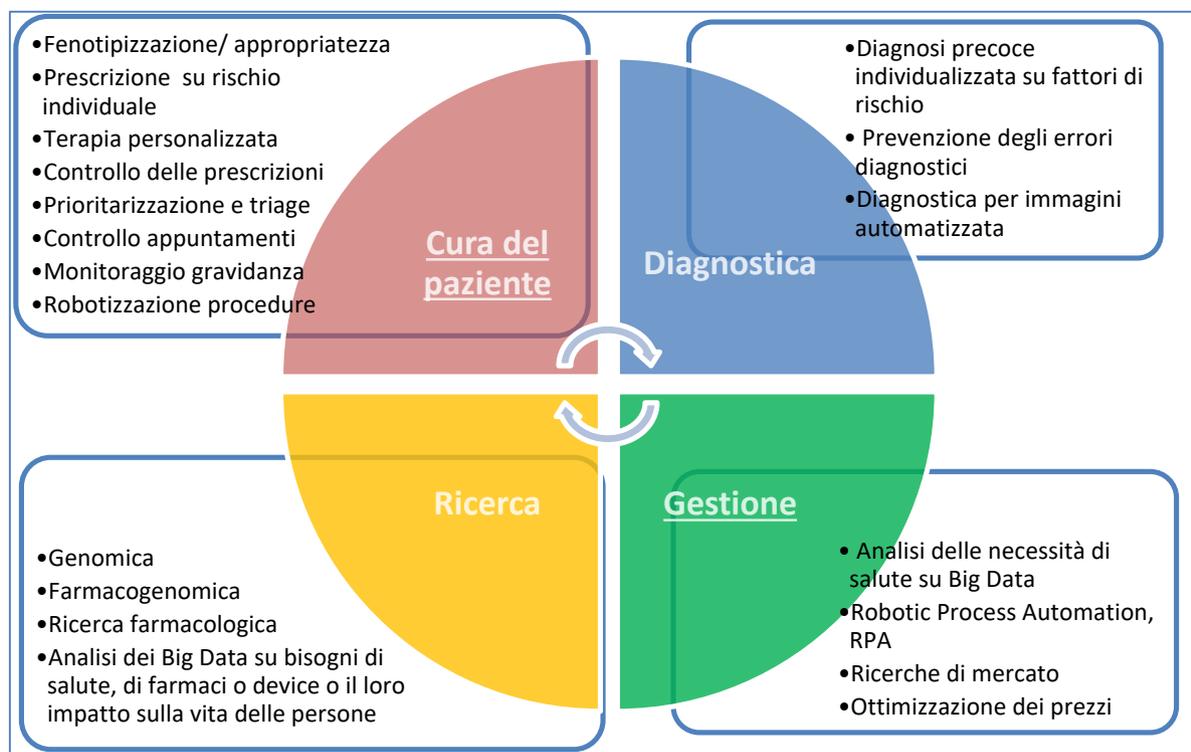


Figura 1 | Ambiti di utilizzo dell'intelligenza artificiale in Medicina (modificata da rif. 12).

Tabella 1 | Ambiti di utilizzo dell'Intelligenza Artificiale in medicina.

<p>CURA DEL PAZIENTE</p> <ul style="list-style-type: none"> • Prescrizione di analisi appropriate per tipologia e tempistiche sulla base delle caratteristiche dei pazienti, attraverso l'individuazione di fenotipi specifici nell'ambito della malattia • Valutazione di rischi individuali o di fattori prognostici per prescrivere azioni ad hoc • Prescrizione di farmaci personalizzata sulla base delle caratteristiche del paziente • Prescrizione assistita ed automatizzata con sistemi di audit automatizzato capaci di controllare e ridurre gli errori di prescrizione • Prioritarizzazione e triage in tempo reale: soluzioni per il triage capaci di valutare sintomi clinici multipli, creare scale di priorità e quindi affidare allo specialista più appropriato • Individuazione dei pazienti in cura cronica a maggiore rischio di non presentarsi agli appuntamenti di follow up • Monitoraggio di particolari situazioni cliniche come la gravidanza per la diagnosi precoce di possibili complicanze • Robotizzazione delle procedure interventistiche 	<p>DIAGNOSTICA</p> <ul style="list-style-type: none"> • Diagnosi precoce individualizzata sulla base dei fattori di rischio • Diagnosi assistita od automatizzata per una maggiore sensibilità e per la prevenzione degli errori diagnostici • Diagnostica per immagini automatizzata (indagini radiologiche, screening dermatologici, esame del fundus oculi)
<p>GESTIONE</p> <ul style="list-style-type: none"> • Analisi di dati clinici, epidemiologici ed amministrativi su grandi numeri per gestire la salute della popolazione, utilizzare efficacemente le risorse e ridurre i costi • Analisi delle necessità di salute per programmare l'offerta di servizi efficienti ed efficaci • Automatizzazione intelligente delle procedure amministrative e del reporting (Robotic Process Automation, RPA) con capacità dei sistemi di adeguarsi in modo adattivo ai cambiamenti dell'ambiente in cui operano • Ricerche di mercato • Ottimizzazione dei prezzi sulla base delle richieste del mercato e della concorrenza 	<p>RICERCA E SVILUPPO</p> <ul style="list-style-type: none"> • Analisi del genoma, ricerca sulla radice genetica delle malattie per ottimizzare il trattamento e la prevenzione • Ricerca sui farmaci e sulla connessione fra farmaci e malattie • Analisi dei Big Data per comprendere i bisogni di salute, di farmaci o device o il loro impatto sulla vita delle persone

CRITICITÀ

Esiste una condivisa fiducia che l'utilizzo dei Big Data da fonti diverse porterà vantaggi per la cura dei pazienti, per la ricerca di terapie innovative, per l'attività degli operatori sanitari e per i gestori della salute pubblica⁽²⁰⁾.

Vi sono tuttavia alcuni aspetti che presentano delle criticità:

- La qualità dei dati, spesso provenienti da database nati per altri fini e non per la cura e la ricerca scientifica.
- La validità ed il corretto utilizzo dei dati.
- La disponibilità di dati clinici informatizzati ancora largamente incompleta (nel Regno Unito si calcola che gli ospedali saranno totalmente "senza carta" non prima del 2027).
- Il rispetto delle normative sulla privacy nella condivisione e trasmissione dei dati dei pazienti.
- La difficoltà tecnica ed anche formale/ autorizzativa (proprietà del dato vs. trasparenza) di raccogliere dati da fonti diverse.
- La gestione delle grandi modificazioni nella pratica clinica e nella formazione professionale dei medici e degli altri operatori sanitari indotte dall'utilizzo dell'Intelligenza Artificiale, ad esempio per diagnosi e screening automatizzati⁽²¹⁾.

- I problemi etici legati all'utilizzo dell'Intelligenza Artificiale.
- Il rapporto fra l'utilizzo dell'Intelligenza Artificiale ed il consolidato approccio della Evidence-Based Medicine (EBM) nella Medicina Clinica.

LA QUALITÀ DEI DATI

La ricerca scientifica tradizionale si è sempre basata su dati raccolti in modo rigoroso su campioni piccoli, mentre i Big Data forniscono, con bassi costi e facile accessibilità, dati meno puliti ma su tutta la popolazione⁽¹⁴⁾. In questa prospettiva pertanto il volume dei dati potrebbe "superare il loro disordine", anche se i ricercatori tradizionali possono trovare eretico questo approccio epistemologico⁽²²⁾. In altri termini, poiché la popolazione in studio si approssima al 100% anziché essere un 1% pur altamente rappresentativo e accuratamente raccolto, il potere predittivo potrebbe essere maggiore.

I Big Data possono essere analizzati con tecniche di Intelligenza Artificiale, capaci di generare ipotesi dall'analisi stessa dei dati e delle loro associazioni, da testare con analisi ulteriori perché possano confermare il loro eventuale valore predittivo. Sottolineiamo tuttavia come l'associazione statistica non significhi necessariamente un nesso causale.

Restano comunque alcune perplessità:

- I dati devono essere realmente rappresentativi di una popolazione e non di una sottopopolazione, come per esempio i soggetti seguiti da una particolare organizzazione, in cui, inoltre, la raccolta dei dati potrebbe non essere omogenea nelle diverse sedi assistenziali; la qualità della imputazione e della codificazione deve essere garantita; i criteri diagnostici delle singole condizioni devono essere precisi ed inequivocabili ed anche omogenei nei diversi studi⁽²³⁾.
- I database amministrativi forniscono dati raccolti per scopi non clinici ed i dati delle cartelle elettroniche non sempre sono strutturati per una raccolta adeguata ai fini della ricerca scientifica e talora non sono raccolti con completezza.
- Gli algoritmi non possono riconoscere se modelli o associazioni trovate in assenza di un costruito teorico di base sono vere, spurie o falsate da errori sistematici. A differenza della EBM, il Machine Learning non ha un sistema per valutare la possibilità di errore sistematico o la qualità delle prove. Inoltre non può distinguere variazioni di procedura giustificate o ingiustificate o analizzare le sequenze di decisioni cliniche¹⁵

Grandi moli di dati possono essere analizzate con tecniche di Intelligenza Artificiale, capaci di creare le ipotesi dall'analisi degli stessi e dalle loro associazioni. L'associazione statistica non significa necessariamente un nesso causale.

VALIDITÀ ED UTILIZZO CORRETTO DEI DATI

Esistono fondamentalmente due ordini di problemi.

- I sistemi di supporto decisionale fondati sull'apprendimento automatico associano dati empirici ad una interpretazione categorica. Potenziali conseguenze non intenzionali di questo approccio possono derivare dalla formalizzazione in un modello di decisione della mappatura tra i segni fisici che un medico può valutare e la loro "corretta" classe, identificata dagli osservatori. Nella pratica medica spesso gli osservatori non concordano sui segni diagnostici e sulla valutazione dei risultati e ciò non è solo soggettivo, ma legato ad una intrinseca ambiguità nei fenomeni clinici osservati. Quindi i dati di input per l'apprendimento artificiale non sono necessaria-

mente univoci e precisi e ciò influenza inevitabilmente i risultati prodotti⁽²⁴⁾.

Ulteriore ricerca deve essere finalizzata a sviluppare e convalidare degli algoritmi di apprendimento automatico adatti a dati in entrata non precisi ed univoci come le informazioni mediche, anziché imporre un'idea di accuratezza e completezza dei dati non presente nei registri medici, in cui la qualità dei dati non è ottimale⁽²⁵⁾.

- Affidarsi a sistemi di supporto decisionale fondati sull'apprendimento automatico potrebbe voler dire accettare che un dato digitale possa essere la rappresentazione completa ed affidabile di un fenomeno clinico; ciò diviene problematico se i medici perdono la coscienza della esistenza di elementi clinici non inclusi in una registrazione digitale. Ciò può portare ad una cattiva interpretazione delle risposte provenienti da un sistema automatico ed alla perdita di un approccio olistico al paziente nei suoi aspetti psicologici, relazionali e sociali ed anche ad una decontestualizzazione dagli aspetti organizzativi nella interpretazione dei dati⁽²⁵⁾.

Ad esempio, un modello prognostico realizzato con l'utilizzo dell'Intelligenza Artificiale ha dato l'indicazione che i malati asmatici ricoverati per polmonite muoiono meno dei non asmatici⁽²⁶⁾, dato chiaramente poco accettabile per i clinici. In vero, il modello utilizzato non aveva considerato che i malati con asma sono direttamente ricoverati in terapia intensiva per prevenire complicazioni e questa modalità organizzativa non poteva essere inserita nel modello prognostico, in cui l'asma diveniva quindi un elemento protettivo.

In queste prospettive un'agenzia indipendente che certifichi i modelli di previsione prima di diffonderli nella pratica clinica potrebbe assicurare che le analisi predittive mantengano le loro promesse di migliori risultati per singoli pazienti e l'intero sistema sanitario⁽²⁷⁾.

I dati di input per l'apprendimento artificiale, come le informazioni mediche, non sono univoci e precisi e ciò influenza i risultati prodotti. Affidarsi a sistemi di supporto decisionale fondati sull'apprendimento automatico potrebbe essere problematico se i medici perdono la coscienza della esistenza di elementi clinici non inclusi in una registrazione digitale.

IL RISPETTO DELLA PRIVACY

L'utilizzo dei Big Data necessariamente pone dei problemi per il rispetto della privacy. Gli individui vivono liberamente la propria vita nell'ipotesi che alcune informazioni personali non siano conosciute da chi non è autorizzato. La crescente disponibilità e lo scambio di informazioni sulla salute sostengono i progressi nella cura dei singoli e nella sanità pubblica, ma facilitano anche pratiche di marketing invasivo e discriminazione al di fuori della legge.

È difficile conciliare le due esigenze e sono necessarie regole e leggi ad hoc, anche perché all'evolvere delle tecniche di deidentificazione segue una altrettanto rapida evoluzione delle tecniche di reidentificazione attraverso dati su database meno regolati⁽²⁸⁾.

Il prezzo dell'innovazione non può essere l'erosione del diritto fondamentale alla privacy assicurato dalla legge^(20,29).

L'utilizzo dei Big Data necessariamente pone dei problemi per il rispetto della privacy; sono necessarie regole e leggi ad hoc.

LA PRATICA CLINICA: LA POSSIBILE

RIDUZIONE DELLE ABILITÀ

PROFESSIONALI DEI MEDICI

L'utilizzo nella clinica di sistemi di supporto decisionale fondati sull'apprendimento automatico può portare ad un eccesso di fiducia sulle capacità dell'automazione e quindi ad un suo elevato utilizzo. Questo in prospettiva potrebbe ridurre nei professionisti il livello di capacità necessarie nel portare a termine un compito in larga parte automatizzato, con gravi conseguenze in caso di malfunzionamento del sistema stesso⁽³⁰⁾.

È necessaria ulteriore ricerca per capire meglio se l'eccessivo affidamento su sistemi di supporto decisionale fondati sull'apprendimento automatico, che potrebbero dare prestazioni anche migliori di un professionista, possa causare una sottile perdita di fiducia in se stessi e influenzare la volontà di un medico a fornire l'interpretazione o la diagnosi definitiva⁽²⁵⁾.

PROBLEMI ETICI

Le criticità sopra riportate hanno anche ovvi risvolti etici.

- Gli algoritmi potrebbero rispecchiare gli stessi pregiudizi umani nelle scelte decisionali o diven-

tare il “magazzino” dell'opinione medica collettiva. Ad esempio un algoritmo predittivo di un rischio su base genetica può portare ad errori se applicato a popolazioni in cui non esistono studi genetici o l'analisi di patologie in cui venga sistematicamente sospesa la cura perché ritenute ad esito infausto può portare alla conclusione che siano comunque incurabili: una profezia che si autoconferma⁽⁴⁸⁾.

- Gli algoritmi potrebbero perseguire obiettivi non etici: ad esempio guidare verso pratiche mediche che soddisfino gli obiettivi amministrativi ma non la reale qualità della cura, creare sistemi di supporto decisionale che perseguano l'aumentato consumo di farmaci o presidi diagnostici, senza che gli utenti clinici lo percepiscano. In altri termini, il conflitto etico potrebbe crearsi per la differenza di intenti fra chi finanzia e realizza un algoritmo e chi lo utilizza, per il diverso obiettivo economico o medico⁽³¹⁾.
- Lo spazio dell'apprendimento automatico nella pratica medica potrebbe divenire eccessivo. Nell'attuale pratica medica i medici spesso non seguono l'evoluzione di una malattia dall'esordio alla fine e la cultura medica si fonda più sui dati della letteratura e sui dati raccolti nei sistemi di cura che sull'esperienza. Questo dà giusto spazio ai sistemi di apprendimento automatico⁽³²⁾, ma rischia di dar loro troppo spazio. In altri termini l'Intelligenza Artificiale rischia di “stabilire” diagnosi e terapia, andando al di là del suo ruolo di strumento di supporto, guidato dai principi universali del rispetto e del beneficio per il paziente.
- Un ulteriore problema etico è la confidenzialità dei dati. L'utilizzo dei dati per costruire algoritmi implica che nessun dato possa essere omesso, pena la perdita di validità. Questo problema comporta risvolti notevoli sul rapporto di fiducia fra medico e paziente.

Occorre creare linee guida etiche per l'utilizzo dell'Intelligenza Artificiale e soprattutto i medici devono conoscerne meccanismi e limiti.

APPRENDIMENTO AUTOMATICO

ED EVIDENCE BASED MEDICINE

L'apprendimento automatico, che converte dati complessi in algoritmi, è una sfida al tradizionale approccio dell'epidemiologia clinica sulla quale si basa la EBM⁽⁴⁵⁾. Comprendere la distinzione tra dati, informazioni e conoscenza è necessario per colmare il

divario tra Big Data e medicina basata sulle prove⁽²²⁾. EBM e machine learning hanno obiettivi comuni sul miglioramento delle conoscenze, della cura e dell'assistenza, ma il machine learning ha l'obiettivo di raggiungerli più rapidamente utilizzando dati già disponibili. Tuttavia l'apprendimento automatico ha un limitato potere esplicativo: l'algoritmo può identificare le correlazioni fra migliaia di variabili, ma non evidenzia il nesso di causalità. Le differenze fra i due approcci sono schematizzate nella tabella 2.

EBM e machine learning si integrano nel miglioramento delle conoscenze, della cura e dell'assistenza.

Proprio le differenze fra EBM e Machine learning rappresentano un'opportunità di integrazione e arricchimento reciproco:

- Gli algoritmi possono facilitare una stima più precisa del rischio individuale, con implicazioni per la scelta tra test diagnostici o terapie che possono successivamente essere confrontate in uno studio randomizzato controllato.
- I modelli di regressione con prestazioni superiori prodotti dall'Intelligenza Artificiale potrebbero essere applicati a studi clinici che utilizzano la biostatistica tradizionale.
- I trial clinici possono verificare le ipotesi generate dall'apprendimento automatico.

La Evidence Based Medicine ha bisogno della potenza di calcolo dei Big Data, e il rigore epistemologico della EBM ha bisogno di grandi quantità di dati. La combinazione di questi due approcci alla conoscenza offre il percorso migliore per allargare e rafforzare la base di conoscenze della medicina clinica⁽²²⁾.

La figura 2 sintetizza la tassonomia degli studi tradizionali o utilizzanti l'Intelligenza Artificiale.

L'apprendimento automatico per affermarsi ed essere compreso deve sviluppare nomenclature condivise, standard di valutazione e di presentazione, analisi comparative ed esplicative dei diversi algoritmi, programmi di formazione per i medici. Questo percorso culturale e formativo è già stato fatto negli ultimi trenta anni dall'EBM, che può aiutare quindi l'apprendimento automatico con la sua esperienza⁽¹⁵⁾.

È necessario sviluppare nomenclature condivise, standard di valutazione e di presentazione, analisi comparative ed esplicative dei diversi algoritmi, programmi di formazione per i medici, affinché l'apprendimento automatico possa affermarsi in ambito sanitario

LA NECESSITÀ DI APRIRE LA SCATOLA NERA DELL'APPRENDIMENTO AUTOMATICO

Come si è già detto nella prospettiva medica non è accettabile un modello di "black box" in cui il razionale

Tabella 2 | Raffronto fra Medicina basata sulle prove e apprendimento automatico (machine learning) (da: rif.15, Scott et al. modificata).

MEDICINA BASATA SULLE PROVE	APPRENDIMENTO AUTOMATICO (MACHINE LEARNING)
Basata su sperimentazioni con protocolli sequenziali ben definiti per verificare una ipotesi	Scoperta guidata dai dati che non utilizza nessun protocollo e opera in parallelo o contemporaneamente
Esamina le relazioni tra un numero limitato di variabili pre-spezifcate con bassa eterogeneità e dimensioni	Esamina le relazioni tra molte variabili non pre-spezifcate con elevata eterogeneità e dimensioni
Utilizza dati di volumi inferiori (megabyte o gigabyte), meno partecipanti (centinaia o migliaia) e una più piccola gamma di fonti (RCT o studi di coorte prospettici)	Utilizza dati, anche non strutturati, di maggiori volumi (terabyte o petabyte), più partecipanti (da migliaia a centinaia di migliaia) e una più vasta gamma di fonti (cartelle elettroniche, insiemi di dati amministrativi, sensori indossabili, banche genomiche e proteomica, social media)
Metodi analitici basati sulla teoria, con assunzioni dichiarate o confermate riguardo a completezza di dati, precisione, classificazione e indipendenza	Gli algoritmi sono agnostici e basati sui dati, con poche ipotesi intorno a completezza di dati, precisione, classificazione e indipendenza
Si basa sul confronto tra gruppi per dedurre il nesso di causalità	Si basa su correlazioni tra variabili all'interno di insiemi di dati per dedurre il nesso di causalità
Utilizza una gerarchia delle prove che riflette il rischio di errori sistematici dei diversi disegni di studio	Non utilizza alcuna gerarchia per valutare il rischio di errori sistematici di diversi algoritmi
La fiducia nelle prove aumenta con risultati costantemente replicati in studi multipli	La fiducia in algoritmi sviluppati in modelli di prova aumenta con risultati costantemente replicati in più modelli di verifica

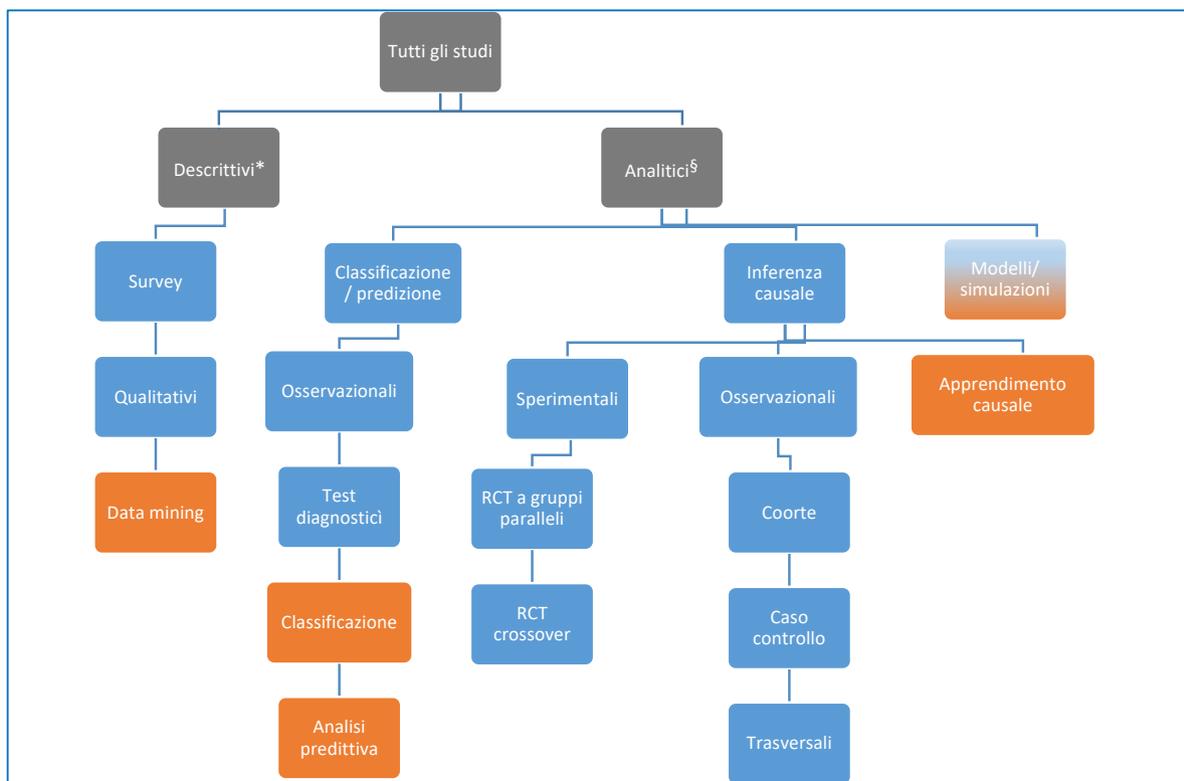


Figura 2 | Tassonomia degli studi con metodologia tradizionali o con Big Data/IA (modificata da Centre for Evidence-Based Medicine⁽³³⁾). * Descrivono una situazione. § Quantificano una relazione. In blu i disegni di studi eseguiti con metodologia classica, in arancione i disegni di studi eseguiti con metodologia Big Data. Modelli e simulazioni possono essere realizzati con entrambe le metodologie.

dell’input generato sia imperscrutabile o comunque esplicitabile solo a fronte di costose e lunghe analisi matematiche. Data la diffusione di queste nuove metodiche di analisi, sarà sempre più importante poter osservare e raffrontare in modo efficiente i differenti modelli aperti in modo da identificare i fattori di confondimento e scegliere il modello migliore.

In questa prospettiva occorre progredire verso sistemi di apprendimento artificiale capaci di proporre in automatico spiegazioni ed offrire ai medici strumenti di visualizzazione interattiva per esplorare le implicazioni delle diverse variabili di esposizione. Questa esigenza è presente in modo pervasivo in tutti i settori che utilizzano l’Intelligenza Artificiale, non solo la medicina, e infatti l’orientamento tecnologico sta cercando di offrire sempre più modelli predittivi “trasparenti”, un esempio dei quali è rappresentato dalla Logic Learning Machine.

La disponibilità di potentissimi strumenti tecnologici non solleva i medici da acquisire competenze più forti nel pesare e giudicare il valore degli strumenti basati sull’Intelligenza Artificiale nella pratica e nella ricerca clinica⁽²⁵⁾. In questa prospettiva è necessario avviare una formazione accademica ad hoc^(34,35).

CONSIDERAZIONI CONCLUSIVE

L’apprendimento automatico non è una bacchetta magica che può trasformare i dati in oro, anche se molte informazioni diffuse tendono a farlo credere. Esso è invece una naturale estensione degli approcci statistici tradizionali, uno strumento prezioso e sempre più necessario per il sistema sanitario moderno. Considerando la grande quantità di informazioni che per un medico è oggi necessario valutare – come storia personale del paziente, malattie familiari, sequenze genomiche, farmaci, attività sui social media, ricoveri in altri ospedali – la decisione clinica può diventare un compito eccessivo per qualsiasi persona. Si può affermare che la complessità della medicina oggi supera la capacità della mente umana⁽³⁵⁾.

L’apprendimento automatico non è una bacchetta magica che può trasformare i dati in oro, ma uno strumento prezioso, sempre più necessario per il sistema sanitario moderno. La complessità della medicina oggi supera la capacità della mente umana.

Nel momento in cui un maggiore controllo delle scelte viene ceduto agli algoritmi, è importante notare che questi nuovi strumenti decisionali non sono dotati di alcuna garanzia di correttezza, equità o persino veridicità. Anche con i migliori algoritmi di apprendimento il giudizio sulla efficacia rimane indispensabile e le migliori pratiche analitiche devono essere utilizzate per garantire che il risultato finale sia robusto e valido. Questo è particolarmente vero nel settore sanitario, dove gli algoritmi hanno il potenziale di influenzare la vita di milioni di pazienti⁽³⁶⁾. L'utilizzo quotidiano dei sistemi di Intelligenza Artificiale aiuterà a conoscere meglio la loro utilità ed i loro limiti, raffrontati con la realtà clinica e l'efficacia delle scelte organizzative che potranno guidare⁽³⁷⁾.

La qualità di qualsiasi sistema di supporto decisionale fondato sull'apprendimento automatico e le decisioni normative sulla sua adozione non dovrebbero essere valutate solo su metriche di performance, ma misurarsi in confronto alla cura abituale su end points solidi come i miglioramenti clinicamente importanti, insieme anche alla soddisfazione dei pazienti e dei medici.

Un atteggiamento prudente e la necessaria ricerca sulle conseguenze impreviste potrebbero contribuire a ridurre le probabilità di errori e danni. Inoltre, se questi si verificassero nonostante questi sforzi, ulteriore ricerca potrebbe aiutare a gestirli e ridurli⁽²⁵⁾.

Allo stesso modo in cui l'automatizzazione della misurazione della pressione sanguigna o della conta delle cellule del sangue ha liberato i medici da alcune attività ripetitive richiedenti tempo, l'Intelligenza Artificiale potrebbe liberare da molte altre e dare un significato nuovo alla pratica della medicina, fornendo nuovi livelli di efficienza e precisione.

La qualità di qualsiasi sistema di supporto decisionale fondato sull'apprendimento automatico e le decisioni normative sulla sua adozione non devono essere valutate solo su metriche di performance, ma in confronto alla cura abituale su end points solidi, come i miglioramenti clinicamente importanti, insieme anche alla soddisfazione dei pazienti e degli operatori sanitari.

I medici, in modo proattivo, devono guidare, sorvegliare e monitorare l'adozione dell'Intelligenza Artificiale come partner nella cura del paziente. Certamente l'intelligenza umana che lavora con l'Intelligenza Artificiale – cioè un clinico ben in-

formato, empatico, dotato di rigorosi strumenti predittivi e di guida nelle scelte e alleggerito dal lavoro ingrato e ripetitivo – può portare ad avvicinarsi maggiormente alla realizzazione della vera “cura per il paziente”⁽³⁸⁾.

L'intelligenza umana che lavora con l'Intelligenza Artificiale – cioè un clinico ben informato, empatico, dotato di rigorosi strumenti predittivi e di guida nelle scelte e alleggerito dal lavoro ingrato e ripetitivo – può portare ad avvicinarsi maggiormente alla realizzazione della vera “cura per il paziente”

GLOSSARIO

Algoritmo: indica una serie di passi computazionali eseguibili in un tempo finito. Prende origine dal nome del matematico arabo Al-Khowarizimi che nel IX secolo definì un metodo per calcolare la somma di 2 numeri nel sistema numerico Hindu. È un concetto fondamentale dell'informatica, perché si riferisce alla sequenza d'istruzioni elementari che devono essere svolte da un programma per risolvere un problema più complesso.

IOT o Internet delle Cose: sono gli oggetti di uso comune in grado di raccogliere dati dal contesto in cui si trovano e mandarli in rete (es. smartphone, componenti di domotica, tecnologie installate sugli autoveicoli, capi d'abbigliamento intelligenti). In medicina: dispositivi dotati di sensori che possono rilevare informazioni dall'organismo e trasmetterle sul web; ad esempio: battito cardiaco, livelli di glucosio, pressione arteriosa, saturazione di ossigeno, temperatura corporea...)

Medicina di precisione o Precision medicine: è un approccio innovativo alle cure mediche che considera le caratteristiche “uniche” di ciascun individuo (patrimonio genetico, stile di vita, ambiente in cui vive, etc). La medicina di precisione, sfruttando le nuove scoperte biotecnologiche, consente di superare il tradizionale concetto di “paziente standard” a cui tradizionalmente si rivolgono le terapie, per arrivare a curare il “singolo individuo”. L'oncologia ne è stato il primo campo di applicazione: molte terapie antitumorali vengono già applicate in base al profilo genetico del paziente.

Medicina dei sistemi, Systems Medicine o Biologia dei Sistemi: studia gli esseri viventi in quanto strutture in evoluzione, la cui natura è definita dall'interazione dinamica delle parti di cui sono

composte. Parte dall'identificazione dei geni e delle proteine presenti in un organismo e verifica la loro evoluzione nel tempo. Sfrutta tecniche di bioinformatica e di matematica-statistica per disegnare dei modelli di funzionamento dei sistemi biologici.

4P Medicine: Significa “Medicina Preventiva-Predittiva-Personalizzata-Partecipativa”. È un termine coniato dallo scienziato Leroy Hood, convinto assertore del passaggio culturale dalla “cura delle malattie” alla “gestione complessiva del benessere degli individui”. Hood sostiene che il concetto di paziente standard fra qualche anno verrà sostituito da quello di “individuo geneticamente unico”, su cui cucire su misura l'approccio più corretto per preservare la sua salute o, meglio, il suo benessere. Questo nuovo, rivoluzionario concetto di medicina, è favorito da 3 tendenze: lo sviluppo della medicina dei sistemi, la disponibilità di dati generati dagli individui (IOT, social network) e la rivoluzione digitale⁽³⁹⁾.

Descriptive analytics o analisi descrittiva dei dati: si basa su tecniche di aggregazione dei dati e visualizzazione tramite grafici per rappresentare la fotografia della situazione contenuta nei dati. Fornisce informazioni su “ciò che è successo in passato”.

Predictive analytics o analisi predittiva dei dati: utilizza modelli statistici e algoritmi matematici predittivi; si basa su dati generati in passato per cercare di prevedere “che cosa potrebbe accadere in futuro?”.

Prescriptive analytics o analisi prescrittiva dei dati: si basa su algoritmi di simulazione e ottimizzazione che forniscono informazioni su possibili esiti (outcome), per provare a rispondere alla domanda “che cosa potremmo fare?”.

Data Mining: riguarda le tecniche informatiche necessarie ad analizzare i dati (anche da più fonti, come nel caso dei Big Data), con l'obiettivo d'individuare “pattern” o correlazioni nascoste. La fase di data mining è un passaggio tecnico imprescindibile per poter ricavare il vero valore dell'analisi dei dati: l'estrazione di conoscenza, attraverso cui è possibile effettuare analisi di tipo predittivo o prescrittivo.

Intelligenza Artificiale: riguarda la capacità di sistemi informatici di risolvere problemi o svolgere attività tipici della mente e delle abilità umane. Alcune importanti branche di questa scienza riguardano: la comprensione del linguaggio, la capacità di auto-apprendere, l'abilità di pianificare, il problem solving.

Machine Learning o Apprendimento Automatico: è una disciplina volta ad aumentare la conoscenza umana attraverso l'uso della tecnologia. Le tecniche di machine learning consentono ai computer, grazie

ad appositi algoritmi di programmazione, di gestire situazioni nuove attraverso l'analisi dei dati, il self-training, l'osservazione e l'esperienza.

Support Vector Machines o Macchine a Vettori di Supporto: sono modelli matematici di machine learning e servono a classificare i dati in modo automatico. Consentono cioè di identificare i criteri per distinguere gli elementi di insiemi disgiunti. La SVM “impara” da un campione di dati esemplificativi in cui gli elementi siano stati classificati in “n” categorie; costruisce così un modello che potrà essere utilizzato per attribuire dei nuovi dati in input ad una delle “n” categorie. Rappresentano una tecnica alternativa alle Reti Neurali.

Reti Neurali o Neural Networks: si tratta di modelli matematici che afferiscono al mondo del machine learning e, in particolare, all'ambito degli strumenti “predittivi”, ma anche di classificazione. Si basano su algoritmi che, analizzando dei dati, apprendono in modo automatico e “adattivo” (ovvero, si modificano con l'esperienza). Vengono utilizzate per simulazioni, per le analisi in campo ingegneristico, nell'intelligenza artificiale e in ambito informatico. Le reti neurali simulano il comportamento dei neuroni del cervello umano, in cui le informazioni vengono elaborate parallelamente nei tanti “nodi” della rete: non vi è una gerarchia centralizzata. Rappresentano una tecnica alternativa alle SVM.

Deep learning: è una sottocategoria del machine learning basata sulle reti neurali, quindi algoritmi che funzionano in modo simile ai neuroni del cervello umano, che lavorano con meccanismi di apprendimento su più livelli. Data la natura di tipo “black box” di questi algoritmi, gli ambiti di applicazione più utili sono la comprensione del linguaggio e la classificazione delle immagini, dove la cosa importante è che l'algoritmo fornisca il risultato, non il fatto che venga esplicitato “come il sistema sia arrivato a una determinata conclusione” (cosa invece fondamentale, per esempio, in ambito diagnostico, dove è cruciale sapere in base a quali valori di determinati parametri, un certo individuo potrebbe sviluppare una specifica patologia: non è sufficiente sapere solo che quella persona sia a rischio di malattia).

BIBLIOGRAFIA

1. <https://foreignpolicy.com/2012/10/08/big-data-a-short-history/>
2. <https://forum.huawei.com/enterprise/en/data/attachment/forum/dm/ecommunity/uploads/2016/0322/17/56f1129356a58.png>
3. <https://www.emc.com/leadership/digital-universe/2014/view/executive-summary.htm>

4. <https://www.igi-global.com/chapter/electronic-health-record-ehr-diffusion-and-an-examination-of-physician-resistance/184077>
5. Hasis Y, Seldin M, Lusic A. 'Multi-omics approaches to disease' - *Genome Biol.* 2017; 18:83. Published online doi: 10.1186/s13059-017-1215-1, 2017
6. <https://www.techemergence.com/machine-learning-in-pharma-medicine/>
7. <http://clinicalml.org/research.html>
8. https://en.wikipedia.org/wiki/Logic_learning_machine
9. https://en.wikipedia.org/wiki/Right_to_explanation
10. <https://www.youtube.com/watch?v=O8OfhvUJ66A>
11. https://www.corriere.it/salute/18_maggio_04/anche-intelligenza-artificiale-puo-sbagliare-ma-colpa-uomo-a5e3dof2-4f79-11e8-8934-320f-886fd76.shtml
12. <https://blog.appliedai.com/healthcare-ai/>
13. Boyd D, Crawford K. Critical questions for Big Data. *Information, Communication & Society.* 15:662-79, 2012.
14. Hansen MM, Miron-Shatz T, Lau AY, Paton C. Big Data in Science and Healthcare: A Review of Recent Literature and Perspectives. *Contribution of the IMIA Social Media Working Group. Yearb Med Inform* 15:21-6, 2014.
15. Scott IA. Machine Learning and Evidence-Based Medicine. *Ann Intern Med.* 169:44-46, 2018.
16. Lo-Ciganic WH, Donohue JM, Thorpe JM, Perera S, Thorpe CT, Marcum ZA, Gellad WF. Using machine learning to examine medication adherence thresholds and risk of hospitalization. *Med Care.* 53:720-8, 2015.
17. Alanazi HO, Abdullah AH, Qureshi KN, Ismail AS. Accurate and dynamic predictive model for better prediction in medicine and healthcare. *Ir J Med Sci.* 187:501-513, 2018.
18. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. *N Engl J Med.* 378:981-983, 2018.
19. Hood L. Systems biology and p4 medicine: past, present, and future. *Rambam Maimonides Med J* 4: e0012, 2013.
20. Armstrong S. Data, data everywhere: the challenges of personalised medicine. *BMJ* 359: j4546, 2017.
21. Beam AL, Kohane IS. Translating Artificial Intelligence Into Clinical Care. *JAMA* 316:2368-2369, 2016.
22. Sim I. Two Ways of Knowing: Big Data and Evidence-Based Medicine. *Ann Intern Med* 164:562-3, 2016.
23. Asghari S, Mahdavian M. Secondary analysis of electronic databases: potentials and limitations. *Diabetologia* 56:2096-7, 2013.
24. Svensson CM, Hübler R, Figge MT. Automated Classification of Circulating Tumor Cells and the Impact of Interobserver Variability on Classifier Training and Performance. *J Immunol Res.* 2015: 573165, 2015.
25. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. *JAMA* 318:517-518, 2017.
26. Caruana R, Lou Y, Gehrke J et al. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Cham, Switzerland: Springer International Publishing AG 1721-1730, 2015.
27. Shah ND, Steyerberg EW, Kent DM. Big Data and Predictive Analytics: Recalibrating Expectations. *JAMA* 320:27-28, 2018.
28. Cohen IG, Mello MM. HIPAA and Protecting Health Information in the 21st Century. *JAMA* 320:231-232, 2018.
29. Cross M. Show us the data: why clinical outcomes matter. *BMJ.* 344:e66, 2012.
30. Hoff T. Deskillling and adaptation among primary care physicians using two work innovations. *Health Care Manage Rev.* 36:338-48, 2011.
31. Lenzer J. Big data's big bias: bringing noise and conflicts to US drug regulation. *BMJ* 358:j3275, 2017.
32. Longhurst CA, Harrington RA, Shah NH. A 'green button' for using aggregate patient data at the point of care. *Health Aff (Millwood)* 33:1229-35, 2014.
33. Centre for Evidence-Based Medicine. Study Designs. www.cebm.net/study-designs
34. Moskowitz A, McSparron J, Stone DJ, Celi LA. Preparing a New Generation of Clinicians for the Era of Big Data. *Harv Med Stud Rev* 2:24-27, 2015.
35. Obermeyer Z, Lee TH. Lost in Thought - The Limits of the Human Mind and the Future of Medicine. *N Engl J Med* 377:1209-1211, 2017.
36. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA* 319:1317-1318, 2018.
37. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. *N Engl J Med* 376:2507-2509, 2017.
38. Verghese A, Shah NH, Harrington RA. What This Computer Needs Is a Physician: Humanism and Artificial Intelligence. *JAMA* 319:19-20, 2018.
39. http://www.diabete-rivistamedia.it/wp-content/uploads/2017/05/02_SeZ_FORMATIVA_Agg_Zilich-1.pdf