

**24. febbraio**

## **Intelligenza artificiale, accuratezza diagnostica ed errore medico**

*La verità sta da qualche parte tra i nostri errori.*

Esistono alcuni modi in cui l'intelligenza artificiale (AI) sta emergendo per fare la differenza in termini di accuratezza diagnostica.

Nell'era del deep learning supervisionato con reti neurali convoluzionali addestrate a interpretare immagini mediche, ci sono stati numerosi studi che mostrano che la precisione può essere migliorata con il supporto dell'intelligenza artificiale oltre ai medici esperti che lavorano da soli.



Il team del *Division of Diagnostic Radiology, Department of Translational Medicine, Lund University, Malmö*, diretto da **Kristina Lång** in un ampio studio randomizzato sulla mammografia condotto su oltre 80.000 donne sottoposte a screening per il cancro al seno, con o senza il supporto dell'intelligenza artificiale ai radiologi,

*Lång K et al*

**Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study.**

*Lancet Oncol. 2023 Aug;24(8):936-944.*

ha mostrato un miglioramento dell'accuratezza con una notevole riduzione del **44%** del carico di lavoro di lettura dello schermo.



Il *Department of Colorectal Surgery, Harbin Medical University Cancer Hospital, Heilongjiang*, coordinato da **Shengha Lou** nel report :

*Lou S et al.*

**Artificial intelligence for colorectal neoplasia detection during colonoscopy: a systematic review and meta-analysis of randomized clinical trials.**

*EClinicalMedicine. 2023 Nov 30;66:102341..*

Riporta una analisi sistematica di 33 studi randomizzati di colonscopia, con o senza visione artificiale con intelligenza artificiale in tempo reale che indica una riduzione di oltre il 50% nei polipi e adenomi mancanti e che il tempo di ispezione aggiunto dall'intelligenza artificiale per ottenere questa maggiore precisione era in media solo di 10 S.

Questi studi hanno utilizzato modelli di rete neurale profonda unimodali, basati su immagini. Ora, con i progressi compiuti con i modelli di trasformatori, che consentono input multimodali, sostiene



**Eric Topol** (una delle 12 "Rock Stars of Science" secondo *Gentlemen's Quarterl e classificato come* come il dirigente medico più influente negli Stati Uniti) che esiste un potenziale maggiore per l'intelligenza **artificiale generativa** per facilitare l'accuratezza della diagnostica medica. Ciò equivale alla capacità di inserire tutti i dati di un individuo, comprese le cartelle cliniche elettroniche con testo non strutturato, file di immagini, risultati di laboratorio e altro ancora.

*Topol EJ. As artificial intelligence goes multimodal, medical applications multiply. Science. 2023 Sep 15;381(6663):adk6139.*

Non molto tempo dopo il rilascio di ChatGPT, sono comparsi aneddoti sul suo potenziale per risolvere diagnosi sfuggenti e mancate.

Ad esempio, la storia Alex (vedi allegato) un ragazzo con dolore grave crescente, mal di testa, andatura anormale e arresto della crescita, esaminato da 17 medici per tre anni senza una diagnosi. La diagnosi corretta di spina bifida occulta è stata infine fatta quando sua madre ha inserito i suoi sintomi in ChatGPT, il che ha portato alla neurochirurgia per sciogliere il midollo spinale e un netto miglioramento.

[\*Da TODAY: A boy saw 17 doctors over 3 years for chronic pain. ChatGPT found the diagnosis\*](#)

Allo stesso modo, una donna ha visitato diversi medici di base e neurologi e le è stata assegnata una diagnosi di Long Covid per la quale non esiste un trattamento validato. Ma il suo parente ha inserito i suoi sintomi e i test di laboratorio nella ChatGPT e ha ottenuto la diagnosi **encefalite limbica** successivamente confermata dal test degli anticorpi, e per la quale esiste un trattamento noto (immunoglobulina endovenosa) che è stato utilizzato con successo.

Tali casi aneddotici non cambieranno il modo in cui viene praticata la medicina e potrebbero essere distorti per ottenere risultati positivi, con meno probabilità che le diagnosi errate da parte di ChatGPT ricevano attenzione.

In alternativa a ChatGPT potrebbe essere utile, che ne dici di utilizzare i Case Records del Massachusetts General Hospital (consultabili in rete) che propongono sfide diagnostiche complesse presentate a medici esperti.

Hanno una storia di 100 anni e sono pubblicati ogni due settimane sul *New England Journal of Medicine* come conferenze clinicopatologiche



Un recente studio randomizzato (pre stampa) [\*\*\*Towards Accurate Differential Diagnosis with Large Language Models\*\*\*](#) ha avuto come obiettivo quello di elaborare una diagnosi differenziale, che includesse la diagnosi corretta, per oltre 300 CPC, confrontando le prestazioni di 20 medici internisti esperti (tempo medio di pratica medica di 9 anni) con quelle di un modello linguistico di grandi dimensioni (LLM). Il LLM era quasi due volte più accurato di quello dei medici per quanto riguarda l'accuratezza della diagnosi, rispettivamente 59,1 contro 33,6%. I medici hanno mostrato miglioramenti quando hanno utilizzato una ricerca, e ancora di più con l'accesso al LLM.

Questo lavoro ha confermato ed esteso il precedente confronto LLM con i medici per l'accuratezza diagnostica, incluso uno studio pre stampato di 69 CPC che utilizzano GPT-4v

***Accuracy of a Vision-Language Model on Challenging Medical Cases***

e la pubblicazione di uno studio che valuta 70 CPC con GPT-4

[\*\*\*Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge\*\*\*](#)

Ma i CPC sono casi diagnostici estremamente difficili e generalmente non rappresentativi della pratica medica. Potrebbero, tuttavia, essere un utile indicatore per la corretta diagnosi di condizioni rare, come è stato osservato con le malattie rare (pre stampa)

[\*\*\*ChatGPT as a medical doctor? A diagnostic accuracy study on common and rare diseases\*\*\*](#)

In particolare, il pregiudizio dei medici nei confronti dell'intelligenza artificiale può andare in entrambe le direzioni. Un recente studio randomizzato su 180 radiologi con o senza supporto della rete neurale convoluzionale, ha valutato l'accuratezza dell'interpretazione delle radiografie del torace. Sebbene l'intelligenza artificiale abbia sovraperformato i radiologi nell'analisi complessiva, è stata riscontrata una marcata eterogeneità, con alcuni radiologi che mostravano "negligenza nell'automazione", altamente fiduciosi nella propria lettura e che ignoravano le interpretazioni dell'intelligenza artificiale.

**Nel complesso, le prove fino ad oggi suggeriscono che esiste un reale potenziale per l'intelligenza artificiale generativa per migliorare l'accuratezza delle diagnosi mediche, ma è necessario affrontare le preoccupazioni sulla propagazione dei bias.**

Ben prima che venisse preso in considerazione l'uso aggiuntivo dell'IA, c'erano ampie prove che i pregiudizi dei medici contribuivano a errori diagnostici medici, come la diagnosi errata di attacchi di cuore al pronto soccorso in persone di età inferiore ai 40 anni.

*Behaving Discretely: Heuristic Thinking in the Emergency Department*

Modelli di base come **GPT-4 e Llama** e, più recentemente, **GEMINI**, si allenano con questi pregiudizi sui contenuti umani e pochi, se non nessuno, LLM hanno una messa a punto specializzata per migliorare le diagnosi mediche, non meno il corpus di conoscenze mediche aggiornate. È facile dimenticare che nessun medico è in grado di tenere il passo con tutta la letteratura medica sui circa 10.000 tipi di malattie umane.



**Geoffrey Hinton** uno dei padri dell'intelligenza artificiale ha proposito delle prospettive delle prospettive in campo medico ha detto:

*"Io faccio sempre riferimento alla medicina come esempio di tutto il bene che può fare perché quasi tutto ciò che farà lì è andrà bene. ...Avremo un medico di famiglia che ha visitato cento milioni di pazienti e sarà un medico di famiglia molto migliore".*



**Ilya Sutskever** cofondatore di OpenAI più enfaticamente:

*"Se avete un computer intelligente, un'AGI [intelligenza artificiale generale], costruito per essere un medico, avrà una conoscenza completa ed esauriente di tutta la letteratura medica, avrà miliardi di ore di esperienza clinica".*

Sicuramente non siamo ancora arrivati a questo punto. Ma negli anni a venire, man mano che realizzeremo l'aspirazione e il potenziale di costruire modelli di IA più capaci e dedicati dal punto di vista medico, diventerà sempre più probabile che l'IA svolgerà un ruolo inestimabile nel fornire seconde opinioni con il pensiero automatico del Sistema 2 automatizzato, per aiutare ci muoviamo verso l'obiettivo irraggiungibile ma meritevole di sradicare gli errori diagnostici.

# FILIPPINA



Filippina 82 anni affetta da malattia di Alzheimer, allergie stagionali, fibrillazione atriale e diabete di tipo 2 ha sperimentato un declino subacuto delle sue funzioni cognitive.

Secondo i suoi familiari, è diventata più confusa, spesso dimentica di prendere i farmaci e ha difficoltà a riconoscere le persone familiari.

La paziente e la sua famiglia hanno notato un lieve miglioramento delle sue capacità cognitive quando ha iniziato a prendere donepezil 5 mg al giorno un anno fa. Gli altri suoi farmaci sono atenololo, apixaban, lisinopril e metformina.

Non si è ammalata di recente, ma un aumento delle allergie stagionali ha portato recentemente all'inizio della terapia con difenidramina.

Il suo punteggio al **Mini-Mental Status Examination** di Folstein è sceso da 23/30 a 19/30, quest'ultimo che rappresenta una gamma moderata di demenza, con un ritardo di ricordo e orientamento molto scarso.

Per il resto, i suoi esami generali e neurologici sono normali.

La sua pressione sanguigna è **106/74 mm Hg** e la sua frequenza cardiaca è di **62 battiti al minuto**.

L'analisi delle urine tramite stick è normale.

**Quale dei seguenti cambiamenti terapeutici è più appropriato per questo paziente?**

- 1- Interrompere l'atenololo
- 2- Aggiungi risperidone a basso dosaggio
- 3- Interrompere la difenidramina
- 4- Interrompere il donepezil
- 5- Aumentare la dose di donepezil