



Intelligenza artificiale e auscultazione cardiaca in Medicina Generale: cosa insegna lo studio tricorder

Artificial intelligence and cardiac auscultation in General Practice: what the tricorder study teaches us

Valeria Zurlo

SIMG -componente direttivo nazionale

Conflitto di interessi

L'Autore dichiara nessun conflitto di interessi.

How to cite this article:

Intelligenza artificiale e auscultazione cardiaca in Medicina Generale: cosa insegna lo studio tricorder
Rivista SIMG 2026; 33 (02):46-48.

© Copyright by Società Italiana dei Medici di Medicina Generale e delle Cure Primarie.



OPEN ACCESS

L'articolo è open access e divulgato sulla base della licenza CC-BY-NC-ND (Creative Commons Attribuzione - Non commerciale - Non opere derivate 4.0 Internazionale). L'articolo può essere usato indicando la menzione di paternità adeguata e la licenza; solo a scopi non commerciali; solo in originale. Per ulteriori informazioni: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.it>

Il fonendoscopio rappresenta, da oltre due secoli, uno degli strumenti simbolo della professione medica. Introdotto da René Laennec nel 1816, esso ha consentito per la prima volta di trasformare l'auscultazione cardiopolmonare in un atto clinico sistematico e riproducibile. Tuttavia, nonostante la sua centralità nella pratica clinica, l'auscultazione tradizionale presenta limiti intrinseci legati alla soggettività dell'interpretazione e alla variabilità dell'esperienza clinica dell'operatore.

Negli ultimi anni l'intelligenza artificiale (IA) applicata ai segnali fisiologici ha mostrato risultati promettenti nel miglioramento delle capacità diagnostiche degli strumenti clinici tradizionali e negli studi di validazione algoritmica. Tra questi, il fonendoscopio digitale rappresenta uno dei dispositivi più interessanti, poiché consente di acquisire e analizzare simultaneamente segnali acustici cardiaci e tracciati elettrocardiografici mediante algoritmi di machine learning.

In questo contesto si inserisce lo studio TRICORDER (Triple Cardiovascular Disease Detection with an Artificial Intelligence-Enabled Stethoscope)¹, recentemente pubblicato su *The Lancet*, che rappresenta uno dei primi trial pragmatici cluster-randomizzati che valuta l'impatto clinico reale di un fonendoscopio supportato da IA in medicina generale.

Nonostante le elevate performance algoritmiche documentate negli studi preliminari di validazione²⁻⁵

il trial pragmatico TRICORDER non ha dimostrato un aumento statisticamente significativo delle nuove diagnosi di insufficienza cardiaca nella pratica clinica reale. Questo apparente paradosso rappresenta uno degli aspetti più interessanti e metodologicamente rilevanti dello studio.

Infatti, nella letteratura sull'IA applicata alla medicina, la maggior parte delle evidenze deriva da studi di validazione algoritmica condotti in contesti altamente controllati. In tali studi^{6,7}, gli algoritmi vengono testati su dataset selezionati di casi patologici, spesso derivati da database clinici o da popolazioni ad alta prevalenza di malattia, e le loro prestazioni vengono misurate attraverso indicatori di accuratezza diagnostica quali sensibilità, specificità o area sotto la curva ROC. Sebbene questi parametri siano fondamentali per dimostrare la validità tecnica di un modello predittivo, essi non garantiscono automaticamente che l'algoritmo produca un beneficio clinico quando viene integrato nella pratica quotidiana.

Lo studio TRICORDER evidenzia con chiarezza questa distanza tra performance algoritmica ed efficacia clinica: l'introduzione di uno stetoscopio digitale dotato di algoritmi di IA non si è tradotta automaticamente in un incremento delle diagnosi di scompenso cardiaco nella popolazione generale assistita dai MMG.

Tale risultato suggerisce che il valore clinico di una tecnologia non dipende esclusivamente dalla sua capacità di identificare pattern patologici nei dati fisiologici,

ma anche dalla sua capacità di inserirsi in modo fluido nei processi decisionali e organizzativi della pratica medica quotidiana.

Uno dei fattori determinanti emersi dallo studio riguarda il livello di adozione della tecnologia da parte dei medici; infatti, nel corso del trial una quota rilevante di medici ha progressivamente ridotto o abbandonato l'utilizzo del dispositivo individuando le ragioni principali di questo fenomeno nella scarsa integrazione dello strumento con i sistemi di cartella clinica elettronica, nell'aumento del tempo necessario per

completare la visita e nella difficoltà di incorporare il dispositivo nel flusso operativo della consultazione clinica. In altre parole, il limite principale non è risultato essere l'algoritmo, ma il contesto organizzativo in cui l'algoritmo veniva utilizzato.

Questa osservazione rimanda a un concetto sempre più discusso nella medicina digitale: il cosiddetto "implementation gap", ovvero la distanza tra la dimostrazione di efficacia tecnologica e la reale trasformazione della pratica clinica. Numerosi dispositivi basati su IA hanno dimostrato presta-

zioni elevate in ambito sperimentale, ma pochi studi hanno documentato un miglioramento concreto degli esiti clinici quando tali strumenti vengono implementati su larga scala.

Nel caso dello studio TRICORDER, questo gap emerge in modo particolarmente evidente. L'analisi intention-to-treat, che riflette l'effetto della tecnologia sull'intera popolazione di studi partecipanti, non ha evidenziato benefici significativi.

Tuttavia, l'analisi per protocol, limitata ai medici che hanno utilizzato il dispositivo

DISEGNO DELLO STUDIO

Lo studio TRICORDER è stato progettato come trial pragmatico cluster-randomizzato, condotto in 205 studi di medicina generale nel Regno Unito.

Studi medici randomizzati: 96 nel gruppo intervento, 109 nel gruppo controllo

Partecipazione complessiva: 972 professionisti sanitari, 1.553.175 pazienti registrati

Gruppo intervento: i medici avevano a disposizione uno stetoscopio digitale dotato di algoritmi di IA, capace di registrare simultaneamente fonocardiogramma ed ECG (singola derivazione - durata di 15 sec). Gli algoritmi analizzavano automaticamente i segnali generando una previsione sulla presenza di insufficienza cardiaca FE \leq 40%, fibrillazione atriale, valvulopatia significativa

Gruppo controllo: i medici hanno proseguito la pratica clinica abituale, basata su auscultazione tradizionale e percorsi diagnostici standard, senza l'ausilio del dispositivo digitale con algoritmi di IA.

Eseguiti: 12.725 esami con il dispositivo su pazienti selezionati durante la pratica clinica quotidiana, nel gruppo di controllo non è disponibile un numero direttamente comparabile di esami in quanto la pra-

tica clinica usuale non prevede una procedura diagnostica standardizzata e tracciabile come quella introdotta nel gruppo intervento; pertanto, il confronto tra i gruppi è basato sugli esiti clinici e non sul volume di test eseguiti.

Endpoint e risultati principali: l'endpoint primario dello studio era rappresentato dall'incidenza di nuove diagnosi di insufficienza cardiaca nei due gruppi, endpoint secondari sono stati incidenza di nuove diagnosi di fibrillazione atriale e di valvulopatie.

L'analisi *intention-to-treat* include tutti i medici randomizzati, indipendentemente dal livello di utilizzo del dispositivo, comprendendo quindi anche coloro che lo hanno utilizzato in modo discontinuo, non corretto o nullo. Tale approccio consente di valutare l'efficacia dell'intervento nella pratica reale, rispondendo alla domanda se l'introduzione della tecnologia nel sistema sanitario determini un beneficio clinico complessivo.

Nel trial TRICORDER, questa analisi non ha evidenziato differenze statisticamente significative tra i gruppi: IRR 0,94 (IC 95% 0,86-1,02) per l'insufficienza cardiaca, con assenza di differenze significative anche per fibrillazione atriale e valvulopatie. In altre parole, l'introduzione del fonendoscopio sup-

portato da IA non ha determinato un aumento significativo delle diagnosi di scompenso cardiaco nella popolazione generale seguita in medicina generale. Analogamente, anche per gli endpoint secondari non sono state osservate differenze significative nell'analisi primaria.

Tuttavia, l'analisi per protocol include esclusivamente i medici che hanno utilizzato il dispositivo in modo continuativo e conforme al protocollo, escludendo i soggetti con bassa aderenza o utilizzo occasionale, ha evidenziato risultati opposti: l'utilizzo del dispositivo è risultato associato a un aumento significativo della diagnosi di insufficienza cardiaca (IRR 2,33; IC 95% 1,28-4,26), fibrillazione atriale (IRR 3,45; IC 95% 2,24-5,32) e valvulopatie (IRR 1,92; IC 95% 1,09-3,40), rispetto ai controlli.

Un ulteriore elemento dei più significativi emersi dallo studio riguarda il tasso di utilizzo del dispositivo: circa il 40% dei medici assegnati al gruppo intervento ha smesso di utilizzare lo strumento prima della fine dello studio a causa di mancata integrazione con i sistemi di cartella clinica elettronica, aumento del tempo necessario per completare la visita, difficoltà nell'inserimento del dispositivo nel *workflow* clinico quotidiano. Questo aspetto rappresenta probabilmente la chiave interpretativa principale dei risultati del trial.

con maggiore continuità, ha suggerito un aumento della capacità di identificare patologie cardiovascolari rilevanti, risultato interpretabile con cautela poiché non riflette pienamente la randomizzazione originale dello studio. Ad ogni modo, questo dato indica che la tecnologia possiede un potenziale diagnostico reale, ma che può esprimersi solo quando lo strumento viene utilizzato in modo sistematico e integrato nella pratica clinica.

Per la medicina generale questo risultato assume una rilevanza particolare poiché opera in un contesto caratterizzato da tempo limitato per visita, elevata complessità clinica e gestione simultanea di numerosi problemi di salute e quindi, in questo scenario, l'introduzione di nuove tecnologie diagnostiche non può prescindere da una progettazione attenta dei flussi operativi e dei sistemi informativi: una tecnologia che richiede passaggi aggiuntivi, inserimenti manuali di dati o interfacce non integrate rischia di essere progressivamente abbandonata, indipendentemente dalle sue prestazioni teoriche.

Pertanto, il messaggio principale che emerge dal TRICORDER non riguarda tanto l'efficacia dell'IA nella diagnosi cardiovascolare – che rimane promettente – quanto piuttosto la necessità di ripensare il modo in cui queste tecnologie debbano essere introdotte nella pratica clinica.

L'impatto reale dell'IA nella medicina territoriale dipenderà probabilmente meno dalla sola sofisticazione degli algoritmi e molto di più dalla capacità dei sistemi sanitari di progettare strumenti che si integrino naturalmente nei processi decisionali e organizzativi della pratica clinica.

Gli strumenti diagnostici basati su IA devono essere considerati interventi "socio-tecnici" in quanto il loro impatto clinico è determinato non solo dalla performance algoritmica, ma anche dal livello di adozione, dall'integrazione nei workflow assistenziali e dalle dinamiche organizzative e decisionali che caratterizzano la pratica clinica reale.

In questa prospettiva, lo studio TRICORDER rappresenta non tanto una dimostrazione dei limiti dell'IA, quanto piuttosto una lezione metodologica fondamentale sulla complessità dell'innovazione in sanità: un algoritmo può essere eccellente, ma senza un'adeguata integrazione nel contesto clinico il suo impatto sulla salute dei pazienti rimane inevitabilmente limitato.

Bibliografia

1. Kelshiker M, et al. Triple cardiovascular disease detection with an artificial intelligence-enabled stethoscope (TRICORDER) in the UK: a cluster-randomised controlled implementation trial. *The Lancet* 2026;407:704-15.
2. Attia ZI, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* 2019;25:70-4.
3. Attia ZI, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm. *The Lancet* 2019;394:861-7.
4. Ko WY, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *JACC* 2020;75:722-33.
5. Goto S, et al. Artificial intelligence-enabled phonocardiography for detection of valvular heart disease. *Eur Heart J - Digital Health* 2021.
6. Hannun AY, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019;25:65-9.
7. Rajpurkar P, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med* 2018;15:e1002686.