La Intelligenza Artificiale nella diagnostica dermatologica ci delude

Data 05 giugno 2022
Categoria Medicinadigitale

La validazione degli algoritmi nella pratica clinica, ovvero la verifica nella consueta attività professionale del medico non è ancora una prassi consolidata, ma piuttosto una eccezione riservata a studi controllati.
I risultati di una competizione internazionale tra algoritmi e dermatologi, organizzata dall'International Skin Imaging Collaboration, hanno un particolare interesse per la medicina pratica in quanto ridimensionano le aspettative e confermano l'importanza della sistematica verifica degli algoritmi in setting clinici e non solo sperimentali

Gli algoritmi sono stati valutati su due dataset di immagini, riguardanti 8 malattie cutanee (nevi, melanomi, cheratosi benigne, dermatofibromi, basaliomi, epiteliomi, lesioni vascolari e cheratosi attinica). Una ulteriore categoria riguardava immagini diagnostiche non incluse nel dataset di training.
Sono state considerate, diversamente dai comuni studi sperimentali, anche immagini che meglio riflettono la pratica clinica, cioè caratterizzate dalla presenza di "artefatti" come croste o ulcerazioni sopra la lesione, peli o capelli, segni di penna, assenza di pigmentazione. **Il miglior algoritmo ha ottenuto in media un'accuratezza dell'82% sul dataset "ideale" e del 60% su quello "reale", tratto dalla pratica comune, con una differenza del 22%.**
L'accuratezza scende ancora se si considerano le immagini che raffigurano un tipo di lesione non contenuta nelle immagini di allenamento, in massima parte lesioni infiammatorie benigne e cicatrici.

Tra i primi 25 algoritmi testati, la categoria non presente nelle immagini di addestramento è stata etichettata correttamente solo nell'11% dei casi. In quasi la metà dei casi non correttamente interpretati le immagini sono state assegnate alla categoria delle neoplasie, in gran parte basaliomi (32,4%) ma anche melanomi (7,8%) ed epiteliomi spino-cellulari(6,9%). **Nella realtà clinica tali predizioni di falsa positività porterebbero ad un aumento di biopsie non necessarie con relativo carico di preoccupazioni ed ansietà per I pazienti e le loro famiglie.**
Gli autori dell'articolo scrivono che «questi risultati evidenziano i problemi di sicurezza legati all'impiego di algoritmi automatizzati in ambito clinico e la necessità di progettare metodi migliori per identificare immagini al di fuori dell'area di competenza di un algoritmo, al fine di evitare biopsie non necessarie od errate interpretazioni diagnostiche particolarmente drammatiche per malattie quali i melanomi , che si sarebbero verificate se fossero stati impiegati gli algoritmi testati in questo lavoro».

Concludono affermando di aver identificato «carenze specifiche e problemi di sicurezza nei sistemi diagnostici di intelligenza artificiale per il cancro della pelle che dovrebbero essere affrontate nei futuri protocolli di valutazione diagnostica per migliorare la sicurezza e l'affidabilità nella pratica clinica».

Riflessioniconclusive

Una delle limitazioni che pregiudicano in parte, a oggi, l'impiego di sistemi di AI in medicina e, in particolare, lo sviluppo di strumenti diagnostici sono i bias di selezione nella scelta dei campioni su cui tali sistemi vengono istruiti e che rendono l'algoritmo non applicabile alla pratica quotidiana. Un grande volume di dati non corrisponde automaticamente ad una migliore qualità delle inferenze e delle applicazioni che da queste derivano. Il valore dei dati non è nella loro ampiezza ma nella validità del percorso che ha portato alla loro misurazione, base della stima della validità interna ed esterna di qualsiasi ricerca, al fine di non giungere ad inferenze causali sbagliate che potrebbero determinare una scarsa generalizzabilità dei risultati e quindi potenziali gravi errori diagnostici oltre che sottrazione di risorse nei confronti di interventi di dimostrata efficacia.

Giampaolo Collecchia e Riccardo De Gobbi

Bibliografia: Combalia M et al. Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: the 2019 International Skin Imaging Collaboration Grand Challenge
Lancet Digit Health 2022; 4: e330–39

Per Approfondire:

Giampaolo Collecchia e Riccardo De Gobbi: Intelligenza Artificiale e Medicina Digitale Il Pensiero Scientifico Ed. Roma 2020

pensiero.it/catalogo/libri/pubblico/intelligenza-artificiale-e-medicina-digitale