Intelligenza artificiale negli ospedali, i cinque rischi per la privacy dei pazienti

Data: 9 maggio 2026 | Autore: Redazione

Intelligenza artificiale negli ospedali, i cinque rischi per la privacy dei pazienti

Una revisione sistematica mette in evidenza le vulnerabilità legate all'impiego dell'IA in ambito sanitario. Anche esami apparentemente anonimi possono rivelare l'identità di una persona, mentre manipolazioni e accessi non autorizzati rischiano di compromettere diagnosi e sicurezza

L'intelligenza artificiale cambia la sanità ma aumenta i rischi

L'**intelligenza artificiale negli ospedali**sta assumendo un ruolo sempre più importante nell'analisi delle immagini mediche, nell'elaborazione dei dati clinici e nel supporto alle decisioni dei professionisti.

Questi sistemi possono velocizzare alcune procedure e aiutare a riconoscere anomalie difficili da individuare. La loro diffusione, tuttavia, solleva interrogativi rilevanti sulla **privacy dei pazienti**, sulla sicurezza informatica e sull'affidabilità dei risultati prodotti.

Una revisione sistematica ha analizzato **7.285 record**provenienti da sei banche dati scientifiche, ai quali si sono aggiunti altri 205 lavori individuati attraverso le citazioni. Al termine della selezione sono stati esaminati **22 studi empirici**, nove dei quali dedicati all'imaging medico. (sciencedirect.com)

Dall’indagine emergono cinque principali categorie di rischio:

- **re-identificazione dei pazienti**;
- attacchi di **membership inference**;
- accessi non autorizzati e attacchi informatici;
- alterazione dei dati sottoposti all’algoritmo;
- uso improprio o interpretazione eccessiva dei risultati dell’IA.

Primo rischio la re-identificazione del paziente

La cancellazione di nome, cognome e codice fiscale potrebbe non essere più sufficiente per rendere realmente anonimo un dato sanitario. Le informazioni cliniche possono infatti contenere caratteristiche biometriche invisibili a un osservatore, ma riconoscibili da un algoritmo.

Un sistema di **deep learning**può individuare elementi distintivi all’interno di un elettrocardiogramma, di una radiografia toracica, di una Tac o di una scansione della retina. Questi segnali possono permettere di collegare più esami alla stessa persona e, in determinate condizioni, di risalire alla sua identità.

La re-identificazione è la minaccia maggiormente studiata: riguarda **12 dei 22 lavori**selezionati. Le ricerche coinvolgono diversi settori, tra cui Cardiologia, Radiologia, Oftalmologia e Patologia digitale.

Tra i materiali analizzati figurano:

- elettrocardiogrammi;
- risonanze magnetiche;
- immagini Tac;
- radiografie del torace;
- fotografie del fondo oculare;
- tomografie a coerenza ottica;
- immagini istopatologiche digitali.

Alcuni sistemi sono riusciti a riconoscere le persone attraverso il loro **tracciato ECG**. Le risonanze magnetiche della testa, inoltre, possono conservare informazioni sufficienti per ricostruire caratteristiche del volto e confrontarle con fotografie disponibili pubblicamente.

Le tecniche di mascheramento facciale possono ridurre il rischio, ma la protezione non è sempre assoluta.

Le immagini mediche possono diventare identificatori biometrici

La capacità di riconoscere un paziente non riguarda soltanto gli esami del volto. Alcuni algoritmi hanno identificato persone partendo da immagini tridimensionali ottenute tramite Tac, anche quando gli esami erano stati realizzati in strutture differenti o con apparecchiature prodotte da aziende diverse.

Nella **Patologia digitale**, uno studio ha raggiunto un’accuratezza fino all’80 per cento nell’individuazione del paziente dal quale provenivano le immagini dei tessuti.

Risultati significativi sono stati ottenuti anche con le immagini del fondo oculare, le scansioni OCT e le radiografie toraciche. I tassi di re-identificazione rilevati con alcuni modelli di base sono risultati compresi tra il 26 e il 46 per cento. I sistemi progettati per compiti specifici hanno raggiunto valori ancora più elevati, arrivando fino al **94 per cento nelle immagini OCT**.

In un’altra sperimentazione è stata ottenuta una re-identificazione quasi completa utilizzando radiografie del torace considerate anonime. Il riconoscimento è stato possibile anche confrontando esami effettuati a distanza di oltre dieci anni.

Questi risultati dimostrano che alcune immagini cliniche possono funzionare come vere e proprie **impronte biometriche sanitarie**.

Secondo rischio scoprire se un paziente è presente nel database

Il secondo pericolo è rappresentato dal **membership inference attack**. In questo caso, l’aggressore non cerca necessariamente di ricostruire il nome della persona, ma tenta di scoprire se i suoi dati siano stati utilizzati per addestrare un sistema di intelligenza artificiale.

La semplice presenza in un archivio sanitario può già rivelare un’informazione riservata. Potrebbe indicare, per esempio, che il paziente:

- ha partecipato a una sperimentazione clinica;
- è stato sottoposto a un determinato esame;
- presenta una particolare condizione genetica;
- è stato curato per una specifica patologia.

Tre studi hanno analizzato questo rischio nei settori della **Genomica**, della Neurologia e delle banche dati sanitarie.

In una ricerca condotta su informazioni genomiche, una tecnica di protezione ha ridotto l’efficacia dell’attacco da un valore AUC di **0,71 a 0,50**, portandola quindi vicino al livello di una previsione casuale.

L’introduzione della cosiddetta **privacy differenziale**, basata sull’aggiunta controllata di rumore ai dati, ha invece ridotto le prestazioni del modello senza riuscire sempre a neutralizzare completamente il rischio.

Anche i dati sanitari sintetici possono essere vulnerabili

La creazione di dati artificiali viene spesso presentata come una possibile soluzione per proteggere i pazienti. Il principio consiste nel produrre informazioni statisticamente simili a quelle reali, ma non direttamente collegate a persone esistenti.

La sicurezza, tuttavia, dipende dal modo in cui questi dati vengono generati. Nei dataset parzialmente sintetici, alcune caratteristiche originali possono rimanere riconoscibili.

In una delle ricerche analizzate, gli aggressori hanno ottenuto una precisione superiore a **0,9 per una quota fino al 44 per cento dei pazienti**presenti nel database esaminato. I dati completamente sintetici hanno invece mostrato una maggiore resistenza agli attacchi.

Anche i sistemi basati sull’elettroencefalogramma e sulle **interfacce cervello computer**sono risultati vulnerabili. Ciò conferma che qualsiasi informazione biologica complessa può conservare elementi utili a distinguere una persona dalle altre.

Terzo rischio accessi abusivi e attacchi ai sistemi ospedalieri

Gli strumenti di IA sanitaria hanno bisogno di grandi quantità di dati e spesso comunicano con cartelle cliniche elettroniche, archivi diagnostici e infrastrutture connesse alla rete. Questa

integrazione amplia il numero dei possibili punti di accesso per un aggressore.

Un attacco informatico potrebbe consentire di:

- consultare dati sanitari riservati;
- sottrarre immagini ed esami diagnostici;
- modificare le informazioni cliniche;
- interrompere il funzionamento dei servizi;
- compromettere il modello utilizzato dall'ospedale.

La protezione non può quindi limitarsi alla piattaforma di intelligenza artificiale. Deve coinvolgere l'intero ambiente digitale, compresi dispositivi medici, server, credenziali di accesso, sistemi cloud e collegamenti tra reparti.

Quarto rischio la manipolazione degli esami

Un'altra minaccia riguarda gli **attacchi avversariali**, attraverso i quali vengono introdotte alterazioni quasi impercettibili nei dati analizzati dall'algoritmo.

Una radiografia o una scansione medica può apparire normale all'occhio umano, ma contenere modifiche capaci di ingannare il sistema. L'IA potrebbe così non riconoscere una lesione realmente presente oppure segnalare una patologia inesistente.

La manipolazione degli input può interessare:

- immagini radiologiche;
- risultati di laboratorio;
- segnali cardiaci;
- parametri provenienti da dispositivi medici;
- informazioni presenti nella cartella clinica.

In campo sanitario un errore di questo tipo non produce soltanto un problema tecnico. Può influenzare la diagnosi, ritardare una cura o indirizzare il professionista verso una decisione non appropriata.

Quinto rischio affidarsi troppo all'intelligenza artificiale

Il pericolo non deriva esclusivamente dagli attacchi esterni. Anche un uso scorretto del sistema da parte degli operatori può creare conseguenze importanti.

Un risultato espresso con un'elevata percentuale di probabilità potrebbe essere interpretato come una certezza. Il medico rischierebbe così di attribuire all'algoritmo un'autorità superiore a quella realmente posseduta.

I modelli possono commettere errori quando incontrano pazienti o condizioni differenti rispetto ai dati utilizzati durante l'addestramento. Possono inoltre risentire di dataset incompleti, poco rappresentativi o caratterizzati da squilibri legati all'età, al sesso o all'origine geografica.

L'**IA applicata alla medicina**deve quindi rimanere uno strumento di supporto. La valutazione clinica finale, l'interpretazione del contesto e la responsabilità della decisione devono restare affidate ai professionisti sanitari.

Perché la sola anonimizzazione non è più sufficiente

Le evidenze disponibili suggeriscono la necessità di superare un modello di protezione basato esclusivamente sulla rimozione dei dati anagrafici.

Occorre adottare una strategia multilivello che comprenda:

- controllo rigoroso degli accessi;
- cifratura delle informazioni;
- verifica delle persone autorizzate;
- tracciamento delle operazioni;
- valutazione periodica delle vulnerabilità;
- tecniche avanzate di anonimizzazione;
- test contro gli attacchi avversariali;
- supervisione umana delle decisioni;
- formazione continua del personale.

Deve essere valutato anche il rischio che un dato apparentemente innocuo venga collegato ad altre informazioni disponibili e consenta di ricostruire l'identità del paziente.

La sicurezza deve continuare dopo l'ingresso in ospedale

I modelli di nuova generazione possono cambiare nel tempo attraverso aggiornamenti, nuove informazioni o modifiche del software. Per questo motivo, una valutazione effettuata prima dell'adozione potrebbe non essere sufficiente a garantire la sicurezza durante l'intero ciclo di utilizzo.

Un sistema inizialmente affidabile potrebbe comportarsi diversamente dopo un aggiornamento oppure mostrare vulnerabilità non emerse nei test iniziali. Servono quindi **controlli continui**, procedure per segnalare gli incidenti e verifiche successive all'introduzione nella pratica clinica.

La sfida consiste nel permettere agli ospedali di beneficiare dell'innovazione senza trasformare i pazienti in soggetti inconsapevolmente esposti a rischi informatici.

Innovazione e tutela del paziente devono procedere insieme

L'intelligenza artificiale può offrire un contributo importante alla diagnosi e all'organizzazione dell'assistenza sanitaria. Le sue potenzialità non devono però far sottovalutare la natura estremamente delicata dei dati utilizzati.

La **sicurezza dell'IA in ospedale**dipenderà dalla qualità dei sistemi, dalla solidità delle infrastrutture informatiche e dalla capacità delle strutture sanitarie di mantenere il controllo umano sulle decisioni.

Proteggere la privacy non significa ostacolare l'innovazione, ma creare le condizioni necessarie affinché le nuove tecnologie possano essere utilizzate con responsabilità, trasparenza e reale beneficio per i pazienti.

Articolo scaricato da www.infooggi.it
https://www.infooggi.it/articolo/intelligenza-artificiale-negli-ospedali-i-cinque-rischi-per-la-privacy-dei-pazienti/155088