Uno studio, frutto della collaborazione tra l’Istituto di fisica applicata “Nello Carrara” del Cnr e l’Università di Pisa, ha evidenziato come il “machine learning”, che utilizza gli algoritmi per l’analisi delle immagini cliniche, può essere utilizzato anche per modificarle, creando i cosiddetti “attacchi avversi”, in grado di ingannare gli stessi sistemi di analisi. La ricerca è stata pubblicata sulla rivista European Journal of Nuclear Medicine and Molecular Imaging del gruppo Springer Nature

Roma,
18 giugno 2020 - Un articolo, pubblicato sulla rivista *European Journal of Nuclear Medicine and Molecular Imaging* del gruppo Springer Nature da Andrea Barucci dell’Istituto di fisica applicata “Nello Carrara” del Consiglio nazionale delle ricerche (Cnr-Ifac) e dal radiologo dell’Università di Pisa Emanuele Neri, analizza attraverso lo strumento del machine learning[1] la possibilità di modificare le immagini radiologiche, pilotando l’esito di una diagnosi. Un rischio che solo lo studio dell’Intelligenza artificiale (AI) può consentire di fronteggiare, sventando errori o azioni compiute in malafede.

“Nel mondo digitale, la sanità 4.0 si muove veloce verso una nuova visione, fondata su dati e integrazione di informazioni. Le analisi basate sull’AI costituiscono uno strumento sempre più diffuso in tutti gli ambiti clinici, suscitando grandi aspettative. Un esempio è la Radiomica, ossia l’estrazione di parametri quantitativi dalle immagini radiologiche, con cui creare modelli diagnostici e predittivi: uno strumento ormai ampiamente utilizzato e, negli ultimi anni, rafforzato dall’introduzione delle reti neurali, dando origine alla deep-radiomics”, dichiara Andrea Barucci del Cnr-Ifac.

“Il machine learning, su cui principalmente si fonda la potenza di queste analisi informatiche delle immagini, può tuttavia essere usato anche in modo negativo, per creare attacchi avversi ai sistemi di analisi delle immagini, cioè modifiche ad hoc delle immagini, impercettibili anche all’occhio umano esperto, studiate per ingannare gli stessi algoritmi e pilotare l’esito di una diagnosi”, prosegue Barucci.

Lo studio dei due ricercatori definisce questo fenomeno in ambito di imaging radiologico come “Adversarial Radiomics”: “Un’analogia con il più ampio campo di ricerca dell’adversarial machine learning, in cui il fenomeno è studiato da anni, per esempio, nella cyber-security e nella guida autonoma”, aggiunge Neri.

“Gli esempi avversi sono un problema relativamente recente nello studio del machine learning, ma la loro applicazione all’imaging clinico è un ambito ancor più nuovo e con risvolti sociali importanti - prosegue Barucci - ad esempio nelle frodi assicurative. D’altronde proprio lo studio di questi esempi avversi è estremamente utile per approfondire la comprensione di algoritmi complessi come le reti neurali e migliorare lo sfruttamento degli strumenti informatici a disposizione”.

“L’intuizione del radiologo è ancora essenziale nel controllo e nell’integrazione delle complesse analisi fornite dagli algoritmi di intelligenza artificiale e il futuro impone una sempre maggiore armonizzazione fra l’analisi informatica e quella umana - conclude Neri - Lo studio vuole mettere in luce come i nuovi

strumenti di analisi digitale basati sull’AI dovranno essere sempre più volti proprio a migliorare quest’interazione (human in the loop)”.

[1] *Gli algoritmi di machine learning usano metodi matematico-computazionali per apprendere informazioni direttamente dai dati, senza la necessità di modelli matematici predefiniti*