UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN
INGEGNERIA BIOMEDICA

Etica dell'intelligenza artificiale in medicina e chirurgia

Relatore: Prof. Giovanni Sparacino

Correlatore: Prof. Simone Del Favero

Laureando: Lorenzo Ferrante

ANNO ACCADEMICO 2022/2023

Data di laurea 19/07/2023

ABSTRACT

L'intelligenza artificiale sta trasformando rapidamente il campo della medicina, offrendo supporto per potenziali miglioramenti nelle diagnosi, nelle terapie e nella gestione dei pazienti. Tuttavia, la sua implementazione introduce anche nuovi rischi per paziente medico e solleva importanti questioni etiche.

Questo elaborato si propone di esaminare le implicazioni etiche dell'applicazione dell'intelligenza artificiale in ambito medico sanitario. La trattazione che segue vuole essere una panoramica ad alto livello dei temi trattati, che ne fornisca una visione generale senza addentrarsi eccessivamente nei dettagli tecnici, trattando comunque adeguatamente e con il dovuto grado di approfondimento gli aspetti fondamentali.

INDICE

Capitolo I

L'intelligenza artificiale

1.1 Concetti generali

L'intelligenza artificiale (abbreviato "IA") rappresenta l'abilità di una macchina di riprodurre parzialmente l'attività intellettuale propria dell'uomo (con particolare riguardo ai processi di apprendimento, di riconoscimento e di scelta) attraverso l'elaborazione di modelli ideali o, in altri casi, utilizzando elaborati elettronici per perseguire tale fine. [2]

Come suggerito dalla definizione, l'impiego dell'intelligenza artificiale è una disciplina molto ampia che lascia spazio a molteplici applicazioni e implementazioni. Se si considera il solo ambito medico sanitario, le tecniche di intelligenza artificiale possono esser utilizzate per l'analisi di grandi quantità di dati medici al fine di migliorare le diagnosi, o anche per l'analisi di immagini mediche, lo sviluppo di trattamenti personalizzati basati sui dati del paziente, il monitoraggio di pazienti cronici, la chirurgia robotica e molte altre applicazioni.

Per lo sviluppo di queste applicazioni in campo medico, trattandosi spesso di problemi molto complessi, vengono utilizzati modelli di Machine Learning (una branca dell'AI) ed in particolare di Deep Learning (una branca del Machine Learning).

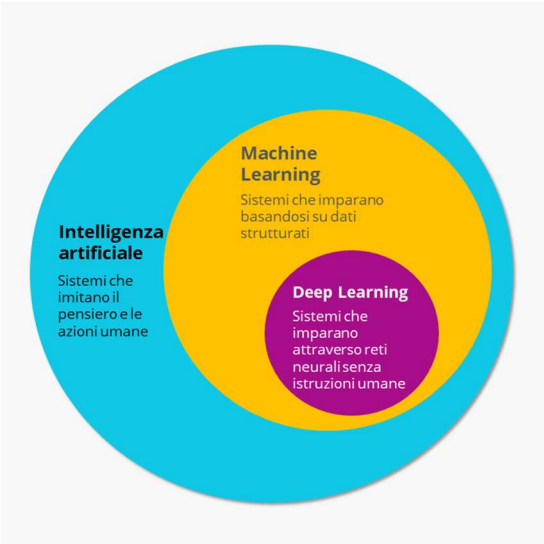

Figura 1: Relazione tra Intelligenza Artificiale, Machine Learning e Deep Learning [11]

1.2 Machine Learning

Il Machine Learning (in italiano “apprendimento automatico”) è una disciplina dell’intelligenza artificiale che consiste nello sviluppo di algoritmi e modelli in grado di apprendere dai dati come svolgere un determinato compito senza esser esplicitamente programmati per farlo e di migliorare le proprie performance nel tempo.

L'obiettivo principale del Machine Learning è quello di consentire ai computer di analizzare grandi quantità di dati, identificare pattern e relazioni nascoste, e utilizzare tali informazioni per prendere decisioni o fornire previsioni in modo autonomo. Ciò può essere ottenuto attraverso l'addestramento (“training”) di algoritmi di ML su una grande mole di dati di input, al fine di estrarre informazioni rilevanti e creare modelli predittivi.

Una volta addestrato, un algoritmo di ML è in grado di generalizzare il problema ed elaborare nuovi dati appartenenti allo stesso dominio applicativo, fornendo previsioni o decisioni basate sui modelli appresi durante la fase di addestramento.

La complessità del Machine Learning ha portato a dover sviluppare differenti tecniche di apprendimento per la fase di addestramento, a seconda delle capacità e dei compiti che si richiedono alla macchina. Ad esempio, si hanno tra le più comuni:

- Apprendimento supervisionato: vengono forniti alla macchina dei dati già etichettati, in maniera tale da guidarla per imparare a riconoscere determinati schemi o gruppi. Questa tecnica viene utilizzata nel caso si vogliano addestrare algoritmi di classificazione o regressione.
- Apprendimento non supervisionato: vengono forniti alla macchina dati non etichettati e spetta all’algoritmo determinare i pattern e le strutture nascoste presenti tra i dati in input. Questa tecnica viene utilizzata nel caso si vogliano addestrare algoritmi per problemi di clustering.

Bisogna comunque considerare il Machine Learning una disciplina molto ampia nella quale rientrano diversi tipi di modelli di calcolo, ognuno con uno specifico approccio per apprendere dai dati e generare previsioni o decisioni. Tra i modelli più noti ci sono il Random Forest, le Support Vector Machine (SVM), il K-Nearest-Neighbors e le reti neurali, le quali in particolare saranno approfondite nei paragrafi successivi.

1.3 Reti Neurali

Una rete neurale artificiale è un modello computazionale che si ispira alla struttura e agli aspetti funzionali dei neuroni biologici nel cervello umano.

Per riprodurre artificialmente tale modello biologico, le reti neurali artificiali hanno un'architettura stratificata, composta da diversi strati di nodi collegati tra loro per elaborare i dati. I nodi, o neuroni artificiali, sono le unità di elaborazione delle informazioni e sono interconnessi tra i diversi strati da linee di connessione pesate, o sinapsi, che ne mediano l'interazione.

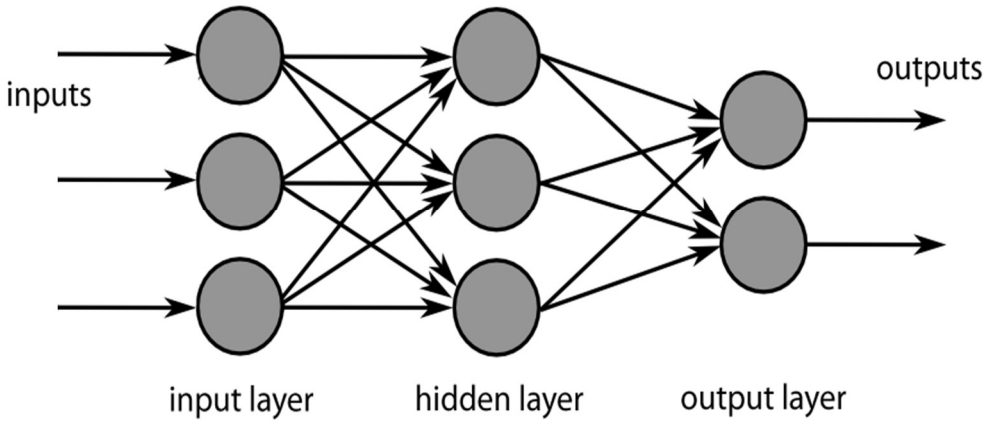

Figura 2: Esempio di rete neurale [2]

I diversi strati della rete neurale si dividono in:

- Input layer: è lo strato che riceve i dati grezzi, deve esser composto un numero di neuroni pari alla dimensionalità dei dati in input;
- Hidden layer: è lo strato in cui i dati vengono elaborati;
- Output layer: è lo strato che restituisce l'output. Deve esser composto da un numero di neuroni pari alla dimensionalità dell'output richiesto.

Il singolo neurone artificiale non è altro che una funzione matematica cui, dato in input un insieme di variabili indipendenti $x = (x_1,\ x_2,\ \dots, x_m)$ ed i relativi pesi $w = (w_1,\ w_2,\ \dots, w_m)$, restituisce come output $y = \varphi(v + b)$ dove:

- $v = \sum_{i=1}^{m} x_i w_i$ è la sommatoria pesata dei dati in input al neurone;

- b (“bias” o “treshold”) è un parametro esterno, solitamente con valore negativo, che viene applicato per gestire la sensibilità della risposta del neurone ai dati in input;
- φ è la funzione di attivazione che determina l’output o l’attivazione del neurone.

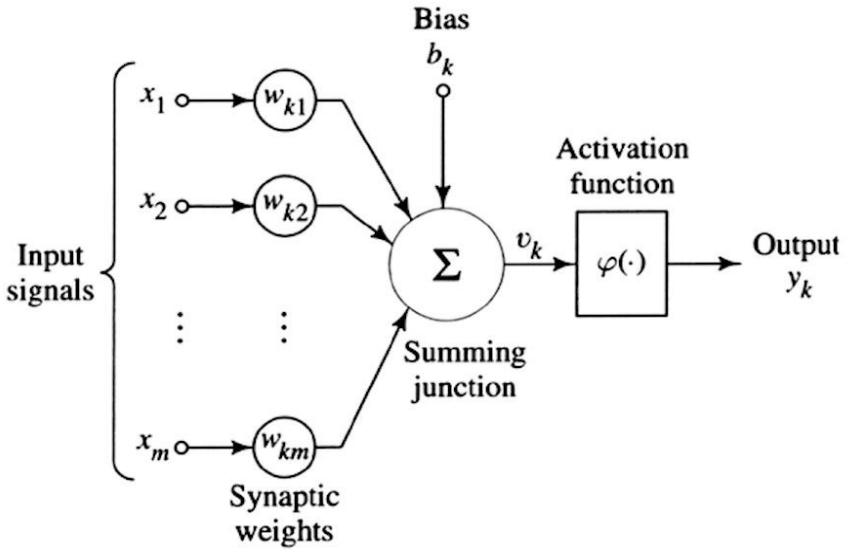

Figura 3: Modello matematico di neurone artificiale [12]

La funzione di attivazione viene scelta in base al tipo di dati che si devono elaborare e al tipo di output che si desidera. In generale, vengono usate funzioni non lineari poiché permettono alla rete di modellare relazioni complesse presenti tra i dati in input, e di affrontare quindi problemi più complessi.

Durante il processo di addestramento di una rete neurale, l'algoritmo di apprendimento ha l’obiettivo di regolare i pesi delle connessioni neurali in modo che l'output prodotto dalla rete sia il più vicino possibile all'output desiderato. Ciò viene fatto attraverso un processo di feedback chiamato “error backpropagation”, in cui l'errore tra l'output atteso e quello effettivo viene propagato all'indietro attraverso la rete per aggiornare i pesi.

Le reti neurali, rispetto ad altre tecniche di Machine Learning, sono in grado di apprendere ed elaborare strutture complesse di dati senza che questi vengano preelaborati. Questo le rende adatte per affrontare problemi in cui le features rilevanti dei dati possono essere difficili da definire in modo esplicito.

1.4 Reti Neurali Profonde e Deep Learning

Il deep learning (in italiano "apprendimento profondo") è una sottobranca del machine learning che si basa sull'utilizzo di reti neurali formate da più hidden layer, dette reti neurali profonde (DNNs, "Deep Neural Networks").

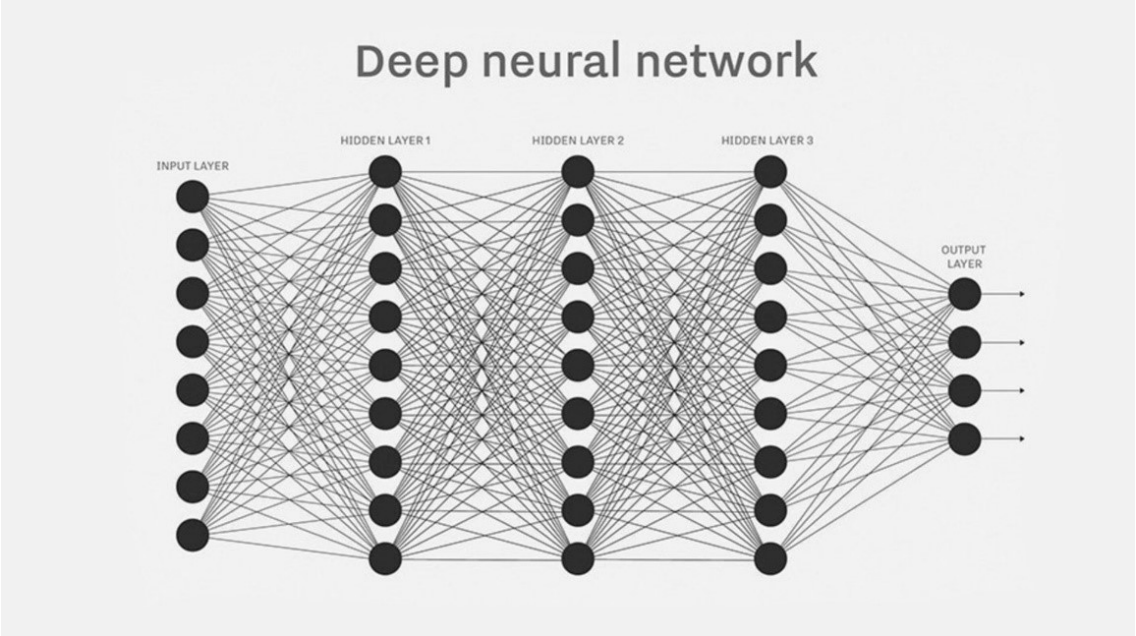

Figura 4: Esempio di rete neurale profonda [12]

Il numero di hidden layer definisce le prestazioni e la complessità dei problemi che la rete può svolgere. In generale, all'aumentare del numero di strati della rete aumentano la capacità della rete di elaborare dati più complessi e il livello di astrazione ma, allo stesso tempo, aumenta notevolmente il costo computazionale necessario alla rete per processare tutti i calcoli.

Ad oggi trovano maggiore applicazione DNNs con 10-50 strati, ma per svolgere problemi più complessi vengono progettate anche reti neurali con oltre 150 strati.

I neuroni tra due strati contigui possono essere connessi in modo diverso a seconda del tipo di rete. Ad esempio, si possono avere:

- neuroni "fully-connected" (come in Figura 4), dove ogni neurone di uno strato è collegato a tutti i neuroni dello strato successivo. In questa configurazione ogni input è in grado di influenzare ogni output;

- neuroni localmente connessi, dove ogni neurone di uno strato è connesso solo ad un sottoinsieme locale dei neuroni dello strato successivo. Le DNNs con i neuroni così configurati (come le Reti Neurali Convoluzionali) sono più efficaci per catturare caratteristiche locali dei dati ed elaborare immagini;
- neuroni con connessioni ricorrenti, dove l'output di un neurone viene rimandato tra gli input dello stesso neurone al passo temporale successivo. Le reti con questa configurazione, dette Reti Neurali Ricorrenti, hanno quindi una sorta di "memoria" delle informazioni passate, il che le rende particolarmente adatte per l'elaborazione di dati sequenziali come testi o serie temporali.

Dunque, le reti neurali differiscono per tipo di interconnessioni e numero di layer; inoltre, ciascuna di esse processa i dati in maniera molto differente dalle altre ed è adatta per svolgere un determinato compito.

1.5 L'importanza dei dati

L'obiettivo dei modelli di Machine Learning e di Deep Learning è di consentire alla macchina di raggiungere un certo livello di accuratezza nella produzione dei risultati, a partire dai dati che le si forniscono nella fase di addestramento. L'accuratezza che si ottiene è incrementale con il numero di dati a disposizione, per cui al fine di raggiungere ottimi livelli di accuratezza è necessario che il modello abbia accesso ad una quantità enorme di dati.

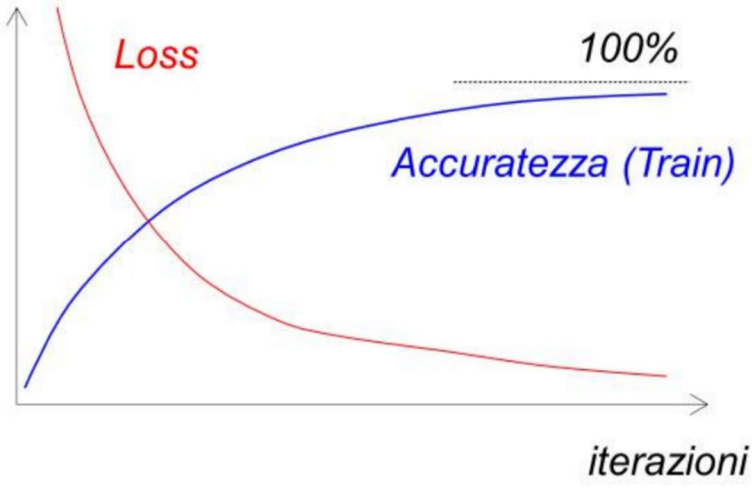

Figura 5: Variazione dell'accuratezza e dell'errore nella fase di addestramento [2]

Un importante aspetto da valutare è innanzitutto la qualità dei dati forniti al modello. Infatti, bisogna tener presente che i dati forniti vengono analizzati in modo acritico dal modello, per cui vale il principio GIGO (letteralmente 'Garbage In, Garbage Out'). Vale a dire che, se i dati in input sono di scarsa qualità o errati, non ci si può aspettare che la macchina li corregga ma, al contrario, li userà apprendendo in modo errato.

Inoltre, è necessario che i dati a disposizione per la fase di addestramento siano rappresentativi della popolazione in esame; cioè, devono essere distribuiti equamente per ciascuna categoria della popolazione.

Se invece i dati utilizzati per l'addestramento non rappresentano equamente tutte le categorie, il modello di AI non riesce a gestire correttamente nuovi dati facenti parte delle categorie sottorappresentate nella fase di training. In questo caso di parla di 'bias' nei dati di training, che si riflette in una disparità nelle decisioni prese dal modello di AI riguardo le varie categorie.

1.6 Black Box vs Explainable AI

Con "Black Box" ci si riferisce alla mancanza di comprensione completa del funzionamento interno degli algoritmi di intelligenza artificiale. Questo termine viene utilizzato per descrivere un modello di intelligenza artificiale di cui l'input e l'output sono noti, ma le operazioni intermedie rimangono oscure o non comprensibili per l'operatore umano.

La complessità delle tecniche di apprendimento automatico, ed in particolar modo delle reti neurali profonde, rende difficile comprendere il ragionamento o la logica utilizzata dagli algoritmi per produrre i risultati.

Se, ad esempio, si pensa ad una comune DNN, questa può esser composta da migliaia neuroni e da decine di migliaia di interconnessioni. Ogni qualvolta si fornisce un input alla rete, non è possibile per un umano ripercorrere tutti i calcoli effettuati dalla rete per arrivare all'output e neanche tener traccia dell'aggiornamento dei pesi.

Inoltre, mentre per le tecniche di ML i dati in input vengono preelaborati in modo da estrarre le features di interesse, ai modelli di Deep Learning vengono solitamente forniti direttamente dati grezzi. Questo aggiunge un ulteriore grado di opacità riguardo al

funzionamento delle DNNs, in quanto non è possibile comprendere a quali features la rete neurale abbia dato maggior rilevanza nell'elaborazione della risposta.

Questa mancanza di trasparenza solleva preoccupazioni, specialmente nel campo medico, dove le decisioni prese potrebbero avere un impatto significativo sulla vita umana. La mancanza di comprensibilità degli algoritmi può rendere difficile valutare la loro affidabilità. Se, ad esempio, gli algoritmi vengono addestrati su dati che contengono pregiudizi o disuguaglianze, questi pregiudizi possono essere amplificati nell'output dell'algoritmo stesso, senza che gli operatori umani ne siano consapevoli.

Per affrontare il problema Black Box è nato un filone di ricerca, l'eXplainable AI (XAI), il cui obiettivo è di trovare soluzioni in grado di aumentare il livello di trasparenza e interpretabilità dei modelli AI.

Ciò consentirebbe di valutare l'affidabilità delle decisioni dell'AI e di riconoscere facilmente potenziali problemi e correggerli. Inoltre, favorirebbe una maggiore fiducia generale nell'uso dell'AI e una crescita nell'adozione.

Capitolo II
Aspetti etici dell'AI in medicina

2.1 L'etica

L'etica può essere intesa sia come lo studio dei valori e delle regole morali, che come l'indagine filosofica delle motivazioni e delle ragioni per le quali gli individui e le società seguono certe norme e principi.

Al centro del concetto di etica c'è la nozione di bene e male, di cosa è giusto e cosa sbagliato. Questi concetti possono variare da una cultura all'altra, da un individuo all'altro, e anche all'interno di un singolo individuo nel corso della vita.

L'etica si preoccupa di capire e di analizzare queste differenze, cercando di tracciare linee guida per la condotta morale e di identificare i principi fondamentali che dovrebbero guidare il comportamento umano.

Esistono diverse teorie e scuole di pensiero etico, che possono essere applicata a una vasta gamma di contesti: dalla vita personale, alle questioni sociali, politiche ed economiche.

Nel seguito verranno introdotti i principi a fondamenta dell'etica in campo medico. In particolare, si analizzeranno gli aspetti etici dell'introduzione e dello sviluppo dell'intelligenza artificiale in medicina.

2.2 L'etica medica ed i suoi principi

L'etica medica è un aspetto rilevante dell'etica applicata. Nella pratica medica, i professionisti devono affrontare una serie di dilemmi etici complessi, che spesso riguardano questioni di vita o di morte.

La storia dell'etica medica ha radici molto antiche. Il primo documento significativo ritrovato a riguardo è il Giuramento di Ippocrate, risalente al IV secolo a.C.. Con questo giuramento gli allievi di Ippocrate si impegnavano davanti ad Apollo a visitare i malati e a prescrivere le cure con l'unico scopo di guarirli e senza mai usare la violenza, a non prescrivere mai farmaci mortali o abortivi, anche se richiesti, a non divulgare mai le cose

apprese nell'esercizio dell'arte medicina, delineando così i principi morali alla base della professione e ponendo in primo piano il benessere ed il rispetto del paziente.

Ovviamente, negli anni, questi principi basilari si sono evoluti per adattarsi meglio alle esigenze della società ed alle nuove sfide etiche poste dall'avanzare della tecnologia, pur rimanendo molto simili al loro significato originale.

Ad oggi, il WHO ("World Health Organization") racchiude i principi dell'etica medica in quattro concetti principali:

- Autonomia: sostiene che i pazienti hanno il diritto di prendere decisioni informate riguardo alla propria salute, ai propri trattamenti ed ai propri dati personali. Questo può includere il diritto di rifiutare un trattamento, anche se il medico ritiene che sia nel miglior interesse del paziente;
- Beneficienza: afferma che il medico deve sempre agire per il bene del paziente;
- Non-Maleficenza: deriva dal criterio "primum non nocere, neminem laedere" di Ippocrate. Sostiene che il medico deve sempre cercare di non arrecare danno al paziente.
- Giustizia: sostiene che deve essere garantita un'equa distribuzione delle risorse mediche, dei benefici e dei rischi, garantendo ad ogni paziente il diritto alla cura che gli è necessaria.

Tali concetti etici sono strettamente collegati tra loro. Infatti, la libertà di scelta e di autonomia del paziente dovrebbero essere maggiori quando viene utilizzato un sistema di intelligenza artificiale, in quanto maggiore è il potenziale di causare danni e minore la possibilità di controllare il processo.

Allo stesso tempo anche i concetti di beneficienza e non-maleficenza sono interdipendenti. Infatti, il principio di non-maleficenza non è un principio assoluto e, ad esempio, è comunque etico ledere in modo ragionevole e adeguato un paziente al fine di curarlo (si pensi ad un'operazione chirurgica o alla somministrazione di trattamenti chemioterapici).

È bene tener presente che sebben questi quattro principi siano universali, la loro implementazione può variare a seconda del contesto culturale, religioso o sociale.

2.3 Principio di Beneficienza ed AI

Fin dall'antichità, le professioni mediche vengono svolte con l'intento di apportare un beneficio ai pazienti. Con l'integrazione di sistemi di intelligenza artificiale, l'obiettivo che la medicina si pone non è cambiato e, finora, sono stati apportati molteplici benefici; di seguito ne verranno discussi alcuni.

Una delle applicazioni più promettenti dell'AI in medicina è l'analisi dei dati. La sanità è un settore che genera un'enorme quantità di dati, tra i quali risultati di laboratorio, immagini mediche, dati genetici, dati dei dispositivi wearable e cartelle cliniche elettroniche. Inoltre, la capacità dell'AI di analizzare anche testi e dati non strutturati, frequenti in campo medico, rende disponibili dati altrimenti di difficile utilizzo. L'analisi e l'interpretazione di queste enormi quantità di dati in modo efficiente e accurato può migliorare notevolmente la precisione della diagnosi e la pianificazione del trattamento.

Un altro importante utilizzo dell'AI in medicina è nell'imaging medico. L'AI è in grado di analizzare immagini radiologiche o patologiche, identificando segni di malattie che possono essere difficili da rilevare per l'occhio umano e possono sfuggire all'occhio del medico. Questo può portare a diagnosi più rapide e accurate, riducendo il tempo di attesa per i pazienti e migliorando i risultati del trattamento.

La chirurgia robotica è un altro campo in cui l'AI sta pian piano inserendosi, seppur ancora in modo molto limitato. I robot chirurgici assistiti da intelligenza artificiale possono eseguire interventi con una precisione che supera quella dei chirurghi umani, riducendo i rischi associati agli interventi. Questi robot possono anche essere guidati a distanza, permettendo ai chirurghi specializzati di eseguire interventi su pazienti in luoghi remoti.

L'intelligenza artificiale può anche esser applicata nel telemonitoraggio e nella gestione delle malattie croniche. Gli algoritmi possono monitorare i dati dei pazienti in tempo reale, permettendo ai medici di identificare i cambiamenti nelle condizioni dei pazienti e di modificare il trattamento di conseguenza. Questo può aiutare a prevenire le complicanze impreviste e a migliorare la qualità della vita dei pazienti.

Inoltre, l'AI può essere utilizzata per la personalizzazione dei trattamenti medici. Gli algoritmi di Machine Learning possono analizzare i dati dei pazienti per creare piani di

trattamento personalizzati basati sulle specifiche esigenze di ciascun individuo. Questo può migliorare l'efficacia del trattamento e ridurre il rischio di effetti collaterali.

Infine, l'AI è fondamentale per la ricerca medica. Gli algoritmi possono analizzare grandi quantità di dati di ricerca per identificare nuove potenziali terapie e farmaci, e per simulare gli effetti delle terapie, riducendo la necessità di test su animali e accelerando il processo di sviluppo di nuovi trattamenti.

2.4 Principio di Non-Maleficenza ed AI

Nonostante l'applicazione dei sistemi di intelligenza artificiale in campo medico offra una vasta gamma di benefici, espone anche a dei rischi. Supponendo ragionevolmente che questi sistemi non siano di certo programmati con l'intenzione di nuocere o ledere in alcun modo il paziente, il loro utilizzo è comunque correlato a dei rischi intrinseci dei sistemi di AI.

Uno dei principali rischi associati all'AI in campo medico è il rispetto della privacy e la sicurezza dei dati. Gli algoritmi di AI dipendono fortemente dall'accesso ad enormi quantità di dati medici, che rappresentano dati sensibili dei pazienti e che, se non protetti adeguatamente, potrebbero essere bersaglio di attacchi informatici con una conseguente violazione della privacy. La divulgazione di dati sensibili potrebbe arrecare gravi disagi ai pazienti come stress emotivo, imbarazzo, paranoia, sfiducia nel sistema sanitario e potrebbero anche essere presi di mira da atti discriminatori.

In secondo luogo, esiste il rischio di errori di diagnosi o di trattamento causati da algoritmi di AI non accurati o affetti da bias. Poiché l'AI si basa sull'apprendimento automatico dai dati, se i dati utilizzati sono incompleti, sbagliati o non rappresentativi della popolazione, gli algoritmi possono produrre risultati inaccurati o distorti.

Ad esempio, se l'addestramento di un algoritmo diagnostico viene eseguito principalmente su dati provenienti da un particolare gruppo demografico, potrebbe esserci un rischio di discriminazione nei confronti di altri gruppi etnici.

Pertanto, è fondamentale garantire che i dati utilizzati per l'addestramento siano rappresentativi e di alta qualità, al fine di cercare di evitare bias e discriminazioni.

In ogni caso, è bene tenere a mente che l'AI non è infallibile e, seppur in alcune applicazioni riesce ad ottenere un'accuratezza molto elevata, può comunque commettere errori, con conseguenze potenzialmente gravi se ciò avviene in un contesto medico.

2.5 Principio di Autonomia ed AI

L'autonomia nell'etica medica è un principio fondamentale che sostiene il diritto di un individuo di fare scelte informate e volontarie riguardo alla propria assistenza sanitaria. Il principio di autonomia è strettamente correlato a concetti come l'indipendenza e l'autodeterminazione.

Il principio di autonomia riconosce il diritto di una persona di avere il controllo sul proprio corpo e sulle decisioni relative alla propria salute ed ai propri dati personali.

Nella pratica medica il concetto di autonomia viene affrontato attraverso il consenso informato. Prima di procedere ad un qualsiasi intervento, trattamento o procedura il medico e il personale sanitario hanno l'obbligo di fornire al paziente tutte le informazioni rilevanti (rischi, complicazioni, probabilità di successo, possibili alternative, etc.) per consentire al paziente di prendere in autonomia una decisione consapevole e informata riguardo la propria salute.

Per quanto riguarda le implicazioni che l'introduzione dell'AI ha sul principio di autonomia, va analizzato questo principio sia per l'utilizzo di sistemi AI a supporto del personale medico, sia per il trattamento dati dei pazienti per l'addestramento dei modelli di ML e Deep Learning.

Capitolo III
Uso dell'AI in ambito medico e problematiche etiche

3.1 Autonomia ed utilizzo dell'AI da parte del personale medico

Nei capitoli precedenti si è parlato di come l'intelligenza artificiale può rappresentare un valido supporto alla diagnosi medica, in quanto già in diversi campi di applicazione ha dimostrato di raggiungere un ottimo livello di accuratezza nelle sue previsioni. Un medico che si avvale di sistemi di intelligenza artificiale come aiuto nella sua diagnosi, necessita di informare il paziente riguardo al loro utilizzo?

Ad oggi, si ritiene impossibile per le macchine (e quindi anche per i sistemi di AI) avere responsabilità legale e dunque, la responsabilità ricadrebbe comunque sul medico indipendentemente dal processo decisionale che lo ha condotto alla diagnosi. Si potrebbe quindi pensare che il medico non debba necessariamente informare il paziente se siano stati utilizzati sistemi di AI.

Tuttavia, non essendo l'utilizzo dell'AI esente da rischi, dovrebbe esser parte delle informazioni che il medico fornisce al paziente, in modo che il paziente possa esprimere il proprio consenso informato anche considerando gli eventuali rischi/benefici aggiuntivi che ne derivano.

Come affermato però dal CNB ("Comitato Nazionale Bioetica") e dal CNBBSV ("Comitato Nazionale per la Biosicurezza, le Biotecnologie la Scienza della Vita) [7], può non essere facile per il paziente comprendere appieno i rischi associati all'utilizzo dell'AI essendo questo un campo molto complesso e in continua evoluzione. Il consenso viene quindi dato più sulla fiducia verso il medico che non in base all'effettiva comprensione.

È necessario, dunque, che i pazienti che si sottopongono a trattamenti sanitari in cui venga utilizzata l'AI siano informati dal medico nelle modalità più consone e comprensibili, a cui quindi spetta un ruolo da 'mediatore' tra il paziente ed i potenziali rischi legati all'AI.

Inoltre, un aspetto importante che viene sottolineato anche dal WHO [8] è che i sistemi di intelligenza artificiale, seppur in alcuni contesti siano già in grado di poter prendere decisioni in modo autonomo, non debbano mai minare l'autorità del medico nel processo

decisionale. Spetta quindi sempre al medico la decisione finale, con gli strumenti di AI che possono solo fornire da supporto.

3.2 Autonomia ed utilizzo dei dati sanitari

I dati sono un elemento fondamentale per le AI. La disponibilità di grandi quantità di dati permette di poter addestrare adeguatamente gli algoritmi, che altrimenti non sarebbero in grado di funzionare.

I dati utilizzati per addestrare le AI in applicazioni mediche, come ad esempio i dati clinici, le immagini mediche, i segnali biologici e i dati genetici (lista non esaustiva), rientrano nella categoria dei dati personali.

Nell'art. 4 del GDPR ("General Data Protection Regulation"), ovvero la normativa comunitaria di riferimento in materia di gestione e protezione dei dati, i dati personali vengono definiti come "*qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale".*

In particolare, i dati relativi alla salute ed i dati genetici rientrano nelle categorie particolari di dati personali e sono quindi sottoposti a regolamentazioni più stringenti.

La raccolta iniziale dei dati avviene nel momento in cui il paziente che deve sottoporsi a un trattamento medico presta il consenso alla raccolta dati per le finalità del trattamento stesso, senza il quale non potrebbe usufruirne. Il riutilizzo dei dati, ad esempio allo scopo di addestrare un modello AI, rappresenta un uso secondario del quale i pazienti non sono tenuti ad esser informati.

Infatti, il titolare del trattamento (dall'art.4 del GDPR definito come "*la persona fisica o giuridica, l'autorità pubblica, il servizio o altro organismo che, singolarmente o insieme ad altri, determina le finalità e i mezzi del trattamento di dati personali […]*") può utilizzare i dati personali anche senza il diretto consenso del paziente.

Per l'art.9 comma 2 del GDPR, il trattamento di categorie particolari di dati personali è possibile se si verifica uno dei seguenti casi (si elencano gli esempi più rilevanti ai fini dell'elaborato):

- *"l'interessato ha prestato il proprio consenso esplicito al trattamento di tali dati personali per una o più finalità specifiche";*
- *"il trattamento è necessario per finalità di medicina preventiva o di medicina del lavoro, valutazione della capacità lavorativa del dipendente, diagnosi, assistenza o terapia sanitaria o sociale ovvero gestione dei sistemi e servizi sanitari o sociali";*
- *"il trattamento è necessario per motivi di interesse pubblico nel settore della sanità pubblica";*
- *"il trattamento è necessario a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici";*

purché, secondo l'art.5, siano però "*trattati in maniera da garantire un'adeguata sicurezza dei dati personali, compresa la protezione, mediante misure tecniche e organizzative adeguate, da trattamenti non autorizzati o illeciti e dalla perdita, dalla distruzione o dal danno accidentali («integrità e riservatezza»)*".

Ne consegue che, il consenso del paziente ("*l'interessato*") al trattamento dei propri dati sanitari non è un requisito fondamentale perché i dati vengano utilizzati, ma è solo uno dei diversi casi per cui il trattamento è consentito. In questo modo viene meno il principio di autonomia del paziente che non ha quindi il pieno controllo sui propri dati personali.

A mitigare questa mancanza, vi è il "diritto all'oblio" (art.17 del GDPR) con il quale il paziente ha il diritto, previa richiesta, di far cancellare tutti i propri dati personali dal titolare del trattamento. Tuttavia, molti pazienti sono inconsapevoli o ignorano che il titolare può riutilizzare i propri dati ed è molto improbabile che richiedano la loro cancellazione. Così facendo, si lascia di fatto al titolare del trattamento la possibilità di utilizzare i dati, esponendo però a dei rischi aggiuntivi i pazienti.

Infatti, le banche dati possono essere bersaglio di attacchi informatici e, per evitare dunque la fuoriuscita di dati sensibili, necessitano di esser protette attraverso opportune tecniche di pseudonimizzazione (attraverso le quali i dati identificativi vengono sostituiti da pseudonimi) o anonimizzazione (attraverso le quali i dati identificativi vengono rimossi irreversibilmente).

È importante tener presente però, che qualsiasi tecnica riduce la possibilità di re-identificazione, ma non la azzera.

Soprattutto quando si hanno grandi quantità di dati a disposizione, come nel caso dell'AI, è possibile la re-identificazione, anche di dati anonimizzati, attraverso l'incrocio con altre banche dati. È fondamentale quindi che vengano sempre adottate misure di sicurezza all'avanguardia, per minimizzare il rischio e garantire al meglio la protezione dei dati personali.

3.3 Principio di Giustizia ed AI

Il principio di giustizia nell'etica medica sostiene che le risorse, i diritti e i servizi sanitari devono essere distribuiti in modo equo e imparziale tra le persone, indipendentemente dal sesso, l'età, l'etnia, il reddito o altre caratteristiche. [8]

Questo principio può essere interpretato in diversi modi, ma generalmente include due aspetti chiave: la giustizia distributiva e la giustizia sociale.

La giustizia distributiva riguarda la distribuzione equa di beni e servizi, mentre la giustizia sociale riguarda la rimozione di ingiustizie sistemiche che creano disparità nel trattamento medico.

Per quanto riguarda la giustizia distributiva, l'intelligenza artificiale può avere impatti diversi.
Da una parte potrebbe aumentare il divario digitale a livello globale, in particolar modo tra paesi ad alto e basso reddito. Infatti, mentre in un mondo che segue ciecamente i principi etici l'accesso e la diffusione delle nuove tecnologie dovrebbero essere garantiti a tutti, nella realtà gli interessi economici di aziende e sviluppatori, la mancanza di figure professionali qualificate e infrastrutture mancanti o non adeguate, ostacolano l'ingresso di queste tecnologie nei paesi a basso reddito.

A livello locale invece, in particolare nei paesi ad alto reddito, l'AI potrebbe migliorare la qualità dell'assistenza sanitaria per i pazienti che vivono in località remote attraverso l'integrazione con applicativi di telemedicina.

Per quanto concerne invece la giustizia sociale, potrebbe accadere che l'intelligenza artificiale in alcuni casi assuma atteggiamenti discriminatori.

Sebbene solitamente si sia portati a pensare che le macchine siano imparziali, bisogna tener conto che gli algoritmi di intelligenza artificiale sono addestrati a partire da dati reali, nei quali si possono nascondere disuguaglianze nel trattamento sanitario tra generi ed etnie diverse.
Infatti, è possibile che gli insiemi di dati attraverso cui vengono addestrate le AI siano affetti da bias, e quindi non siano rappresentativi in egual modo di minoranze o gruppi socialmente emarginati.

Nonostante l'accuratezza generale dei modelli di AI addestrati attraverso dati affetti da bias sia elevata, nell'elaborazione dei dati di pazienti appartenenti alle categorie sottorappresentate l'accuratezza può diventare molto bassa, o addirittura negativa, rendendo di fatto impossibile utilizzare tali modelli per determinati pazienti.

Tuttavia, non è semplice accorgersi immediatamente di queste distorsioni. Infatti, il carattere black-box dei modelli di intelligenza artificiale rende difficile per gli operatori sanitari, ma anche per gli sviluppatori, rendersi conto di quando l'AI sia affetta da bias di alcun tipo.

Per evitare che queste distorsioni possano danneggiare o arrecare danno ai pazienti, potrebbe essere utile verificare, prima della messa in uso, l'accuratezza del modello AI in questione su sottoinsiemi specifici di dati contenenti soli esempi di categorie sottorappresentate, senza limitarsi a valutare la sola accuratezza globale.

Inoltre, per cercare di evitare in principio dei bias nel modello AI, si potrebbe pensare di ribilanciare l'insieme di dati a disposizione per la fase di addestramento attraverso tecniche di *over-sampling*.

Questo tipo di tecniche, tra le quali la più comune è SMOTE ("Synthetic Minority Over-Sampling Technique"), permette di generare nuovi dati sintetici appartenenti alle classi sottorappresentate e di avere quindi un insieme dati più equilibrato tra le varie categorie.

Conclusioni

Come anticipato nel sommario, lo scopo di questo elaborato era di fornire una panoramica ad alto livello delle implicazioni etiche dell'integrazione dei sistemi di intelligenza artificiale in campo medico.

Tra i temi trattati, di primaria importanza e a cui bisogna prestare maggiore attenzione, ci sono la sicurezza dei dati ed il rischio di discriminazioni poiché se non gestisti correttamente possono apportare danni considerevoli al paziente.

A chiusura, si vuole evidenziare come diverse volte nel corso della trattazione è emerso il tema del carattere black-box dell'intelligenza artificiale. Un'intelligenza artificiale etica non può infatti prescindere da una maggiore trasparenza, per garantire un maggior controllo nel processo decisionale ed ottenere la fiducia dei pazienti e del personale medico.

BIBLIOGRAFIA

[1]. F. Rigobello, "*Neural Networks for Medical Decision*", Tesi di Laurea Triennale, Università degli Studi di Padova, 2012, G.M. Toffolo.

[2]. L. Nanni, "*slides del corso di Fondamenti di Intelligenza Artificiale",* Corso di Laurea Triennale, Università degli Studi di Padova*,* 2022.

[3]. Ș. Busnatu et al., "*Clinical Applications of Artificial Intelligence—An Updated Overview*." Journal of Clinical Medicine, vol. 11, 2022. Available: http://dx.doi.org/10.3390/jcm11082265.

[4]. Y. I. Abdullah et al., "*Ethics of Artificial Intelligence in Medicine and Ophthalmology.*", Asia-Pacific journal of ophthalmology (Philadelphia, Pa.), vol. 10, 2021, pp. 289-298. doi:10.1097/APO.0000000000000397

[5]. B. R. Jackson et al., "*The Ethics of Artificial Intelligence in Pathology and Laboratory Medicine: Principles and Practice*.", Academic pathology, vol. 8, 2021, doi:10.1177/2374289521990784

[6]. G. Corbellini, "Dall'etica medica alla bioetica", Treccani. Available: https://www.treccani.it/enciclopedia/dall-etica-medica-alla-bioetica_%28Storia-della-civilt%C3%A0-europea-a-cura-di-Umberto-Eco%29/

[7]. CNB, CNBBSV, "Intelligenza artificiale e medicina: aspetti etici". Available: https://bioetica.governo.it/it/pareri/pareri-gruppo-misto-cnbcnbbsv/intelligenza-artificiale-e-medicina-aspetti-etici/

[8]. World Health Organization, "*Ethics and governance of artificial intelligence for health: WHO guidance*", 2021. Available: https://www.who.int/publications/i/item/9789240029200

[9]. Parlamento Europeo, Consiglio dell'Unione Europea, "*Regolamento (UE) 2016/679 del Parlamento Europeo e del Consiglio*", Gazzetta Ufficiale dell'Unione Europea, 2016. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504

[10]. C. Rauccio, "*Consenso al trattamento dei dati sanitari: ecco perché col GDPR è cambiato tutto*", Agenda Digitale, 2020. Available: https://www.agendadigitale.eu/sicurezza/privacy/consenso-al-trattamento-dei-dati-sanitari-ecco-perche-col-gdpr-e-cambiato-tutto/#:~:text=Con%20riferimento%20al%20consenso%20al,anche%20i%20suoi%20dati%20personali.

[11]. "Deep Learning vs Machine Learning: qual è la differenza?", IONOS, 2020. Available: https://www.ionos.it/digitalguide/online-marketing/marketing-sui-motori-di-ricerca/deep-learning-vs-machine-learning/

[12]. "La nuova rivoluzione digitale, il Deep Learning", DifesaOnline, 2022. Available: https://www.difesaonline.it/evidenza/cyber/la-nuova-rivoluzione-digitale-il-deep-learning