



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



**DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE**

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA MAGISTRALE IN

BIOINGEGNERIA INDUSTRIALE

Applicazioni di IA nella medicina di genere

Relatore: Prof. Rodà Antonio

Correlatrice: Prof.ssa Badaloni Silvana

**Laureanda: Aquino Raffaella
mat. 2026681**

ANNO ACCADEMICO 2022-2023

Data di laurea 03/04/2023

Sommario

L'Intelligenza Artificiale (IA) è sempre più utilizzata in ambito medico per la diagnosi, il trattamento e la prevenzione di patologie, ma presenta alcune problematiche legate alla discriminazione di genere, di etnia, ecc.

La presente tesi di laurea analizza i bias di genere presenti in sistemi basati sull'Intelligenza Artificiale applicati alla medicina.

Viene innanzitutto analizzata la presenza di dati sbilanciati tra i due generi all'interno dei dataset utilizzati per l'addestramento dell'IA; questo può portare a una mancanza di rappresentatività dei casi femminili o di altri gruppi minoritari nella valutazione di diagnosi e terapie.

Si evidenzia poi come i sistemi di IA possano riprodurre stereotipi di genere presenti nella società, come ad esempio la convinzione che alcune patologie siano più comuni in uno specifico genere e ciò possa influenzare la precisione della diagnosi e del trattamento per i pazienti di entrambi i sessi.

Non sono da sottovalutare gli effetti dei bias di genere sull'accesso alle cure mediche: in alcuni casi l'IA può escludere erroneamente alcuni pazienti dall'accesso a determinate cure a causa della mancanza di rappresentatività dei loro casi all'interno dei dataset.

Infine, vengono discusse alcune possibili soluzioni per ridurre i bias di genere nell'IA applicata alla medicina, come ad esempio l'utilizzo di dataset bilanciati per genere, l'analisi critica dei dati e l'implementazione di algoritmi di IA in grado di correggere gli eventuali stereotipi di genere presenti.

In conclusione, la tesi mostra come i bias di genere possano influenzare l'accuratezza dell'IA applicata alla medicina e, di conseguenza, l'accesso alle cure per i pazienti di entrambi i sessi.

La consapevolezza di questo problema è fondamentale per sviluppare soluzioni innovative che possano garantire un accesso equo alle cure mediche a tutti i pazienti, indipendentemente dal loro genere.

Indice

Introduzione	7
1 STEREOTIPI DI GENERE	9
1.1 SESSO E GENERE.....	9
1.2 IL CERVELLO NON HA SESSO	14
1.3 BIAS DESIDERABILI O INDESIDERATI	14
1.4 GENDER GAP E GENDER BIAS.....	17
1.5 BIAS DI GENERE	18
1.6 SESSO E GENERE NELL'APPRENDIMENTO AUTOMATICO	19
2 L'INTELLIGENZA ARTIFICIALE	22
2.1 INTELLIGENZA ARTIFICIALE E MACHINE LEARNING	22
2.2 RETI NEURALI ARTIFICIALI	26
2.3 INTELLIGENZA ARTIFICIALE SPIEGABILE	29
2.4 IDENTIFICAZIONE DEI PREGIUDIZI PER L'EQUITA'	31
3 MEDICINA DI GENERE E INTELLIGENZA ARTIFICIALE	32
3.1 LA MEDICINA DI GENERE.....	32
3.2 LA MEDICINA DI GENERE IN ITALIA	34
3.3 INTELLIGENZA ARTIFICIALE IN CAMPO MEDICO	36
3.4 IA E MEDICINA PREDITTIVA.....	40
3.5 LIMITI IA IN MEDICINA	41
3.6 CONOSCENZA DI GENERE E IA.....	44
4 L'INTELLIGENZA ARTIFICIALE E LO STUDIO DELLE MALATTIE CARDIOVASCOLARI	45
4.1 SOMMARIO DEL CAPITOLO	45
4.2 <i>SCREENING DELLE MALATTIE CARDIOVASCOLARI NELLE DONNE</i>	46
4.3 INTELLIGENZA ARTIFICIALE PER LO SCREENING DI CVD NELLE DONNE.....	49
4.4 MODELLI DI APPRENDIMENTO AUTOMATICO PER LA PREVISIONE DELLE MALATTIE CARDIOVASCOLARI.....	55
4.5 BIAS	58

4.6 BIAS CONTESTUALE NEI SISTEMI DI INTELLIGENZA ARTIFICIALE	59
4.7 LA SOTTOVALUTAZIONE DEL GENERE NEI SISTEMI DI MACHINE LEARNING IN CARDIOLOGIA.....	61
4.8 DIVARIO DI GENERE PERMANENTE NEL RISCHIO DI INFARTO MIOCARDICO	62
4.9 IMPLICAZIONI SULLA SALUTE PUBBLICA.....	64
5 Conclusioni	67
5.1 PROSPETTIVE FUTURE NELLA PRATICA GENERALE	68
Bibliografia	70

Introduzione

L'intelligenza artificiale (IA) è una tecnologia in rapido sviluppo che sta rivoluzionando molti settori, compresa la medicina, in cui il suo utilizzo offre la possibilità di migliorare la qualità dell'assistenza sanitaria, rendendo i processi di diagnosi e trattamento più precisi e personalizzati.

Tuttavia, l'efficacia e l'accuratezza dell'IA dipendono dalla sua formazione e dalle informazioni di cui dispone.

Numerosi studi hanno dimostrato che l'IA, in particolare la componente dell'Apprendimento Automatico – Machine Learning, può catturare bias di genere, cioè una preferenza per certi risultati o categorie basata sul genere del paziente o del medico che le utilizza; questi bias possono portare a errori diagnostici, prescrizioni di farmaci errate o l'omissione di sintomi importanti, con conseguenze potenzialmente gravi per la salute dei pazienti.

I bias di genere dell'IA sono il risultato della mancanza di rappresentatività dei dati utilizzati per addestrare l'algoritmo: in altre parole, se i dati utilizzati per addestrare sistemi di Apprendimento Automatico sono sbilanciati rispetto al genere, l'algoritmo può apprendere degli stereotipi e dei pregiudizi che influenzeranno le sue decisioni future e, ad esempio, un'IA addestrata su dati prevalentemente maschili potrebbe non rilevare alcuni sintomi specifici delle malattie nelle donne.

La realizzazione di sistemi di IA equi ed efficaci in medicina richiede un'adeguata rappresentatività dei dati utilizzati per addestrare l'algoritmo, che deve includere una varietà di pazienti di diversi generi e origini etniche.

Lo scopo principale di questa tesi è di analizzare le cause e le conseguenze di questi bias e di esaminare le possibili soluzioni per ridurli: verranno considerati gli studi più recenti e le ricerche sul tema fornendo una panoramica approfondita della letteratura scientifica sull'argomento.

Nel primo capitolo verranno introdotti i concetti di sesso e genere e le relative differenze, per poi analizzare il concetto di bias di genere e di come esso influenzi l'Apprendimento Automatico.

Nel secondo capitolo viene introdotto il concetto di Intelligenza Artificiale e di come essa si debba applicare per evitare che i pregiudizi possano inficiarne la validità.

Il terzo capitolo presenta l'applicazione dell'Intelligenza Artificiale alla medicina, e quindi lo sviluppo della medicina predittiva; si analizzano anche i limiti che questa comporta e come le conoscenze di genere possano migliorare l'applicazione dell'IA.

Nel quarto capitolo sono descritti i metodi basati sull'Intelligenza Artificiale applicati allo studio delle malattie cardiovascolari, sottolineando la validità di tali metodi senza trascurare i possibili bias presenti e le relative conseguenze. Si sottolinea come ci sia un divario tra uomini e donne e come la considerazione di esso possa avere degli impatti sulla salute pubblica e sulla gestione di diagnosi e trattamenti.

Il lavoro di tesi si conclude con alcune prospettive future dell'applicazione di metodi basati sull'Apprendimento Automatico nella medicina nella pratica generale.

Verrà esaminato il ruolo dell'IA in medicina, con particolare attenzione all'uso di algoritmi di Machine Learning per la diagnosi e il trattamento di malattie; si esamineranno i bias di genere dell'IA e le loro conseguenze in campo medico, con esempi di casi in cui l'IA ha prodotto diagnosi errate o prescrizioni inadeguate a causa di questi bias.

Verranno poi discusse le cause dei bias di genere dell'IA in medicina, con particolare attenzione al ruolo dei dati di addestramento nell'indurre questi bias e, infine, gli approcci di rimozione dei bias di genere a posteriori e di addestramento equo dell'IA, nonché l'importanza dell'etica e della responsabilità sociale nell'uso dell'IA in medicina.

È bene sottolineare l'importanza della raccolta di dati rappresentativi e dell'addestramento equo dell'IA, nonché la necessità di promuovere l'etica e la responsabilità sociale nell'uso dell'IA in campo medico.

La presente tesi di laurea si propone di contribuire alla comprensione degli effetti dei bias di genere dell'IA in medicina e di fornire un contributo fondamentale per lo sviluppo di sistemi di intelligenza artificiale più equi, sicuri ed efficaci.

1 STEREOTIPI DI GENERE

1.1 SESSO E GENERE

Il termine sesso viene associato al genere assegnato alla nascita da cromosomi, ormoni e genitali; coincide quindi col sesso anatomico e può essere maschile, femminile o intersessuale.

Il genere, l'identità di genere, è l'esperienza soggettiva di un individuo e dei legami con fattori sociali e culturali. Attualmente, il genere comprende cisgender, transgender, genere neutro, non binario, agender, pangender e genderqueer.

L'orientamento sessuale è considerato come categoria speciale di dati nel regolamento sulla protezione dei dati dell'UE (GDPR). La figura [1] riporta una schematizzazione della relazione esistente tra sesso e genere.

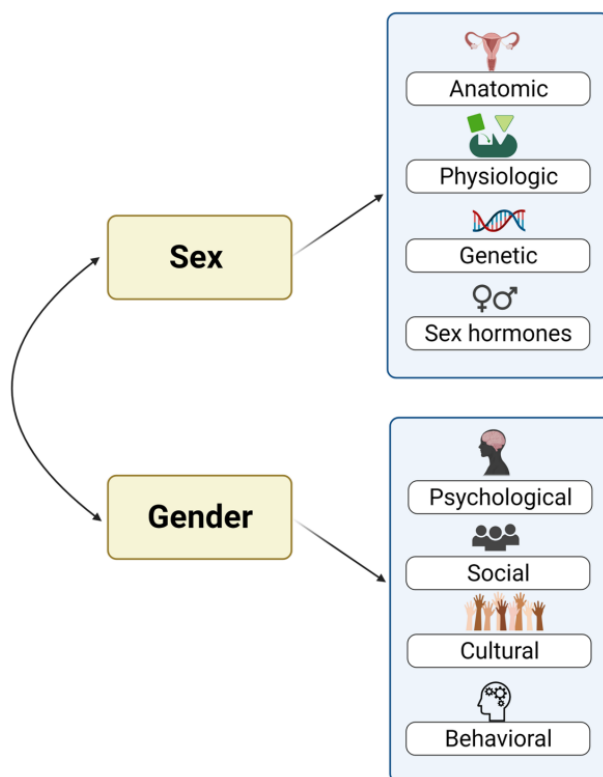


Fig. [1] Relazione tra sesso e genere. (Adedinsewo et al., 2022)

Gli algoritmi, in genere, considerano il sesso come riferimento principale concentrandosi su maschi e femmine ignorando gli individui intersessuali.

La concezione che il genere sia radicato fisiologicamente nel corpo di ciascun individuo sin dalla nascita nuoce la comunità non binaria che, in questo modo, non viene classificata con precisione (Fergus, 2020); ciò “potrebbe esacerbare le disparità sanitarie esistenti, facilitare comportamenti sanitari a rischio e portare a morti prevenibili.” (Barbee et al., 2022).

I sistemi di deep learning sono addestrati, tramite la somministrazione di numerosi set di dati, a riconoscere i modelli, ma non riescono ad integrare le considerazioni di genere risultando scarsi nel riconoscere segnali legati a fattori sociali e culturali (Fosch, 2021).

Il genere non si limita alla classificazione binaria uomo-donna, ma è una costruzione sociale che comprende molte tipologie interiori dell'individuo; è così che si può definire lo stereotipo di genere come “l'attribuire a un individuo caratteristiche o ruoli specifici solo a causa della loro appartenenza al gruppo sociale ‘donne’ o ‘uomini’” (ONU, 2022).

In ambito medico è fondamentale la comprensione delle differenze individuali in termini di salute e di malattie inerenti a fattori genetici e ambientali, e risulta indispensabile implementare tecnologie basate sull'IA sempre più precise (Cirillo et al., 2021).

Senza ignorare i miglioramenti degli ultimi anni, risultano ancora troppo poco diffuse le tecnologie di Intelligenza Artificiale che considerano adeguatamente il sesso e il genere (Cirillo et al., 2021).

L'exasperazione di disuguaglianze preesistenti può essere favorita dai pregiudizi incorporati in diversi algoritmi; questo avviene quando nei set di dati forniti in fase di sviluppo non si includono le minoranze (Topol, 2019).

Le disparità tra donne e uomini in molte malattie sono influenzate proprio dal sesso e dal genere che influenzano anche le principali cause di morte a livello globale (Mauvais-Jarvis et al., 2020).

L'ambiente, compresi lo stress e il comportamento, ha un impatto drammatico sulla biologia del nostro corpo; stabilire una netta distinzione tra gli effetti del sesso e del genere è quasi impossibile, costringendo la medicina ad includere tutte le diverse

dimensioni del genere perché la distribuzione degli attributi legati al genere all'interno delle popolazioni di uomini e donne può influenzare la salute in modo diverso dal sesso biologico (Mauvais-Jarvis et al., 2020).

Uno dei motivi per cui le caratteristiche legate al sesso sono state sottovalutate nella conduzione degli esperimenti è che si tende sempre a ridurre la variabilità per avere dati più facilmente analizzabili; o ci si concentrava solo su individui di sesso maschile, oppure benché si arruolassero soggetti femminili non si distinguevano poi i risultati in fase di analisi (Viola, 2022). La comprensione della differenza fisiologica che sussiste tra uomini e donne è fondamentale per tutelare la salute femminile.

L'influenza del genere sulla salute delle donne si manifesta in quei comportamenti, ruoli e opportunità che la società reputa adatti per uomini e donne.

I determinanti del genere, che hanno un ruolo predominante per tutta la vita di un individuo, agiscono sulla salute di una persona in diversi modi (Viola, 2022):

- pratiche, norme, valori e comportamenti discriminatori;
- maggiore esposizione a malattie o disabilità;
- sistema sanitario discriminatorio;
- ricerca scientifica discriminatoria.

Nell'ambito dell'IA per la medicina è chiara la complessità dell'elaborazione di grandi moli di dati, e gli studi rivelano che l'emarginazione e l'esclusione sociale sono i principali fattori che causano l'evitamento e la disattivazione dell'assistenza sanitaria da parte delle minoranze, che conseguentemente risultano essere meno sani della popolazione generale (Vermeir et al., 2018); la tendenza dei sistemi di Intelligenza Artificiale ad apprendere da modelli distorti, che riproducono gli stereotipi sociali è molto rischiosa nel contesto dell'assistenza sanitaria (Larrazabal et al., 2020).

I dati di addestramento dei software devono essere il più rappresentativi possibili del mondo reale, perché la loro accuratezza determina la successiva qualità del processo decisionale comportando, se tarate in modo scorretto, danni ai pazienti.

Numerosi studi empirici sulle prestazioni degli algoritmi basati sull'Intelligenza Artificiale addestrati su set di dati distorti hanno evidenziato che quando le popolazioni femminili sono sottorappresentate nella fase di formazione, si creano divari di genere nel modo in cui l'algoritmo si comporta, rendendole così una minoranza nel campo dell'assistenza sanitaria (Larrazabal et al., 2020).

La carenza di un campione rappresentativo nel set di dati utilizzato per addestrare l'algoritmo comporta una selettività irragionevole; i pregiudizi di genere ereditati dal sistema di IA possono addirittura essere amplificati perpetuando così gli stereotipi. (Caliskan et al., 2017).

L'IA in medicina è una "lama a doppio taglio" (Cirillo et al., 2020) perché da un lato rischia di tramandare i pregiudizi già esistenti, dall'altro potrebbe assumere un ruolo significativo nell'attenuare tali disuguaglianze.

La questione dei pregiudizi dell'intelligenza in medicina può essere quindi affrontata da due prospettive: si può in un certo senso volere una sorta di discriminazione, ovvero fare in modo che l'algoritmo tenga conto delle differenze di sesso e di genere tra gli individui al fine di fornire una diagnosi quanto più personalizzata possibile; risulta al contempo doveroso, però, evitare discriminazioni indesiderate e potenzialmente dannose (Cirillo et al., 2020).

Un'Intelligenza Artificiale sensibile alle differenze di genere è indispensabile per riuscire a fornire una buona assistenza; ignorare tali differenze genererà errori e risultati discriminatori con potenziali conseguenze dannose nell'assistenza sanitaria (Cirillo et al., 2020).

Le comunità transgender e intersex risultano quelle maggiormente discriminate e danneggiate (Barbee et al., 2022) e le norme decisionali non sembrano trasparenti e selettive; i danni del misgender variano da fisici specifici per un paziente, fino a problemi più vari come il rafforzamento del binarismo di genere con conseguenti pregiudizi algoritmici (Fosch-Villaronga et al., 2021).

Negli ultimi anni c'è stata una sempre più crescente attenzione alla creazione e condivisione di set di dati per la ricerca che, quando applicati in sanità, includono variabili quali sesso, genere, età ed etnia.

Le complesse relazioni tra sesso e genere, che non sempre sono univoche, talvolta risultano ignorate quando i dati vengono estrapolati dalle cartelle cliniche elettroniche; "questo può portare a peggiori risultati di salute per chiunque non si adatti alla norma." (Albert e Delano, 2022).

Nel mondo medico, tutt'oggi, è ancora radicata l'idea che il sesso e il genere siano binari, statici e sempre concordanti; quando si ha a che fare con i concetti di sesso e di genere bisogna considerare tre fattori:

- *sex/gender slippage*: la sostituzione del sesso e dei termini ad esso correlati col genere e viceversa;
- *sex confusion*, si verifica quando una variabile di sesso abbia molteplici significati;
- *sex obsession*, la convinzione che il genere di un individuo coincida col sesso con cui viene identificato alla nascita.

A partire dall'inizio degli anni 2000 le agenzie di finanziamento, ad esempio il National Institutes of Health (NIH), richiedono l'inclusione delle donne e di altri gruppi minoritari in tutte le ricerche da loro finanziate e che gli studi clinici forniscano informazioni sulle differenziazioni in base al sesso e al genere e all'etnia (Ghassemi et al., 2020).

Il sesso e il genere non devono essere confusi con l'orientamento sessuale che indica il "desiderio [o mancanza di desiderio] di avere relazioni sessuali con qualcuno della stessa o diversa identità di genere e/o sesso anatomico." (Lau, et al., 2020).

Alcune stime ipotizzano che le persone intersessuali rappresentino l'1.7-4.3 % della popolazione; tra queste ci sono quelle con un genere diverso dal sistema binario maschio-femmina, e quelle senza genere, le cosiddette agender (Richardson, 2022). È indubbio che il sesso assegnato alla nascita e l'identità di genere si ritrovano nella cartella clinica del paziente, e questi sono ottimali per la maggior parte delle persone cisgender (Ding et al., 2019).

L'attuale approccio di progettazione delle cartelle cliniche elettroniche prevede gli utenti di default escludendo, di fatto, i pazienti transgender e appartenenti ad altre minoranze (Costanza-Chock, 2020); includendo anche le persone che sono tutt'oggi ai margini, si possono sviluppare sistemi in grado di funzionare per tutti i potenziali pazienti, senza alcuna discriminazione.

Ad oggi, il metodo preferito per tentare di ottenere rappresentazioni più complete è l'uso di inventari di organi (Kronk et al., 2021).

1.2 IL CERVELLO NON HA SESSO

Un campo in cui sono disponibili molti dati è quello della ricerca di differenze tra il cervello femminile e quello maschile: quello di una donna è, in media, l'11% più piccolo di quello dell'uomo.

I ricercatori si sono chiesti se, oltre a questa differenza di dimensioni, il sesso biologico modifichi l'attività e le funzioni del nostro cervello; nessuno studio ha mai stabilito che le donne siano meno intelligenti o motivate degli uomini, né che il loro cervello le renda meno propense alla matematica e più inclini alle scienze sociali, per citare uno dei più comuni stereotipi (Viola, 2022).

Nel momento in cui si analizza il cervello umano, però, risulta impossibile distinguere il contributo del sesso da quello del genere.

Il cervello maschile e femminile si somigliano enormemente, e la letteratura scientifica è enorme in questo senso, sia strutturalmente che funzionalmente.

Il cervello è un organo plastico e quindi i circuiti neuronali, ovvero le connessioni tra le varie regioni del cervello, non sono statici ma si creano e modificano col tempo in base agli stimoli ricevuti; sono quindi le esperienze che viviamo, o i traumi che subiamo, a modificare il funzionamento del nostro cervello.

Le differenze che possono eventualmente essere riscontrate tra un cervello maschile e uno femminile non sono altro che il frutto di fattori culturali e socioeconomici.

È innegabile che il sesso e il genere agiscano sulla plasticità del nostro cervello, ma in assenza di patologie i limiti alle capacità mentali degli individui non dipendono dalla biologia ma dalla società che, in alcuni casi, pone dei limiti.

1.3 BIAS DESIDERABILI O INDESIDERATI

Bias è un termine inglese di difficile traduzione e racchiude diversi significati quali preconcetto, pregiudizio, errore sistematico, distorsione, discriminazione (Viola, 2022).

Il *pregiudizio* ha una connotazione negativa a causa dell'associazione a pregiudizi ingiusti, ma in realtà la considerazione e il trattamento differenziati in specifici aspetti biomedici sono necessari per una valida medicina di precisione.

I pregiudizi di sesso e genere posso quindi dividersi in due categorie (Hay et al., 2021):

- bias desiderabili;
- bias indesiderabili.

La figura seguente schematizza le distorsioni desiderabili e indesiderabili nell'Intelligenza Artificiale per la salute.

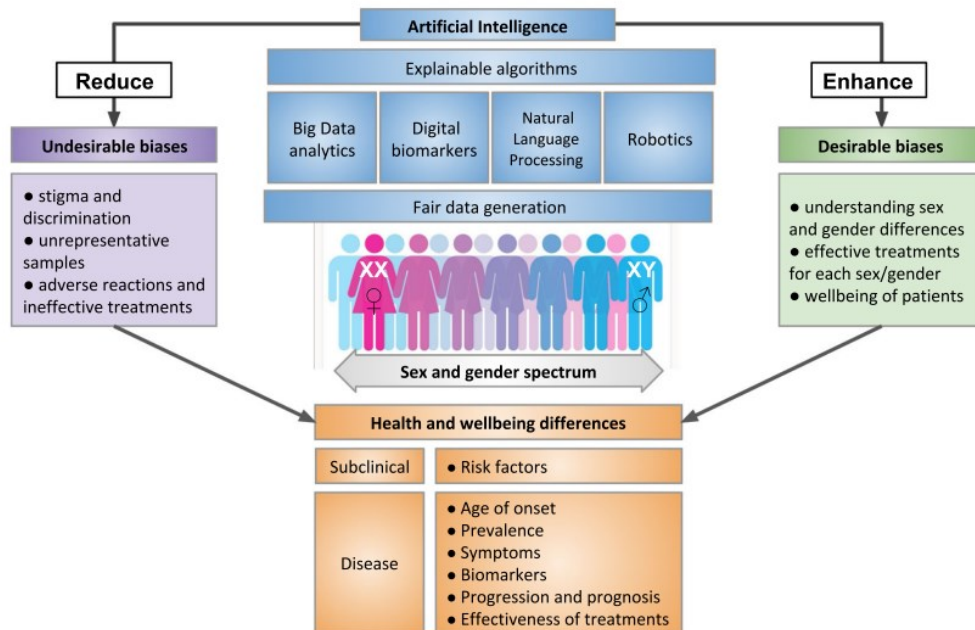


Fig. [2] Distorsioni desiderabili e indesiderabili. (Adedinsewo et al., 2022)

La corretta generazione di dati e gli algoritmi spiegabili sono requisiti fondamentali per la progettazione e l'applicazione dell'Intelligenza Artificiale la salute e il benessere di ogni individuo, indipendentemente dal suo sesso e genere.

Un bias desiderabile implica la considerazione delle differenze per formulare una diagnosi precisa e proporre un trattamento efficace per quel determinato individuo (Shansky et al., 2019).

Un bias indesiderabile comporta una discriminazione sessuale e di genere, e questo avviene quando mancano delle conoscenze certe sul sesso e il genere del paziente.

La falsa rappresentazione della popolazione bersaglio, escludendo quindi le minoranze, è un'altra importante forma di pregiudizio che è necessario evitare;

diversi studi certificano come questo errore comporti, ad esempio, un'insufficiente rappresentazione delle donne incinte nella ricerca psichiatrica (Hay et al., 2019). La tabella seguente (Cirillo et al., 2020) illustra alcuni esempi di condizioni cliniche e tecniche biomediche in cui i pregiudizi desiderabili risultano utili per la ricerca clinica, la diagnosi e il trattamento.

Clinical conditions and studies	Current status without the desirable bias	Utility of the desirable bias
Autistic spectrum disorder	There is a current lack of consideration of the demonstrated age-dependent sex differences in the symptomatology related with impairments in social communication and interaction, expressive behaviour, reciprocal conversation, non-verbal gestures for diagnostic purposes ¹²³ .	Differential diagnostic criteria for males and females could facilitate the identification of the clinical diagnosis leading to appropriate treatment.
Cardiovascular disorders	Although it has been documented that men and women respond differently to many cardiovascular medications such as statins, angiotensin-converting enzyme inhibitors and β -Blockers among others, adopted treatments do not consider sex differences ¹²⁴ . Despite the fact that Coronary heart disease (CHD) is the leading cause of death among women ¹²⁵ , the majority (67%) of patients enrolled in clinical trials for cardiovascular devices are male ¹²⁶ .	Making prescriptions according to the sex of the patient could lead to improved health benefits. The application of a desirable bias towards women would lead to a more accurate representation of sex differences in clinical research.
Genome-wide association studies (GWAS)	Most of genome-wide association studies (GWAS) focus on white male subjects ¹²⁷ and those that explore sex differences in complex traits are scarce ¹²⁸ .	The introduction of desirable biases to deliberately include female subjects and other ethnicities in GWAS could lead to better account for potential sex differences in disease that are currently unknown because of being overlooked.
Human immunodeficiency virus (HIV)	The observed lower female representation in HIV clinical trials depends, among other factors, from the disadvantaged awareness about treatment and enrolment options compared with men ¹²⁹⁻¹³¹ .	Promoting empowerment initiatives in those patients with disadvantages will increase their exposure to treatment options and clinical trial enrolment.

Fig. [3] Condizioni cliniche e tecniche biomediche con pregiudizi desiderabili utili. (Cirillo et al., 2020)

Le fonti dei pregiudizi indesiderati, che possono essere inseriti accidentalmente in un algoritmo, sono numerose e la più comune è sicuramente la mancanza di un adeguato campione rappresentativo della popolazione nel set di dati proposti nella fase di formazione.

Un'accortezza per progettare e applicare un'Intelligenza Artificiale ottimizzata per il benessere e la salute di tutti gli individui indipendentemente dall'appartenenza alla comunità cisgender è generare e utilizzare dati ed algoritmi che non contengano pregiudizi in partenza; in questo modo si limita la diffusione di pregiudizi indesiderabili che comportano la conseguente iniquità e discriminazione.

All'interno della seguente tabella (Cirillo et al., 2020) si trovano alcuni esempi delle fonti di bias indesiderabili.

Source of bias in artificial intelligence	Description
Historical bias	Arises even if the data is perfectly measured and sampled, when the world as it is leads a model to produce outcomes that are not desired. e.g. incorrectly assuming that HIV is inherently linked to homosexual and bisexual men as its prevalence is higher in this population ¹³² .
Representation bias	Occurs when certain parts of the input space are underrepresented. e.g. European male populations are the primary focus in genomics research and its derived clinical findings, neglecting other ethnicities and populations ¹³³ .
Measurement bias	Occurs when measured data are often proxies for some ideal features and labels. e.g. the use of clinical, social, and cognitive variables to detect the prodromal phase in schizophrenia and other psychotic disorders despite of observed sex differences in the expression of those symptoms and their associated risk for psychosis ¹³⁴ .
Aggregation bias	Arises when a one-size-fits-all model is used for groups with different conditional distributions. e.g., for the diagnosis and monitoring of diabetes, haemoglobin A1c (HbA1c) levels are routinely used, despite of differences associated with ethnicities ¹³⁵ and gender ¹³⁶ .
Evaluation bias	Occurs when the evaluation and/or benchmark data for an algorithm does not represent the target population. e.g. underperformance of commercial facial recognition algorithm in dark-skinned female faces as most benchmark face image datasets come from white men ¹³⁷ .
Algorithmic Bias	Occurs when bias is introduced in the algorithm consciously or unconsciously in ad-hoc solutions. e.g. by using health care cost as a proxy feature for health status without correcting for existing inequalities in health access, a commercial algorithm to predict health care needs was found to exhibit significant racial discrimination ¹³⁸ .

Fig. [4] Bias indesiderabili. (Cirillo et al., 2020)

1.4 GENDER GAP E GENDER BIAS

Il gender gap e il gender bias sono due problematiche strettamente legate alla questione della discriminazione di genere, che ancora oggi persiste in diverse aree della società; il gender gap si riferisce alle disuguaglianze tra uomini e donne in ambiti come l'istruzione, l'occupazione, la politica e la retribuzione economica, mentre il gender bias riguarda l'insieme delle preconcizioni e degli stereotipi che influenzano le decisioni e le azioni degli individui in base al loro genere.

L'equità di genere è una questione importante in tutti i settori della società, incluso il mondo della medicina e per questo motivo il gender gap e il gender bias hanno un impatto significativo sulla salute delle donne.

Il gender gap si riferisce alla differenza nella salute, nell'accesso alle cure mediche e nella partecipazione alla ricerca tra uomini e donne; tale differenza è influenzata da una serie di fattori sociali, culturali ed economici. Le donne spesso affrontano discriminazioni nell'accesso alle cure mediche e sono sottorappresentate nella ricerca clinica e questo comporta impatti negativi sulla loro salute e sulla loro capacità di ricevere la giusta assistenza medica (Bahous, 2023).

Le differenze di genere nella partecipazione agli studi clinici sono questioni affrontate da diversi studi clinici, e gli autori hanno rilevato che le donne sono sottorappresentate nella maggior parte dei casi, compresi quelli che riguardano malattie specifiche delle donne come il cancro al seno e all'ovaio; ciò significa che le donne spesso non ricevono il trattamento più efficace per le loro malattie (Huang et al., 2020).

Il gender bias si riferisce alle percezioni errate e alle stereotipizzazioni dei ruoli di genere che influenzano la pratica medica: le donne spesso affrontano pregiudizi e discriminazioni nel sistema sanitario, che possono influenzare negativamente la qualità dell'assistenza sanitaria che ricevono provocando diagnosi errate, trattamenti inadeguati e peggiori esiti di salute.

Nel 2019 uno studio pubblicato sulla rivista *Journal of Women's* ha esaminato il gender bias nella diagnosi e nel trattamento della malattia di Lyme: gli autori hanno rilevato che le donne hanno maggiori probabilità di essere mal diagnosticati e di ricevere un trattamento inadeguato rispetto agli uomini, e questo accade perché spesso è una malattia considerata tipicamente maschile, e di conseguenza non vengono considerate come un gruppo ad alto rischio (Rebman et al., 2019).

Uno studio condotto nel 2016 e pubblicato sulla rivista "JAMA Internal Medicine" ha rilevato che i medici spesso sottovalutano i sintomi delle donne e li attribuiscono a fattori psicologici o emotivi, piuttosto che a cause fisiche ottenendo come risultato diagnosi errate o ritardate che spesso influenzano negativamente la salute delle donne (Kreatsoulas, 2016).

Il gender gap e il gender bias sono due fenomeni strettamente legati alla discriminazione di genere che possono portare ad una maggiore disuguaglianza tra uomini e donne in ambiti diversi: studiando e comprendendo questi fenomeni, si può cercare di contrastare la discriminazione di genere e promuovere una maggiore parità tra i generi.

1.5 BIAS DI GENERE

Il concetto di pregiudizio è molto ampio e i bias di genere sono un tema fondamentale dell'equità (Leavy et al., 2018).

I bias di genere sono profondamente radicati nella nostra società, di conseguenza è elevato il rischio che i set di dati utilizzati contengano pregiudizi legati al genere (Badaloni e Rodà, 2022).

I bias di genere sono una questione da risolvere dal punto di vista informatico e, contemporaneamente, anche da quello sociale e culturale, il fatto che siano diffusi fa sì che risulti relativamente facile ottenere set di dati su cui sperimentare tecniche di analisi e debiasing.

La molteplicità di genere è ormai largamente riconosciuta, ma gli studi di genere confermano comunque l'esistenza di due polarità opposte e ben definite, maschio e femmina, facilitando la definizione di protocolli sperimentali.

Una revisione scientifica recente, *Gender Bias in AI: A Review of Contributing Factors and Mitigating Strategies* (Nadeem et al., 2020), ha identificato alcuni fattori che contribuiscono a creare e tramandare bias di genere: la mancanza di diversità nei dati di formazione, la presenza di stereotipi generici nella società e i pregiudizi del programmatore che, consciamente o inconsciamente, si insinuano nell'algoritmo.

Lo sviluppo di un'Intelligenza Artificiale affidabile richiede un approccio multidisciplinare poiché i fattori da considerare sono molteplici e vari (Badaloni e Rodà, 2022).

1.6 SESSO E GENERE NELL'APPRENDIMENTO AUTOMATICO

La *sex gender slippage* e la convinzione che il sesso e il genere concordino sempre possono avere conseguenze decisamente negative soprattutto in quelle situazioni in cui le persone transgender affrontano maggiori rischi di salute rispetto agli individui cisgender (Kundu et al., 2021).

Diversi studi sul tema del suicidio hanno evidenziato come l'identità di genere e le esperienze di vita influiscano in maniera decisa le tendenze suicidarie, e difatti è emerso che le persone transgender e non conformi al genere sono esposte maggiormente al rischio di suicidio; escludere le questioni di sesso e genere da queste considerazioni potrebbe comportare un fallimento nel riconoscimento di segnali di probabili istinti suicidi in persone trans (Budge et al., 2020).

In quei contesti in cui l'Apprendimento Automatico viene impiegato per gestire l'assistenza clinica è di fondamentale importanza evitare la *sex confusion* per evitare errori di valutazione che potrebbero portare anche alla morte dei pazienti (Lundberg, 2019).

Determinare se le assunzioni alla base dello sviluppo dell'algoritmo sono corrette per uno specifico paziente è fondamentale per considerare e quindi prevenire i rischi a cui vanno incontro tutti quei pazienti che deviano dalle aspettative e dai principi di coloro che hanno sviluppato i modelli.

Per garantire che le persone transgender non siano lasciate fuori dalle ricerche e dalle cure mediche è necessaria una profonda comprensione del sesso e del genere; i documenti che meglio considerano le prestazioni dei sistemi con attenzione alle differenze di genere sono quelli che estrapolano i dati da centri di riferimento a cui si rivolgono tutti quei pazienti che non si riconoscono nella definizione binaria e statica di sesso e conseguente genere.

Le cartelle cliniche elettroniche sono state concepite col presupposto che il sesso e il genere siano caratteristiche binarie, statiche e concordanti, ma queste assunzioni hanno spesso come conseguenza i problemi di *sex gender slippage*, *sex confusion* e *sex obsession*.

A partire dal 2016, il mondo della medicina è quello in cui l'Intelligenza Artificiale ha avuto la maggior crescita rispetto ad altri settori.

I metodi statistici, solitamente, affrontano il processo decisionale medico caratterizzando i modelli con equazioni matematiche; tramite il Machine Learning, l'Apprendimento Automatico, l'IA riesce a cogliere complesse relazioni che non possono essere ridotte ad un'equazione.

I sistemi ML sono quindi in grado di affrontare problemi complessi nello stesso modo in cui agirebbe un medico, riuscendo ad analizzare i dati giungendo poi a conclusioni ragionate. Tali sistemi, a differenza di un umano, hanno la capacità di analizzare ed elaborare un enorme quantità di input arrivando anche ad imparare da ogni caso incrementale.

Le nuove tecnologie che si utilizzano in campo medico devono integrarsi in modo proficuo con le pratiche attuali ed essere soggette a delle specifiche approvazioni normative.

L'Intelligenza Artificiale deve essere intesa come supporto al lavoro del medico e non come sostituto; le macchine non possiedono qualità umane quali l'empatia o la compassione, ed è necessario che i pazienti percepiscano che alla base ci sono i medici umani (Chen et al., 2017).

L'impatto e il potenziale sviluppi della medicina di precisione la rendono una delle iniziative sanitarie più promettenti nella società (Ginsburg et al., 2019).

Gli aspetti biologici (sesso) e quelli socioculturali (genere) risultano determinanti per le variazioni di diverse condizioni cliniche riuscendo ad influenzare i fattori di rischio, prevalenza, età di insorgenza, manifestazione sintomatica, prognosi, biomarcatori e addirittura efficacia del trattamento (Regitz-Zagrosek, 2021).

La scarsità di studi dedicati ad individui intersessuali, transgender e non binari comporta la presenza di molte questioni aperte sulle differenze di salute nello spettro di genere completo (Scandurra et al., 2019); alcune iniziative, come la *Global Trans Research Evidence Map* hanno l'obiettivo di promuovere l'accesso alla ricerca per colmare questi vuoti (Marshall et al., 2019).

Le differenze di sesso e di genere osservate nella salute e nel benessere sono influenzate da complessi legami tra fattori biologici e socioeconomici, che sono spesso circondati da variabili confondenti come lo stigma, gli stereotipi e la falsa rappresentazione dei dati; di conseguenza, la ricerca e le pratiche sanitarie possono essere intrecciate con le disuguaglianze e i pregiudizi di sesso e genere.

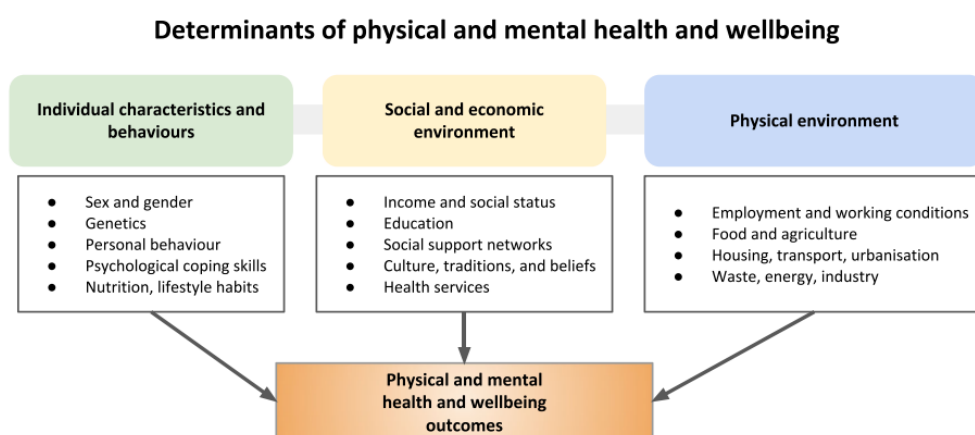


Fig. [5] La salute e il benessere degli individui e delle comunità sono influenzati da diversi fattori: le caratteristiche e i comportamenti individuali della persona e l'ambiente socio-economico e fisico. (www.who.int/hia/evidence/doh/en/)

2 L'INTELLIGENZA ARTIFICIALE

2.1 INTELLIGENZA ARTIFICIALE E MACHINE LEARNING

“L'Intelligenza Artificiale è una disciplina scientifica che mira a definire e sviluppare programmi o macchine (software o hardware) che mostrano un comportamento che verrebbe definito intelligente se fosse esibito da un essere umano.” (Rossi, 2019)

“The goal of AI is to develop machines that behave as though they were intelligent.”
(John McCarthy, 1955)

“Artificial intelligence is the study of how to make computers do things at which, at the moment, people are better.” (Elaine Rich, 1983)

Il termine *Artificial Intelligence* venne coniato da John McCarty nel 1956 durante la conferenza di Dartmouth; venne istituita come disciplina accademica e l'obiettivo fu quello di fare in modo che i computer svolgessero delle attività considerate proprie degli esseri umani.

L'IA è quel ramo dell'informatica utilizzato per la programmazione e la progettazione di sistemi hardware e software col fine di dotare le macchine di caratteristiche ritenute proprie dell'uomo: la capacità decisionale o le percezioni visive e spazio-temporali.

È il concetto di intelligenza a mutare: non è più intesa come la capacità di calcolo o la mera conoscenza di dati; l'obiettivo ultimo è dotare le macchine di intelligenza quanto più simile a quella umana.

Il computer viene quindi a riprodurre il comportamento umano senza che questo debba intervenire, o comunque l'intervento è ridotto al minimo indispensabile.

L'Intelligenza Artificiale non deve essere confusa con l'Apprendimento Automatico.

Un sistema di Intelligenza Artificiale, nonostante siamo consci del fatto che il funzionamento del cervello umano non è tuttora pienamente compreso, deve garantire alcune caratteristiche:

- agire umanamente, in modo indistinto dall'essere umano;
- pensare umanamente, risolvendo i problemi con funzioni cognitive;
- pensare razionalmente, ovvero utilizzare la logica;

- agire razionalmente, in modo da ottenere il miglior risultato possibile considerando le condizioni di partenza.

Ma come si determina se una macchina è in grado di pensare in modo umano? Un criterio largamente utilizzato è il *Test di Turing* (Turing, 1950) che venne pubblicato per la prima volta sulla rivista *Mind*. Alan Turing a soli 24 anni era stato in grado di definire la computazione e fornì un modello del calcolatore valido ancora oggi.

Il computer supererà il test solo se un esaminatore umano, a seguito della somministrazione di alcune domande scritte, non sarà in grado di comprendere se le risposte provengano da una persona o da un programma.

L'Intelligenza Artificiale è presente nella quotidianità delle nostre vite ma, come sostiene Francesca Rossi, global leader dell'IBM per l'etica dell'IA, "L'ipotesi che le macchine un giorno si sveglino e cambino idea su quello che devono fare non è realistica."

Il Machine Learning può essere considerato come un sottoinsieme dell'IA ed è costituito da tecniche che permettono ai computer di apprendere dai dati ricevuti.

I sistemi biologici hanno sviluppato, per poter sopravvivere, dei sofisticati sistemi di comportamento adattivi attraverso apprendimento ed evoluzione.

L'apprendimento può essere definito come "l'acquisizione della conoscenza del mondo" (I. Kupferman, 1985) e come "un cambiamento adattivo nel comportamento causato dall'esperienza" (G. Sherperd, 1988).

Il Machine Learning si diffuse soprattutto negli anni '80, quando la complessità dei problemi da risolvere aumentava; si è poi arrivati al deep learning, sottoinsieme del Machine Learning capace di risolvere problemi molto complessi tramite l'utilizzo di reti neurali artificiali. Col passare degli anni vennero fornite diverse definizioni di Machine Learning, sempre più esaustive e complete: "Si dice che un programma per computer impari dall'esperienza E rispetto a qualche classe di compiti T e misura della prestazione P, se la sua prestazione nei compiti in T, come misurato da P, migliora con l'esperienza E." (T. Mitchell, 1997).

I sistemi a codifica fissa mostrarono difficoltà nel riconoscimento di immagini e nell'estrazione di significato dal testo; la soluzione fu cercata in un sistema che simulasse la capacità degli umani di apprendere, e che quindi non ne imitasse solo il

comportamento. Per comprendere meglio questo concetto si può pensare al bambino che si avvicina alla lettura: egli non dovrà preliminarmente imparare la grammatica, ma inizierà a leggere dapprima le lettere, poi le parole ed infine i libri. Risulta quindi possibile imparare dall'elaborazione di dati, ed è questo il concetto alla base dell'Apprendimento Automatico che ha lo scopo di far sì che la macchina impari ad apprendere in modo autonomo e a migliorare le proprie prestazioni e conoscenze. L'idea alla base del Machine Learning è mettere l'algoritmo in condizione di imparare dai dati forniti; sviluppandosi grazie ai nuovi dati forniti, gli algoritmi riuscivano a risolvere problemi sempre più complessi.

La figura [6] schematizza lo sviluppo dei sistemi basati sull'IA nel corso degli anni.

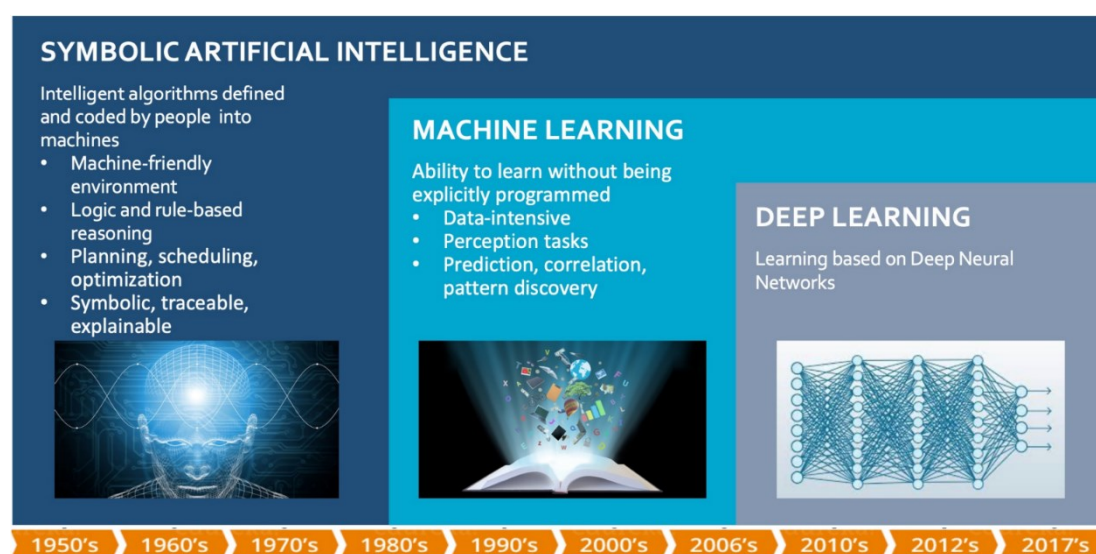


Fig. [6] Sviluppo dei sistemi basati su IA (Francesca Rossi, IBM)

Questioni come la classificazione di esempi, la regressione, il calcolo delle probabilità e la conoscenza di nuove nozioni sono compiti che ben si addicono al Machine Learning.

Il Machine Learning si basa sulla combinazione di tre elementi:

1. Task: ovvero la descrizione degli obiettivi, delle questioni che devono essere risolte;
2. Model: una rappresentazione matematica dei problemi;
3. Feature: rappresentazione degli oggetti coinvolti nei modelli.

L'insieme dei dati forniti per eseguire l'addestramento viene detto training set; un training set è formato da coppie $x, f(x)$ dove x è una possibile istanza descritta tramite i suoi attributi mentre $f()$ rappresenta la reale funzione che assegna i giusti valori e che l'algoritmo tenta di approssimare.

Un buon training set deve essere in grado di generalizzare lo scenario fornendo così gli strumenti che permettono alla macchina di lavorare correttamente sui nuovi dati, sconosciuti, che le verranno forniti.

È necessario che il training set sia adeguatamente rappresentativo affinché l'algoritmo possa procedere con le giuste inferenze.

La bontà del training set risulta quindi di fondamentale importanza; è ciò che l'algoritmo utilizza per costruire un proprio modello di comprensione del dominio ed è ciò da cui dipendono le sue scelte future.

Un training set non adeguatamente vario e rappresentativo potrebbe causare un apprendimento errato. La mancanza di complessità di un modello troppo semplificato impedisce di stimare i dati con adeguata precisione; un modello troppo complesso, d'altra parte, si adatta troppo ai dati risultando penalizzato nel momento in cui le istanze si discostano anche di poco.

La figura seguente presenta una schematizzazione dell'architettura dell'Intelligenza Artificiale.

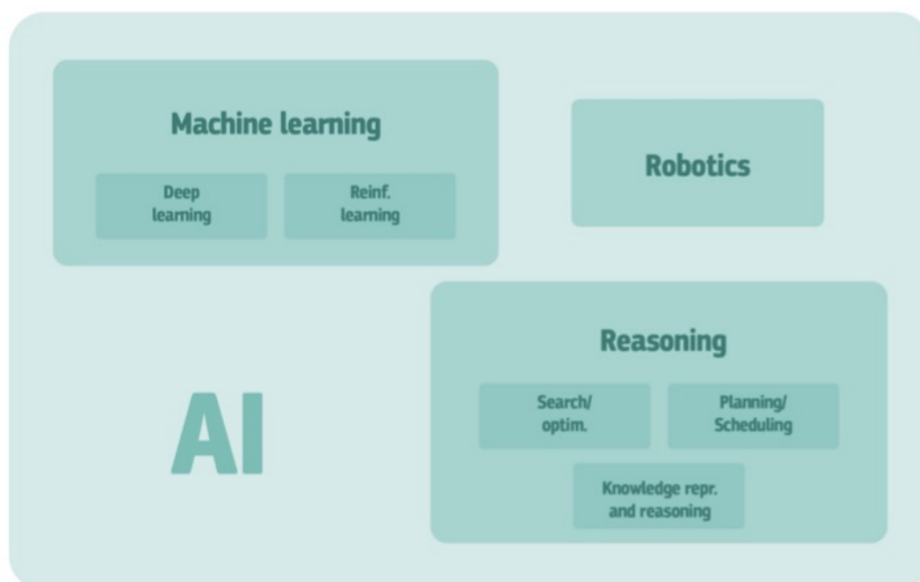


Fig. [7] Architettura dell'Intelligenza Artificiale. (Badaloni, 2023)

2.2 RETI NEURALI ARTIFICIALI

Una rete neurale è un processore massivo strutturato in parallelo e formato da tanti semplici processori unitari, con una naturale propensione ad immagazzinare informazioni e renderle disponibili per essere utilizzate. Assomiglia al cervello per due aspetti: 1) Le informazioni sono acquisite dall'ambiente attraverso processi di apprendimento. 2) Le forze delle connessioni tra i neuroni, anche chiamate "pesi sinaptici", vengono utilizzate per immagazzinare le informazioni acquisite. (Parfitt, 1991).

Le reti neurali artificiali (ANN: Artificial Neural Network) sono modelli matematici che simulano i comportamenti del cervello umano (Hinton, 1992).

La storia del loro sviluppo delle reti neurali artificiali parte già dai primi anni '40 del XX secolo, e il primo neurone artificiale fu sviluppato da McCulloch e Pitts nel 1943 (McCulloch et al., 1943).

Per evidenziare come il neurone fosse l'unità logica di base del cervello, cercarono di dimostrare che una macchina di Turing potesse essere realizzata con una rete finita di neuroni.

Il modello in questione, il cosiddetto *modello di rete deep* viene rappresentato tramite un grafo multilivello, si veda la figura seguente, in cui è possibile riconoscere i dati di input che vengono propagati in avanti attraverso tutta la rete: si parla di rete feedforward, cioè dal livello di input si procede a quello di output.

La topologia di una rete, dal punto di vista matematico, è ciò che definisce lo spazio di ipotesi in cui trovare delle rappresentazioni valide dei dati che si vuole analizzare; in figura [8] è rappresentata una generica rete multistrato.

Il grafo si compone di nodi che sono collegati tra loro da archi; Il nodo è l'unità della rete neurale dove viene effettuata l'operazione.

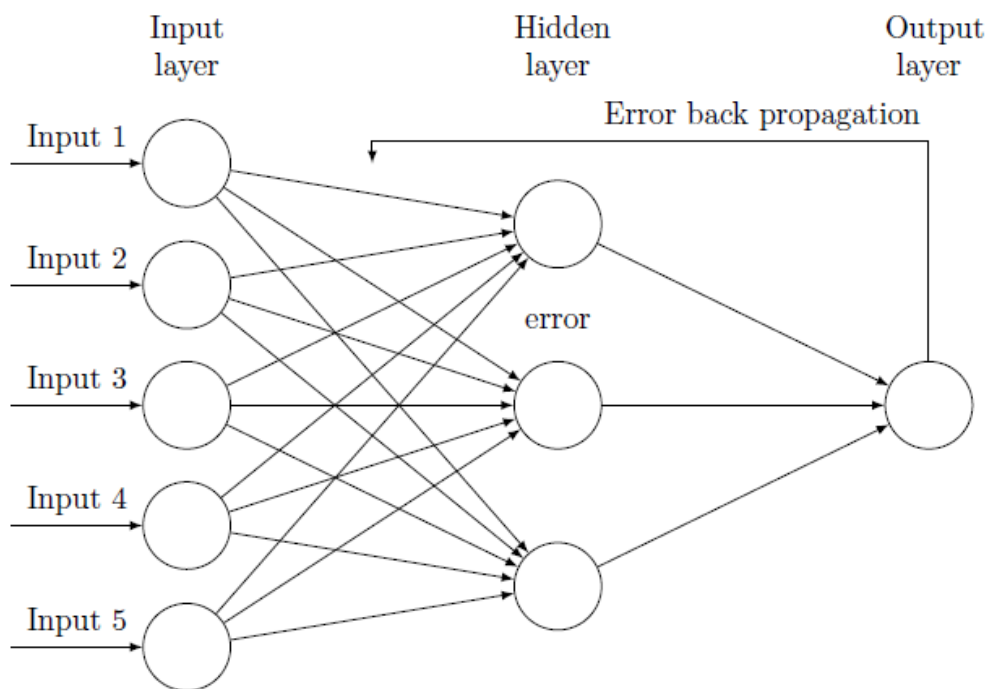


Fig. [8] Struttura di una rete neurale (Researchgate.net)

Le reti neurali sono formate da tre strati che possono coinvolgere migliaia di neuroni e decine di migliaia di connessioni:

- *I-Input*, lo strato degli ingressi, che riceve ed elabora i segnali in ingresso adattandoli alle richieste dei neuroni della rete;
- *H-hidden*, lo strato nascosto, è quello in cui avviene il processo di elaborazione vero e proprio. La struttura di questo strato può contenere più livelli di neuroni;
- *O-output*, lo strato di uscita,

Si parla di reti neurali semplici se c'è un solo strato intermedio, in caso contrario di deep learning.

L'obiettivo ultimo delle reti neurali artificiali è quello di riconoscere e rispondere in modo non programmato a stimoli non codificati in precedenza.

L'attività cerebrale umana è proprio il risultato di quest'attività svolta quotidianamente dalle cellule nervose.

L'obiettivo è quello di fare in modo che le reti neurali artificiali, utilizzando opportuni algoritmi, simulino le complesse reti neuronali biologiche rispondendo coerentemente a input non codificati in precedenza.

È bene notare come esse funzionino come una sorta di scatola nera in cui il funzionamento non è di immediata.

Il processo di Apprendimento Automatico è semplificato rispetto a quello delle reti biologiche e gli algoritmi di apprendimento utilizzati per istruire le reti neurali sono divisi in tre categorie, ciascuna utilizzata in un determinato campo di applicazione.

I tre tipi di algoritmi sono:

- supervisionato;
- non supervisionato;
- di rinforzo.

L'apprendimento supervisionato è quello in cui si fornisce alla rete un insieme di input a cui corrisponde il training set, ovvero degli output noti.

Il compito della rete è quello di comprendere il nesso che li unisce e successivamente creare nuove corrette associazioni input-output ricevendo in ingresso input non appartenenti al training set.

L'apprendimento non supervisionato consiste nel fornire alla rete solo un insieme di variabili in input; analizzandolo, la rete le categorizza generando dei cluster rappresentativi.

L'apprendimento per rinforzo, infine, permette alle reti di imparare unicamente dall'interazione con l'ambiente.

L'algoritmo orienta la rete neurale all'obiettivo prefissato utilizzando degli incentivi nel caso di azioni positive e dei disincentivi in caso di azioni negative.

Non è azzardato definire creative le soluzioni che le macchine trovano utilizzando tale algoritmo.

Basandosi su dati esterni e informazioni interne, il sistema è adattivo, cioè in grado di modificare la propria struttura.

Le reti neurali artificiali presentano delle caratteristiche che ne rendono vantaggioso l'utilizzo:

- tolleranza al rumore, ovvero la capacità di lavorare in modo corretto nonostante input imprecisi;
- elevato parallelismo che permette di processare un'elevata quantità di dati in tempi rapidi;
- tolleranza ai guasti garantita dall'architettura parallela;

- evoluzione adattiva che garantisce l'autoaggiornamento della rete neurale in presenza di modifiche ambientali.

La struttura dei neuroni presenta dendriti, sinapsi ed assoni.

Un neurone riceve i segnali provenienti da altre cellule attraverso i dendriti ed emette impulsi elettrici attraverso l'assone; la sua estremità è il luogo in cui la sinapsi trasforma l'attività dell'assone in fenomeni elettrici che eccitano o inibiscono i neuroni collegati.

La fase dell'apprendimento si verifica nel momento in cui le sinapsi trasmettono i segnali da un neurone all'altro.

Le unità interconnesse che compongono le reti artificiali richiamano la funzione dei neuroni, mentre il peso modificabile associato ad ogni connessione rimanda al ruolo delle sinapsi.

I dati forniti in input vengono moltiplicati per il peso di ciascuna connessione e infine sommati; è così che il neurone si specializza, nel tempo, a riconoscere determinati stimoli e ad apprenderli.

I neuroni artificiali utilizzati nelle reti neurali artificiali sono i perceptron e i sigmoidei.

Il funzionamento del perceptrone prevede in ingresso diversi valori binari $x_1, x_2, x_3, \dots, x_n$ producendo in uscita un singolo valore binario; l'output viene valutato combinando i pesi (w_j) corrispondenti ad ogni ingresso (x_j).

Nel caso in cui la somma pesata degli input risulti maggiore di un valore soglia dipendente dal neurone, il valore binario assunto è 1, in caso contrario è 0.

2.3 INTELLIGENZA ARTIFICIALE SPIEGABILE

La possibilità di spiegare in che modo l'IA perviene ad una certa decisione è una questione molto delicata che in Europa è particolarmente sentita.

La legge GDPR dà diritto a ogni persona di chiedere una spiegazione riguardo le decisioni prese da un algoritmo se queste hanno un impatto significativo sulla propria vita (Rossi, 2019).

Considerando i sistemi di Intelligenza Artificiale basati su regole e simboli, poiché sono metodi che descrivono in modo chiaro la sequenza di azioni necessarie, è

abbastanza semplice capire per quale motivo il programma abbia agito in un determinato modo arrivando alla causa che ha originato quella specifica decisione.

I sistemi di Machine Learning, di contro, non descrivono i passaggi necessari per risolvere un dato problema ma apprendono come risolverlo in base all'osservazione di una grande mole di dati.

I modelli di Intelligenza Artificiale forniscono previsioni di rischio e diagnosi con una rilevante influenza sulla vita e sulla salute delle persone; tuttavia, la complessità delle loro strutture interne ha messo in luce un serio problema tecnologico, il cosiddetto problema *black-box* (Holzinger et al., 2017).

Nei modelli di Apprendimento Automatico non esistono rappresentazioni dichiarative esplicite, e risulta quindi impossibile fornire spiegazioni e interpretazioni del *come* e *perché* operino in un determinato modo.

La comprensione degli specifici parametri utilizzati per giungere a determinate conclusioni cliniche è sempre più una necessità, così da giustificare quelle scelte che hanno un impatto rilevante sulla vita e sulla salute dei pazienti; la direttiva UE 2016/680 sancisce il "diritto ad una spiegazione" sull'output di un algoritmo (Towards trustable Machine Learning. Nat. Biomed. Eng. 2, 709–710, 2018).

La spiegabilità nell'IA aiuterebbe a giustificare le previsioni e le raccomandazioni cliniche ricavate dagli algoritmi in quei casi in cui sono discordanti per i pazienti con sesso e genere diversi.

La spiegazione del processo decisionale renderebbe possibile trovare potenziali conclusioni errate derivanti da un addestramento insufficiente dell'algoritmo, facilitando l'identificazione dei pregiudizi indesiderati che solitamente sono presenti in quei dati clinici provenienti da rappresentazioni squilibrate di sesso e genere.

Allo stesso tempo sarà più semplice scoprire le differenze di sesso e genere nei dati clinici rappresentativi, promuovendo i pregiudizi desiderati per personalizzare gli interventi preventivi e terapeutici (Fauw et al., 2018).

L'interpretabilità e la completezza dell'Intelligenza Artificiale sono ritenuti requisiti fondamentali per contribuire alla fiducia, alla sicurezza, all'etica e all'equità della medicina; gli algoritmi che soddisfano tali requisiti sono identificati col termine XAI.

2.4 IDENTIFICAZIONE DEI PREGIUDIZI PER L'EQUITA'

Una delle principali sfide per sviluppare un'Intelligenza Artificiale affidabile è definire il significato di equità nella pratica dell'Apprendimento Automatico ottenendo un algoritmo decisionale equo (Zemel et al., 2017).

Uno degli approcci maggiormente utilizzati per garantire l'equità è rimuovere alcuni dati sensibili quali il sesso, il genere e le caratteristiche correlate; questa è però una procedura indesiderabile perché vanifica le differenze di sesso nella prevalenza della patologia, rendendo i risultati non equi per alcune porzioni di popolazione.

La consapevolezza e la conoscenza delle differenze di sesso e genere e dei conseguenti aspetti socioeconomici sono aspetti fondamentali per raggiungere l'equità nello sviluppo algoritmico; solo così si potranno avere modelli imparziali (Holzinger et al., 2017).

Esaminando le informazioni fornite da diversi lavori, sono state fornite alcune raccomandazioni per garantire la contemplazione delle differenze tra sesso e genere nelle implementazioni dell'Intelligenza Artificiale (Cirillo et al., 2020):

- Distinguere nettamente tra pregiudizi desiderabili e indesiderabili rappresentando solo i primi;
- Aumentare la consapevolezza dei pregiudizi non intenzionali;
- Implementare algoritmi spiegabili;
- Incorporare considerazioni etiche durante ogni fase dello sviluppo tecnologico.

3 MEDICINA DI GENERE E INTELLIGENZA ARTIFICIALE

3.1 LA MEDICINA DI GENERE

La medicina di genere, come ha affermato la ricercatrice Giovannella Baggio durante un convegno del 2019, è definita come una dimensione innovativa della medicina tradizionale; permette di studiare in che misura il sesso e il genere influenzino la fisiologia, fisiopatologia e patologia umana.

Non si tratta dunque di una nuova disciplina, ma è doveroso considerarla in quanto tutte le malattie manifestano marcate differenze nell'uomo e nella donna che possono essere riscontrate nei parametri di riferimento, sintomi, terapie ed effetti collaterali dei farmaci.

In letteratura si osserva come molti studi descrivano le malattie focalizzandosi su casistiche di un solo sesso: l'infarto del miocardio rappresenta la prima causa di morte della donna, ma le malattie cardiovascolari vengono studiate prevalentemente sull'uomo. (Leavy, 2018)

Non è solo il sesso ad essere discriminante nelle malattie, ma anche il genere che porta a ripercussioni sulle abitudini e gli stili di vita.

Una svolta importante per lo sviluppo della medicina di genere, in Italia, è stata l'approvazione della legge 3/2018 atta a introdurre l'attenzione alle differenze di genere nelle pratiche sanitarie.

Questa legge, conosciuta come *Decreto Lorenzin*, dispone il riassetto e la riforma delle norme vigenti in materia di sperimentazione clinica dei medicinali per uso umano introducendo un'apposita delega al Governo che dovrà adottare uno o più decreti legislativi che dovranno essere coordinati con le norme dell'Unione europea (Regolamento (UE) n. 536/2014), le convenzioni internazionali, il rispetto degli standard internazionali per l'etica nella ricerca medica sugli esseri umani, in conformità alle previsioni dalla Dichiarazione di Helsinki dell'Associazione medica mondiale del 1964 ed alle sue successive revisioni. Inoltre, un decreto del Ministro della salute, emanato entro dodici mesi dall'entrata in vigore della Legge, disporrà la predisposizione di un piano organico volto alla diffusione della medicina attenta alle

differenze per sesso e genere per garantire la qualità e l'appropriatezza delle prestazioni del SSN da erogarsi in modo omogeneo sul territorio nazionale.

Eliana Ferroni, dirigente medico in Igiene, Epidemiologia e Sanità Pubblica, in un documento approvato dall'osservatorio dedicato alla medicina di genere approvato il 17/01/2023, sostiene che il genere come variabile deve essere sempre considerato, soprattutto negli studi di farmacoepidemiologia che servono per misurare efficacia e sicurezza dei trattamenti.

Secondo uno studio del 2021 condotto dalla Fondazione Veronesi, considerando i trial clinici randomizzati si nota come le donne risultino sottorappresentate, finendo quindi per ottenere risultati basati su una popolazione quasi esclusivamente maschile.

La medicina genere-specifica è dunque una dimensione innovativa delle scienze biomediche che studia l'influenza delle caratteristiche biologiche, il sesso, e di quelle socioculturali, il genere, sullo stato di salute e di malattia.

I primi accenni ad una, in un certo senso, questione di genere in ambito medico si possono riscontrare già nel 1932 quando gli studiosi Barrow e Nicholas notarono che la dose ipnoinducente di barbiturici nelle ratte femmine era inferiore del 50% rispetto a quella dei maschi (Haley et al., 1991).

Fu solo nel 1991, però, che in medicina si attenzionò quella che può essere definita come *questione femminile*; la cardiologa americana B. Healy evidenziò le differenze nella gestione delle patologie coronariche negli uomini e nelle donne (Haley et al., 1991). Haley sottolineò un numero inferiore di interventi diagnostici prima e terapeutici poi effettuati sulle donne e, di conseguenza, un approccio discriminatorio e insufficiente.

Nel 2020 l'OMS decide di inserire la medicina di genere nel documento *Equity act* quasi per sottolineare come il principio di equità non miri solo a garantire parità di accesso alle cure per uomini e donne, ma anche e soprattutto l'adeguatezza di cure secondo il proprio genere.

Ed è nel 2002 che viene istituito *The Department of Gender Woman's Health* attraverso cui l'OMS ammette l'esistenza di significative differenze, tra uomini e donne, nei fattori che determinano la salute; le differenze di genere e di sesso devono quindi essere tenute in considerazione nei programmi sanitari.

La salute della donna non viene più studiata solo in riferimento alle patologie considerate solamente femminili, ma tiene conto che la donna non è una mera copia dell'uomo; solo così si potrà garantire la "centralità del paziente e la personalizzazione delle terapie." (Gazzetta ufficiale della repubblica italiana, 2018)

In definitiva si può affermare che la medicina di genere evidenzia come alcune patologie, in diagnosi e terapia, si differenzino tra donna e uomo a livello di sesso biologico e di genere.

È noto che le donne abbiano un'aspettativa di vita più alta rispetto agli uomini, ma risultano più sensibili a malattie cardiovascolari e, ad esempio, meno ai melanomi; è quindi importante tenere in considerazione le differenze di genere al fine di sviluppare un percorso di cura adatto alle esigenze di ciascun paziente, su misura.

Il presidente della società internazionale di medicina di genere, Glezerman, nel 2013 dichiarò che la medicina era ad un punto di svolta simile a quello che attraversò quando venne definita la pediatria (Lipton et al., 1965).

Il punto nodale fu comprendere che il bambino non è un adulto in miniatura¹ e che il suo corpo è radicalmente differente.

Egli crede che in futuro si potrà in un certo senso parlare di due medicine, una rivolta agli individui di sesso maschile e l'altra a quelli di sesso femminile; saranno così sottolineate somiglianze e differenze.

3.2 LA MEDICINA DI GENERE IN ITALIA

Fu nel 1998 che in Italia, per la prima volta, si pose l'attenzione sulla medicina di genere; i ministeri della Salute e delle Pari Opportunità diedero vita al progetto *Una salute a misura di donna*.

Il tavolo di lavoro *Salute delle donne e farmaci per le donne* fu istituito nel 2005 presso il Ministero della Salute e vide la partecipazione dell'Istituto Superiore di Sanità (ISS), dell'Agenzia Italiana del Farmaco (AIFA), dell'Agenzia per i Servizi Sanitari Regionali (ASSR) e della Società Italiana di Farmacologia (SIF).

È del 2008, a cura del Comitato nazionale di bioetica, il rapporto *La sperimentazione farmacologica sulle donne* e nel 2010 vennero istituiti gruppi di lavoro per delineare linee guida di genere nelle malattie respiratorie e cardiovascolari.

L'ISTAT, per individuare caratteristiche differenti in base a specifici indicatori di genere, utilizza la metodologia della *Sesso-stratificazione* per l'analisi dei dati.

Giovannella Baggio, presidente del Centro Studi Nazionale su salute e medicina di genere, organizzò a Padova, nel 2009, il primo congresso nazionale sulla medicina di genere.

Nel 2018 il ministro della salute promosse un disegno di legge approvato poi dalla Camera dei deputati con il quale si delegava il governo al *riassetto della normativa in materia di sperimentazione clinica e per la riforma degli ordini delle professioni sanitarie, nonché disposizioni in materia sanitaria* e contestualmente l'OMS riconobbe il genere come determinante nelle programmazioni sanitarie.

A seguito della legge promulgata nel 2018, divenne obbligatorio, nei percorsi clinici, l'analisi delle differenze tra uomini e donne.

Il decreto emanato dal ministro della salute nel 2019 prevedeva l'adozione del *Piano per l'applicazione e la diffusione della medicina di genere*.

Nel 2007 presso l'Istituto Superiore di Sanità è stata creata una struttura ad hoc che si occupa delle differenze biologiche e ha coordinato un progetto strategico "Salute Donna" che ha coinvolto 25 unità operative sparse sul territorio nazionale, su 5 aree di azione prioritaria (Malattie dismetaboliche e cardiovascolari – Immunità ed endocrinologia– Ambiente di lavoro – Malattie iatrogene e reazioni avverse – Determinanti della salute della donna).

Condizione necessaria per ottenere validi risultati è formare una rete che coordini le iniziative a supporto dello sviluppo della medicina di genere.

Il Veneto, al pari di altre regioni quali Toscana, Emilia-Romagna e Piemonte, ha inserito nel proprio piano sociosanitario regionale la medicina di genere. Nel 2012 all'università di Padova è stata istituita la prima cattedra italiana in materia detenuta dalla professoressa Giovannella Baggio.

La medicina di genere deve pervadere tutte le discipline, a partire dalle materie precliniche fino ad arrivare a tutte le specialità.

3.3 INTELLIGENZA ARTIFICIALE IN CAMPO MEDICO

È negli ultimi decenni che l'Intelligenza Artificiale ha trovato ampio uso in ambito medico grazie alla precisione con cui opera in specifici contesti clinici; alcuni sistemi informatici, sfruttando il deep learning, hanno raggiunto un'accuratezza diagnostica paragonabile a quella di medici esperti (Cameron e Stinson, 2019).

Applicando il Machine Learning alla medicina si può ottenere un aumento della produttività del sistema, una maggiore accuratezza diagnostica e perfino il riuscire ad offrire la possibilità di accesso a sistemi diagnostici anche in luoghi che non potrebbero beneficiarne a causa di impedimenti geografici o economici.

In sanità l'IA viene utilizzata soprattutto per:

- l'analisi delle immagini radiologiche al fine di individuare tumori;
- rilevazione della retinopatia diabetica;
- predizione di patologie cardiovascolari tramite appositi algoritmi.

L'Apprendimento Automatico medico può essere catalogato in due diversi modi (Cameron e Stinson, 2019):

- sistemi che generano dati relativi al sesso e al genere, come ad esempio l'elaborazione del linguaggio naturale per identificare pazienti transgender;
- sistemi che utilizzano i dati di sesso, genere ed orientamento raccolti precedentemente.

Nelle applicazioni mediche del Machine Learning si fornisce in input l'immagine diagnostica e ricevendo in output, solitamente sotto forma di punteggio o valore percentuale, la corrispondente predizione.

La fase di addestramento serve per presentare al sistema numerosi casi catalogati servendosi di una diagnosi fornita a maggioranza da un gruppo di specialisti, il cosiddetto *gold standard*.

Durante la fase di test si forniscono al modello alcune immagini senza specificarne la preventiva classificazione: si può così osservare la sua accuratezza diagnostica e successiva capacità predittiva.

I test vengono ripetuti fin tanto che non si raggiungono risultati soddisfacenti, ovvero livelli di sicurezza molto elevati.

I primi studi, iniziati negli anni '70, si focalizzavano sullo sviluppo di software per simulare il ragionamento umano basandosi sulle conoscenze medico-cliniche, e si trattava quindi di uno strumento di supporto nelle decisioni cliniche.

Negli ultimi anni, grazie allo sviluppo dell'informatica e alla massiccia diffusione di internet che rende disponibili grandi moli di dati (Combi et al., 2010) e quindi di conoscenze, ci si è concentrati sullo sviluppo di applicazioni, i *sistemi esperti, capaci* di fornire le informazioni rilevanti ai medici in modo semplice e veloce.

Il costante e continuo sviluppo dell'Intelligenza Artificiale porta a una sempre più diffusa interazione tra il mondo fisico, quello digitale e l'ambito biologico; quest'ultimo si presta sempre di più all'elaborazione algoritmica grazie ai big data biomedici e ai processi di fenotipizzazione resi possibili dai dispositivi digitali personali.

Secondo Klaus Schwab, ingegnere ed economista tedesco, i repentini cambiamenti della vita quotidiana apportati dalla diffusione dell'IA a cui abbiamo assistito negli ultimi anni, altro non sono che lo sfondo della quarta rivoluzione industriale (Schwab, 2016).

Le attività svolte da questi sistemi di IA vengono chiamate *data mining*, e consistono nell'estrarre informazioni implicite dai dati che si hanno a disposizione (Patel et al., 2009).

Le principali funzioni del data mining sono:

- classificazione: consiste nella collocazione di oggetti all'interno di un sistema predefinito di classi;
- clusterizzazione: porta all'individuazione di gruppi omogenei di oggetti;
- derivazione di regole di associazione: usate per l'estrapolazione di associazioni ricorrenti;
- predizione: sulla base di dati a disposizione, permette la stima di eventi futuri;
- analisi what-if: consente ragionamenti ipotetici rispetto a differenti situazioni;
- estrazione di pattern dalle moli di dati che si hanno a disposizione.

Il medico di medicina generale ha a disposizione dati biomedici quali cartelle cliniche, database e monitoraggi eterogenei derivanti da fonti tra loro distinte e spesso non

comunicanti; le tecniche di IA risultano fondamentali per la gestione e l'interpretazione di questi dati.

Il monitoraggio e la gestione dei pazienti risultano rivoluzionati grazie all'utilizzo di dispositivi comuni quali smartphones, pc e dispositivi indossabili largamente diffusi tra la popolazione che misurano, controllano e memorizzano i cosiddetti *biomarkers digitali*.

Un esempio di quanto detto si può ricercare nello studio in cui vengono analizzati i valori ottenuti monitorando costantemente, tramite dispositivi indossabili, pulsossimetria, frequenza cardiaca e respiratoria, passi e temperatura corporea di soggetti affetti da broncopneumopatia cronica ostruttiva (Esteban et al., 2015).

Il sistema è in grado di predire con elevata accuratezza (AUC 0.87) l'insorgenza di riacutizzazione con tre giorni di anticipo; le variabili più predittive sono risultate la frequenza cardiaca e il numero di passi giornalieri (Malva et al., 2019).

Le tecnologie basate sull'IA offrono la possibilità di utilizzare le procedure diagnostiche tipiche di branche specialistiche nelle cure primarie in una sorta di democratizzazione di accesso alle cure, portando al miglioramento delle cure territoriali.

Essendo anch'essa una tecnologia sanitaria, l'IA applicata alla medicina per poter essere utilizzata deve superare una fase preclinica, una clinica pilota ed infine, per poter poi essere utilizzata nella vita reale, la fase clinica allargata.

La valutazione formale, *Health technology assessment*, deve approvare la procedura tenendo conto anche delle implicazioni etiche, legali e sociali.

L'espressione *medicina digitale* è spesso intesa in modo generico poiché non esiste una definizione universalmente riconosciuta; generalmente si intende la *digital health* come quell'ambito che racchiude le applicazioni legate ai dispositivi indossabili, l'informatica medica e la medicina personalizzata (Vayena et al., 2018).

Al centro di tutti questi modelli c'è il flusso circolare di dati dal paziente ai dispositivi, ai medici, ai sistemi sanitari e, infine, di nuovo ai pazienti (Fossa et al., 2019).

Le tecnologie digitali applicate alla salute umana si possono classificare in quattro ambiti principali (Elenk et al, 2015):

- monitoraggio della salute;
- diagnosi della malattia;

- algoritmi predittivi applicati a prognosi e trattamento;
- salute pubblica.

Alcuni studi hanno dimostrato come la fenotipizzazione digitale, agevolata dall'acquisizione di dati relativi all'attività quotidiana monitorando in tempo reale lo stato di salute, risulta utile nel monitoraggio di condizioni cliniche quali depressione e ansia o disturbo bipolare (Huckvale, 2019).

I biomarcatori digitali sono ormai parte integrante della pratica clinica e della ricerca biomedica, e sono approvati dall'FDA come device e risultano soggetti alle medesime classificazioni e modalità di utilizzo di quelli non digitali.

Rispetto ai biomarcatori non digitali hanno il vantaggio di poter fornire dati qualitativi e quantitativi in modo drasticamente meno invasivo e più economico.

Gli algoritmi basati sull'Apprendimento Automatico, il Machine Learning, sono in grado di diagnosticare problematiche relative al declino delle facoltà cognitive causato dalle demenze (Dagum, 2018).

I biomarcatori digitali misurano i pattern nell'interazione tra utente e dispositivo, come la velocità di scrittura, il tempo di reazione a stimoli cognitivi e gli errori commessi.

L'azienda Fitbit e il National Institute of Health (NIH) hanno stipulato un accordo nell'ambito della medicina di precisione; lo scopo di queste ricerche è quello di studiare i dati fenotipici digitali assieme ai dati genomici interpolandoli con quelli ambientali e socioeconomici per poter adattare la diagnostica alle caratteristiche di ogni specifico paziente.

La diagnostica è sicuramente l'ambito in cui la medicina digitale trova maggiore applicazione; si prospetta la possibilità di automatizzare i processi diagnostici aumentandone l'accuratezza e, allo stesso tempo, diminuendo il carico di lavoro degli operatori sanitari.

Un algoritmo capace di analizzare la storia clinica di un paziente tenendo in considerazione i dati relativi ad una patologia, può essere impiegato per prevedere in maniera attendibile il decorso clinico della malattia (Fossa et al., 2021).

I temi che aprono scenari di dibattito per queste procedure sono principalmente tre:

1. privacy e protezione dei dati personali;
2. questioni sull'Apprendimento Automatico;

3. trasparenza degli algoritmi di IA.

Il ruolo dell'IA nella medicina e nell'assistenza sanitaria sarà sempre più importante anche perché la biologia umana è estremamente complessa e la potenza di calcolo e degli algoritmi, unita alla disponibilità di big data proveniente dai registri sanitari, dimostra come la medicina digitalizzata permetta di colmare il vuoto di comprensione dei complessi sistemi biologici (Ahuja, 2019).

L'IA negli ultimi anni ha anche avuto un importante impatto nella gestione di pazienti cronici e nei processi decisionali (Bresnick, 2016) ed è sempre più impiegata in genomica, imaging e diagnosi, medicina di precisione e scoperta di nuovi farmaci più efficaci e sicuri.

L'implementazione di tecnologie di Intelligenza Artificiale sempre più potenti permettono progressi medici incredibili e senza precedenti, sia in termini qualitativi che quantitativi.

L'Intelligenza Artificiale in medicina può essere catalogata in due sottotipi (Amisha et al., 2019):

- IA virtuale, consiste in cartelle cliniche elettroniche e reti neurali utilizzate per le decisioni di trattamento;
- IA fisica, ovvero i robot impiegati negli interventi chirurgici e le protesi intelligenti.

3.4 IA E MEDICINA PREDITTIVA

La medicina predittiva può essere intesa come la capacità di determinare in termini probabilistici i fattori che possono facilitare l'insorgenza di una determinata malattia in uno specifico paziente; lo scopo è dunque individuare i fattori che predispongono una persona sana a sviluppare la malattia (Di Benedetto, 2020).

Gli strumenti abilitati all'IA possono identificare relazioni significative nei dati non elaborati e possono essere applicati in quasi ogni campo della medicina, tra cui lo sviluppo di farmaci, le decisioni sui trattamenti, la cura del paziente e le decisioni finanziarie e operative.

La differenziazione dei fattori di rischio ambientali da quelli individuali permette lo sviluppo di piani personalizzati per la prevenzione.

Un algoritmo è in grado di elaborare milioni di immagini in poco tempo, riuscendo a distinguere similarità e differenze proprio come farebbe un medico radiologo, ma molto più velocemente e con un costo minore.

L'Intelligenza Artificiale applicata alla medicina predittiva riesce addirittura a riconoscere pattern all'interno dei big data arrivando a sviluppare una nuova conoscenza.

L'IA fornisce un importante impulso alla medicina predittiva, permettendo l'elaborazione di elevate quantità di dati, individuando correlazioni non percepibili dal cervello umano.

I sistemi di Intelligenza Artificiale risultano un valido supporto dei medici nel riuscire a formulare diagnosi più rapide, precise ed accurate.

3.5 LIMITI IA IN MEDICINA

Nell'ambito delle ricerche sanitarie e nella medicina, considerare il sesso e il genere del paziente è fondamentale poiché sono proprio questi due elementi a determinare importanti differenze nella salute e nelle potenziali malattie del soggetto.

La maggior parte degli algoritmi sviluppato in ambito medico, però, non tengono conto di questi aspetti che quindi non hanno poi evidenza negli eventuali bias.

Il non soffermarsi sull'importanza di questi aspetti comporta l'ottenimento di risultati falsati con conseguenti errori nell'analisi dei pazienti e cadendo, potenzialmente, in discriminazioni.

L'Apprendimento Automatico e l'elaborazione del linguaggio naturale sono due aspetti che l'IA sfrutta per rivoluzionare quei settori ad alta intensità di conoscenza come, ad esempio, l'assistenza sanitaria (Leen e Yoon, 2021).

I sistemi basati sull'Intelligenza Artificiale risultano i più adatti all'applicazione in medicina, portando allo sviluppo di cure innovative, migliorando lo standard e l'accessibilità dei trattamenti e per migliorare le condizioni di salute di numerosi pazienti (Drukarch, 2022).

L'Intelligenza Artificiale risulta performante anche nell'organizzazione di processi ripetitivi riuscendo a gestire grandi quantità di dati fornendo, inoltre, un supporto decisionale migliorando i risultati positivi e riducendo i costi della terapia (Frost et al, 2016).

L'IA assicura la comprensione e l'utilizzo delle complesse relazioni che sono alla base del modo in cui l'uomo lavora e come egli cerca di migliorare le terapie scoprendo nuovi trattamenti e ipotesi scientifiche; il supporto della tecnologia è fondamentale, in quanto l'uomo, da solo, non riuscirebbe a cogliere il legame (Price e Nicholson, 2019).

Ad oggi la maggior parte degli algoritmi implementati ed utilizzati nel contesto sanitario non tengono in considerazione le differenze di genere e il conseguente rilevamento di eventuali pregiudizi (Cirillo et al., 2020).

Nell'ambito della medicina le conseguenze della mancanza di attenzione alle questioni di sesso e genere negli algoritmi decisionali sono state per molto tempo sottovalutate (Saddler et al., 2021) anche se gli errori commessi possono comprometterne la sicurezza giungendo a diagnosi errate.

In letteratura si possono trovare diversi studi in cui ci si sofferma su come l'IA possa dedurre il genere di un soggetto considerando alcuni tratti specifici (Nieuwenhuis et al., 2018); una questione aperta è, però, il fatto che questi sistemi identifichino oggettivamente il genere di un individuo e questo si scontra col principio per cui il genere sia una considerazione assolutamente soggettiva.

Talvolta i risultati possono portare a situazioni di mis-gendering con conseguenze negative per una parte della popolazione quali la comunità transgender, inter-sex e non binaria (Fosch-Villaronga et al., 2021).

In termini di sesso, genere, razza e background socioeconomico le differenze tra gli individui sono ovvie e in medicina, influenzando lo stato di salute e di eventuale malattia, diventano cruciali; anche la risposta a farmaci e trattamenti è diversa, e questo spesso comporta esiti sanitari dannosi e aumento dei costi sanitari (Weiner et al., 2020).

Ignorare gli aspetti della diversità in medicina solleva dubbi sulla rappresentazione dei pazienti (Carnevale et al., 2021), sulla discriminazione (Rotenstein e Jena, 2018),

sull'autonomia e sulla sicurezza potendo, in realtà, comportare esiti dannosi e addirittura la morte (Muñoz et al., 2020).

Tali nefaste conseguenze sono il frutto della tradizionale considerazione di sesso e genere che solitamente vengono ridotte a opposti risultati binari, maschi contro femmine (Nielsen, 2021), spesso però confusi e considerati in modo indistinto dalla comunità medica (The EUGenMed et al., 2015).

Prendendo in considerazione una recente applicazione di IA in medicina, l'*IBM Watson Health's cancer AI algorithm*, si evidenziano alcune problematiche relative all'applicazioni di IA in medicina (Bender et al., 2021).

Fornendo a questo algoritmo i dati effettivi, ovvero alimentandolo con input limitati, si osserva un aumento di potenziali errori e danni ai pazienti conseguenti a negligenza medica imputabile ad un algoritmo difettoso.

Grandi quantità di dati, però, possono essere inglobati in schemi egemonici predefiniti su genere, razza ed etnia avendo conseguenze negative sulle comunità ai margini (Bender et al., 2021).

Si evidenziano così le preoccupazioni derivanti dall'utilizzo dei *blackbox algorithms* sollevando sempre maggiori richieste di trasparenza e spiegazioni degli ambienti soggetti ad IA prima che l'algoritmo venga effettivamente impiegato nei trattamenti dei pazienti (Felzmann et al., 2021).

In realtà è stata chiesta cautela in relazione alle richieste di maggiore trasparenza dei sistemi ML e IA, spiegando come un'eccessiva spiegabilità di tali sistemi potrebbe diminuire le prestazioni, rendendoli dipendenti dall'ontologia formale facendoli quindi risultare inadeguati se applicati nella pratica clinica (Smith, 2019).

Le relazioni tra sesso, genere e sexuality sono molteplici ed intricate, e questo si rispecchia anche nel modo in cui il genere viene compreso ed utilizzato dagli algoritmi e dalla legge; anch'essa si è progressivamente evoluta nell'integrare tali elementi in norme che contrastino la discriminazione e consentano la protezione dei dati (Fosch, 2021).

Gli algoritmi funzionano in termini binari, riducendo il mondo a mere classificazioni da risolvere; il mondo, però, non è semplificabile in *si-no*, *bianco-nero*, ma è costituito da complessità tra i vari concetti che non possono essere ignorate.

È importante sottolineare come non tutti i concetti siano facili da descrivere, soprattutto in quei campi in cui si devono intersecare leggi e nuove tecnologie (Fosch-Villaronga e Drukarch, 2022).

3.6 CONOSCENZA DI GENERE E IA

I pregiudizi legati al genere, in contesti come l'Italia in cui c'è tradizionalmente una visione patriarcale della società, assumono una notevole importanza (Badaloni e Rodà, 2022).

Per sviluppare un'IA affidabile, è necessario porre attenzione all'analisi del genere, dell'etnia e dello sviluppo personale e sociale; le linee guida etiche per un'Intelligenza Artificiale affidabile elencano sette requisiti imprescindibili che devono essere soddisfatti:

- Azione e supervisione umana;
- Robustezza tecnica e sicurezza;
- Privacy e governance dei dati;
- Trasparenza e diversità;
- Non discriminazione ed equità;
- Benessere sociale e ambientale;
- Responsabilità.

La dimensione del genere non è inclusa negli orientamenti sessuali comunitari, “per sfruttare il potere creativo del sesso, del genere e dell'analisi intersezionale per l'innovazione e la scoperta” è necessario affrontare l'innovazione di genere nel campo dell'Intelligenza Artificiale (Schiebinger, 2021).

Col termine bias si intende un sistema di conoscenza condivisa nella società a favore o contro qualcosa, e gli algoritmi di Apprendimento Automatico, in modo analogo a quanto accade per le persone, risultano vulnerabili a queste distorsioni (Mehrabi et al., 2021).

4 L'INTELLIGENZA ARTIFICIALE E LO STUDIO DELLE MALATTIE CARDIOVASCOLARI

4.1 SOMMARIO DEL CAPITOLO

Le malattie cardiovascolari sono tra le principali cause di morbidità e mortalità in tutto il mondo; la diagnosi precoce e accurata di queste patologie è fondamentale per prevenire complicanze e migliorare la prognosi dei pazienti affetti.

L'Intelligenza Artificiale rappresenta una promettente area di ricerca per lo sviluppo di strumenti di supporto alla diagnosi delle malattie cardiovascolari.

Questa metodologia si basa su algoritmi e tecniche di Apprendimento Automatico per analizzare grandi quantità di dati e fornire previsioni o decisioni basate su modelli, e nel contesto delle malattie cardiovascolari, l'IA può essere utilizzata per analizzare i dati provenienti da vari test diagnostici e fornire informazioni utili per la diagnosi e la gestione della patologia.

Tra le tecniche di IA più utilizzate per la diagnosi delle malattie cardiovascolari si trovano:

- Reti neurali artificiali: sono in grado di imparare dai dati attraverso il processo di Apprendimento Automatico e di fornire previsioni sulla base di modelli matematici complessi; hanno la capacità di riconoscere pattern e correlazione nei dati.

Le reti neurali artificiali sono state utilizzate con successo per la diagnosi di malattie cardiovascolari come l'infarto miocardico, la fibrillazione atriale e la stenosi aortica (Adedinsewo et al., 2022);

- Support Vector Machines (SVM): queste tecniche di Apprendimento Automatico sono state utilizzate per la diagnosi delle malattie cardiovascolari attraverso l'analisi di dati provenienti da test diagnostici come l'elettrocardiogramma (ECG) e l'ecocardiogramma (Adedinsewo et al., 2022);
- Alberi decisionali: questi algoritmi sono in grado di analizzare grandi quantità di dati e fornire decisioni sulla base di una serie di regole logiche. Gli alberi decisionali sono stati utilizzati per la diagnosi di patologie come la cardiomiopatia e la stenosi mitralica (Adedinsewo et al., 2022).

L'IA può essere utilizzata anche per la gestione delle malattie cardiovascolari, ad esempio attraverso la predizione del rischio di complicanze o la personalizzazione dei trattamenti analizzando i dati clinici dei pazienti come l'età, il sesso, la pressione sanguigna e il profilo lipidico, calcolando in modo preciso il rischio di sviluppare una malattia cardiovascolare, identificando così i pazienti ad alto rischio.

L'utilizzo di tecniche di Apprendimento Automatico e di analisi dei dati può migliorare l'accuratezza della diagnosi e la gestione delle patologie, consentendo una maggiore precisione e personalizzazione dei trattamenti.

4.2 SCREENING DELLE MALATTIE CARDIOVASCOLARI NELLE DONNE

Le malattie cardiovascolari rimangono la principale causa di morte nelle donne, e considerando le numerose evidenze delle differenze basate su sesso e genere nel loro sviluppo ed esito, risulta necessario un approccio più efficace allo screening dei fattori di rischio e dei fenotipi nelle donne.

I sistemi di sorveglianza della salute pubblica e di erogazione dell'assistenza sanitaria ora generano continuamente enormi quantità di dati che potrebbero essere sfruttati per consentire sia lo screening del rischio cardiovascolare sia l'attuazione di interventi preventivi su misura per tutta la durata della vita di una donna; tuttavia, gli operatori sanitari, i comitati per le linee guida cliniche e gli esperti di politica sanitaria non sono ancora sufficientemente attrezzati per ottimizzare la raccolta di dati sulle donne e sviluppare approcci per mirare agli interventi (Adedinsewo et al., 2022).

I dati sulle malattie cardiovascolari (CVD) delle donne sono i seguenti: si arriva al 28% delle morti negli USA e al 35% se consideriamo tutto il mondo; sono anche la causa principale di decesso durante la gravidanza, arrivando a circa il 34% (Mehta et al., 2020).

Sebbene circa il 90% delle donne negli USA ha almeno un fattore di rischio cardiovascolare, nella consapevolezza, nel riconoscimento e nella gestione opportuna del rischio di CVD resta un grande divario rispetto a quanto accade per gli uomini (Brown et al., 2018).

Il sesso è un costrutto biologico che fa riferimento a caratteristiche anatomiche e fisiologiche classificate come maschili o femminili; il genere è un costrutto

multidimensionale che comprende identità e ruoli psicologici, sociali e culturali (Connelly et al., 2021).

Sesso e genere, quindi, sono due identità distinte che influenzano le CVD nelle donne: ci sono chiari effetti sesso-specifici sul rischio cardiovascolare, come un più basso livello di estrogeni associati ad un più alto livello di rischio nelle donne giovani (Morselli et al., 2017).

Gli effetti specifici di genere includono, ad esempio, l'associazione tra stress coniugale ed eventi cardiaci tra le donne con cardiopatia ischemica e uno stato socioeconomico inferiore indipendentemente dalle malattie cardiovascolari (Connelly et al., 2021).

Identificare ed evidenziare tali differenze è essenziale per comprendere le CVD nelle donne e per pianificare interventi mirati, come ad esempio lo screening per migliorare i risultati.

Sebbene sia chiaro che la CVD aterosclerotica nelle donne aumenta sia con i fattori tradizionali che con i fattori di aumento del rischio come età, menopausa, diabete e fumo, questi fattori che aumentano il rischio per le donne non sono stati inclusi in nessuno strumento di valutazione del rischio cardiovascolare (Garcia et al., 2016).

Le donne hanno maggiori probabilità rispetto agli uomini di sviluppare ischemia con arterie coronarie non ostruttive e sintomi atipici che vengono ignorati o diagnosticati solo nella fase avanzata del decorso della malattia (Van Oosterhout et al., 2020); presentano anche un rischio maggiore di sviluppare esiti avversi in seguito all'ictus (Carcel et al., 2019).

Le istituzioni sanitarie dovrebbero quindi offrire valutazioni individualizzate dei fattori di rischio cardiovascolare incentrate sulle donne, uno screening dedicato e strategie di intervento personalizzate e quindi più efficaci.

La raccolta continua dei dati sanitari, combinata con il rapido sviluppo dell'Apprendimento Automatico e con gli strumenti di Intelligenza Artificiale per l'analisi, rende possibile una più completa valutazione del rischio di CVD nelle donne; l'ampiezza e la profondità dei dati e degli strumenti analitici possono facilitare gli approcci per la prevenzione primordiale, primaria e secondaria delle CVD nelle donne. (Adedinsewo et al., 2022)

La figura seguente illustra le fasi di screening e prevenzione delle malattie cardiovascolari nelle donne.

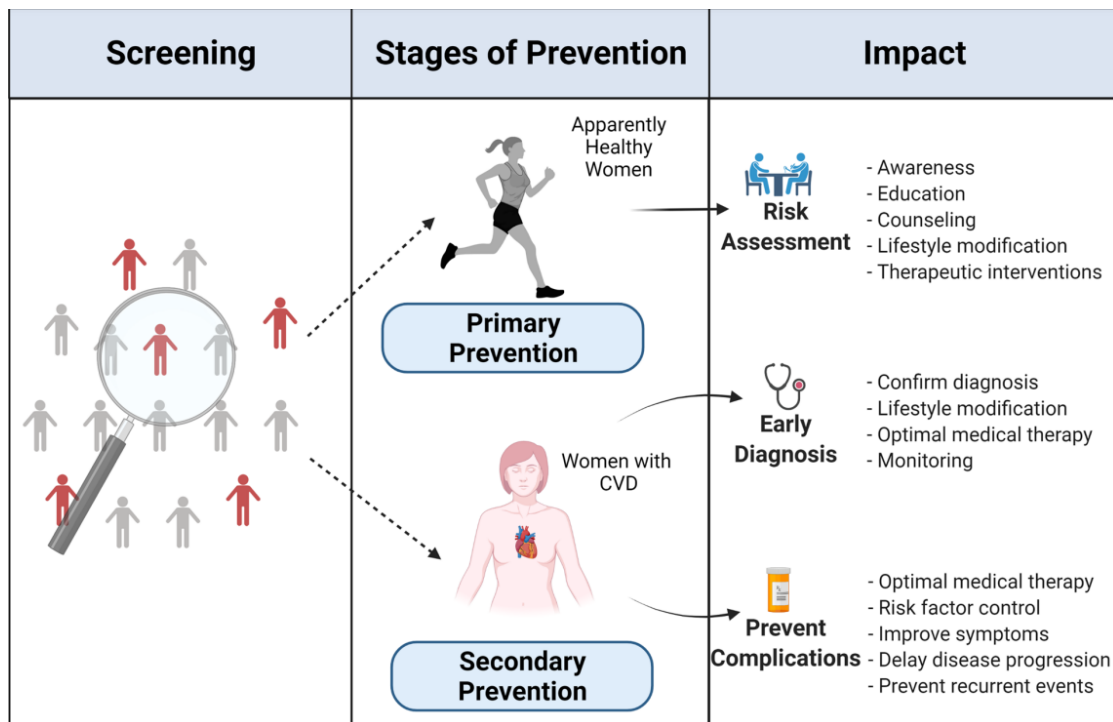


Fig. [9] Screening e fasi della prevenzione delle malattie cardiovascolari nelle donne. (Roger et al., 2020)

L'Intelligenza Artificiale, integrata con l'evoluzione dei sistemi di acquisizione dei dati sanitari, ha un ruolo fondamentale nell'analisi dei dati che sono intrinsecamente complessi e in rapida e costante espansione.

Gli algoritmi possono supportare la prevenzione in tutte le sue fasi:

- **Prevenzione primordiale:** l'Intelligenza Artificiale può espandere l'uso delle applicazioni per il controllo della salute personale incorporando dati provenienti da più sensori per prevedere in modo preciso le donne a rischio di sviluppare CVD. Fattori di rischio possono essere ipertensione, obesità e ipercolesterolemia;
- **Prevenzione primaria:** gli strumenti digitali possono essere usati per lo screening di fattori noti ma non diagnosticati come sedentarietà, obesità e stress psicosociale;

- Prevenzione secondaria: gli algoritmi di IA possono identificare le donne che hanno già ricevuto una diagnosi di CVD conclamata che potrebbero beneficiare di terapie mediche più intensive (Trimboli et al., 2019).

4.3 INTELLIGENZA ARTIFICIALE PER LO SCREENING DI CVD NELLE DONNE

“Tutti i modelli sono sbagliati, ma alcuni sono utili.” (Box, 1987)

È noto che i modelli statistici forniscono una soluzione per mettere in relazione un insieme di variabili a un risultato di interesse, ma la scelta delle variabili può influenzare l'adattamento del modello ai dati.

Il Machine Learning riesce a creare uno spazio di ricerca più ampio per un modello attuale sfruttando valide capacità di calcolo; il deep learning crea una rete di complesse operazioni di estrazione e riepilogo dei dati in modo da fornire meccanismi adatti a identificare i pattern nei set di dati.

L'Intelligenza Artificiale, combinando gli elementi propri del Machine Learning e del deep learning, fornisce strumenti statistici che consentono di svolgere attività computazionali che, di norma, richiederebbero la percezione umana per essere affrontate.

L'uso del ML nella medicina cardiovascolare si sta diffondendo sempre di più anche grazie all'espansione dell'uso di segnali di dati grezzi, come l'ECG o l'elaborazione delle immagini per i modelli basati sulle immagini dell'ecocardiogramma (Lopez et al., 2020).

Le reti neurali convoluzionali sono uno strumento efficace nella modellazione di dati grezzi ottenuti dall'ECG; sono utilizzate, ad esempio, per il rilevamento della disfunzione ventricolare sinistra (Attia et al., 2019), della cardiomiopatia ipertrofica (Ko et al., 2019) e della fibrillazione atriale silente da un ECG con un normale ritmo sinusale (Attia et al., 2019).

L'algoritmo per la disfunzione ventricolare sinistra è stato convalidato è stato addestrato e validato in contesti clinici e non clinici e in sottogruppi etnici e razziali, ed è stato associato ad un miglior risultato in uno studio clinico programmato (Yao et al., 2021).

L'architettura di rete di questo modello è specifica per estrapolare modelli ripetuti dagli ECG attraverso kernel convoluzionali e combinarli utilizzando reti neurali connesse.

Il kernel convoluzionale lo si può intendere come una sorta di lente d'ingrandimento che scorre l'ECG alla ricerca di corrispondenze; se l'input corrisponde al kernel convoluzionale verrà generato un grande impulso numerico.

Attraverso l'addestramento, gli algoritmi apprendono più kernel diversi e gli impulsi risultanti consentono poi la classificazione degli ECG in categorie discrete come, ad esempio, la disfunzione ventricolare sinistra.

Uno dei vantaggi di questo modello è che l'analista deve fornire solo la struttura di base, e l'algoritmo apprende i kernel con poche specifiche umane: questo, comunque, non significa che il processo sia automatico perché è necessaria la specifica giudiziosa dell'architettura.

In queste condizioni è da scongiurare il rischio di ottenere un modello finale non generalizzabile o soggetto a sovrallenamento perché in questo caso si avrebbe uno scarso compromesso tra bias e varianza.

Il bias, in questo caso, indica quanto bene le previsioni del modello si allineano con la verità, la varianza misura la robustezza del modello in relazione a piccole differenze nei dati.

Una delle sfide nello sviluppo di modelli statistici comuni è l'adattamento a ciò che è considerato ottimismo del modello; se i dati cambiano, i risultati del modello dovrebbero cambiare e, per questo, gli studi di convalida sono una parte essenziale del processo di sviluppo (Iba et al., 2021).

Quei modelli basati sul ML programmati per apprendere tutte le caratteristiche uniche in un set di dati corrono il rischio di pregiudizi nelle previsioni, mentre un quadro di modelli più complessi permette un adattamento personalizzato; il rischio è che le funzionalità apprese non si generalizzino a sei di dati aggiuntivi, e ciò è sottovalutato dal quadro statistico più comune.

Oltre a garantire una rappresentazione equilibrata tra le popolazioni di formazione e test, un approccio chiave per migliorare robustezza delle prestazioni del modello è considerare e rappresentare il modo in cui vengono prese le decisioni cliniche: le decisioni si basano su prove e giudizi (Adedinsewo et al., 2022).

La diagnosi precoce delle cardiopatie congenite può avere un impatto positivo sull'esito clinico, e l'anamnesi familiare e i test cardiovascolari sono strumenti essenziali per una buona riuscita (Peters et al., 2020).

Un algoritmo di deep learning basato sull'ECG è risultato efficace nell'identificare la cardiomiopatia ipertrofica nei bambini al di sotto dei 18 anni, con l'area sotto la curva di 0.98 (Siontis et al., 2021) e questi strumenti potrebbero essere fondamentali nello screening in pazienti di età inferiore ai 35 anni, per la quale la morte improvvisa è una grande preoccupazione (Harmon et al., 2015).

Molti giovani pazienti hanno contatti limitati con l'assistenza medica e potrebbero quindi non essere a conoscenza di un eventuale rischio di CVD attuale o futuro a cui sono soggetti; i modelli di Intelligenza Artificiale possono affrontare questo problema con il *cognitive computing*, come il riconoscimento di specifici pattern e l'elaborazione del linguaggio naturale per poter facilitare la cura individualizzata del paziente (Krittanawong et al., 2017).

L'analisi dei fonocardiogrammi, la registrazione dei suoni cardiaci all'auscultazione, basata sull'IA è risultata valida per l'individuazione di cardiopatie congenite e valvolari (Alkhodari et al., 2021).

Quando i fattori di rischio comuni associati alle CVD - ipertensione, obesità e ridotta glicemia a digiuno - si manifestano nell'infanzia e nella prima età adulta, l'impatto sulla salute cardiovascolare è molto più marcato; il riconoscimento precoce, il monitoraggio e il trattamento di questi pericoli può migliorare la qualità di vita e la salute a lungo termine, specialmente per le donne (McSweeney et al., 2016).

Considerata l'evidenza di maggiori esiti cardiovascolari associati al tabagismo nelle donne rispetto a quelli rilevati negli uomini, smettere di fumare è un obiettivo fondamentale per le donne di tutte le età: tuttavia anche bambini e giovani adulti sono estremamente vulnerabili a causa della loro suscettibilità all'uso delle sigarette elettroniche che sono note per causare danni soprattutto durante la gravidanza o nei contesti di malattie delle vie aeree preesistenti (Kavousi et al., 2020).

La gravidanza e il peripartum sono stati identificati come i periodi ideali per valutare i fattori di rischio e i comportamenti delle donne e quindi ridurre il rischio cardiovascolare futuro; i fattori di rischio preesistenti, se non rilevati e trattati,

determinano un aumento del rischio di eventi avversi durante la gravidanza (Smith et al., 2019).

Anche quando la donna non presenta fattori di rischio precedentemente alla gravidanza, i cambiamenti fisiologici ad essa dovuti come la resistenza all'insulina, la sovraregolazione delle vie infiammatorie e l'alterazione dipendente dall'endotelio nella funzione vascolare possono far emergere importanti rischi.

La cardiomiopatia è difficilmente diagnosticabile in gravidanza a causa della sovrapposizione dei sintomi, rimanendo così un'esigenza insoddisfatta dello screening cardiovascolare delle donne nel periodo della gravidanza (Davis et al., 2020).

Esistono modelli ML specifici per rilevare o prevedere il rischio di disturbi collegati alla gravidanza ed ECG potenziati dall'Intelligenza Artificiale in grado di evidenziare le cardiomiopatie nel periodo postparto: i risultati sono promettenti, ma la validazione esterna risulta ancora limitata (Davidson et al., 2020, Adedinsewo et al., 2021).

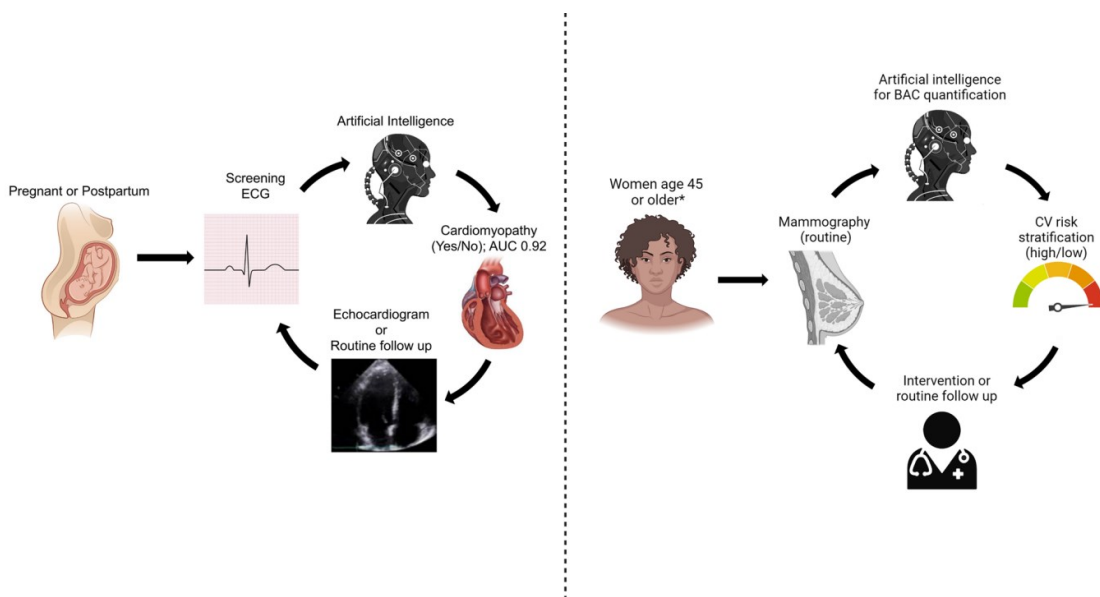


Fig. [10] Esempi di potenziali applicazioni di Intelligenza Artificiale per identificare le malattie cardiovascolari (CVD) nelle donne. (Adedinsewo et al., 2021).

È noto che le donne che vivono un esito negativo di una gravidanza hanno un aumentato rischio di futuri eventi cardiovascolari avversi (Parikh et al., 2021) perché gli aborti spontanei sono essenzialmente manifestazioni di adattamenti fisiologici

anomali ai cambiamenti metabolici e vascolari necessari per una gravidanza in salute (Smith et al., 2019).

I fattori di rischio delle CVD aumentano, sia per gli uomini che per le donne, con l'età. Le strategie di prevenzione sono tuttora sotto prescritte nelle donne rispetto agli uomini con gradi simili di rischio aterosclerotico: considerate le maggiori complessità associate alla valutazione del rischio nelle donne e la mancanza di uno strumento che incorpori i fattori di rischio esclusivi delle donne, è evidente il bisogno di nuove strategie per affrontare e superare il divario nella preparazione dei medici per valutare il rischio cardiovascolare nelle donne (Isakadze et al., 2019).

All'interno della gestione del sistema sanitario, gli algoritmi di IA offrono la possibilità di sfruttare le cartelle cliniche elettroniche per perfezionare i criteri di inclusione ed esclusione per lo screening mirato, in modo tale da sottoporre ai trattamenti solo i pazienti che ne trarrebbero effettivamente benefici.

In questo ambito viene applicata la diagnostica per immagini e i biomarcatori; le calcificazioni arteriose mammarie osservate durante le mammografie di routine sono state associate al rischio di morte e ad esiti cardiovascolari, e le reti neurali ne consentono il rilevamento automatico e la quantificazione (Trimboli et al., 2019).

Per quanto riguarda le modalità di imaging focalizzate sulle malattie cardiovascolari, l'IA può anche aumentare la qualità e la quantità di informazioni che possono essere ottenute per affinare l'efficacia e accuratezza degli sforzi di screening: ad esempio, l'analisi dell'imaging tomografico computerizzato del tipo e della posizione delle placche calcificate, aggiunta alle variabili cliniche, è un predittore di mortalità migliore rispetto alle metriche cliniche esistenti (Rogers et al., 2019).

Un sempre più crescente numero di studi e di evidenze supporta l'uso dell'Intelligenza Artificiale in tutte le modalità di imaging cardiaco per l'acquisizione e l'interpretazione dei dati e per lo sviluppo degli algoritmi di previsione del rischio (Lin et al., 2020).

Dato che le donne hanno maggiori probabilità di presentare fattori di rischio cardiovascolare non tradizionali, nonché caratteristiche uniche specifiche date dal sesso nell'imaging, è probabile che i modelli, integrando immagini e dati clinici, migliorino la valutazione di

rischio cardiovascolare nelle donne; è probabile che i modelli così intesi siano particolarmente efficaci per le donne (Alaa et al., 2019).

I progressi negli strumenti digitali per l'acquisizione dei dati e negli algoritmi di Intelligenza Artificiale per l'analisi possono essere utilizzati per stratificare il rischio nella prevenzione secondaria e potenzialmente abbinare i profili di rischio individuali con interventi appropriati.

Le strategie mediche basate sull'Apprendimento Automatico e l'Intelligenza Artificiale possono persino aumentare l'accesso all'assistenza alle donne e agli uomini appartenenti a gruppi razziali ed etnici tradizionalmente svantaggiati, in quanto anche in luoghi non presidiati da strutture mediche all'avanguardia si potrebbero effettuare diagnosi basate sulle conoscenze ottenute e sviluppate in altre parti del mondo (Yuan et al., 2021).

In seguito alla diagnosi di un problema cardiovascolare, le differenze negli esiti della malattia sono correlate alle disparità delle terapie orientate alle linee guida (Peters et al., 2018).

Usufruire delle potenzialità dell'Intelligenza Artificiale interfacciata con i sistemi di cartelle cliniche elettroniche sviluppando algoritmi e strumenti di elaborazione del linguaggio naturale, può portare alla rapida valutazione delle più appropriate prescrizioni di terapie mediche ottimali per le donne migliorando la conformità dei farmaci e interagendo con i pazienti tramite applicazioni basate sull'Intelligenza Artificiale.

Un'altra straordinaria potenzialità di intervento è la riabilitazione cardiaca, e nonostante sia chiaro che porti vantaggio in termini di mortalità per gli individui che si sottopongono a questi trattamenti, le donne hanno meno probabilità di essere indirizzate, partecipare e completare tale strategia rispetto agli uomini (Ritchey et al., 2020).

L'esigenza di parità di trattamento delle malattie cardiovascolari tra uomini e donne trova compimento, ad esempio, nelle cliniche specializzate nella salute del cuore dove i vantaggi delle tecnologie digitali si integrano con la formazione specialistica (Garcia et al., 2016).

L'applicazione dell'IA e degli strumenti digitali ha un enorme potenziale per incrementare la raccolta di fattori specifici per lo studio delle CVD in relazione al sesso e al genere, catalizzando questi dati in linee guida di pratica clinica cardiovascolare. I fattori di rischio sesso-specifici per le malattie cardiovascolari, come la storia cardio-ostetrica comprensiva della menopausa, parto pretermine o disturbi ipertensivi gestazionali o diabetici (che possono aumentare il rischio di CVD in una fase successiva della vita di una donna), non sono rilevati in modo coerente dal *Global Burden of Disease collaborative Network Study* che ha permesso la generazione del rapporto della *Lancet Commission* sulle malattie cardiovascolari e le donne (Vogel et al., 2021).

È fondamentale che tali studi comprendano un'analisi dei dati disaggregati per sesso includendo anche fattori socioculturali, di genere, geografici e generazionali fornendo informazioni che risultino immediatamente applicabili nella pratica; raccogliendo ed analizzando tali dati in modo appropriato si potrebbero sviluppare algoritmi di gestione specifici per sesso e genere contribuendo a migliorare la prevenzione cardiovascolare per le donne.

4.4 MODELLI DI APPRENDIMENTO AUTOMATICO PER LA PREVISIONE DELLE MALATTIE CARDIOVASCOLARI

I modelli predittivi per la valutazione del rischio sono fondamentali per identificare i pazienti ad alto rischio ed attuare una prevenzione precoce; risultano utili anche nell'apprendimento di dati sanitari direttamente dalle cartelle cliniche elettroniche (Johnson et al., 2021).

Bisogna sempre tenere conto che un set di dati utilizzato per addestrare i modelli può comportare bias sistematici e di campionamento, mancanze differenziali o errori di stima statistica attraverso la raccolta e preelaborazione; la preoccupazione è che questi modelli amplifichino il pregiudizio producendo decisioni distorte verso alcuni gruppi di persone, poiché le iniquità prodotte influenzerebbero l'equità sanitaria (Park et al., 2021).

Diversi ricercatori hanno proposto metodi per valutare e mitigare i bias per varie applicazioni che potrebbero influenzare negativamente i gruppi sottorappresentati,

ma la ricerca sistematica sulla distorsione risulta ancora scarsa (McCradden et al., 2020).

Uno studio intitolato *Evaluating and mitigating bias in Machine Learning models for cardiovascular disease prediction* e pubblicato sul *Journal of biomedical informatics* 138 nel 2023 affronta una questione importante nell'uso di modelli di Machine Learning in medicina: la presenza di bias nei modelli di Apprendimento Automatico utilizzati per la predizione delle malattie cardiovascolari.

Descriverò dettagliatamente questo lavoro di ricerca.

Il primo aspetto che l'articolo affronta è la natura dei dati utilizzati per addestrare i modelli di Machine Learning.

Il bias si verifica quando un algoritmo di Apprendimento Automatico produce risultati sbagliati o non equilibrati a causa di un pregiudizio incorporato nel set di dati di addestramento e questo può portare ad una ridotta accuratezza nella previsione delle malattie cardiovascolari e ad una mancata identificazione dei fattori di rischio importanti.

In molti casi questi dati possono essere pregiudiziali, ad esempio perché non riflettono la diversità della popolazione che si vuole servire: ad esempio, se un dataset utilizzato per addestrare un modello di previsione della malattia cardiovascolare contiene principalmente dati di pazienti maschi bianchi, il modello potrebbe avere difficoltà a fare previsioni accurate per le donne o per le persone di altre etnie.

Per mitigare questo problema gli autori suggeriscono di utilizzare dati più rappresentativi della popolazione che si vuole servire cercando, ad esempio, di includere dati su pazienti di diverse etnie e di entrambi i sessi prestando attenzione alla selezione dei dati e assicurarsi che non ci siano pregiudizi o errori nei dati utilizzati per addestrare il modello.

L'articolo affronta anche il rischio di pregiudizio nei modelli di Machine Learning: anche se i dati utilizzati per addestrare il modello sono rappresentativi della popolazione, il modello stesso potrebbe ancora essere pregiudiziale.

Un modello potrebbe essere addestrato a dare maggiore peso ai fattori che influenzano la malattia cardiovascolare nei pazienti maschi rispetto alle donne, anche se i dati di addestramento includono pazienti di entrambi i sessi.

Per mitigare questo problema, l'articolo suggerisce di esaminare attentamente i modelli di Machine Learning e di identificare eventuali pregiudizi o errori valutando i modelli utilizzando metriche di performance adeguate che tengano conto della diversità della popolazione che si vuole servire.

Il terzo aspetto che l'articolo affronta è l'impatto che i pregiudizi nei modelli di Machine Learning possono avere sulla pratica medica poiché se un modello fosse pregiudiziale, potrebbe portare a diagnosi errate o a decisioni terapeutiche inesatte per alcune popolazioni: un modello di previsione della malattia cardiovascolare pregiudiziale nei confronti delle donne potrebbe portare a diagnosi errate o a sottovalutare i rischi di malattia cardiovascolare in esse.

In questo articolo si discute anche delle tecniche utilizzate per mitigare il bias nei modelli di Apprendimento Automatico.

Una delle tecniche più comuni è la raccolta di dati diversificati, che rappresentano diverse categorie di pazienti; in questo modo si cerca di evitare il bias dovuto ad un campione di dati non rappresentativo. Un'altra tecnica consiste nel rilevare il bias all'interno dei dati di addestramento e nel correggerlo attraverso la riformulazione dei dati di addestramento o attraverso l'aggiunta di nuovi dati di addestramento.

È importante anche valutare il modello di Apprendimento Automatico una volta che è stato addestrato analizzando i risultati per verificare che il modello stia producendo previsioni accurate e non influenzate da bias; si dovrebbero anche valutare i fattori di rischio utilizzati dal modello, per verificare che siano rappresentativi di tutte le categorie di pazienti.

Questo articolo esplora anche alcune delle sfide future che il campo dell'Intelligenza Artificiale dovrà affrontare per ridurre ulteriormente il bias nei modelli di Apprendimento Automatico come l'implementazione di nuove tecniche di addestramento, l'identificazione di nuovi fattori di rischio per le malattie cardiovascolari e la valutazione continua dei modelli di Apprendimento Automatico per garantire che rimangano accurati e non influenzati da bias.

4.5 BIAS

La sfida principale dell'applicazione delle tecnologie basate sull'Intelligenza Artificiale applicate alla medicina cardiovascolare è evitare i pregiudizi intrinseci ed estrinseci che influenzano la qualità e gli esiti degli studi clinici.

La prima fonte di eterogeneità nella qualità degli esami diagnostici e delle modalità di attuazione si riscontra nelle tecnologie che forniscono immagini, come ad esempio ecografie e risonanze magnetiche, sono soggette sia all'operatore che le studia che all'interprete che analizza poi i risultati; le informazioni che l'Intelligenza Artificiale ricava da questi dati dipende quindi dalle interpretazioni umane (Lopez-Jiminez et al., 2020).

Recenti studi che utilizzano l'IA nell'imaging cardiaco e nella diagnostica mostrano enormi potenzialità, ma bisogna considerare che potrebbero valere limitatamente alle dimensioni del campione analizzato (Rabbat et al., 2020).

Nei modelli di Apprendimento Automatico in cui gli algoritmi sono progettati per riconoscere i modelli, si osserva la tendenza a sovradimensionare il set di dati a causa dell'incapacità di distinguere un fatto di contributo effettivo da uno fasullo; un modello sovradimensionato può mostrare una valida performance nei dati forniti in allenamento, ma non avere pari efficacia se alle prese con un nuovo set di dati (Kagiyama et al., 2019).

L'elaborazione dei dati richiede un costante intervento del medico preposto, che deve annotare regolarmente per contribuire al controllo qualità: c'è però il rischio che l'annotatore trasferisca il proprio pregiudizio nell'algoritmo (Kagiyama et al., 2019).

Un'altra preoccupazione per quanto riguarda l'uso di modelli predittivi basati sull'IA è il rischio di peggiorare le disparità di salute attraverso pregiudizi: le donne e i gruppi minoritari sono storicamente sottorappresentati in cardiologia, e la maggior parte della medicina basata sull'evidenza potrebbe non essere applicata correttamente a queste popolazioni (Tahhan et al., 2020).

Fermo restando che la stessa percentuale di uomini e donne possano presentare dolore toracico, gli individui di sesso maschile hanno, rispetto alle donne, più del

doppio di probabilità di essere indirizzati ad un cardiologo per la gestione del caso (Clerc et al., 2018).

Uno studio condotto da Obermeyer ha evidenziato che l'uso di un algoritmo di previsione commerciale ampiamente diffuso comporta un significativo pregiudizio razziale nella previsione dei risultati (Obermeyer et al., 2019).

L'algoritmo ha identificato i pazienti bianchi con punteggio di rischio più elevato e sono quindi stati selezionati più spesso per ricevere cure aggiuntive rispetto ai pazienti neri altrettanto malati.

In uno studio di Nordling un algoritmo di Apprendimento Automatico ha identificato il codice postale della residenza del paziente come il principale predittore per definire la degenza ospedaliera prolungata; se questo programma fosse progettato per ottimizzare le risorse ospedaliere, i pazienti bianchi ad alto reddito potrebbero essere selezionati come destinatari di maggiori risorse ospedaliere, amplificando ulteriormente il divario nell'accesso alle cure per i gruppi minoritari e sottoserviti (Nordling, 2019).

Per garantire che i modelli predittivi rappresentino l'intera popolazione è fondamentale il coinvolgimento di medici, etici e rappresentanti dei gruppi minoritari durante la fase di sviluppo; nella fase di test gli algoritmi dovrebbero essere convalidati anche su gruppi minoritari e a basso reddito.

Nella fase in cui gli algoritmi vengono integrati nell'ambiente sanitario risultano necessarie linee guida, normativa e norme etiche per garantire che i benefici apportati dall'Intelligenza Artificiale siano equamente distribuiti (Tat et al., 2020).

4.6 BIAS CONTESTUALE NEI SISTEMI DI INTELLIGENZA ARTIFICIALE

L'Intelligenza Artificiale ha il potenziale adatto per rivoluzionare la medicina cardiovascolare sfruttando i dati provenienti da studi clinici, cartelle cliniche o dati estrapolati da sensori, ma ci sono preoccupazioni relative alla probabilità di bias nelle previsioni per le donne o le minoranze svantaggiate.

Questi gruppi sono sottorappresentati e la maggior parte delle evidenze cliniche esclude queste categorie di persone (Tat et al., 2020).

Alcuni studi hanno evidenziato il potenziale ruolo dell'IA nel propagare pregiudizi razziali e di genere, rafforzare le disuguaglianze e amplificare esponenzialmente questi schemi, chiamati *bias algoritmici* (Obermeyer et al., 2021).

Joy Buolamwini, ricercatrice presso il Massachusetts Institute of Technology, ha valutato gli algoritmi di riconoscimento facciale disponibili in commercio (Microsoft, Face++ e IBM) dimostrando come questi siano distorti con tassi di errore che arrivano al 35% nelle donne di carnagione scura (Buolamwini et al., 2021).

Le raccomandazioni per affrontare i pregiudizi negli algoritmi di IA comprendono la garanzia che le donne e le minoranze siano ben rappresentate nei set di dati, test rigorosi e valutazioni approfondite dei rischi: questa è definita *Intelligenza Artificiale responsabile* (West et al., 2021). L'organizzazione mondiale della sanità ha istituito un gruppo di lavoro con lo scopo di sviluppare dei parametri standard di valutazione per i modelli di Intelligenza Artificiale applicati alla medicina e diversi studi hanno identificato l'accesso agli strumenti digitali e all'alfabetizzazione digitale quali determinanti sociale della salute (Siek et al., 2021).

La figura seguente riassume i principi fondamentali dell'IA.

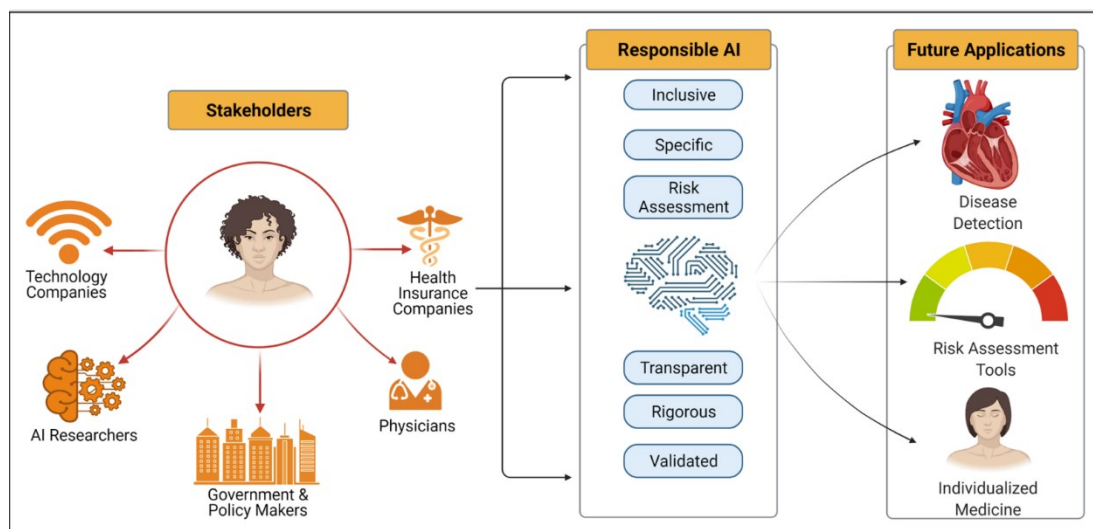


Fig. [11] Intelligenza Artificiale responsabile e principi fondamentali (Researchgate.net)

4.7 LA SOTTOVALUTAZIONE DEL GENERE NEI SISTEMI DI MACHINE LEARNING IN CARDIOLOGIA

L'utilizzo di algoritmi di Machine Learning (ML) nella cardiologia ha portato a importanti progressi nella diagnosi e nella prevenzione delle malattie cardiovascolari. Tuttavia, l'uso di questi sistemi può anche portare a bias di genere, ovvero a discriminazione basata sul genere dei pazienti.

Uno studio pubblicato nel 2020 su *JAMA Cardiology*, *Artificial Intelligence and sex bias in cardiovascular disease risk prediction*, ha esaminato il rischio di bias di genere nei sistemi di Intelligenza Artificiale utilizzati per prevedere il rischio di malattie cardiovascolari (Nandakumar et al., 2020).

Sono stati analizzati quattro studi che utilizzavano algoritmi di ML per prevedere il rischio di malattie cardiovascolari in pazienti maschi e femmine e hanno scoperto che solo il 17% dei sistemi di ML ha considerato il genere come variabile di input e solo il 4% come variabile di output.

Inoltre, i sistemi di ML erano meno precisi nella previsione del rischio per le donne rispetto agli uomini.

Gli autori suggeriscono che è necessario sviluppare sistemi di ML che considerino il genere come variabile per migliorare la precisione nella previsione del rischio di malattie cardiovascolari.

Un altro studio pubblicato nel 2016 su *JAMA Cardiology*, *Association of sex with reclassification of coronary heart disease risk using coronary artery calcium and other risk factors*, ha esaminato il ruolo del genere nella previsione del rischio di malattie cardiovascolari utilizzando il calcio nelle arterie coronarie come fattore di rischio (Blaha et al., 2016).

Gli autori hanno analizzato dati di 6.698 pazienti maschi e femmine per valutare se il genere influenzasse la previsione del rischio di malattie cardiovascolari utilizzando il calcio nelle arterie coronarie come fattore di rischio.

Hanno scoperto che i pazienti maschi presentavano un rischio maggiore di malattie cardiovascolari rispetto alle donne con lo stesso livello di calcio nelle arterie

coronarie; hanno suggerito che le attuali scale potrebbero sottostimare il rischio cardiovascolare delle donne.

Questi studi dimostrano che i sistemi di ML utilizzati nella cardiologia possono essere soggetti a bias di genere, con conseguenze negative sulla precisione nella previsione del rischio di malattie cardiovascolari per le donne ed è pertanto necessario sviluppare sistemi di ML che considerino il genere come variabile di input e di output per garantire una diagnosi e una prevenzione corrette e precise delle malattie cardiovascolari per entrambi i sessi.

4.8 DIVARIO DI GENERE PERMANENTE NEL RISCHIO DI INFARTO MIOCARDICO

A causa dei pochi studi che considerano le differenze di sesso presenti in letteratura non è chiaro in che misura la maggior incidenza di malattia coronarica (CHD) negli uomini rispetto alle donne è spiegata da fattori di rischio dissimili (Albreksten et al., 2016).

I primi rapporti basati sul confronto dei tassi di mortalità grezzi standardizzati per età per cardiopatia ischemica hanno rilevato tassi da 3 a 4 volte più alti negli uomini rispetto alle donne (Uemura et al., 1988) e non era chiaro se questo contrasto fosse dovuto ad una diversa esposizione a fattori di rischio accertati; già uno studio del 1997 (Barrett-Connor, 1997) concluse che il divario di genere non poteva essere spiegato da differenze nei livelli di fattori di rischio noti.

La scarsità di stime del rischio calibrate sul sesso si riferisce in parte al fatto che gli studi precedenti hanno esaminato principalmente gli uomini poiché quelli che includevano le donne si concentravano sugli effetti specifici del sesso sui fattori di rischio di CHD (Yahagi et al., 2015).

Nonostante le limitate prove empiriche, sono state proposte diverse ipotesi sui meccanismi biologici per spiegare una potenziale differenza di sesso capace di influenzare il rischio di malattia coronarica, e si suggerisce che il più basso rischio di CHD nelle donne in premenopausa sia correlato ad un effetto protettivo degli ormoni endogeni femminili (Yahagi et al., 2015).

Nella tabella seguente sono indicate le stime del rischio specifico per sesso ed età di infarto miocardico.

Lo studio di Albreksten e dei suoi colleghi mostra che il rischio di infarto miocardico aumenta con l'età in entrambi i sessi, seppur rimanendo più basso per le donne.

Il divario di genere persiste per tutta la vita, ma mentre le stime di rischio relativo diminuiscono con l'età, le differenze assolute di rischio aumentano.

Complessivamente, gli uomini hanno circa il doppio del rischio di IM rispetto alle donne, un contrasto che non può essere spiegato dalle differenze nei livelli di altri fattori di rischio di CHD.

Gli uomini hanno circa il doppio del rischio di infarto miocardico rispetto alle donne e questo contrasto non può essere spiegato da fattori di rischio CHD accertati.

La differenza di sesso persiste per tutta la vita ma diminuisce con l'età, e i piccoli cambiamenti negli indici di rischio quando ci si avvicina alla menopausa rende improbabile che i cambiamenti nei livelli ormonali femminili influenzino il rischio di infarto del miocardio. Una migliore comprensione dei meccanismi alla base del divario di genere, riassunto nella tabella seguente, osservato nel rischio di IM è importante sia per la prevenzione che per il trattamento della malattia sia nelle donne che negli uomini.

Age, y	Men			Women			Men vs Women	
	Incident MI, No.	Person-years	IR ^a	Incident MI, No.	Person-years	IR ^a	IRD ^b	IRR ^c
35-39	29	44 519.6	6.5	2	48 241.0	0.4	6.1	15.7
40-44	79	50 865.3	15.5	9	53 670.2	1.7	13.9	9.3
45-49	173	49 456.5	35.0	23	51 248.4	4.5	30.4	7.8
50-54	261	44 296.2	58.9	47	43 576.0	10.8	48.1	5.5
55-59	310	36 624.4	84.6	74	33 653.1	22.0	62.7	3.8
60-64	281	26 305.8	106.8	84	23 994.7	35.0	71.8	3.1
65-69	190	16 506.4	115.1	78	16 515.9	47.2	67.9	2.4
70-74	182	10 362.6	175.6	119	12 976.4	91.7	83.9	1.9
75-79	165	7 060.1	233.7	137	10 570.3	129.6	104.1	1.8
80-84	141	4 035.7	349.4	140	7 778.0	180.0	169.4	1.9
85-89	74	1 679.2	440.7	104	4 263.8	243.9	196.8	1.8
90-94	20	391.0	511.5	54	1 512.6	357.0	154.5	1.4
≥95	2	57.1	350.0	15	291.3	515.0	-165.0	0.7

Fig. [12] Rischio di CVD in base all'età dei pazienti. (Researchgate.net)

4.9 IMPLICAZIONI SULLA SALUTE PUBBLICA

L'ambiente dei big data è in continua crescita e supera la capacità di interpretazione cognitiva umana: Intelligenza Artificiale e Machine Learning sono tecniche utilizzate per gestire set di dati creando algoritmi che stabiliscono connessioni tra elementi di dati apparentemente disparati.

L'espansione delle applicazioni di IA e ML suscita molto entusiasmo ma, allo stesso tempo, preoccupa per il rigore scientifico, l'equità, la privacy e l'uso appropriato dei dati: la vera sfida è costruire una base solida in modo che i risultati possano essere utilizzati in modo diffuso e sicuro.

Al fine di migliorare la salute della popolazione è necessario ottimizzare la comprensione dei modelli integrando i dati spaziando dai biomarcatori fino ai fattori ambientali (Khoury et al., 2018).

Poiché la sanità pubblica è radicata in un ecosistema di big data, i sistemi basati sull'Intelligenza Artificiale richiedono ampio accesso a set di dati su larga scala e fonti di informazioni affidabili e pertinenti; per garantire applicazioni eque è necessario un impegno esplicito per la convalida degli algoritmi e il controllo dei pregiudizi sviluppati dall'IA.

La sorveglianza della salute della popolazione è un prerequisito fondamentale per attuare interventi efficaci e l'utilizzo di nuovi dati digitalizzati sarà una pratica che si espanderà sempre di più rendendo ancora più sicura e precisa la gestione delle CVD. Un report del 2019 ha sintetizzato gli approcci pragmatici che vengono utilizzati per affrontare e superare le sfide dell'applicazione dell'Intelligenza Artificiale negli algoritmi utilizzati per l'assistenza sanitaria (Matheny et al., 2019).

La tabella [13] riassume le raccomandazioni principali che sono emerse nel rapporto della National Academy of Medicine.

È indubbio che non sarà facile adottare tutte le raccomandazioni fornite dalla National Academy of Medicine, e addirittura si pensa che "L'adozione dell'Intelligenza Artificiale nell'assistenza sanitaria sarà lenta e difficile" (Kuan, 2019) e non bisogna dimenticare, visto l'impatto sulla salute pubblica e sulla medicina clinica, che la regolamentazione è fondamentale (Adedinsewo et al., 2022).

Promoting population-representative data with accessibility, standardization, and quality
Contextualizing the dialogue of transparency and trust requires accepting differential needs
Prioritize ethical, equitable, and inclusive health care AI while addressing explicit and implicit bias
Near-term focus is needed on augmented intelligence vs AI autonomous agents
Develop and deploy appropriate training and educational programs to support health care AI
Leverage frameworks and best practices for learning health care systems, human factors, and implementation science to address the challenges in operationalizing health care AI
Balance innovation with safety via regulation and legislation to promote trust

Fig. [13] Raccomandazioni per applicazioni valide e sicure (Matheny et al., 2019)

Il successo degli ultimi anni delle applicazioni basate sull'Intelligenza Artificiale e sul Machine Learning è stato possibile soprattutto grazie ai progressi nelle tecnologie informatiche e alla crescita esponenziale della quantità di dati digitali disponibili.

Il campo cardiologico è stato protagonista di un'enorme crescita forse anche perché le decisioni relative ai trattamenti applicati in questo settore si basano su molteplici dati digitalizzati e test diagnostici dei pazienti (Quer et al., 2021).

Gli algoritmi di Intelligenza Artificiale possono sintetizzare e interpretare rapidamente grandi quantità di dati clinici in maniera molto più performante di quanto sia possibile basandosi sulle sole capacità umane.

Appurato che le donne sono soggette a molteplici malattie cardiovascolari e, allo stesso tempo, hanno minori possibilità di ricevere una diagnosi tempestiva e quindi trattamenti adeguati (Vogel et al., 2021), l'Intelligenza Artificiale e gli strumenti digitali ad essa collegata offrono la possibilità di modificare l'attuale situazione dell'assistenza cardiovascolare nelle donne.

È indubbio che non sarà facile adottare tutte le raccomandazioni fornite dalla National Academy of Medicine, e addirittura si pensa che "L'adozione dell'Intelligenza Artificiale nell'assistenza sanitaria sarà lenta e difficile" (Kuan, 2019) e non bisogna dimenticare, visto l'impatto sulla salute pubblica e sulla medicina clinica, che la regolamentazione è fondamentale (Adedinsewo et al., 2022).

Queste tecnologie potrebbero ridurre il divario tra uomini e donne, programmando gli algoritmi in modo che incorporino in modo preciso fattori specifici di sesso e genere (Cirillo et al., 2020).

Nonostante le potenziali opportunità per l'IA nello screening delle CVD screening, sono necessarie ricerche future per esaminare se le tecnologie digitali mantengono la loro promessa di migliorare i risultati specifici per le terapie personalizzate per ciascun paziente riducendo le disparità di genere.

Lo sviluppo e l'implementazione dei supporti decisionali basati sull'IA e degli strumenti incorporati nelle cartelle cliniche elettroniche comportano un costo elevato e l'efficacia di tali strategie deve essere attentamente valutata prima dell'eventuale adozione di queste tecniche innovative (Tseng et al., 2021).

Nella figura [14] sono schematizzate le tecniche di screening che utilizzano l'Intelligenza Artificiale e che sono impiegate per la diagnosi di CVD nelle donne.

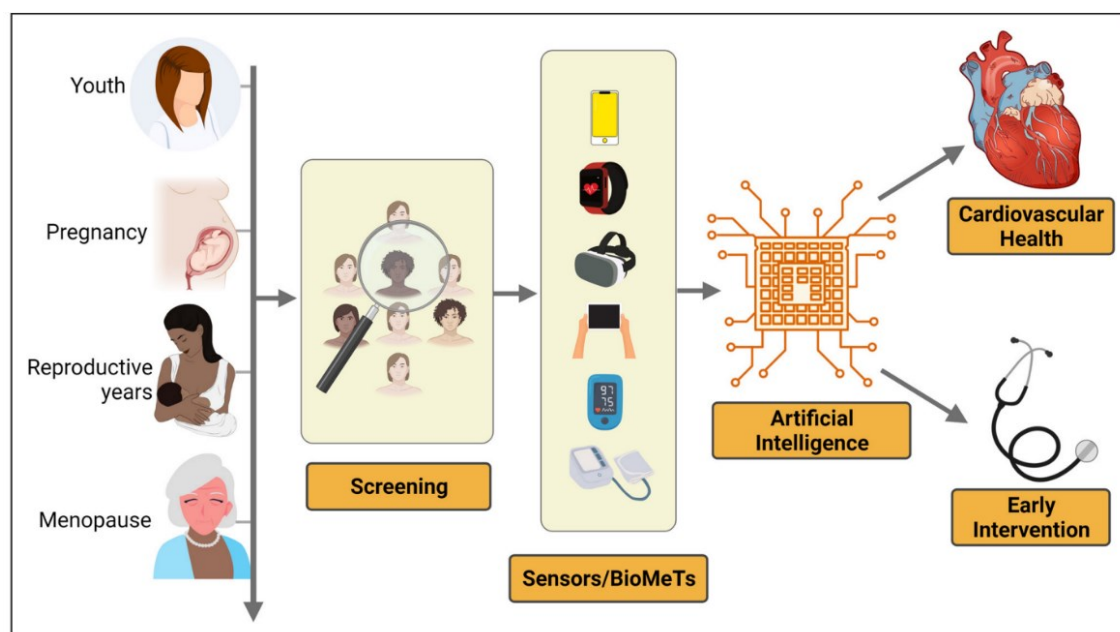


Fig. [14] Applicazioni di IA nello screening delle CVD nelle donne. (Researchgate.net)

5 Conclusioni

La comunicazione in medicina, in particolare l'uso e la traduzione dei termini, è fondamentale.

Sesso e genere vengono confusi e si confondono realmente nell'origine delle differenze, ma usare un termine e dimenticare l'altro non aiuta a comprendere questa visione della medicina così trasversale e integrativa.

I Centri che se ne occupano sono versatili e seguono una linea fortemente legata al contesto in cui lavorano: alcuni operano in favore dell'abolizione del complesso e poliedrico concetto di genere affinché non vi siano filtri interpretativi sul territorio e vi sia una diversificazione tra sesso maschile e femminile nella visione della patologia, e del suo trattamento, solo quando questa è effettivamente diagnosticata.

Le riviste scientifiche pubblicano studi sulle differenze "sex-related" e "gender-related" utilizzando non sempre la terminologia appropriata e pertanto si vuole sensibilizzare i ricercatori a ricordare i termini corretti e ad effettuare le sperimentazioni su animali che siano del sesso del quale si vogliono studiare le patologie e i comportamenti.

Tornare allo scopo originale della "sex and gender medicine", capirne le ragioni e riconsiderarla per quello che è, una ampia e tridimensionale valutazione dei due sessi e dei generi nella più ampia accezione dei termini, può aiutare a cambiare le regole ed abolire i preconcetti.

Nel corso della ricerca presentata in questa tesi, sono state analizzate le cause e gli effetti dei bias di genere dell'IA in medicina, mettendo in luce le conseguenze negative per le donne e le minoranze sottorappresentate nella diagnosi, nel trattamento e nella cura delle malattie.

Inoltre, la tesi ha esaminato le sfide e le opportunità associate alla raccolta di dati rappresentativi e all'addestramento di algoritmi imparziali ed equi, nonché le attuali soluzioni proposte per ridurre i bias di genere dell'IA in medicina.

Si tratta un problema complesso e multifattoriale che richiede un approccio interdisciplinare e un impegno condiviso da parte di tutti gli attori coinvolti, inclusi ingegneri biomedici, medici, ricercatori, responsabili politici e aziende tecnologiche.

Per affrontare questo problema, è necessario promuovere la raccolta di dati rappresentativi e l'addestramento equo dell'IA, nonché sensibilizzare sulle implicazioni etiche e sociali dell'uso dell'IA in medicina.

L'Intelligenza Artificiale è, senza ombra di dubbio, una grande opportunità per migliorare la diagnosi e la cura delle malattie, ma l'uso in campo medico deve essere affrontato con attenzione, poiché possono sorgere dei bias di genere, cioè delle discriminazioni verso un genere, causati dalla natura stessa dell'algoritmo o dal modo in cui i dati vengono raccolti e analizzati.

È importante che le aziende, i ricercatori e i professionisti della salute adottino queste soluzioni e promuovano l'etica e la responsabilità sociale nell'uso dell'IA in medicina, al fine di garantire un futuro più equo ed efficace per la diagnosi e la cura delle malattie.

5.1 PROSPETTIVE FUTURE NELLA PRATICA GENERALE

L'Intelligenza Artificiale potrà estrarre informazioni importanti dall'impronta elettronica di un paziente. Inizialmente questa pratica farà risparmiare tempo e migliorerà l'efficienza, ma dopo un test adeguati guiderà anche direttamente la gestione del paziente.

Man mano che questi sistemi saranno meglio convalidati avranno più responsabilità. Inoltre, poiché l'IA è in grado di monitorare contemporaneamente milioni di input, avrà un ruolo significativo nella medicina preventiva, suggerendo gli interventi nel momento in cui il paziente corre il rischio di sviluppare complicanze.

I sistemi basati sull'Apprendimento Automatico e l'Intelligenza Artificiale porteranno anche competenze diagnostiche specialistiche nelle cure primarie: i pazienti identificati come a basso rischio riceverebbero una rassicurazione immediata, mentre i pazienti ad alto rischio potrebbero sperimentare tempi di attesa di riferimento più bassi in quanto le cliniche riceverebbero solo casi selezionati.

La futura ricerca sull'Intelligenza Artificiale dovrebbe essere diretta verso compiti accuratamente selezionati: l'integrazione di questi sistemi nella pratica clinica richiede la costruzione di una relazione reciprocamente vantaggiosa tra l'IA e i medici, in cui essa offre ai medici una maggiore efficienza ed efficacia in termini di costi; di

contro, i medici offrono all'IA l'esposizione clinica essenziale di cui ha bisogno per apprendere una complessa gestione dei casi clinici.

Uno scenario futuro potrebbe essere quello di utilizzare l'IA per la diagnosi precoce del cancro al seno nelle donne; secondo un articolo pubblicato sulla rivista *Breast Cancer Research* (Amritha e Shanthi, 2022), gli algoritmi di Apprendimento Automatico possono essere utilizzati per analizzare le immagini mammografiche e identificare i tumori al seno con una precisione del 90%.

Inoltre, l'IA può essere utilizzata per migliorare la diagnosi e il trattamento di disturbi mentali come la depressione e l'ansia nelle donne: secondo uno studio pubblicato sulla rivista *Nature Medicine*, l'IA può essere utilizzata per analizzare i dati clinici e biologici delle donne con disturbi mentali e identificare i trattamenti più efficaci in base al loro sesso biologico.

In generale, l'utilizzo dell'IA nella medicina di genere offre la possibilità di personalizzare i trattamenti e migliorare la salute delle donne in modo significativo, ma è importante notare che non deve sostituire la valutazione e il giudizio clinico dei medici, ma deve essere utilizzata come un supporto per migliorare la cura dei pazienti.

Durante tutto il processo sarà fondamentale garantire che l'IA non oscuri il volto umano della medicina perché il più grande ostacolo all'adozione diffusa dell'IA sarà l'esitazione del pubblico ad abbracciare una tecnologia sempre più controversa.

Bibliografia

1. Adedinsewo, Demilade A, Patrick W Johnson, Erika J Douglass, Itzhak Zachi Attia, Sabrina D Phillips, Rohan M Goswami, Mohamad H Yamani, et al. 2021. «Detecting Cardiomyopathies in Pregnancy and the Postpartum Period with an Electrocardiogram-Based Deep Learning Model». *European Heart Journal - Digital Health* 2 (4): 586–96.
<https://doi.org/10.1093/ehjdh/ztab078>.
2. Adedinsewo, Demilade A., Amy W. Pollak, Sabrina D. Phillips, Taryn L. Smith, Anna Svatikova, Sharonne N. Hayes, Sharon L. Mulvagh, et al. 2022. «Cardiovascular Disease Screening in Women: Leveraging Artificial Intelligence and Digital Tools». *Circulation Research* 130 (4): 673–90.
<https://doi.org/10.1161/CIRCRESAHA.121.319876>.
3. Adedinsewo, Demilade, Rickey E. Carter, Zachi Attia, Patrick Johnson, Anthony H. Kashou, Jennifer L. Dugan, Michael Albus, et al. 2020. «Artificial Intelligence-Enabled ECG Algorithm to Identify Patients With Left Ventricular Systolic Dysfunction Presenting to the Emergency Department With Dyspnea». *Circulation: Arrhythmia and Electrophysiology* 13 (8): e008437. <https://doi.org/10.1161/CIRCEP.120.008437>.
4. Albert, Kendra, e Maggie Delano. 2022. «Sex Trouble: Sex/Gender Slippage, Sex Confusion, and Sex Obsession in Machine Learning Using Electronic Health Records». *Patterns* 3 (8): 100534.
<https://doi.org/10.1016/j.patter.2022.100534>.
5. Albrektsen, Grethe, Ivar Heuch, Maja-Lisa Løchen, Dag Steinar Thelle, Tom Wilsgaard, Inger Njølstad, e Kaare Harald Bønaa. 2016. «Lifelong Gender Gap in Risk of Incident Myocardial Infarction: The Tromsø Study». *JAMA Internal Medicine* 176 (11): 1673.
<https://doi.org/10.1001/jamainternmed.2016.5451>.
6. Badaloni Silvana e Francesca A. Lisi., 2020, “Towards a Gendered

Innovation in AI (short paper).”, pubblicato in DP@AI*IA 2020. <https://ceur-ws.org/Vol-3319/invited1.pdf>

7. Badaloni, Silvana, e Antonio Rodà. 2022 «Gender Knowledge and Artificial Intelligence». <https://ceur-ws.org/Vol-3319/paper12.pdf>
8. Baggio, Giovannella. 2015. «Dalla medicina di genere alla medicina genere-specifica». *Ital J Gender-Specific Med* 2015 1 (1): 3–5.
9. Buch, Varun H, Irfan Ahmed, e Mahiben Maruthappu. 2018a. «Artificial Intelligence in Medicine: Current Trends and Future Possibilities». *British Journal of General Practice* 68 (668): 143–44. <https://doi.org/10.3399/bjgp18X695213>.
10. Buch, Varun H, Irfan Ahmed, e Mahiben Maruthappu. 2018 «Future Possibilities». *British Journal of General Practice* 68 (668): 143–44. <https://doi.org/10.3399/bjgp18X695213>.
11. Commissione Europea, La sperimentazione clinica di medicinali nell'Unione Europea: analisi degli aspetti di genere (Bruxelles: Ufficio delle pubblicazioni dell'Unione europea, 2010).
12. Calise, Fulvio, Gaya Spolverato, e Micaela Piccoli. 2021a. «Gender Gap or Gender Bias? That Is the Question». *Updates in Surgery* 73 (1): 3–5. <https://doi.org/10.1007/s13304-021-00985-w>.
13. Cirillo, Davide, Silvina Catuara-Solarz, Czuee Morey, Emre Guney, Laia Subirats, Simona Mellino, Annalisa Gigante, et al. 2020. «Sex and Gender Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare». *Npj Digital Medicine* 3 (1): 81. <https://doi.org/10.1038/s41746-020-0288-5>.
14. «Combi C, Keravnou-Papailiou E, Shahar Y. «Temporal information systems in medicine», Springer Science & Business Media, 2010, pp. 301-350

15. Connelly, Paul J., Zahra Azizi, Pouria Alipour, Christian Delles, Louise Pilote, e Valeria Raparelli. 2021. «The Importance of Gender to Understand Sex Differences in Cardiovascular Disease». *Canadian Journal of Cardiology* 37 (5): 699–710. <https://doi.org/10.1016/j.cjca.2021.02.005>.

16. Ebersole. 2019. «Prediction of Cardiovascular Disease Risk Factors Using Machine Learning: An Analysis of NHANES Data». doi: 10.1109/EMBC46164.2021.9630119. PMID: 34891625.

17. Ertel, Wolfgang. 2017. *Introduction to Artificial Intelligence*. Undergraduate Topics in Computer Science. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-58487-4>.

18. Fosch-Villaronga, Eduard, Hadassah Drukarch, Pranav Khanna, Tessa Verhoef, e Bart Custers. 2022. «Accounting for Diversity in AI for Medicine». *Computer Law & Security Review* 47 (novembre): 105735. <https://doi.org/10.1016/j.clsr.2022.105735>.

19. Fossa, Fabio, Viola Schiaffonati, e Guglielmo Tamburrini. 2021. *Automi e persone Introduzione all'etica dell'Intelligenza Artificiale e della robotica*. Carocci Editore.

20. *GE 2018: 2018 ACM/IEEE 1st International Workshop on Gender Equality in Software Engineering : proceedings : 28 May 2018, Gothenburg, Sweden*. 2018. New York, New York], [Los Alamitos, California: The Association for Computing Machinery ; IEEE Computer Society.

21. Ginsburg GS, Wu RR, Orlando LA. Family health history: underused for actionable risk assessment. *Lancet*. 2019 Aug 17;394(10198):596-603. doi: 10.1016/S0140-6736(19)31275-9. Epub 2019 Aug 5. PMID: 31395442; PMCID: PMC6822265.

22. Grego, Susanna, Elena Pasotti, Tiziano Moccetti, e Aldo P Maggioni. 2020 «“Sex and gender medicine”»: il principio della medicina di genere», *Giornale italiano di cardiologia*, vol. 21, n.8
23. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? *Heart*. 2018 Jul;104(14):1156-1164. doi: 10.1136/heartjnl-2017-311198. Epub 2018 Jan 19. PMID: 29352006.
24. Holzinger, Andreas, Chris Biemann, Constantinos S. Pattichis, e Douglas B. Kell. 2017. «What Do We Need to Build Explainable AI Systems for the Medical Domain?» arXiv. <http://arxiv.org/abs/1712.09923>.
25. Huang, Junming, Alexander J. Gates, Roberta Sinatra, e Albert-László Barabási. 2020. «Historical Comparison of Gender Inequality in Scientific Careers across Countries and Disciplines». *Proceedings of the National Academy of Sciences* 117 (9): 4609–16. <https://doi.org/10.1073/pnas.1914221117>.
26. Kagiya, Nobuyuki, Sirish Shrestha, Peter D. Farjo, e Partho P. Sengupta. 2019. «Artificial Intelligence: Practical Primer for Clinical Research in Cardiovascular Disease». *Journal of the American Heart Association* 8 (17): e012788. <https://doi.org/10.1161/JAHA.119.012788>.
27. Karatzia, Loucia, Aung Nay, e Dunja. 2022. «Artificial Intelligence in Cardiology: Hoper for the Future and Power for the Present». *Natoinal Library of Medicine*
28. Khan, Nadia A. 2013. «Sex Differences in Acute Coronary Syndrome Symptom Presentation in Young Patients». *JAMA Internal Medicine*, settembre. <https://doi.org/10.1001/jamainternmed.2013.10149>.
29. Kidambi, Ananth. 2021. «Artificial Intelligence in Medicine», Academic Press, cap. 19 «Artificial Intelligence in Cardiovascular Imaging», pag. 383-

30. Kormilitzin, Andrey, Nenad Tomasev, Kevin R. McKee, e Dan W. Joyce. 2023. «A Participatory Initiative to Include LGBT+ Voices in AI for Mental Health». *Nature Medicine* 29 (1): 10–11. <https://doi.org/10.1038/s41591-022-02137-y>.
31. Kretsoulas, Catherine, Mary Crea-Arsenio, Harry S Shannon, James L Velianou, e Mita Giacomini. 2016. «Interpreting Angina: Symptoms along a Gender Continuum». *Open Heart* 3 (1): e000376. <https://doi.org/10.1136/openhrt-2015-000376>.
32. Kuan R. 2021. «Adopting AI in Health Care Will Be Slow and Difficult», Harvard Business Review. <https://hbr.org/2019/10/adopting-ai-in-health-care-will-be-slow-and-difficult>
33. Leavy Susan. 2018 «Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning». *IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE)*, pp. 14-16
34. Li, Fuchen, Patrick Wu, Henry H. Ong, Josh F. Peterson, Wei-Qi Wei, e Juan Zhao. 2023. «Evaluating and Mitigating Bias in Machine Learning Models for Cardiovascular Disease Prediction». *Journal of Biomedical Informatics* 138 (febbraio): 104294. <https://doi.org/10.1016/j.jbi.2023.104294>.
35. Lopez-Jimenez, Francisco, Zachi Attia, Adelaide M. Arruda-Olson, Rickey Carter, Panithaya Chareonthaitawee, Hayan Jouni, Suraj Kapa, et al. 2020a. «Artificial Intelligence in Cardiology: Present and Future». *Mayo Clinic Proceedings* 95 (5): 1015–39. <https://doi.org/10.1016/j.mayocp.2020.01.038>.
36. Malva, Alberto, e Valeria Zurlo. 2019. «La medicina nell'era dell'Intelligenza Artificiale: applicazioni in Medicina Generale». *Intelligenza Artificiale* 26 (4).

37. Marshall, Stuart M., Cole Mathis, Emma Carrick, Graham Keenan, Geoffrey J. T. Cooper, Heather Graham, Matthew Craven, et al. 2021. «Identifying Molecules as Biosignatures with Assembly Theory and Mass Spectrometry». *Nature Communications* 12 (1): 3033. <https://doi.org/10.1038/s41467-021-23258-x>.
38. McCulloch, Warren S., e Walter Pitts. 1943. «A logical calculus of the ideas immanent in nervous activity». *The bulletin of mathematical biophysics* 5 (4): 115–33. <https://doi.org/10.1007/BF02478259>.
39. McSweeney, Jean C., Anne G. Rosenfeld, Willie M. Abel, Lynne T. Braun, Lora E. Burke, Stacie L. Daugherty, Gerald F. Fletcher, et al. 2016. «Preventing and Experiencing Ischemic Heart Disease as a Woman: State of the Science: A Scientific Statement From the American Heart Association». *Circulation* 133 (13): 1302–31. <https://doi.org/10.1161/CIR.0000000000000381>.
40. Minh, Nguyen, e Villemain. 2022. «Artificial Intelligence in the Pediatric Echocardiography Laboratory: Automation, Physiology and Outcomes», *Frontiers in Radiology*, vol. 2/2022, <https://www.frontiersin.org/articles/10.3389/fradi.2022.881777>
41. Obermeyer, Z, Nissan, e Stern. 2021. «Algorithmic bias playbook. » *The Center for Applied Artificial Intelligence*, Chicago Booth. <https://www.chicagobooth.edu/research/center-for-applied-artificial-intelligence/research/algorithmic-bias/playbook>
42. Peltola, Timo. 2020. «Machine Learning for Risk Prediction in Atrial Fibrillation», *Frontiers in Cardiovascular Medicine*, vol. 9/2022
43. Rabbat, Mark, Jonathon Leipsic, Jeroen Bax, Brian Kauh, Rina Verma, Demetrios Doukas, Sorcha Allen, et al. 2020. «Fractional Flow Reserve

Derived from Coronary Computed Tomography Angiography Safely Defers Invasive Coronary Angiography in Patients with Stable Coronary Artery Disease». *Journal of Clinical Medicine* 9 (2): 604.
<https://doi.org/10.3390/jcm9020604>.

44. Roger, Véronique L., Stephen Sidney, Amy L. Fairchild, Virginia J. Howard, Darwin R. Labarthe, Christina M. Shay, A. Colby Tiner, Laurie P. Whitsel, Wayne D. Rosamond, e On behalf of the American Heart Association Advocacy Coordinating Committee. 2020. «Recommendations for Cardiovascular Health and Disease Surveillance for 2030 and Beyond: A Policy Statement From the American Heart Association». *Circulation* 141 (9). <https://doi.org/10.1161/CIR.0000000000000756>.
45. Rossi Francesca. 2019. *Il confine del futuro*. Feltrinelli.
46. Russell, Stuart, e Peter Norvig. 2021. *Intelligenza Artificiale. Un approccio moderno*. Vol. 1. 2 voll. Pearson.
47. Signani, Fulvia. 2021 «Medicina di genere: a che punto è l'Italia?». *The Italian journal of gender-specific medicine*. Med2021
48. Strodthoff, Nils, e Claas Strodthoff. 2019. «Detecting and Interpreting Myocardial Infarction Using Fully Convolutional Neural Networks». *Physiological Measurement* 40 (1): 015001. <https://doi.org/10.1088/1361-6579/aaf34d>.
49. Surakka, Ida, Brooke N. Wolford, Scott C. Ritchie, Whitney E. Hornsby, Nadia R. Sutton, Maiken Elvenstad Gabrielsen, Anne Heidi Skogholt, et al. 2023. «Sex-Specific Survival Bias and Interaction Modeling in Coronary Artery Disease Risk Prediction». *Circulation: Genomic and Precision Medicine* 16 (1). <https://doi.org/10.1161/CIRCGEN.121.003542>.
50. Tat, Emily, Deepak L Bhatt, e Mark G Rabbat. 2020a. «Addressing Bias: Artificial Intelligence in Cardiovascular Medicine». *The Lancet Digital*

Health 2 (12): e635–36. [https://doi.org/10.1016/S2589-7500\(20\)30249-1](https://doi.org/10.1016/S2589-7500(20)30249-1).

51. Taylor, J. G. 1993. «Introduction to Neural Computing». In *The Promise of Neural Networks*, a cura di J. G. Taylor, 37–72. London: Springer London. https://doi.org/10.1007/978-1-4471-0395-0_4.
52. Trabace L, Ziche M. Sex matters in preclinical studies: application of the 3Rs principles. *Ital J Gender-Specific Med* 2022;8(3):125-126. doi 10.1723/3927.39105
53. Turing, A. M. 1950. «I.—COMPUTING MACHINERY AND INTELLIGENCE». *Mind* LIX (236): 433–60. <https://doi.org/10.1093/mind/LIX.236.433>.
54. Vadiveloo, Maya, Alice H. Lichtenstein, Cheryl Anderson, Karen Aspary, Randi Foraker, Skylar Griggs, Laura L. Hayman, et al. 2020. «Rapid Diet Assessment Screening Tools for Cardiovascular Disease Risk Reduction Across Healthcare Settings: A Scientific Statement From the American Heart Association». *Circulation: Cardiovascular Quality and Outcomes* 13 (9): e000094. <https://doi.org/10.1161/HCQ.0000000000000094>.
55. Viola, Antonella. 2022. *Il sesso è (quasi) tutto Evoluzione, diversità e medicina di genere*. Feltrinelli.
56. Visseren, Frank L J, François Mach, Yvo M Smulders, David Carballo, Konstantinos C Koskinas, Maria Bäck, Athanase Benetos, et al. 2022. «2021 ESC Guidelines on Cardiovascular Disease Prevention in Clinical Practice». *European Journal of Preventive Cardiology* 29 (1): 5–115. <https://doi.org/10.1093/eurjpc/zwab154>.
57. Vogel B, Acevedo M, Appelman Y, Bairey Merz CN, Chieffo A, Figtree GA, Guerrero M, Kunadian V, Lam CSP, Maas AHEM, Mihailidou AS, Olszanecka A, Poole JE, Saldarriaga C, Saw J, Zühlke L, Mehran R. The Lancet women and cardiovascular disease Commission: reducing the global burden by 2030.

Lancet. 2021 Jun 19;397(10292):2385-2438. doi: 10.1016/S0140-6736(21)00684-X. Epub 2021 May 16. PMID: 34010613.

58. Yao, Xiaoxi, David R. Rushlow, Jonathan W. Inselman, Rozalina G. McCoy, Thomas D. Thacher, Emma M. Behnken, Matthew E. Bernard, et al. 2021. «Artificial Intelligence–Enabled Electrocardiograms for Identification of Patients with Low Ejection Fraction: A Pragmatic, Randomized Clinical Trial». *Nature Medicine* 27 (5): 815–19. <https://doi.org/10.1038/s41591-021-01335-4>.