

# Approcci Metodologici basati sul Machine Learning nell'uso dei Big Data in Sanità Pubblica

Dottorando  
Giuseppe Migliara

Tutor  
Prof Corrado De Vito



**SAPIENZA**  
UNIVERSITÀ DI ROMA



**SAPIENZA**  
UNIVERSITÀ DI ROMA

# Approcci Metodologici basati sul Machine Learning nell'uso dei Big Data in Sanità Pubblica

**Dottorato di Ricerca**  
**Malattie Infettive, Microbiologia e Sanità Pubblica**  
**XXXV Ciclo**  
**Coordinatore: Prof. Stefano D'Amelio**

**Dottorando**  
**Giuseppe Migliara**  
**Matricola 1711563**

**Tutor**  
**Prof. Corrado De Vito**

A.A. 2021-2022

Il presente documento è distribuito secondo la licenza Creative Commons CC BY-NC, attribuzione, non usi commerciali.

# Sommario

1	Introduzione.....	3
1.1	Caratteristiche dei Big Data .....	3
1.2	Utilizzo dei Big Data in sanità.....	3
1.3	Uso dei Big Data in Sanità Pubblica.....	5
1.4	Analisi dei Big Data e Machine Learning.....	6
1.5	Fasi dello sviluppo di un algoritmo di Machine Learning Supervisionato .....	9
1.6	Obiettivo .....	10
2	Metodi.....	11
2.1	Revisione della letteratura.....	11
2.2	Estrazione dei dati.....	14
3	Risultati .....	17
3.1	Studi sulla predizione del rischio .....	19
3.1.1	Caratteristiche degli studi inclusi.....	19
3.1.2	Metodologia degli studi .....	23
3.1.3	Approccio metodologico nell'utilizzo degli algoritmi.....	29
3.2	Studi sul disegno di interventi di sanità pubblica .....	36
3.3	Studi sulla sorveglianza delle malattie/modelli di previsione.....	40
4.	Discussione.....	43
5.	Conclusioni.....	51
6.	Bibliografia .....	54



# 1 Introduzione

## 1.1 Caratteristiche dei Big Data

Sebbene i Big Data siano ormai estesamente utilizzati in ambito accademico e lavorativo, non esiste una loro definizione univoca. Una definizione operativa identifica i Big Data come un insieme di dati caratterizzato da volume, velocità e varietà talmente elevati da richiedere tecnologie e metodologie analitiche specifiche per l'estrazione del suo valore [1]. Proprio i concetti di volume, velocità e varietà, che hanno origine nell'ambito dello studio del commercio online agli inizi degli anni 2000, definiscono le caratteristiche fondamentali dei Big Data, le cosiddette 3V. Nello specifico, con volume si intende la dimensione dell'insieme di dati, nell'ordine dei terabyte e oltre, con velocità l'elevata frequenza con cui tali dati vengono generati, e con varietà la diversità di tipologie e fonti dei dati [2]. In seguito altre caratteristiche si sono aggiunte, a partire dal concetto di valore dei Big Data, inteso come l'abilità di prendere decisioni significative sulla base di informazioni generate dai Big Data [3], e di veridicità dei dati raccolti, concetto di particolare rilevanza nell'applicazione dei Big Data in sanità [4].

## 1.2 Utilizzo dei Big Data in sanità

Come altri settori produttivi, anche il settore sanitario è stato interessato da una trasformazione digitale, tramite l'incremento dell'utilizzo dei record

sanitari elettronici (electronic medical records, EMR), l'utilizzo di dati amministrativi collegati ai rimborsi, la produzione e archiviazione di grandi quantità di dati provenienti da diagnostica per immagini e dalle analisi genomiche, proteomiche e metabolomiche, e dall'incremento dell'utilizzo di dispositivi indossabili e applicazioni per dispositivi mobili.. Tuttavia, il settore sanitario fa fatica ad introdurre questa innovazione nella pratica corrente e a ricavare il valore potenziale contenuto in questi dati [5], anche a causa dell'alta eterogeneità dei dati e della loro forma spesso non strutturata, che difficilmente possono essere gestiti dai normali strumenti di analisi [6]. L'utilizzo dei Big Data in sanità si è reso quindi necessario per permettere la gestione di una tale mole di dati. Secondo la Direzione Salute della Direzione Generale per la Ricerca e l'Innovazione della Commissione Europea i Big data in sanità sono costituiti da informazioni biologiche, cliniche, ambientali e sullo stile di vita, sono caratterizzati da un alto volume e da un'alta diversità, e sono raccolti a partire dai singoli individui fino alle grandi coorti, in relazione al loro stato di salute e benessere, in uno o più momenti [7]. Alcuni dei benefici legati all'introduzione dell'analisi dei Big Data in medicina comprendono: il miglioramento della qualità e dall'accuratezza delle decisioni cliniche [8], la medicina personalizzata e di precisione [9], lo sviluppo di farmaci [10], l'analisi genomica [11], la sorveglianza delle malattie [12], il disegno di interventi di sanità pubblica

[13], la diagnosi precoce di patologie [14], la diminuzione dei costi sanitari [8], lo sviluppo sistemi di supporto alle decisioni cliniche [15], lo studio dell'eterogeneità del trattamento [16], la misura della performance [17], la predizione di esiti [18].

### **1.3 Uso dei Big Data in Sanità Pubblica**

L'interesse della Sanità Pubblica nei Big Data è molteplice, visti i suoi potenziali utilizzi. Alcune delle applicazioni dei Big Data di maggior interesse per la Sanità Pubblica sono rappresentate dalla possibilità di individuare in modo maggiormente preciso popolazioni a rischio, di migliorare la sorveglianza di malattie trasmissibili e non-trasmissibili, di facilitare il disegno di strategie e interventi di promozione della salute e di prevenzione efficaci [19]. Inoltre, la recente pandemia causata da SARS-CoV-2 ha messo in luce la necessità di elaborare modelli che possano prevedere in modo accurato l'andamento di epidemie e che possano fornire indicazioni di efficaci di eventuali interventi preventivi, sia farmacologici che non-farmacologici [20].

Alcune fonti di big data che possono specificamente essere utilizzate per la salute pubblica includono[19]:



- Dati strutturati, ad esempio dati provenienti da EMR, da sistemi di sorveglianza partecipata (es. crowdsourcing, crowdmapping), da dispositivi indossabili e applicazioni per smartphone;
- Dati semi-strutturati, ad esempio dati provenienti da dispositivi di monitoraggio della salute;
- Dati non strutturati, che presentano il maggior potenziale di utilizzo di dati sia sanitari che non sanitari per migliorare la salute pubblica. Le fonti di dati non strutturati includono, tra le altre: EMR; social media e dati online; dati geolocalizzati; dati ambientali.

Tuttavia, nonostante le grandi potenzialità, l'applicazione dei Big Data in questo settore appare ancora limitato rispetto ad altre branche della medicina [19], nonostante il sostegno anche da parte di istituzioni pubbliche a iniziative di questo tipo sia in crescita [21].

#### **1.4 Analisi dei Big Data e Machine Learning**

Una parte delle difficoltà nell'introduzione dei Big Data nella Sanità Pubblica, e in particolare per quanto riguarda la branca dell'epidemiologia, è costituita dalla problematicità di abbandonare la classica metodologia statistica, basata spesso sulla verifica d'ipotesi e sull'approccio

controfattuale [22], in favore delle metodologie in uso nell'analisi dei Big Data, basate principalmente su metodiche riconducibili all'apprendimento automatico, o machine learning [23]. Infatti, data la natura complessa dei Big Data, i normali strumenti dell'analisi statistica mal si adattano alla loro analisi. Nonostante siano state sviluppate nel corso degli ultimi anni diverse strategie che permettono la gestione e l'analisi statistica di grandi masse di dati [24], le tecniche più utilizzate nell'analisi dei Big Data in medicina si rifanno all'intelligenza artificiale (IA) e in particolare alla branca del machine learning, come mostrato in una revisione sistematica del 2018 che ha sintetizzato l'utilizzo dei Big Data nell'area medica [25]. La revisione mostra infatti che le metodologie più usate comprendono: cluster analysis; Data Mining; Graph Analytics; Machine Learning; Natural Language Processing; Neural Networks; Pattern Recognition; Spatial Analysis. L'uso dell'IA in ambito sanitario si caratterizza come un fenomeno relativamente recente, sebbene i primi esempi risalgano agli anni 70 [26] e negli anni 80 esistesse un florido mercato per soluzioni basate sull'intelligenza artificiale. Tuttavia, il fallimento di queste prime tecnologie nel soddisfare le altissime attese degli operatori sanitari e degli utenti ha comportato l'avvento del cosiddetto "Inverno dell'Intelligenza Artificiale", caratterizzato da un generale disinteresse da parte del grande pubblico verso queste soluzioni, che è perdurato almeno fino alla fine degli anni 2000 [27]. Il nuovo recente

interesse verso queste metodiche è stato incentivato sia dall'importante aumento delle capacità di calcolo dei processori e della capacità di archiviazione dei dati, che dalla sempre maggiore produzione di dati relativi non solo alla salute delle persone e alle loro abitudini, ma anche relativi all'ambiente in cui essi vivono. Tutto questo ha portato allo sviluppo di nuove metodologie di analisi, in grado di incorporare nei classici modelli di IA i costrutti propri della teoria della probabilità e dell'elaborazione dei segnali, dando origine al cosiddetto machine learning, spesso denominato anche statistical learning [28,29]. Le metodiche di machine learning possono essere raggruppare in tre categorie: (i) apprendimento supervisionato; (ii) apprendimento non supervisionato; (iii) apprendimento con rinforzo. L'apprendimento supervisionato si occupa dell'apprendimento da un insieme di dati "etichettati", ovverosia in cui ad ogni record del dataset è stato attribuito uno specifico esito (spesso dicotomico). Questo tipo di modelli è solitamente utilizzato nella diagnosi di condizioni patologiche e nello sviluppo di modelli predittivi di esiti futuri, ed è adatto a quei contesti in cui sia disponibile chiaramente l'esito attribuibile a ciascun record e siano disponibili una grande quantità di dati. I modelli non supervisionati cercano invece di ricavare raggruppamenti tra i record non "etichettati" contenuti in un dataset, sulla base di tratti comuni delle caratteristiche disponibili. Uno degli esempi di apprendimento non supervisionato più

diffuso sono gli algoritmi di *clustering* (raggruppamento), solitamente utilizzati in indagini esplorative. L'apprendimento rinforzato è rappresentato da algoritmi che apprendono tramite il ricevimento di feedback dall'ambiente circostante oltre che dai dati disponibili. In medicina, gli algoritmi riferibili all'apprendimento supervisionato sono i più utilizzati [30].

## **1.5 Fasi dello sviluppo di un algoritmo di Machine Learning**

### **Supervisionato**

In generale, possono essere identificate sei fasi nello sviluppo di un qualunque modello di machine learning [31]:

1. Raccolta/Preparazione dei dati
2. Selezione delle variabili
3. Scelta dell'algoritmo appropriato
4. Tuning (selezione dei parametri dell'algoritmo)
5. Training (allenamento)
6. Validazione o testing (valutazione della performance)

Lo sviluppo di un algoritmo prevede l'esecuzione di queste fasi e una precisa metodologia di gestione dei dati [32]. In generale, il dataset utilizzato per lo sviluppo di un modello dovrebbe essere suddiviso in modo

casuale in tre subset distinti per le fasi di Tuning, Training e Validazione interna. Nel caso di campioni troppo piccoli per poter essere divisi, esistono delle tecniche che permettono, tramite convalida incrociata k-fold, di assicurare comunque una suddivisione ottimale del dataset di origine. Tali accorgimenti sono necessari per evitare il rischio di overfitting del modello, ovvero che abbia una scarsa generalizzabilità e soffra di scarsa performance in dataset diversi da quelli di origine [33].

## **1.6 Obiettivo**

Vista la novità rappresentata dall'utilizzo del machine learning, l'obiettivo di questa tesi è stato quello di indagare la metodologia basata sull'evidenza utilizzata nell'applicazione di queste metodiche nella Sanità Pubblica, con particolare attenzione alla fase del loro sviluppo e validazione. Per fare questo è stata portata avanti una revisione sistematica della letteratura, di seguito descritta.

## 2 Metodi

### 2.1 Revisione della letteratura

È stata effettuata una revisione sistematica della letteratura medico-scientifica, utilizzando i seguenti database: PubMed, Scopus; Web of Science. Per la costruzione delle stringhe di ricerca sono stati utilizzati termini correlati agli approcci metodologici di data science nell'uso dei Big Data e dei dati del mondo reale (vedi box 1). La strategia di ricerca è stata adattata per soddisfare i criteri di ricerca di ciascun database ed è stata effettuata a partire dal 01/01/2010 fino al 19/01/2022. La ricerca bibliografica è stata inoltre integrata dalla ricerca manuale all'interno delle bibliografie della letteratura selezionata.

I seguenti criteri di inclusione sono stati utilizzati nella selezione degli articoli:

- Uso di metodologie di machine learning nell'analisi, anche parziale, dei dati.
- Applicazione ad un campo di interesse della sanità pubblica, così definiti:

- Elaborazione di predizioni di rischio per lo sviluppo di patologie o esiti di malattia nella popolazione generale o in specifiche popolazioni di maggiore interesse per la sanità pubblica (e.g.: pazienti affetti da malattie croniche);
- Elaborazione di metodologie di sorveglianza delle malattie e degli eventi avversi, inclusi i modelli di previsione;
- Disegno di interventi di sanità pubblica, o ideazione e simulazione di interventi di sanità pubblica;

I criteri di esclusione sono stati i seguenti:

- Utilizzo di metodologie di data science per l'elaborazione di immagini radiodiagnostiche o relative ad analisi istologiche;
- Lingua dell'articolo diversa da Inglese o Italiano
- Articolo non disponibile
- Tipo di articolo: protocolli di ricerca, contributi a convegni, revisioni, lettere, editoriali, commentari.

- Documenti contenenti solo verifiche teoriche di funzionamento e performance delle metodiche descritte (proof of concept)

Box 1 – Termini utilizzati nella stringa di ricerca

("big data" OR "real world data" OR "machine learning" OR "neural network" OR "artificial intelligence" OR "support vector machine" OR "natural language processing" OR "data mining" OR "stream processing" OR "data science" OR "internet of things" OR "IOT" OR "deep learning" OR "ai")

**AND**

("method\*" OR "approach\*" OR "research\*" OR "design" OR "analysis")

**AND**

("evidence-based" OR "proven" OR "hypothesis-driven" OR "evidence based" OR "hypothesis driven")



## 2.2 Estrazione dei dati

I titoli e gli abstract degli studi recuperati con la strategia di ricerca sono stati esaminati indipendentemente da due revisori per identificare tutti gli articoli che potenzialmente soddisfino i criteri di inclusione. Il testo completo degli studi potenzialmente includibili è stato recuperato e valutato in modo indipendente da due membri del gruppo di revisione. Per l'estrazione dei dati è stato utilizzato un modulo standardizzato e sono state recuperate le seguenti informazioni: primo autore, anno di pubblicazione, nazione di affiliazione del primo autore, disegno dello studio, popolazione target, fonte dei dati, tipo di algoritmo di machine learning utilizzato, obiettivo dello studio, risultati principali.

Solo per gli studi classificati come “predizione del rischio” sono state inoltre estratte le seguenti informazioni: tipologia di dati utilizzati, utilizzo dell'algoritmo, trasparenza del modello, gestione dei dati mancanti, modalità di tuning dell'algoritmo, modalità di validazione interna, modalità di validazione esterna, indici di performance riportati, classificazione dell'algoritmo (supervisionato vs non supervisionato). Inoltre, per questi studi è stata elaborata una scala del rischio di overfitting del modello, basata sulle modalità di tuning, training e validazione di un algoritmo supervisionato di machine learning. Nello specifico,

parzialmente sulla base della metodologia riportata da una recente revisione sistematica sull'utilizzo di algoritmi di machine learning in ambito psichiatrico [34], il rischio di overfitting è stato ritenuto elevato a seguito del presentarsi di una delle seguenti situazioni:

- Tuning effettuato su subset di training
- Validazione interna ed esterna non effettuata
- Modalità di validazione interna non valide o non descritte

Il rischio di overfitting è stato ritenuto basso a seguito del presentarsi delle seguenti situazioni:

- Tuning effettuato in modo adeguato (validazione incrociata annidata k-fold o subset separato)
- Modalità di validazione interna valide (validazione incrociata k-fold (stratificata o meno), validazione out-of-bootstrap, subset di training separato)

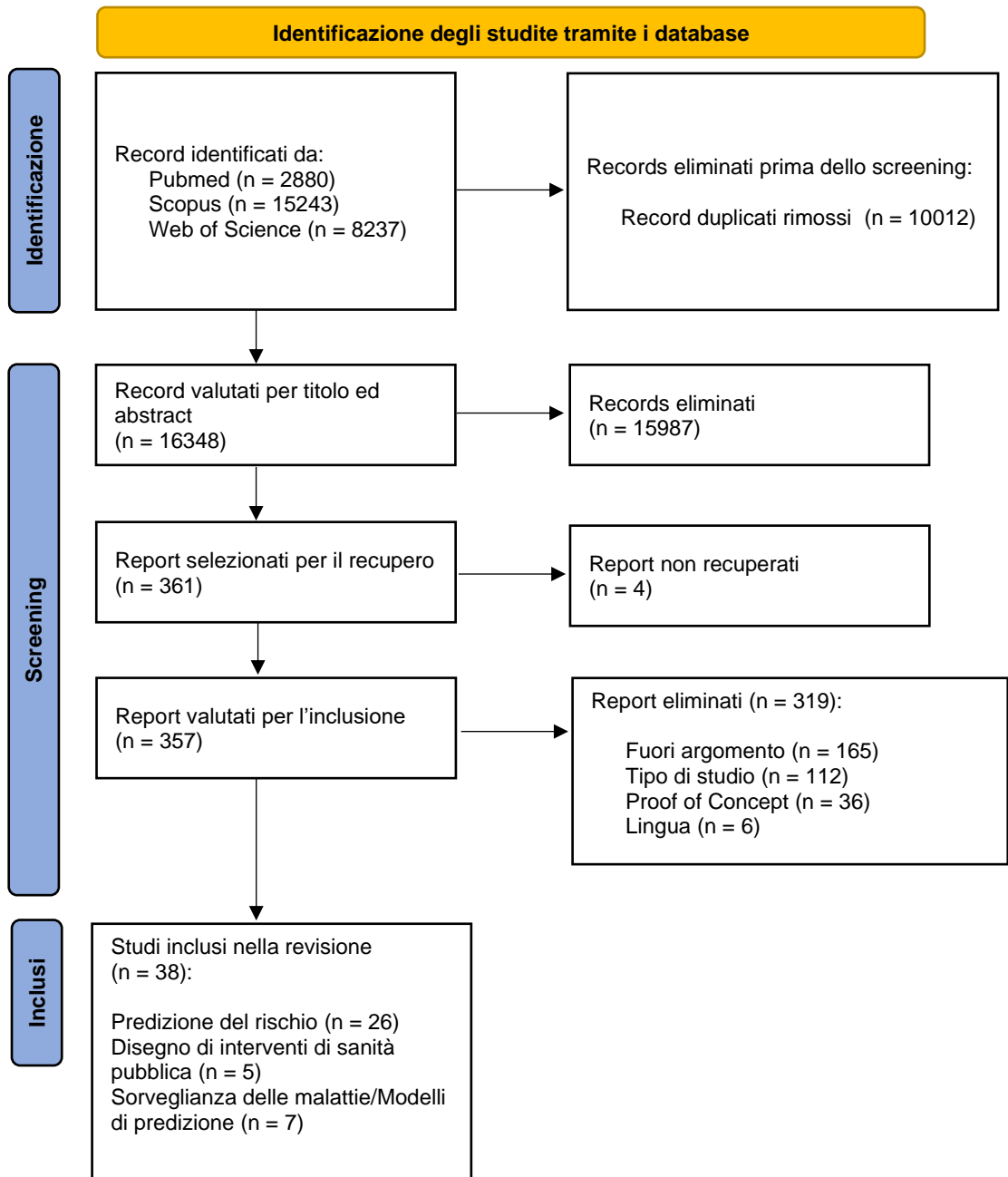
Nel caso unitamente alle precedenti condizioni descritte lo sviluppo degli algoritmi sia stata seguita da modalità di validazione esterna adeguate

(dataset da altra popolazione o da stessa popolazione in altro periodo di tempo), il rischio di overfitting è stato ritenuto minimo.

### 3 Risultati

In tutto, sono stati identificati 26340 record dalla ricerca sui database (Figura 1). Dopo la rimozione dei duplicati e lo screening effettuato su titoli e abstract, 361 documenti sono stati selezionati per l'analisi degli articoli completi, che ha dato infine come risultato l'inclusione di 38 studi [35–72]. Di questi, 26 sono studi sulla predizione del rischio, 5 sono studi sul disegno di interventi di sanità pubblica, e 7 sono studi su sorveglianza delle malattie/modelli di previsione. Le caratteristiche degli studi sono descritte di seguito per tipologia.

Figura 1. Diagramma di flusso PRISMA per revisioni sistematiche



### **3.1 Studi sulla predizione del rischio**

#### **3.1.1 Caratteristiche degli studi inclusi**

I 26 studi che hanno riguardato la predizione del rischio inclusi nella revisione sistematica sono stati pubblicati tra il 2010 e il 2021 e le loro caratteristiche sono riportate in tabella 1 [35–60]. Per quanto riguarda il disegno di studio, 7 sono risultati studi caso-controllo [35,40,44,45,49,52,54], 10 studi di coorte [37,39,41,42,46,48,51,56,57,60], 8 studi trasversali [36,38,43,47,50,53,55,59] e 1 uno studio di tipo ecologico [58]. La metà degli studi è stata condotta negli USA (13 studi) [35,36,38–41,44,47–52,60], 3 in Francia [39,43,58], 2 nei Paesi Bassi [45,56] e 1 ciascuno in Cina [53], Danimarca [57], Emirati Arabi Uniti [46], Giappone [55], Giordania [37], Grecia [54], Italia [42] e Pakistan [59]. 10 studi hanno avuto come riferimento la popolazione adulta [35,37,38,40,52–56,58], 4 studi la popolazione in età pediatrica [45,46,50,59], e 12 studi gruppi specifici di persone o pazienti [36,41–44,47–49,51,57,60]. Le aree di interesse più frequenti degli studi inclusi sono state la sanità pubblica (5 studi) [37,40,46,57,59] e la cardiologia (5 studi) [35,41,43,53,58], seguite dalle malattie metaboliche [38,42,52], dall'oncologia [39,47,56], pediatria [45,48,50] e dalla psichiatria [49,51,54], con 3 studi ciascuna. Ginecologia [36], Malattie Infettive [60], Neurologia [55] e Reumatologia [44] sono state interessate da 1 articolo ciascuno. Solo 7 studi hanno studiato un campione inferiore a 1000 soggetti

[45,47,50,51,55,57,60], mentre 6 studi hanno analizzato campioni superiori a 40000 [42,44,46,53,54,56]. In generale, la mediana dei soggetti analizzati è stata di 3121.5 [IQR 845.5-25745.5]. Le fonti dei dati utilizzati dagli studi sono risultate eterogenee. 7 studi hanno utilizzato dati raccolti appositamente in modo autonomo dai ricercatori [43,45,47,49,53-55], 12 hanno utilizzato dati estratti da cartelle e record sanitari elettronici [37,38,40,42,48,50,52,56-58,60], 3 hanno utilizzato dati diffusi da autorità nazionali [51,58,59], 2 dati da repository disponibili su Internet [35,44], 1 dati raccolti da un'applicazione per smartphone [46], 1 dati provenienti dall'analisi secondaria di uno studio di coorte precedente [39], e infine 1 studio ha utilizzato le informazioni contenute nei foglietti illustrativi dei farmaci [36]. Gli obiettivi dei singoli studi sono riportati in tabella 1.

Tabella 1. Caratteristiche degli studi sulla predizione del rischio inclusi nella revisione sistematica. N=26

Primo autore, anno	Nazione	Disegno dello studio	Campo di interesse	Popolazione target	Fonte dei dati	Obiettivo dello studio
Adams, 2020 [35]	USA	Caso-controllo	Cardiologia	Popolazione generale in trattamento con statine (N = 5892)	Repository	Stratificare il rischio di eventi avversi cardiovascolari maggiori durante terapia con statine basato su predittori epistatici poligenici
Ailes, 2020 [36]	USA	Trasversale	Ginecologia	Donne in età riproduttiva e in gravidanza (N = NA)	Foglietti illustrativi farmaci	Elaborare un elenco di farmaci potenzialmente pericolosi durante la gravidanza
Alyahya, 2016 [37]	Giordania	Coorte	Sanità Pubblica	Popolazione Generale (N = 3962)	Record sanitari elettronici	Identificare i fattori di rischio comportamentali modificabili che sono possibili cause di riammissioni potenzialmente evitabili
Anderson, 2016 [38]	USA	Trasversale	Malattie metaboliche	Popolazione generale (N = 9948)	Record sanitari elettronici	Valutare l'efficacia dell'analisi della fenotipizzazione degli EMR nello screening del diabete mellito di tipo 2 rispetto ai modelli convenzionali
Audureau, 2020 [39]	Francia	Coorte	Oncologia	Pazienti con cirrosi HCV-relata (N = 836)	Precedente studio di coorte	Stimare il rischio di HCC in pazienti con cirrosi HCV-correlata
Banda, 2019 [40]	USA	Caso-controllo	Sanità Pubblica	Popolazione generale ad alto rischio di malattia cardiovascolare aterosclerotica (N = 6787)	Record sanitari elettronici	Identificare potenziali pazienti affetti da ipercolesterolemia familiare tramite analisi degli EMR
Berger, 2020 [41]	USA	Coorte	Cardiologia	Pazienti con ischemia critica degli arti inferiori (N = 3189)	Record sanitari elettronici	Sviluppare un modello predittivo personalizzato per stimare le ospedalizzazioni ad un anno e i costi di cura
Bernardini, 2020 [42]	Italia	Coorte	Malattie metaboliche	Pazienti con Diabete Mellito (N = 120000)	Record sanitari elettronici	Stimare il rischio di sviluppo di Retinopatia diabetica
Bourdès, 2011 [43]	Francia	Trasversale	Cardiologia	Pazienti con pregressa malattia cardiovascolare aterosclerotica (N = 2132)	Dati prodotti autonomamente dai ricercatori	Identificare i fattori chiave che determinano persistenza dell'utilizzo dei farmaci cardiovascolari nei pazienti con sindrome coronarica acuta
Briggs, 2010 [44]	USA	Caso-controllo	Reumatologia	Pazienti con artrite reumatoide (N = 530)	Repository	Identificare fattori di rischio epistatici che interagiscono con PTPN22 che influenzino la suscettibilità all'artrite reumatoide
Cabrera-Quiros, 2021 [45]	Paesi Bassi	Caso-controllo	Pediatria	Neonati pretermine (N = 64)	Dati prodotti autonomamente dai ricercatori	Stimare il rischio di sepsi ad esordio tardivo nei neonati pretermine
Chandir, 2018 [46]	Emirati Arabi Uniti	Coorte	Sanità pubblica	Popolazione in età pediatrica ≤ 2 anni (N = 47554)	Applicazione per smartphone	Identificare i bambini a rischio di non presentarsi ai successivi appuntamenti vaccinali
Chau, 2016 [47]	USA	Trasversale	Oncologia	Pazienti con CCR invasivo e loro familiari (N = 181555)	Dati prodotti autonomamente dai ricercatori	Stima del rischio familiare di CCR
Das, 2017 [48]	USA	Coorte	Pediatria	Pazienti asmatici ≤ 18 anni (N = 2691)	Record sanitari elettronici	Stimare il rischio di accesso frequente ai servizi di emergenza
Dean, 2020 [49]	USA	Caso-controllo	Psichiatria	Soldati veterani (N = 281)	Dati prodotti autonomamente dai ricercatori	Identificare e validare un set di biomarker multi-omici per la diagnosi di disturbo da stress post-traumatico

CCR: carcinoma del colon-retto; EMR: record sanitari elettronici HCC: Epatocarcinoma; HCV: Virus dell'epatite C; NA: Non Applicabile

(Continua)



Tabella 1 (Continua). Caratteristiche degli studi sulla predizione del rischio inclusi nella revisione sistematica.

Primo autore, anno	Nazione	Disegno dello studio	Campo di interesse	Popolazione target	Fonte dei dati	Obiettivo dello studio
Deleger, 2013 [50]	USA	Trasversale	Pediatria	Popolazione in età pediatrica < 18 anni (N = 2100)	Record sanitari elettronici	Stratificare il rischio di appendicite in pazienti pediatrici con dolore addominale
Hastings, 2020 [51]	USA	Coorte	Psichiatria	Pazienti con prescrizione per oppioidi (N = 80768)	Autorità nazionale	per prevedere il rischio di futura dipendenza, abuso o avvelenamento da oppiacei prima di una prescrizione iniziale di oppiacei
Hegde, 2019 [52]	USA	Caso-controllo	Malattie metaboliche	Popolazione generale (N = 41543)	Record sanitari elettronici	sviluppare un modello predittivo utilizzando i dati medico-odontoiatrici per identificare le persone con diabete mellito (DM) non diagnosticato in ambito odontoiatrico.
Huang, 2010 [53]	Cina	Trasversale	Cardiologia	Popolazione generale ≥35 anni (N = 3054)	Dati prodotti autonomamente dai ricercatori	Predire il rischio di ipertensione
Katrakazas, 2020 [54]	Grecia	Caso-controllo	Psichiatria	Popolazione generale (N = 76)	Dati prodotti autonomamente dai ricercatori	Utilizzare i dati microscopici di un simulatore di guida per rilevare le sessioni appartenenti a conducenti depressi
Kimura, 2019 [55]	Giappone	Trasversale	Neurologia	Popolazione generale (N = 855)	Dati prodotti autonomamente dai ricercatori	Studiare la relazione tra stile di vita e funzioni cognitive
Kop, 2016 [56]	Paesi Bassi	Coorte	Oncologia	Popolazione generale ≥30 anni (N = 263879)	Record sanitari elettronici	Predire l'occorrenza di CCR
Pedersen, 2019 [57]	Danimarca	Coorte	Sanità pubblica	Pazienti con patologie croniche	Record sanitari elettronici	Costruire un modello predittivo in grado di identificare i pazienti ad alto rischio di ritirarsi da interventi digitali comportamentali
Pérez, 2021 [58]	Francia	Ecologico	Cardiologia/ Pneumologia	Popolazione Generale (N = NA)	Record sanitari elettronici Autorità nazionale	Studiare la potenziale relazione tra inquinamento atmosferico, status socio-economico e patologie cardiache e pneumologiche
Qazi, 2021 [59]	Pakistan	Trasversale	Sanità pubblica	Popolazione in età pediatrica <5 anni (N = 7153)	Autorità nazionale	Identificare i bambini secondo rischi crescenti di non presentarsi ai successivi appuntamenti vaccinali
Rozenbaum, 2021 [60]	USA	Coorte	Malattie infettive	Pazienti affetti da COVID-19 (N = 764)	Record sanitari elettronici	Costruire un modello predittivo della mortalità e della lunghezza della degenza dei pazienti affetti da COVID-19

CCR: carcinoma del colon-retto

### 3.1.2 Metodologia degli studi

Il numero algoritmi di machine learning utilizzati è variato tra gli studi, e le loro caratteristiche sono riportate in tabella 2. Nello specifico, la maggior parte degli studi ha utilizzato un solo algoritmo (16 su 26, 61.5%) [35–38,40,41,43,44,47,49,50,53–55,58,60], 3 (11.5%) studi hanno applicato 2 algoritmi [39,42,51], 3 (11.5%) studi hanno applicato 3 algoritmi [45,56,57], e infine 4 studi hanno utilizzato 4 algoritmi [46,48,52,59]. Nella totalità degli studi in cui è stato utilizzato più di un algoritmo, questi sono stati comparati tra loro nella performance e non sono stati utilizzati in modo combinato. Per quanto riguarda la tipologia di algoritmi utilizzati, 15 studi hanno fatto utilizzo di modelli di tipo Random Forest [35,38–40,42,44,46,48,49,52,54–57], mentre l'utilizzo di altri tipi di modelli è risultata più sporadica. In particolare, modelli di tipo Decision Tree sono stati utilizzati in 8 studi [39,42,46,48,56,57,59,60], seguiti dai modelli di tipo regressione logistica (7 studi), modelli di tipo Support Vector Machine (6 studi) [36,42,46,48,52,59], modelli di tipo Artificial Neural Network (5 studi) [43,51–53,59], e modelli di tipo Bayesian Network (5 studi) [41,42,45,58,59]. Meno utilizzati sono stati modelli rule-based (2 studi) [37,50], modelli di tipo LASSO (2 studi) [48,51], e di tipo k-means clustering (1 studio) [47].

Notevole è risultata anche la varietà nella tipologia di dati utilizzati dagli studi. In particolare, 7 studi hanno utilizzato una sola tipologia di dati [35,36,43–45,52,54], mentre 6 articoli hanno utilizzato 4 o più tipologie di dati [37,39,48,49,51,55]. I dati anamnestici sono stati in assoluto i più utilizzati (18 studi) [37–43,47,48,50–53,55–58,60], seguiti dai dati demografici e sociodemografici (15 studi) [37–39,41,42,46–49,51,53,55,57,59,60]. Dati “omici” sono stati utilizzati in solo 3 studi [35,44,49]. La tabella 2 riporta le tipologie di dati utilizzati dai singoli studi inclusi nella revisione.

Per quel che riguarda la trasparenza degli algoritmi utilizzati, quasi un terzo degli studi non ha riportato alcuna forma di interpretazione dei risultati ottenuti [36,38,46,50,52–54,59], mentre tutti gli altri studi hanno riportato varie forme di interpretazione (vedi Tabella 2), sebbene solo in minima parte queste fossero associate ad una effettiva quantificazione dell’effetto delle variabili inserite nel modello sull’esito nel senso di Odds Ratios o Rischi Relativi [41,45,47,48,51,56].

Tabella 2. Metodologia e risultati degli studi sulla predizione del rischio inclusi nella revisione sistematica.

Primo autore, anno	Tipologia di dati utilizzati	Tipo di algoritmo utilizzato	Utilizzo dell'algoritmo	Trasparenza del modello	Risultati principali
Adams, 2020 [35]	Dati Genomici	Random Forest	Stima del rischio di eventi avversi cardiovascolari maggiori	Analisi degli alberi decisionali	L'analisi dei pathways ha mostrato che le varianti nei geni correlati alla vasculogenesi (FDR = 0,024), all'angiogenesi (FDR = 0,019) e alla malattia dell'arteria carotidea (FDR = 0,034) sono correlate al rischio di eventi avversi cardiovascolari maggiori durante terapia con statine
Ailes, 2020 [36]	Dati sugli eventi avversi dei farmaci	Support Vector Machine	Stima del rischio di teratogenesi in gravidanza	Modello non interpretato	Classificazione di 80 farmaci sottoposti a prescrizione come di potenzialmente pericolosi durante la gravidanza; oltre la metà erano agenti antineoplastici (n=24), inibitori dell'enzima di conversione dell'angiotensina (n=10), antagonisti del recettore dell'angiotensina II (n=8) e anticonvulsivanti (n=7).
Alyahya, 2016 [37]	Sociodemografici Anamnestici Comportamentali Prescrizioni farmaceutiche	Jrip	Individuazione delle regole (caratteristiche predominanti) nei pazienti riospedalizzati	Estrazione delle regole (caratteristiche predominanti)	I seguenti fattori di rischio per la riospedalizzazione sono stati individuati: abitudine al fumo (OR= 2,318, IC 95% 1,953–2,751); Fattori protettivi: compliance del paziente al trattamento medico; pazienti con un programma di follow-up chiaro e lettere di dimissioni chiare; pazienti adeguatamente dimessi da medici; pazienti coassicurati per il 20% o più
Anderson, 2016 [38]	Demografici Clinici Comportamentali	Random Forest	Validazione di un modello di regressione logistica multivariabile (analisi secondaria)	Modello non interpretato	I modelli contenenti anche la fenotipizzazione degli EMR hanno avuto performance migliori del modello di screening routinario nella rilevazione degli individui a rischio di diabete
Audureau, 2020 [39]	Demografici Clinici Prescrizioni farmaceutiche Comportamentali	Decision tree Random Survival foFest	Stima del rischio di sviluppare cirrosi nei pazienti con HCV	Decision Tree: Analisi dei percorsi decisionali Random Forest: Variable importance	Il modello di Fine & Gray ha identificato 6 predittori indipendenti per HCC prima dell'eradicazione dell'HCV. Il modello basato su Decision Tree ha confermato questi predittori ma ha mostrato interazioni complesse. Il modello basato sulla Random Survival Forest ha dimostrato che i predittori dipendono dallo stato di eradicazione
Banda, 2019 [40]	Clinici Note cliniche non strutturate Prescrizioni farmaceutiche	Random Forest	Identificare possibili casi di Ipercolesterolemia Familiare	Analisi degli alberi decisionali	Il classificatore sviluppato basato su record medici si è dimostrato in grado di identificare i pazienti a rischio di ipercolesterolemia familiare, con un valore predittivo positivo di 0,88, sensibilità di 0,75 e specificità di 0,99.

EMR: record sanitari elettronici; FRD: Tasso di Falsi Positivi (False Detection Rate); HCC: Epatocarcinoma; HCV: Virus dell'epatite C; IC: Intervallo di Confidenza; OR: Odds Ratio

(Segue)

Tabella 2 (Continua). Metodologia e risultati degli studi sulla predizione del rischio inclusi nella revisione sistematica.

Primo autore, anno	Tipologia di dati utilizzati	Tipo di algoritmo utilizzato	Utilizzo dell'algoritmo	Trasparenza del modello	Risultati principali
Berger, 2020 [41]	Demografici Clinici Comportamentali	Bayesian ensemble-model	Stimare il rischio di ospedalizzazione e i costi associati	OR Medio (per l'ospedalizzazione)  Cambio percentuale medio (per i costi associati)	Il modello ha per i ricoveri da tutte le cause ha mostrato una AUC di 0.63. Il modello per i costi associati ha mostrato un R <sup>2</sup> di 0.12
Bernardini, 2020 [42]	Demografici Clinici Laboratoristici	Random Forest Gradient Boosted Decision Tree Logistic Regression Decision Tree Support Vector Machine Naive Bayes	Stimare il rischio di Retinopatia Diabetica	Variable importance	I modelli basati su Gradient Boosted Decision Tree e Random Forest hanno ottenuto i valori migliori di performance, con AUPRC di 72.4 e 71.6, rispettivamente, per la predizione dei pazienti a rischio di retinopatia diabetica. Il modello basato su Random Forest ha anche ottenuto i valori migliori (AUPRC 84.38) nella predizione della finestra temporale tra la visita e lo sviluppo di RD.
Bourdès, 2011 [43]	Anamnestici	Artificial neural network	Stima del rischio di terapia persistente  Selezione variabili (analisi secondaria)	Estrazione di regole	I due modelli basati su ANN hanno mostrato un'accuratezza migliore del modello di regressione logistica nel predire il rischio di terapia persistente (AUC 0.78 e 0.80 per ANN, 0.69 per LR)
Briggs, 2010 [44]	Dati Genomici	Random Forest	Selezione delle variabili (analisi secondaria)	Variable importance	Il modello ha individuato 3 varianti SNP in grado di incrementare il l'effetto della mutazione PTPN22 e 3 varianti SNP in grado di diminuirne l'effetto
Cabrera-Quiros, 2021 [45]	Monitoraggio segni vitali	Logistic regression Naive Bayes Nearest mean	Stima del rischio di sepsi ad esordio tardivo nei neonati pretermine	Coefficienti della regressione logistica	Il modello di regressione logistica si è rivelato il migliore nella stima del rischio di sepsi ritardata nei prematuri con una AUC di 0.79
Chandir, 2018 [46]	Sociodemografici Status vaccinale	Recursive Partitioning Support Vector Machine Random Forest C-Forest	Stima del rischio di mancare le vaccinazioni programmate	Modello non interpretato	Il modello basato sulla Recursive Partitioning si è rivelato il più accurato nella stima del rischio di non ricevere le vaccinazioni di richiamo, con una AUC di 0.79
Chau. 2016 [47]	Demografici Anamnestici	K-means clustering	Stima del rischio di CCR	Rischio Relativo Standardizzato	Il modello ha evidenziato la presenza di 5 cluster familiari con rischio diverso di CCR
Das, 2017 [48]	Demografici Anamnestici Prescrizioni farmaceutiche Uso dei servizi sanitari	Logistic Regression LASSO Random Forest CART Support Vector Machine	Stima del rischio di utilizzo frequente dei servizi di emergenza	Coefficienti delle regressioni	Il modello di regressione logistica si è rivelato il migliore nella stima del rischio di uso frequente del pronto soccorso, con una AUC di 0.86

ANN: Artificial Neural Network; AUC: area sotto la curva; AUPRC: area sotto la curva di Precision-Recall; CCR: carcinoma del colon-retto; LR: logistic regression; RD: retinopatia diabetica; SNP: polimorfismo a singolo nucleotide

(Segue)

Tabella 2 (Continua). Metodologia e risultati degli studi sulla predizione del rischio inclusi nella revisione sistematica.

Primo autore, anno	Tipologia di dati utilizzati	Tipo di algoritmo utilizzato	Utilizzo dell'algoritmo	Trasparenza del modello	Risultati principali
Dean, 2020 [49]	Demografici Genomici Proteomici Metabolomici Altri dati biologici	Random Forest	Stima del rischio di PTSD	Variable importance	Il modello ha individuato un set di 28 biomarker in grado di identificare i casi di PTSD tra i veterani, riportando un'accuratezza dell'81%
Deleger, 2013 [50]	Anamnestici Note cliniche non strutturate	Rules-based classifier	Stima del rischio di appendicite	Modello non interpretato	Il modello ha ottenuto un F-score medio di 0.87 nella classificazione dei pazienti a rischio di appendicite.
Hastings, 2020 [51]	Demografici Anamnestici Prescrizioni farmaceutiche Amministrativi	BOLASSO LASSO ensemble Artificial Neural Network	Stima del rischio di dipendenza da oppiacei post-terapia	Odds Ratios (BOLASSO)	Il modello basato sull'ANN si è rivelato il più accurato nella stima del rischio di dipendenza da oppioidi, con una AUC di 0.80
Hegde, 2019 [52]	Anamnestici (clinici e dentistici)	Multilayer Perceptron Random Forests Support Vector Machine Logistic Regression	Stima del rischio di Diabete Mellito non diagnosticato	Modello non interpretato	Il modello basato sulla RF si è rivelato il più accurato nella stima del rischio di diabete mellito non diagnosticato, con una AUC di 0.97
Huang, 2010 [53]	Sociodemografici Anamnestici Comportamentali	Artificial Neural Network Logistic Regression	Stima del rischio di ipertensione	Modello non interpretato	Il modello basato sull'ANN si è rivelato il più accurato nella stima del rischio di depressione, con una AUC di 0.90
Katrakazas, 2020 [54]	Sessioni di guida al simulatore	Random Forest	Stima del rischio di depressione	Modello non interpretato	Il modello basato sulla RF si è rivelato accurato nella stima del rischio di depressione, con un F1-Score di 0.96
Kimura, 2019 [55]	Sociodemografici Anamnestici Comportamentali Prescrizioni mediche Dati raccolti da dispositivi indossabili	Random Forest	Stimare il Mini-mental State Examination Score	Partial dependence plot	Il modello basato sulla RF ha mostrato un R <sup>2</sup> di 0.78 nella stima del Mini-mental State Examination Score
Kop, 2016 [56]	Anamnestici Laboratoristici Prescrizioni farmaceutiche	Random Forest Cart Logistic Regression	Stimare il rischio di CCR	Coefficienti della regressione logistica	Il modello basato sulla RF si è rivelato il più accurato nella stima del rischio di CCR, con una AUC di 0.89
Pedersen, 2019 [57]	Demografici Anamnestici	Logistic Regression Decision Tree Random Forest	Stima del rischio di ritirarsi dall'intervento	Variable importance	Il modello basato sulla RF si è rivelato il più accurato nella stima del rischio di abbandono dell'intervento, con una AUC di 0.92

ANN: Artificial Neural Network; AUC: area sotto la curva; CCR: carcinoma del colon-retto PTSD: disturbo da stress post-traumatico; RF: Random Forest

(Segue)

Tabella 2 (Continua). Metodologia e risultati degli studi sulla predizione del rischio inclusi nella revisione sistematica.

Primo autore, anno	Tipologia di dati utilizzati	Tipo di algoritmo utilizzato	Utilizzo dell'algoritmo	Trasparenza del modello	Risultati principali
Pérez, 2021 [58]	Anamnestici Ambientali Censuari	Bayesian Network	Esplorazione delle relazioni tra variabili	Clustering gerarchico	Lo studio ha rivelato profili di salute differenziati tra diverse aree nella conurbazione di Martigues, con diverse esigenze di assistenza sanitaria e politiche di prevenzione
Qazi, 2021 [59]	Sociodemografici Censuari	Multilayer Perceptron Decision Tree Support Vector Machine Naive Bayes	Stima del rischio di mancare le vaccinazioni programmate	Modello non interpretato	Il modello basato sul Multilayer Perceptron si è rivelato il più accurato nella stima del rischio di mancare vaccinazioni programmate, con una accuratezza di 0.94
Rozenbaum, 2021 [60]	Demografici Anamnestici Laboratoristici	Decision Tree	Stima del rischio di decesso e della lunghezza della degenza in pazienti affetti da COVID-19	Variable importance	I modelli basati sul Decision Tree si sono rivelati accurati nella stima del rischio di mortalità (AUC 0.87 per mortalità a 7 giorni, AUC 0.88 per mortalità a 14 giorni, AUC 0.85 per mortalità a 30 giorni) e della durata della degenza (AUC 0.80 per degenza >7 giorni, AUC 0.82 per degenza >14 giorni)

AUC: area sotto la curva

### **3.1.3 Approccio metodologico nell'utilizzo degli algoritmi**

L'approccio metodologico adottato dagli studi nell'utilizzo degli algoritmi di machine learning è riportato in tabella 3. Data la natura e le fonti dei dati usati dagli articoli selezionati, questi sono particolarmente proni a presentare valori mancanti e diverse tecniche sono state impiegate per la loro gestione. Nello specifico 16 studi su 26 hanno analizzato dataset contenenti dati mancanti. In particolare, 6 studi [41,48,52,57,59,60] hanno semplicemente eliminato i record contenenti dati mancanti, o quelle variabili con una frequenza ritenuta eccessiva degli stessi [60], mentre 5 studi [39,42,46,47,51] hanno utilizzato tecniche di imputazioni e uno studio [56], che ha impiegato dati testuali nell'analisi, ha utilizzato una tecnica di arricchimento semantico. Infine, 4 studi [38,40,50,58] non hanno dichiarato come hanno gestito i valori mancanti, ma è presumibile che abbiano eliminato i record interessati. Per quanto riguarda gli studi che hanno impiegato degli algoritmi di machine learning di tipo supervisionato, 19 studi non hanno effettuato il tuning degli iperparametri. I rimanenti 7 studi hanno utilizzato diversi approcci, quali il tuning sul subset di training (3 studi) [46,56,57], mentre 2 studi hanno utilizzato la tecnica della convalida incrociata sul subset di training [54,60] e 1 studio ha utilizzato la tecnica della convalida incrociata annidata [40]. Un solo studio ha utilizzato una



parte separata del dataset originario per la ricerca degli iperparametri migliori [51].

Per quanto riguarda la validazione interna dei modelli, questa è stata eseguita nella stragrande maggioranza degli studi, con l'eccezione di Adams et al. [35] e di Dean et al [49]. Le modalità di validazione interna sono consistite nella convalida incrociata k-fold (9 studi) [36,37,40–42,50,52,56,57], nella convalida incrociata con bootstrap [38,44,46], e nella validazione con separazione completa dei subset di training e test (Holdout, 7 studi) [48,51,53–55,59,60]. Cabrerias-Quiros et al. hanno invece utilizzato una convalida incrociata *leave-n-out* [45], sebbene il valore di n non sia stato riportato, mentre Bourdes et al. hanno utilizzato una convalida incrociata non meglio definita [43]. Infine, una forma di validazione esterna del modello era presente solo in 5 studi [39,40,49,50,52], anche se questa si configura come effettivamente esterna sono in Audureau et al. [39] e in Banda et al. [40], che hanno utilizzati dati raccolti da popolazioni diverse da quelle in cui l'algoritmo è stato sviluppato, mentre in Dean et al. [49], Deleger et al. [50] e Hedge et al. [52] è stata utilizzata su dati esclusi dal training e dalla validazione interna, ma provenienti dalle stesse popolazioni di origine e raccolti nello stesso periodo.

Nella valutazione della performance degli algoritmi, la misura più utilizzata è stata l'analisi dell'area sotto la curva ROC (AUC), riportata da 16 studi [38–40,42,43,45,46,48,49,51,52,56,57,59,60]. In generale, il rischio di overfitting dei modelli sviluppati nei 24 studi che hanno utilizzato algoritmi supervisionati, basato su modalità di tuning, modalità di validazione interna e modalità di validazione esterna, come riportato nei metodi, è risultato essere elevato in 6 studi [35,43,45,46,56,57], nei quali il tuning degli iperparametri degli algoritmi è stato effettuato nel subset di training [46,56,57] o con modalità di convalida incrociata inadeguate o insufficientemente dettagliate [35,43,45], e basso in 15 studi [36–38,41,42,44,48,50–55,59,60], dove al contrario sono state utilizzate tecniche adeguate alla separazione delle varie fasi di sviluppo degli algoritmi. Ai tre studi che hanno riportato anche delle modalità di validazione esterne appropriate è stato assegnato un rischio di overfitting minimo [39,40,49].

Tabella 3. Approccio metodologico adottato dagli studi nell'utilizzo degli algoritmi di machine learning.

Primo autore, anno	Gestione dei dati mancanti	Modalità di tuning dell'algoritmo	Modalità di validazione interna	Modalità di Validazione esterna	Modalità di valutazione della performance	Approccio	Rischio di overfitting
Adams, 2020 [35]	Nessun dato mancante	Non effettuato	Non effettuata	Non effettuata	Tasso di falsi positivi	Supervisionato	Elevato
Ailes, 2020 [36]	Nessun dato mancante	Non effettuato	Convalida incrociata 5-fold stratificata	Non effettuata	Accuratezza Sensibilità Specificità Valore predittivo positivo	Supervisionato	Basso
Alyahya, 2016 [37]	Nessun dato mancante	Non effettuato	Convalida incrociata 10-fold	Non effettuata	Accuratezza	Supervisionato	Basso
Anderson, 2016 [38]	Non dichiarato	Non effettuato	Convalida incrociata con Bootstrap (30% del database originario riservato alla validazione)	Non effettuata	Accuratezza Sensibilità Specificità Valore predittivo positivo Valore predittivo negativo AUC	Supervisionato	Basso
Audureau, 2020 [39]	Imputazione	Non effettuato	Non effettuata	Effettuata su altra coorte	AUC	Supervisionato	Minimo
Banda, 2019 [40]	Non dichiarato	Effettuato su subset di training tramite convalida incrociata 10-fold annidata	Convalida incrociata 10-fold	Effettuata su record sanitari elettronici da altro ospedale	Sensibilità Specificità Precisione F1-score AUC AUPRC	Supervisionato	Minimo
Berger, 2020 [41]	Rimozione dei record con dati mancanti	Non effettuato	Convalida incrociata 5-fold	Non effettuata	AUC R <sup>2</sup>	Supervisionato	Basso
Bernardini, 2020 [42]	Imputazione	Non effettuato	Convalida incrociata 10-fold	Non effettuata	Accuratezza Precisione Recall F1-score AUC AUPRC	Supervisionato	Basso

AUC: area sotto la curva; AUPRC: area sotto la curva di Precision-Recall

(Segue)

Tabella 3 (Continua). Approccio metodologico adottato dagli studi nell'utilizzo degli algoritmi di machine learning.

Primo autore, anno	Gestione dei dati mancanti	Modalità di tuning dell'algoritmo	Modalità di validazione interna	Modalità di Validazione esterna	Modalità di valutazione della performance	Approccio	Rischio di overfitting
Bourdès, 2011 [43]	Nessun dato mancante	Non effettuato	Convalida incrociata non specificata	Non effettuata	AUC	Supervisionato	Elevato
Briggs, 2010 [44]	Nessun dato mancante	Non effettuato	Bootstrap cross-validation (% database originario usato per validazione non dichiarato)	Non effettuata	Nessuna misura riportata	Supervisionato	Basso
Cabrera-Quiros, 2021 [45]	Nessun dato mancante	Non effettuato	Convalida incrociata Leave-n-out (n non dichiarato)	Non effettuata	Accuratezza Sensibilità Specificità Valore predittivo positivo Valore predittivo negativo Precisione AUC	Supervisionato	Elevato
Chandir, 2018 [46]	Imputazione	Effettuato su subset di training	Convalida incrociata con Bootstrap (25% del database originario riservato alla validazione)	Non effettuata	Accuratezza AUC	Supervisionato	Elevato
Chau, 2016 [47]	Imputazione (Uso di media/mediana)	Non applicabile	Non applicabile	Non applicabile	Non applicabile	Non supervisionato	Non applicabile
Das, 2017 [48]	Rimozione record con dati mancanti	Non effettuato	Validazione con Holdout (% del dataset originario non dichiarata)	Non effettuata	Sensibilità Valore predittivo positivo AUC	Supervisionato	Basso
Dean, 2020 [49]	Nessun dato mancante	Non effettuato	Non effettuata	Effettuata su altri dati prodotti autonomamente dai ricercatori	Accuratezza Sensibilità Specificità AUC	Supervisionato	Minimo

AUC: area sotto la curva

(Segue)

Tabella 3 (Continua). Approccio metodologico adottato dagli studi nell'utilizzo degli algoritmi di machine learning.

Primo autore, anno	Gestione dei dati mancanti	Modalità di tuning dell'algoritmo	Modalità di validazione interna	Modalità di Validazione esterna	Modalità di valutazione della performance	Approccio	Rischio di overfitting
Deleger, 2013 [50]	Non dichiarato	Non effettuato	Convalida incrociata 10-fold	Effettuata sul 10% del dataset originario, separato prima di qualunque intervento	Specificità Precisione Recall F1-score Tabella di contingenza	Supervisionato	Basso
Hastings, 2020 [51]	Imputazione (Uso di media/mediana)	Effettuato su subset separato (n=255)	Validazione con Holdout (50% del dataset originario)	Non effettuata	AUC	Supervisionato	Basso
Hegde, 2019 [52]	Rimozione dei record con dati mancanti/Principal component analysis	Non effettuato	Convalida incrociata 10-fold	Effettuata sul 10% del dataset originario, separato prima di qualunque intervento	Accuratezza bilanciata AUC Tabella di contingenza	Supervisionato	Basso
Huang, 2010 [53]	Nessun dato mancante	Non effettuato	Validazione Holdout (20% del dataset originario)	Non effettuata	Nessuna misura riportata	Supervisionato	Basso
Katrakazas, 2020 [54]	Nessun dato mancante	Effettuato su subset di training tramite validazione incrociata	Validazione Holdout (30% del dataset originario)	Non effettuata	Specificità Precisione Recall F1-score Tasso di falsi positivi	Supervisionato	Basso
Kimura, 2019 [55]	Nessun dato mancante	Non effettuato	Validazione Holdout (34% del dataset originario)	Non effettuata	RMSE (esito continuo)	Supervisionato	Basso
Kop, 2016 [56]	Arricchimento semantico	Effettuato su subset di training	Convalida incrociata 5-fold stratificata	Non effettuata	Precisione Recall F1-score AUC	Supervisionato	Elevato
Pedersen, 2019 [57]	Rimozione variabili con più dell'80% di valori mancanti Rimozione dei record con dati mancanti	Effettuato su subset di training	Convalida incrociata 5-fold stratificata	Non effettuata	Accuratezza Precisione AUC	Supervisionato	Elevato

AUC: area sotto la curva; RMSE: radice dell'errore quadratico medio

(Segue)

Tabella 3 (Continua). Approccio metodologico adottato dagli studi nell'utilizzo degli algoritmi di machine learning.

Primo autore, anno	Gestione dei dati mancanti	Modalità di tuning dell'algoritmo	Modalità di validazione interna	Modalità di Validazione esterna	Modalità di valutazione della performance	Approccio	Rischio di overfitting
Pérez, 2021 [58]	Non dichiarato	Non applicabile	Non applicabile	Non applicabile	Non applicabile	Non supervisionato	Non applicabile
Qazi, 2021 [59]	Rimozione dei record con dati mancanti	Non effettuato	Validazione con Holdout (30% del dataset originario)	Non effettuata	Accuratezza Sensibilità Specificità Precisione Recall F1-score AUC	Supervisionato	Basso
Rozenbaum, 2021 [60]	Rimozione dei record con dati mancanti	Effettuato su subset di training tramite validazione incrociata	Validazione con Holdout (20% del dataset originario)	Non effettuata	AUC	Supervisionato	Basso

AUC: area sotto la curva

### **3.2 Studi sul disegno di interventi di sanità pubblica**

I 5 studi che hanno riguardato il disegno di interventi di sanità pubblica [61–65] sono stati pubblicati tutti nel 2021, ad eccezione dello studio di Chung et al., pubblicato nel 2018 [61]. Le loro caratteristiche sono riportate in tabella 4. I paesi in cui gli studi sono stati condotti sono stati: Australia [61], Canada [64], Cina [65], Turchia [63], USA [62]. 3 studi hanno avuto come riferimento la popolazione adulta [62–64], 1 studio si è occupato dell'accessibilità dei consigli di sanità pubblica da parte di gruppi di popolazione vulnerabili e da parte dei migranti [65] e infine 1 studio si è occupato di organizzazione dei servizi sanitari [61]. Per quanto riguarda le fonti dei dati, due studi hanno utilizzato dati elaborati da autorità nazionali statistiche e amministrative [61,62], uno studio dati sulla mobilità ricavata da smartphone [62], uno studio dati da autorità sanitarie nazionali e sovranazionali [65], e un ultimo studio dati e metadati da ricerca sistematica della letteratura scientifica [64]. Un unico studio ha utilizzato dati provenienti da simulazioni [63].

Per quanto riguarda metodologia, obiettivi e risultati principali, Chung et al. [61] hanno utilizzato una metodologia basata su artificial neural network per sviluppare un sistema di supporto decisionale basato sulle evidenze per l'organizzazione dei servizi di salute mentale, riuscendo in questo modo ad

individuare pattern di utilizzo dei servizi e a perfezionare l'elaborazione di indicatori chiave di prestazione. Katragadda ed al. [62] hanno utilizzato un approccio di tipo association rule mining per studiare l'effetto di diversi interventi non farmaceutici nel contenimento della pandemia di COVID-19, individuando nell'obbligatorietà delle maschere facciali una caratteristica comune a quegli interventi in grado di influenzare il decorso della pandemia. Lo stesso obiettivo è stato perseguito da Tunaligil et al. [63], utilizzando in questo caso una metodologia basata su artificial neural network, e individuando nell'incremento del tasso di vaccinati contro l'influenza la politica più efficace nel limitare la mortalità, e nel limitare i contatti sociali quella più efficace nel ridurre il numero di casi di SARS-CoV-2. Warin et al. [64] hanno utilizzato un approccio di tipo natural language processing nell'analisi dei metadati della letteratura scientifica prodotta su SARS-CoV-2 al fine di individuare elementi che possano informare le decisioni di politica sanitaria, proponendo un framework che permette di identificare i temi chiave della ricerca e le istituzioni di ricerca maggiormente attive. Infine, Xie et al. [65], utilizzando un approccio misto (ensemble) (Natural Language Processing, Support Vector Machine, Naïve Bayes) hanno analizzato l'accessibilità delle raccomandazioni sulla salute pubblica elaborate da diverse autorità sanitarie nazionali e sovranazionali riguardanti la pandemia da SARS-CoV-2, individuandone le caratteristiche



fondamentali ed elaborando un tool per testare l'accessibilità di nuove raccomandazioni.

Tabella 4. Caratteristiche degli studi sul disegno di interventi di sanità pubblica inclusi nella revisione sistematica.

Primo autore, anno	Nazione	Campo di interesse	Target	Fonte dei dati	Metodologia utilizzata	Obiettivo dello studio	Risultati principali
Chung, 2018 [61]	Australia	Sanità Pubblica	Organizzazione dei servizi sanitari	Autorità statistica nazionale Autorità Amministrativa nazionale	Artificial Neural Network	Sviluppare un sistema di supporto decisionale basato sulle evidenze per l'organizzazione dei servizi di salute mentale	La rete di mappe auto-organizzate, combinata con l'analisi collaborativa basata su esperti, ha spiegato efficacemente i pattern globali e locali e ha perfezionamento degli indicatori chiave di prestazione tramite interpretazione analitica.
Katragadda, 2021 [62]	USA	Sanità pubblica	Popolazione generale	Autorità statistica nazionale Autorità amministrativa nazionale Dati di mobilità da smartphone	Association Rule Mining	Comprendere gli effetti di diversi interventi non farmaceutici nel contenimento del COVID-19	L'utilizzo obbligatorio o meno delle maschere facciali sembra essere uno dei fattori comuni degli interventi in grado di influenzare il decorso della pandemia
Tunaligil, 2021 [63]	Turchia	Sanità pubblica	Popolazione generale	Simulazione	Artificial Neural Network	Valutare gli effetti di varie politiche in merito alla gestione della pandemia da SARS-CoV-2	Lo studio ha dimostrato che "incrementare il tasso di vaccinazione contro l'influenza" è la misura più efficace nel ridurre la mortalità, mentre "limitare i contatti sociali" è la misura più efficace nel contenere i casi
Warin, 2021 [64]	Canada	Sanità pubblica	Popolazione generale	Letteratura scientifica sul COVID-19	Natural Language Processing	Utilizzare la ricerca globale sui coronavirus per identificare gli elementi critici che possono aiutare informare le decisioni di politica sanitaria.	Usando il coronavirus come esempio, questo studio ha proposto un framework per identificare i temi chiave e le istituzioni di ricerca che conducono le ricerche più rilevanti.
Xie, 2021 [65]	Cina	Sanità pubblica	Gruppi vulnerabili o stranieri	Autorità sanitarie nazionali e sovranazionali	Natural Language Processing Support Vector Machine Naive Bayes	Identificare le migliori pratiche da adottare per rendere i consigli sulla salute pubblica riguardo il COVID-19 accessibili e inclusivi.	Lo studio ha dimostrato che l'esperienza fatta nell'ideazione di consigli pubblici accessibili può essere traslata in strumenti di supporto per l'elaborazione di consigli di sanità pubblica per eventi e crisi future

### **3.3 Studi sulla sorveglianza delle malattie/modelli di previsione**

I 7 studi che hanno investigato la sorveglianza delle malattie/modelli di previsione [66–72] sono stati pubblicati tutti tra il 2020 e il 2021 e tutti hanno avuto come oggetto di studio l'evoluzione della pandemia da SARS-CoV-2, ad eccezione dello studio di Choi et al. [68], che ha invece riguardato la stima dei decessi settimanali per suicidio. Le loro caratteristiche sono riportate in tabella 5. I paesi in cui gli studi sono stati condotti sono stati: Arabia Saudita [66,71], Bangladesh [69], Francia [70], Spagna [67], e USA [68,72]. Tutti gli studi hanno adottato un disegno di tipo longitudinale. Tutti gli studi riguardanti la pandemia da SARS-CoV-2 hanno utilizzato una metodologia basata su Artificial Neural Network, ad eccezione di Watson et al. [72], dove è invece stato usato un approccio basato su Random Forest. In generale, l'obiettivo di tutti questi studi è stato quello di prevedere le frequenze di nuovi casi giornalieri di positività a SARS-CoV-2. Oltre a questo, obiettivi specifici sono stati: la predizione dei casi critici giornalieri, in Atawneh et al. [66], dei ricoveri giornalieri in Shahin et al. [71], e dei decessi giornalieri in Oshinubi et al. [70], Shahin et al. [71] e Watson et al. [72]. L'obiettivo dell'unico studio che non ha riguardato la pandemia da SARS-CoV-2 è stato quello di costruire un modello predittivo dei decessi per suicidio negli USA [68]. Nel complesso, tutti gli studi hanno dimostrato una buona accuratezza dei modelli nella predizione. Due studi [69,70],

hanno utilizzato dati sulla pandemia presenti in repository pubblici (i.e. [worldometers.info](http://worldometers.info)), mentre i restanti 5 studi hanno utilizzato dati provenienti direttamente dalle autorità sanitarie nazionali di riferimento. Un solo studio, Choi et al. [68], ha integrato i dati sanitari con dati di tipo amministrativo proveniente da vari istituti statistici nazionali e con dati ricavati dall'analisi delle tendenze di ricerca su internet e dall'analisi di social network.

Tabella 5. Caratteristiche degli studi su sorveglianza delle malattie/modelli di previsione inclusi nella revisione sistematica.

Primo autore, anno	Nazione	Disegno dello studio	Campo di interesse	Fonte dei dati	Metodologia utilizzata	Obiettivo dello studio	Risultati dello studio
Atawneh, 2020	Arabia Saudita	Longitudinale	Malattie Infettive	Autorità sanitaria nazionale	Artificial Neural Network	Prevedere la frequenza cumulativa dei nuovi casi e dei casi critici di COVID-19	Lo studio ha dimostrato che il modello è in grado di predire in modo altamente accurato i casi cumulativi confermati e critici di COVID-19
Baltas, 2020	Spagna	Longitudinale	Malattie Infettive	Autorità sanitaria nazionale	Artificial Neural Network	Prevedere la frequenza dei nuovi casi di COVID-19	Il modello predice in modo affidabile i casi di COVID-19, e si è dimostrato capace di generalizzare i risultati anche quando i dati registrati non seguono la tipica risposta SIR
Choi, 2020	USA	Longitudinale	Psichiatria	Autorità sanitaria nazionale Istituti statistici nazionali Tendenze di ricerca Social network	Random Forest	Stimare i decessi settimanali per suicidio	Il modello si è dimostrato in grado di predire con una correlazione elevata il numero di suicidi settimanali e il suo trend
Chowdhury, 2021	Bangladesh	Longitudinale	Malattie Infettive	Repository pubblico ( <a href="https://www.worldometers.info/">https://www.worldometers.info/</a> )	Artificial Neural Network	Identificare il modello che prevedere meglio la frequenza dei nuovi casi di COVID-19	Il modello Long short-term neural network si è rivelato il migliore nella predizione dei nuovi casi di COVID-19 in Arabia Saudita
Oshinubi, 2021	Francia	Longitudinale	Malattie Infettive	Repository pubblico ( <a href="https://www.worldometers.info/">https://www.worldometers.info/</a> )	Artificial Neural Network	Identificare il modello migliore nel prevedere la frequenza dei nuovi casi giornalieri di COVID-19 e delle morti giornaliere alle diverse ondate della pandemia	Il modello Multilayer Perceptron si è rivelato il più accurato nel predire sia i nuovi casi che i decessi giornalieri
Shahin, 2021	Arabia Saudita	Longitudinale	Malattie Infettive	Autorità sanitaria nazionale	Artificial Neural Network	Prevedere la frequenza dei nuovi casi giornalieri, dei ricoveri giornalieri e dei decessi giornalieri per COVID-19	Lo studio ha dimostrato che il modello è in grado di predire in modo altamente accurato i nuovi casi, i ricoveri e i decessi giornalieri per COVID-19
Watson, 2021	USA	Longitudinale	Malattie Infettive	Autorità sanitaria nazionale	Random Forest	Prevedere la frequenza dei nuovi casi giornalieri e dei decessi giornalieri per COVID-19	Lo studio ha dimostrato che il modello è in grado di predire in modo altamente accurato i casi e i decessi giornalieri per COVID-19

## 4. Discussione

I risultati di questa tesi hanno mostrato come l'interesse verso le metodiche di machine learning stia iniziando a coinvolgere anche campi della medicina da sempre interesse della sanità pubblica, come testimoniato dai 26 studi che hanno riguardato l'individuazione di profili di rischio nella popolazione generale e pediatrica, o la gestione di condizioni che interessano ampie fasce della popolazione, come le malattie croniche o la patologia neoplastica. Il fatto che questo tipo di studi costituisca più della metà degli studi inclusi nella revisione sistematica, e che la loro presenza risalga già al 2010, non è sorprendente, essendo storicamente le metodiche di IA molto efficienti in questo campo ed ampiamente utilizzate in altre branche specialistiche della medicina [16]. Gli studi si configurano principalmente come classici studi analitici osservazionali di coorte, caso-controllo, o trasversali, in cui la tradizionale analisi statistica basata sul test di ipotesi è stata sostituita o affiancata da metodiche di IA riferibili alla branca del Machine Learning. È da rilevare come 15 di questi studi hanno utilizzato per le loro analisi dati contenuti in EMR raccolti di routine dalle organizzazioni sanitari o derivate da dati raccolti a livello centrale da varie autorità sanitarie ed amministrative. Il dato è importante perché l'utilizzo efficace di questi dati già disponibili, strutturati e non strutturati, normalmente raccolti principalmente a fini amministrativi, è una

condizione che può facilitare il diffondersi di questo tipo di strumenti nella corrente pratica clinica e di sanità pubblica [73]. La maggiore prevalenza di studi di questo tipo effettuati negli USA potrebbe riflettere la grande diffusione e standardizzazione degli EMR a livello locale, statale e federale. In effetti, sin dal 2004 negli USA è stata incentivata la raccolta in formato elettronico delle informazioni sanitarie e la loro raccolta a livello centrale [74]. Per quanto riguarda la tipologia e la varietà di dati utilizzati dagli studi sulla predizione del rischio, la grande maggioranza degli studi ha utilizzato più di una tipologia di dati. L'integrazione di dati di natura e origine diversa è una dei vantaggi più evidenti dell'utilizzo delle metodiche di machine learning [75], permettendo stime più accurate del rischio rispetto alle tradizionali metodologie statistiche, per le quali la multidimensionalità dei dati è tradizionalmente difficile da gestire [16]. I dati più utilizzati sono risultati quelli tradizionalmente utilizzati in questo tipo di indagini, vale a dire dati anamnestici e sociodemografici, sebbene alcuni studi abbiano integrato nella loro analisi informazioni sul consumo dei farmaci, sull'utilizzo dei servizi sanitari, informazioni "omiche", ambientali e comportamentali. È interessante notare come, nonostante la sempre maggiore ubiquità di sensori indossabili e applicazioni per smartphone in grado di raccogliere numerose tipologie di dati [76], solo uno studio ha utilizzato dati ricavati da dispositivi di questo tipo [55], in grado di

raccogliere informazioni sulla sedentarietà e sull'interazione sociale dei soggetti in studio. La mancanza di questo tipo di dati potrebbe riflettere la difficoltà a far accettare questo tipo di rilevazione in soggetti sani, come quelli che potrebbero essere interessati da studi che investighino fattori di rischio, per i lunghi periodi che spesso richiedono per la loro esecuzione. Sebbene esistano metodi per quantificare la numerosità campionaria appropriata per numerose metodiche di machine learning [77,78], nessuno degli studi ha riportato giustificazioni per le dimensioni del campione utilizzato. In ogni caso, solo sette studi hanno utilizzato un campione inferiore a 1000, mentre ben 6 studi hanno utilizzato dati provenienti da oltre 40000 record, indicando in generale l'utilizzo di grandi campioni negli studi. Tuttavia, le dimensioni del campione dovrebbero sempre essere adeguatamente stimate, anche in considerazione del fatto che le metodiche di machine learning possono richiedere campioni insolitamente grandi per raggiungere livelli di accuratezza stabili [79]. Numerosi studi hanno utilizzato diverse tipologie di algoritmi allo scopo di identificare il modello con la migliore performance. Questo dato potrebbe rivelare una certa immaturità delle metodiche utilizzate, indicando la mancanza di pratiche ben definite e metodologie standardizzate. Questo risulta evidente anche a livello dell'approccio metodologico utilizzato. Quasi un terzo degli studi ha utilizzato approcci non adeguati alle modalità di tuning, training e



validazione degli algoritmi di machine learning. Come già ricordato, la corretta esecuzione di queste azioni è fondamentale per l'elaborazione di un modello la cui performance sia generalizzabile in popolazioni diverse da quella sulla quale è stato costruito. Particolare importanza riveste la separazione netta del dataset di origine in subset indipendenti per ciascuna delle fasi previste dall'apprendimento (tuning, training e validazione), o l'applicazione di appropriate metodiche di partizionamento tramite convalida incrociata, annidata nel caso di esecuzione della fase di tuning [32]. Particolarmente significativo sembra essere il fatto che solo tre studi [39,40,49] hanno applicato tecniche di validazione esterna appropriate [33], che è in effetti il metodo più efficace per stimare la performance degli algoritmi in condizioni diverse da quelle di elaborazione e di evitare la sovrastima della loro performance. La non applicazione dell'approccio metodologico corretto per lo sviluppo di questi sistemi può rallentare la loro adozione nella pratica clinica e di sanità pubblica, anche in considerazione delle difficoltà nel valutare il valore di questi interventi all'interno delle classiche cornici regolatorie [80]. La necessità di valutare rigorosamente queste nuove tecnologie scaturisce anche dalla necessità di controbilanciare le aspettative a volte eccessive e dal carattere quasi fideistico che le circonda, alimentate da convinzioni quali la necessaria

superiorità di un risultato ottenuto da una intelligenza artificiale o l'ineluttabilità di una loro presenza pervasiva nel futuro [81].

La revisione ha inoltre evidenziato che la maggior parte di questi studi ha effettuato una qualche forma di valutazione dell'effetto delle variabili utilizzate nel modello, sotto forma generalmente di score di importanza della variabile e solo raramente nella forma di tradizionali Odds Ratio e Rischi Relativi. La possibilità di studiare l'impatto delle singole variabili all'interno del modello permette comunque di ottenere un certo grado di interpretabilità dell'algoritmo, e di allontanarsi dalla concezione di questi come di black box [82]. L'interpretabilità, e la trasparenza degli algoritmi in generale, è per gli operatori sanitari una tra le condizioni che più incidono sulla fiducia in queste metodiche [83]. L'importanza di rendere interpretabili gli algoritmi e di ottenere stime dell'effetto delle variabili usate per alimentarlo è inoltre considerata fondamentale dall'epidemiologia classica per permettere l'integrazione di queste tecniche nella pratica di sanità pubblica [84]. La mancanza di una più precisa stima dell'effetto delle esposizioni può considerarsi una mancanza di questi studi, anche considerando l'esistenza di metodiche che permettono di individuare all'interno della popolazione in studio sottogruppi per i quali l'effetto dell'esposizione (o del trattamento) è differente dal campione nel suo

complesso, permettendo quindi di individuare specifiche popolazioni a cui applicare particolari politiche di prevenzione o indicazioni terapeutiche [85].

Accanto a questi studi, la revisione sistematica ha individuato applicazioni di queste metodiche a campi maggiormente innovativi e potenzialmente dal maggior impatto sulla salute di popolazione, come è quello del disegno di interventi di sanità pubblica [86]. Il relativamente scarso numero di articoli inclusi nella revisione, pubblicati a partire dal 2018, potrebbe comunque indicare che l'applicazione rigorosa di queste metodologie a questo campo di studio è ancora nella sua fase iniziale. È comunque evidente che la pandemia da SARS-CoV-2, e le successive implementazioni di policy di sanità pubblica che hanno fortemente condizionato la normale vita sociale e lavorativa delle comunità coinvolte [87] e ridotto il benessere mentale di ampie fasce della popolazione [88], ha fornito un forte stimolo a questo ambito di ricerca. In effetti, due studi hanno cercato di stimare l'efficacia di diverse misure non farmaceutiche nella prevenzione della diffusione di SARS-CoV-2, ricavando importanti indicazioni sulle diverse policy applicate. Particolarmente innovativo appare lo studio di Xie et al. [65], che si è concentrato sulla creazione di consigli preventivi da parte delle autorità sanitarie comprensibili e accessibili alle fasce più svantaggiate della

popolazione. L'accessibilità di tali contenuti è una condizione necessaria affinché possano essere efficacemente fruibili, anche in considerazione dell'elevata prevalenza di alfabetizzazione sanitaria inadeguata spesso presente nella popolazione cui si rivolge [89]. Una tematica altrettanto importante, l'organizzazione dei servizi sanitari, è stata affrontata da Chung et al. [61]. La possibilità di utilizzare dati di tipo sanitario e amministrativo nella gestione dell'erogazione dell'offerta sanitaria è una delle promesse più allettanti di queste metodologie, permettendo una più vasta accessibilità ed efficienza dei servizi sanitari [8]. In questi studi, le metodiche maggiormente utilizzate sono state l'Artificial Neural Network e il Natural Language Processing, entrambe in grado di gestire efficacemente le forme di dati non-strutturate o semi-strutturate di dati che sono spesso caratteristici di questa tipologia di analisi.

Infine, la revisione sistematica ha individuato sette studi riguardanti la sorveglianza delle malattie e l'elaborazione di modelli di previsione, anche in questo caso pubblicati recentemente (2020-2021). Come per gli studi sul disegno di interventi di sanità pubblica, appare evidente la spinta fornita dall'attuale pandemia di SARS-CoV-2. La ricerca si è concentrata principalmente sull'elaborazione di modelli in grado di prevedere con elevata precisione i nuovi casi giornalieri e quindi l'andamento futuro

dell'epidemia, con buoni risultati in tutti gli studi. È da evidenziare però come questi modelli abbiano preso in considerazione i soli dati pregressi di contagiati, ricoverati, deceduti e rimessi, senza utilizzare altre tipologie di dati, se non in alcuni casi le caratteristiche sociodemografiche della popolazione. Sebbene tale approccio possa essere considerato sicuramente limitato, tali studi possono contribuire a creare la base di modelli più accurati dei classici modelli SIR (susceptible-infected-recovered), da utilizzare come base per future applicazioni di simulazione di svariati interventi di prevenzione, farmacologica o meno, nonché per l'integrazione di altri dati in grado di migliorare le previsioni, come il grado di mobilità della popolazione, le possibilità di aggregazione e i contenuti dei social media [90]. L'unico studio che si è interessato ad un fenomeno diverso dalla pandemia di SARS-CoV-2, i decessi per suicidio, dimostra che l'integrazione di dati amministrativi e sanitari raccolti di routine, insieme alle attività sui motori di ricerca e sui social network è applicabile a patologie di natura non infettiva, ma non per questo con un impatto non significativo per la salute della popolazione, permettendo di poter individuare gruppi di popolazione o particolari località che possano beneficiare di specifici interventi preventivi [68].

## 5. Conclusioni

I risultati di questa revisione sistematica hanno evidenziato come l'applicazione ai campi di interesse della sanità pubblica delle metodiche di IA sia ancora ad uno stadio iniziale, sebbene l'interesse nei loro confronti sia in continua ascesa. In particolare, mentre gli studi di predizione del rischio compaiono già dal 2010, per gli studi sul disegno degli interventi di sanità pubblica il primo studio risulta pubblicato nel 2018, mentre per gli studi sui modelli di previsione addirittura nel 2020. Tale andamento è probabilmente un riflesso sia dell'affinamento delle metodiche di analisi, sia all'aumento della disponibilità sia materiale che legale di grandi aggregati di dati [91]. Particolarmente per il disegno di interventi di sanità pubblica e per i modelli di previsione, è innegabile la spinta data dall'attuale pandemia da SARS-CoV-2 in questo campo. Tuttavia, come evidenziato, oltre la metà degli studi selezionati riguarda ancora la parte storicamente più consolidata nell'utilizzo di questi strumenti, ovvero la stima del rischio di esiti in una determinata popolazione. Nonostante ci siano grosse aspettative sul contributo che tali metodiche potranno portare nella pratica clinica [30], l'utilizzo di algoritmi basato sull'Intelligenza Artificiale sembra essere tuttora parecchio limitato. In effetti, le applicazioni basate su IA approvate dalla Food and Drug Administration risultavano appena 29 nel 2020 [92]. Sembra evidente che esistano dei problemi di natura

metodologica che possono rallentare la loro approvazione e adozione. In effetti, le agenzie regolatorie nazionali e internazionali si stanno mostrando molto caute nell'approvazione di queste tecnologie, che sono caratterizzate dalla capacità di continuare ad apprendere e modificare la loro performance anche in seguito all'eventuale approvazione e i cui processi interni possono non essere del tutto trasparenti per i valutatori e i decisori [81,93]. Un'ulteriore problematica relativo alla loro implementazione è quella dell'imparzialità di tali algoritmi [94]. Come noto, qualora i dati su cui gli algoritmi di machine learning siano affetti da distorsioni, il modello che ne deriverà, per quanto preciso, non potrà far altro che riproporre ed amplificare tali distorsioni [95], potendo potenzialmente incrementare comportamenti discriminatori nei confronti di determinate fasce della popolazione, disuguaglianze di salute già presenti, differenze nell'accessibilità dei servizi, Per ovviare in parte a queste problematiche, i ricercatori dovrebbero essere quanto più precisi e trasparenti possibili nella descrizione del campione utilizzato nello sviluppo dell'algoritmo e del contesto nel quale è stato testato [96,97]. Inoltre, l'aumento dei costi per i servizi e le organizzazioni che comporta l'applicazione di queste metodiche, che presuppongono l'esistenza di infrastrutture e competenze altamente specializzate, potrebbe impedirne un'equa diffusione [81].

In conclusione, la continua e sempre in aumento disponibilità di dati di varia natura, sanitaria e non, spingerà sempre più all'utilizzo di queste metodiche nella sanità pubblica, ma queste innovazioni dovranno essere attentamente guidate, sia nell'elaborazione di metodologie standardizzate e condivise nel loro sviluppo, che nella valutazione della loro efficacia e dell'impatto che la loro introduzione può avere sui sistemi sanitari e sulla società, per far sì che le loro potenzialità possano apportare un reale beneficio a tutta la popolazione.



## 6. Bibliografia

1. De Mauro, A.; Greco, M.; Grimaldi, M. A Formal Definition of Big Data Based on Its Essential Features. *Libr. Rev.* 2016, *65*, 122–135, doi:10.1108/LR-06-2015-0061.
2. Eaton, C.; Tom, D.; Dirk, D.; George, L.; Paul, Z. *Understanding Big Data*; McGraw-Hill, 2016; ISBN 978-0-07-179053-6.
3. Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C.; Hung Byers, A. *Big Data: The next Frontier for Innovation, Competition, and Productivity*; McKinsey Global Institute, 2011;
4. Reimer, A.P.; Madigan, E.A. Veracity in Big Data: How Good Is Good Enough. *Health Informatics J.* 2019, *25*, 1290–1298, doi:10.1177/1460458217744369.
5. Goli-Malekabadi, Z.; Sargolzaei-Javan, M.; Akbari, M.K. An Effective Model for Store and Retrieve Big Health Data in Cloud Computing. *Comput. Methods Programs Biomed.* 2016, *132*, 75–82, doi:10.1016/j.cmpb.2016.04.016.
6. Costa, F.F. Big Data in Biomedicine. *Drug Discov. Today* 2014, *19*, 433–440, doi:10.1016/j.drudis.2013.10.012.
7. Auffray, C.; Balling, R.; Barroso, I.; Bencze, L.; Benson, M.; Bergeron, J.; Bernal-Delgado, E.; Blomberg, N.; Bock, C.; Conesa, A.; et al. Making Sense of Big Data in Health Research: Towards an EU Action Plan. *Genome Med.* 2016, *8*, 71, doi:10.1186/s13073-016-0323-y.
8. Wang, Y.; Kung, L.; Byrd, T.A. Big Data Analytics: Understanding Its Capabilities and Potential Benefits for Healthcare Organizations. *Technol. Forecast. Soc. Change* 2018, *126*, 3–13, doi:10.1016/j.techfore.2015.12.019.
9. Hulsén, T.; Jamuar, S.S.; Moody, A.R.; Karnes, J.H.; Varga, O.; Hedensted, S.; Spreafico, R.; Hafner, D.A.; McKinney, E.F. From Big Data to Precision Medicine. *Front. Med.* 2019, *6*.
10. Szlezák, N.; Evers, M.; Wang, J.; Pérez, L. The Role of Big Data and Advanced Analytics in Drug Discovery, Development, and Commercialization. *Clin. Pharmacol. Ther.* 2014, *95*, 492–495, doi:10.1038/clpt.2014.29.
11. He, K.Y.; Ge, D.; He, M.M. Big Data Analytics for Genomic Medicine. *Int. J. Mol. Sci.* 2017, *18*, 412, doi:10.3390/ijms18020412.
12. Bansal, S.; Chowell, G.; Simonsen, L.; Vespignani, A.; Viboud, C. Big Data for Infectious Disease Surveillance and Modeling. *J. Infect. Dis.* 2016, *214*, S375–S379, doi:10.1093/infdis/jiw400.
13. Mahamoud, A.; Roche, B.; Homer, J. Modelling the Social Determinants of Health and Simulating Short-Term and Long-Term Intervention Impacts for the City of Toronto, Canada. *Soc. Sci. Med.* 1982 2013, *93*, 247–255, doi:10.1016/j.socscimed.2012.06.036.
14. Chen, G.; Zhang, W. Application of Big Data Information System in Early Diagnosis, Treatment, and Nursing of Cervical Cancer Infected by Human Papillomavirus. *Expert Syst.* 2021, e12791, doi:10.1111/exsy.12791.
15. El Naqa, I.; Kosorok, M.R.; Jin, J.; Mierzwa, M.; Ten Haken, R.K. Prospects and Challenges for Clinical Decision Support in the Era of Big Data. *JCO Clin. Cancer Inform.* 2018, 1–12, doi:10.1200/CCI.18.00002.
16. Wiemken, T.L.; Kelley, R.R. Machine Learning in Epidemiology and Health Outcomes Research. *Annu. Rev. Public Health* 2020, *41*, 21–36, doi:10.1146/annurev-publhealth-040119-094437.

17. Downing, N.S.; Cloninger, A.; Venkatesh, A.K.; Hsieh, A.; Drye, E.E.; Coifman, R.R.; Krumholz, H.M. Describing the Performance of U.S. Hospitals by Applying Big Data Analytics. *PLoS ONE* 2017, *12*, e0179603, doi:10.1371/journal.pone.0179603.
18. Shmueli, G.; Koppius, O.R. Predictive Analytics in Information Systems Research. *MIS Q.* 2011, *35*, 553–572, doi:10.2307/23042796.
19. Health in the 21st Century : Putting Data to Work for Stronger Health Systems - Cap. 5 -Big Data: A New Dawn for Public Health? 2019, doi:doi.org/10.1787/e3b23f8e-en.
20. Rahimi, I.; Chen, F.; Gandomi, A.H. A Review on COVID-19 Forecasting Models. *Neural Comput. Appl.* 2021, doi:10.1007/s00521-020-05626-8.
21. Pastorino, R.; De Vito, C.; Migliara, G.; Glocker, K.; Binenbaum, I.; Ricciardi, W.; Boccia, S. Benefits and Challenges of Big Data in Healthcare: An Overview of the European Initiatives. *Eur. J. Public Health* 2019, *29*, 23–27, doi:10.1093/eurpub/ckz168.
22. Khoury, M.J.; Ioannidis, J.P.A. Big Data Meets Public Health. *Science* 2014, *346*, 1054–1055, doi:10.1126/science.aaa2709.
23. Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine Learning on Big Data: Opportunities and Challenges. *Neurocomputing* 2017, *237*, 350–361, doi:10.1016/j.neucom.2017.01.026.
24. Wang, C.; Chen, M.-H.; Schifano, E.; Wu, J.; Yan, J. Statistical Methods and Computing for Big Data. *Stat. Interface* 2016, *9*, 399–414, doi:10.4310/SII.2016.v9.n4.a1.
25. Mehta, N.; Pandit, A. Concurrence of Big Data Analytics and Healthcare: A Systematic Review. *Int. J. Med. Inf.* 2018, *114*, 57–65, doi:10.1016/j.ijmedinf.2018.03.013.
26. Shortliffe, E.H. A Rule-Based Computer Program for Advising Physicians Regarding Antimicrobial Therapy Selection. *Proc. 1974 Annu. ACM Conf. - Vol. 2* 1974, 739, doi:10.1145/1408800.1408906.
27. Newquist, H.P. *The Brain Makers*; 1st edition.; Sams: Indianapolis, Ind, 1994; ISBN 978-0-672-30412-5.
28. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements Of Statistical Learning: Data Mining, Inference, And Prediction, Second Edition*; 2° edizione.; Springer Nature: New York, NY, 2009; ISBN 978-0-387-84857-0.
29. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT press, 2012; ISBN 978-0-262-01802-9.
30. Matheny, M.E.; Whicher, D.; Thadaney Israni, S. *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril.*; National Academy of Medicine: Washington, DC, 2022; ISBN 978-1-947103-17-7.
31. Alzubi, J.; Nayyar, A.; Kumar, A. Machine Learning from Theory to Algorithms: An Overview. *J. Phys. Conf. Ser.* 2018, *1142*, 012012, doi:10.1088/1742-6596/1142/1/012012.
32. Faes, L.; Liu, X.; Wagner, S.K.; Fu, D.J.; Balaskas, K.; Sim, D.A.; Bachmann, L.M.; Keane, P.A.; Denniston, A.K. A Clinician’s Guide to Artificial Intelligence: How to Critically Appraise Machine Learning Studies. *Transl. Vis. Sci. Technol.* 2020, *9*, 7, doi:10.1167/tvst.9.2.7.
33. Tohka, J.; van Gils, M. Evaluation of Machine Learning Algorithms for Health and Wellness Applications: A Tutorial. *Comput. Biol. Med.* 2021, *132*, 104324, doi:10.1016/j.compbiomed.2021.104324.
34. Sajjadian, M.; Lam, R.W.; Milev, R.; Rotzinger, S.; Frey, B.N.; Soares, C.N.; Parikh, S.V.; Foster, J.A.; Turecki, G.; Müller, D.J.; et al. Machine Learning in the Prediction of Depression Treatment Outcomes: A Systematic Review and Meta-Analysis. *Psychol. Med.* 2021, *51*, 2742–2751, doi:10.1017/S0033291721003871.

35. Adams, S.M.; Feroze, H.; Nguyen, T.; Eum, S.; Cornelio, C.; Harralson, A.F. Genome Wide Epistasis Study of On-Statins Cardiovascular Events with Iterative Feature Reduction and Selection. *J. Pers. Med.* 2020, *10*, 212, doi:10.3390/jpm10040212.
36. Ailes, E.C.; Zimmerman, J.; Lind, J.N.; Fan, F.; Shi, K.; Reefhuis, J.; Broussard, C.S.; Frey, M.T.; Cragan, J.D.; Petersen, E.E.; et al. Using Supervised Learning Methods to Develop a List of Prescription Medications of Greatest Concern during Pregnancy. *Matern. Child Health J.* 2020, *24*, 901–910, doi:10.1007/s10995-020-02942-2.
37. Alyahya, M.; Hijazi, H.; Alshraideh, H.; Alsharman, M.; Abdi, R.; Harvey, H. Integrating the Principles of Evidence Based Medicine and Evidence Based Public Health: Impact on the Quality of Patient Care and Hospital Readmission Rates in Jordan. *Int. J. Integr. Care* 2016, *16*, 12, doi:10.5334/ijic.2436.
38. Anderson, A.E.; Kerr, W.T.; Thames, A.; Li, T.; Xiao, J.; Cohen, M.S. Electronic Health Record Phenotyping Improves Detection and Screening of Type 2 Diabetes in the General United States Population: A Cross-Sectional, Unselected, Retrospective Study. *J. Biomed. Inform.* 2016, *60*, 162–168, doi:10.1016/j.jbi.2015.12.006.
39. Audureau, E.; Carrat, F.; Layese, R.; Cagnot, C.; Asselah, T.; Guyader, D.; Larrey, D.; Lédighen, V.D.; Ouzan, D.; Zoulim, F.; et al. Personalized Surveillance for Hepatocellular Carcinoma in Cirrhosis – Using Machine Learning Adapted to HCV Status. *J. Hepatol.* 2020, *73*, 1434–1445, doi:10.1016/j.jhep.2020.05.052.
40. Banda, J.M.; Sarraju, A.; Abbasi, F.; Parizo, J.; Pariani, M.; Ison, H.; Briskin, E.; Wand, H.; Dubois, S.; Jung, K.; et al. Finding Missed Cases of Familial Hypercholesterolemia in Health Systems Using Machine Learning. *Npj Digit. Med.* 2019, *2*, 1–8, doi:10.1038/s41746-019-0101-5.
41. Berger, J.S.; Haskell, L.; Ting, W.; Lurie, F.; Chang, S.-C.; Mueller, L.A.; Elder, K.; Rich, K.; Crivera, C.; Schein, J.R.; et al. Evaluation of Machine Learning Methodology for the Prediction of Healthcare Resource Utilization and Healthcare Costs in Patients with Critical Limb Ischemia—Is Preventive and Personalized Approach on the Horizon? *EPMA J.* 2020, *11*, 53–64, doi:10.1007/s13167-019-00196-9.
42. Bernardini, M.; Romeo, L.; Mancini, A.; Frontoni, E. A Clinical Decision Support System to Stratify the Temporal Risk of Diabetic Retinopathy. *IEEE Access* 2021, *9*, 151864–151872, doi:10.1109/ACCESS.2021.3127274.
43. Bourdès, V.; Ferrières, J.; Amar, J.; Amelineau, E.; Bonnevey, S.; Berlion, M.; Danchin, N. Prediction of Persistence of Combined Evidence-Based Cardiovascular Medications in Patients with Acute Coronary Syndrome after Hospital Discharge Using Neural Networks. *Med. Biol. Eng. Comput.* 2011, *49*, 947–955, doi:10.1007/s11517-011-0785-4.
44. Briggs, F.B.S.; Ramsay, P.P.; Madden, E.; Norris, J.M.; Holers, V.M.; Mikuls, T.R.; Sokka, T.; Seldin, M.F.; Gregersen, P.K.; Criswell, L.A.; et al. Supervised Machine Learning and Logistic Regression Identifies Novel Epistatic Risk Factors with PTPN22 for Rheumatoid Arthritis. *Genes Immun.* 2010, *11*, 199–208, doi:10.1038/gene.2009.110.
45. Cabrera-Quiros, L.; Kommers, D.; Wolvers, M.K.; Oosterwijk, L.; Arents, N.; van der Sluijs-Bens, J.; Cottaar, E.J.E.; Andriessen, P.; van Pul, C. Prediction of Late-Onset Sepsis in Preterm Infants Using Monitoring Signals and Machine Learning. *Crit. Care Explor.* 2021, *3*, e0302, doi:10.1097/CCE.0000000000000302.
46. Chandir, S.; Siddiqi, D.A.; Hussain, O.A.; Niazi, T.; Shah, M.T.; Dharma, V.K.; Habib, A.; Khan, A.J. Using Predictive Analytics to Identify Children at High Risk of Defaulting From a Routine Immunization Program: Feasibility Study. *JMIR Public Health Surveill.* 2018, *4*, e9681, doi:10.2196/publichealth.9681.
47. Chau, R.; Jenkins, M.A.; Buchanan, D.D.; Ait Ouakrim, D.; Giles, G.G.; Casey, G.; Gallinger, S.; Haile, R.W.; Le Marchand, L.; Newcomb, P.A.; et al. Determining the

- Familial Risk Distribution of Colorectal Cancer: A Data Mining Approach. *Fam. Cancer* 2016, *15*, 241–251, doi:10.1007/s10689-015-9860-6.
48. Das, L.T.; Abramson, E.L.; Stone, A.E.; Kondrich, J.E.; Kern, L.M.; Grinspan, Z.M. Predicting Frequent Emergency Department Visits among Children with Asthma Using EHR Data. *Pediatr. Pulmonol.* 2017, *52*, 880–890, doi:10.1002/ppul.23735.
  49. Dean, K.R.; Hammamieh, R.; Mellon, S.H.; Abu-Amara, D.; Flory, J.D.; Guffanti, G.; Wang, K.; Daigle, B.J.; Gautam, A.; Lee, I.; et al. Multi-Omic Biomarker Identification and Validation for Diagnosing Warzone-Related Post-Traumatic Stress Disorder. *Mol. Psychiatry* 2020, *25*, 3337–3349, doi:10.1038/s41380-019-0496-z.
  50. Deleger, L.; Brodzinski, H.; Zhai, H.; Li, Q.; Lingren, T.; Kirkendall, E.S.; Alessandrini, E.; Solti, I. Developing and Evaluating an Automated Appendicitis Risk Stratification Algorithm for Pediatric Patients in the Emergency Department. *J. Am. Med. Inform. Assoc.* 2013, *20*, e212–e220, doi:10.1136/amiajnl-2013-001962.
  51. Hastings, J.S.; Howison, M.; Inman, S.E. Predicting High-Risk Opioid Prescriptions before They Are Given. *Proc. Natl. Acad. Sci.* 2020, *117*, 1917–1923, doi:10.1073/pnas.1905355117.
  52. Hegde, H.; Shimpi, N.; Panny, A.; Glurich, I.; Christie, P.; Acharya, A. Development of Non-Invasive Diabetes Risk Prediction Models as Decision Support Tools Designed for Application in the Dental Clinical Environment. *Inform. Med. Unlocked* 2019, *17*, 100254, doi:10.1016/j.imu.2019.100254.
  53. Huang, S.; Xu, Y.; Yue, L.; Wei, S.; Liu, L.; Gan, X.; Zhou, S.; Nie, S. Evaluating the Risk of Hypertension Using an Artificial Neural Network Method in Rural Residents over the Age of 35 Years in a Chinese Area. *Hypertens. Res.* 2010, *33*, 722–726, doi:10.1038/hr.2010.73.
  54. Ktrakazas, C.; Antoniou, C.; Yannis, G. Identification of Driving Simulator Sessions of Depressed Drivers: A Comparison between Aggregated and Time-Series Classification. *Transp. Res. Part F Traffic Psychol. Behav.* 2020, *75*, 16–25, doi:10.1016/j.trf.2020.09.015.
  55. Kimura, N.; Aso, Y.; Yabuuchi, K.; Ishibashi, M.; Hori, D.; Sasaki, Y.; Nakamichi, A.; Uesugi, S.; Fujioka, H.; Iwao, S.; et al. Modifiable Lifestyle Factors and Cognitive Function in Older People: A Cross-Sectional Observational Study. *Front. Neurol.* 2019, *10*, 401, doi:10.3389/fneur.2019.00401.
  56. Kop, R.; Hoogendoorn, M.; Teije, A. ten; Büchner, F.L.; Slottje, P.; Moons, L.M.G.; Numans, M.E. Predictive Modeling of Colorectal Cancer Using a Dedicated Pre-Processing Pipeline on Routine Electronic Medical Records. *Comput. Biol. Med.* 2016, *76*, 30–38, doi:10.1016/j.combiomed.2016.06.019.
  57. Pedersen, D.H.; Mansourvar, M.; Sortsø, C.; Schmidt, T. Predicting Dropouts From an Electronic Health Platform for Lifestyle Interventions: Analysis of Methods and Predictors. *J. Med. Internet Res.* 2019, *21*, e13617, doi:10.2196/13617.
  58. Pérez, S.; German-Labaume, C.; Mathiot, S.; Goix, S.; Chamaret, P. Using Bayesian Networks for Environmental Health Risk Assessment. *Environ. Res.* 2022, *204*, 112059, doi:10.1016/j.envres.2021.112059.
  59. Qazi, S.; Usman, M.; Mahmood, A.; Abbasi, A.; Attique, M.; Nam, Y. Smart Healthcare Using Data-Driven Prediction of Immunization Defaulters in Expanded Program on Immunization (EPI). *Comput. Mater. Contin.* 2020, *66*, 589–602, doi:10.32604/cmcc.2020.012507.
  60. Rozenbaum, D.; Shreve, J.; Radakovich, N.; Duggal, A.; Jehi, L.; Nazha, A. Personalized Prediction of Hospital Mortality in COVID-19–Positive Patients. *Mayo Clin. Proc. Innov. Qual. Outcomes* 2021, *5*, 795–801, doi:10.1016/j.mayocpiqo.2021.05.001.

61. Chung, Y.; Salvador-Carulla, L.; Salinas-Pérez, J.A.; Uriarte-Uriarte, J.J.; Iruiñ-Sanz, A.; García-Alonso, C.R. Use of the Self-Organising Map Network (SOMNet) as a Decision Support System for Regional Mental Health Planning. *Health Res. Policy Syst.* 2018, *16*, 35, doi:10.1186/s12961-018-0308-y.
62. Katragadda, S.; Gottumukkala, R.; Bhupatiraju, R.T.; Kamal, A.M.; Raghavan, V.; Chu, H.; Kolluru, R.; Ashkar, Z. Association Mining Based Approach to Analyze COVID-19 Response and Case Growth in the United States. *Sci. Rep.* 2021, *11*, 18635, doi:10.1038/s41598-021-96912-5.
63. Tunaligil, V.; Meral, G.; Dabak, M.R.; Canbulat, M.; Demir, S.S. COVID-19 and the Flu: Data Simulations and Computational Modelling to Guide Public Health Strategies. *Fam. Pract.* 2021, *38*, i16–i22, doi:10.1093/fampra/cmab058.
64. Warin, T. Global Research on Coronaviruses: Metadata-Based Analysis for Public Health Policies. *JMIR Med. Inform.* 2021, *9*, e31510, doi:10.2196/31510.
65. Xie, W.; Ji, M.; Zhao, M.; Lam, K.-Y.; Chow, C.-Y.; Hao, T. Developing Machine Learning and Statistical Tools to Evaluate the Accessibility of Public Health Advice on Infectious Diseases among Vulnerable People. *Comput. Intell. Neurosci.* 2021, *2021*, e1916690, doi:10.1155/2021/1916690.
66. Atawneh, S.H.; Ghaleb, O.A.M.; Hussein, A.M.; Al-Madi, M.; Shehabat, B. A Time Series Forecasting for the Cumulative Confirmed and Critical Cases of the Covid-19 Pandemic in Saudi Arabia Using Autoregressive Integrated Moving Average (ARIMA) Model. *J. Comput. Sci.* 2020, *16*, 1278–1290, doi:10.3844/jcssp.2020.1278.1290.
67. Baltas, G.; Prieto, F.; Frantzi, M.; Garcia-Alonso, C.; Rodriguez, P. Data Driven Modelling of Coronavirus Spread in Spain. *Comput. Mater. Contin.* 2020, *64*, 1343–1357, doi:10.32604/cmc.2020.011243.
68. Choi, D.; Sumner, S.A.; Holland, K.M.; Draper, J.; Murphy, S.; Bowen, D.A.; Zwald, M.; Wang, J.; Law, R.; Taylor, J.; et al. Development of a Machine Learning Model Using Multiple, Heterogeneous Data Sources to Estimate Weekly US Suicide Fatalities. *JAMA Netw. Open* 2020, *3*, e2030932, doi:10.1001/jamanetworkopen.2020.30932.
69. Chowdhury, A.A.; Hasan, K.T.; Hoque, K.K.S. Analysis and Prediction of COVID-19 Pandemic in Bangladesh by Using ANFIS and LSTM Network. *Cogn. Comput.* 2021, *13*, 761–770, doi:10.1007/s12559-021-09859-0.
70. Oshinubi, K.; Amakor, A.; Peter, O.J.; Rachdi, M.; Demongeot, J.; Oshinubi, K.; Amakor, A.; Peter, O.J.; Rachdi, M.; Demongeot, J. Approach to COVID-19 Time Series Data Using Deep Learning and Spectral Analysis Methods. *AIMS Bioeng.* 2022, *9*, 1–21, doi:10.3934/bioeng.2022001.
71. Shahin, A.I.; Almotairi, S. A Deep Learning BiLSTM Encoding-Decoding Model for COVID-19 Pandemic Spread Forecasting. *Fractal Fract.* 2021, *5*, 175, doi:10.3390/fractalfract5040175.
72. Watson, G.L.; Xiong, D.; Zhang, L.; Zoller, J.A.; Shamshoian, J.; Sundin, P.; Bufford, T.; Rimoin, A.W.; Suchard, M.A.; Ramirez, C.M. Pandemic Velocity: Forecasting COVID-19 in the US with a Machine Learning & Bayesian Time Series Compartmental Model. *PLoS Comput. Biol.* 2021, *17*, e1008837, doi:10.1371/journal.pcbi.1008837.
73. Abul-Husn, N.S.; Kenny, E.E. Personalized Medicine and the Power of Electronic Health Records. *Cell* 2019, *177*, 58–69, doi:10.1016/j.cell.2019.02.039.
74. Hesse, B.W.; Hansen, D.; Finholt, T.; Munson, S.; Kellogg, W.; Thomas, J.C. Social Participation in Health 2.0. *Computer* 2010, *43*, 45–52, doi:10.1109/MC.2010.326.
75. Li, Y.; Ngom, A. Data Integration in Machine Learning. *2015 IEEE Int. Conf. Bioinforma. Biomed. BIBM* 2015, 1665–1671, doi:10.1109/BIBM.2015.7359925.

76. Iqbal, S.M.A.; Mahgoub, I.; Du, E.; Leavitt, M.A.; Asghar, W. Advances in Healthcare Wearable Devices. *Npj Flex. Electron.* 2021, *5*, 1–14, doi:10.1038/s41528-021-00107-x.
77. Figueroa, R.L.; Zeng-Treitler, Q.; Kandula, S.; Ngo, L.H. Predicting Sample Size Required for Classification Performance. *BMC Med. Inform. Decis. Mak.* 2012, *12*, 8, doi:10.1186/1472-6947-12-8.
78. Raudys, S.J.; Jain, A.K. Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners. *IEEE Trans. Pattern Anal. Mach. Intell.* 1991, *13*, 252–264, doi:10.1109/34.75512.
79. van der Ploeg, T.; Austin, P.C.; Steyerberg, E.W. Modern Modelling Techniques Are Data Hungry: A Simulation Study for Predicting Dichotomous Endpoints. *BMC Med. Res. Methodol.* 2014, *14*, 137, doi:10.1186/1471-2288-14-137.
80. He, J.; Baxter, S.L.; Xu, J.; Xu, J.; Zhou, X.; Zhang, K. The Practical Implementation of Artificial Intelligence Technologies in Medicine. *Nat. Med.* 2019, *25*, 30–36, doi:10.1038/s41591-018-0307-0.
81. Bélisle-Pipon, J.-C.; Couture, V.; Roy, M.-C.; Ganache, I.; Goetghebeur, M.; Cohen, I.G. What Makes Artificial Intelligence Exceptional in Health Technology Assessment? *Front. Artif. Intell.* 2021, *4*, 736697, doi:10.3389/frai.2021.736697.
82. Zhou, J.; Gandomi, A.H.; Chen, F.; Holzinger, A. Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics. *Electronics* 2021, *10*, 593, doi:10.3390/electronics10050593.
83. Zerilli, J.; Bhatt, U.; Weller, A. How Transparency Modulates Trust in Artificial Intelligence. *Patterns* 2022, *3*, 100455, doi:10.1016/j.patter.2022.100455.
84. Flaxman, A.D.; Vos, T. Machine Learning in Population Health: Opportunities and Threats. *PLoS Med.* 2018, *15*, e1002702, doi:10.1371/journal.pmed.1002702.
85. Wager, S.; Athey, S. Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests. *J. Am. Stat. Assoc.* 2018, *113*, 1228–1242, doi:10.1080/01621459.2017.1319839.
86. Kouser, H.N.; Barnard-Mayers, R.; Murray, E. Complex Systems Models for Causal Inference in Social Epidemiology. *J. Epidemiol. Community Health* 2020, jech-2019-213052, doi:10.1136/jech-2019-213052.
87. Schneiders, M.L.; Naemiratch, B.; Cheah, P.K.; Cuman, G.; Poomchaichote, T.; Ruangkajorn, S.; Stoppa, S.; Osterrieder, A.; Cheah, P.-K.; Ongkili, D.; et al. The Impact of COVID-19 Non-Pharmaceutical Interventions on the Lived Experiences of People Living in Thailand, Malaysia, Italy and the United Kingdom: A Cross-Country Qualitative Study. *PLOS ONE* 2022, *17*, e0262421, doi:10.1371/journal.pone.0262421.
88. Riehm, K.E.; Badillo Goicoechea, E.; Wang, F.M.; Kim, E.; Aldridge, L.R.; Lupton-Smith, C.P.; Presskreischer, R.; Chang, T.-H.; LaRocca, S.; Kreuter, F.; et al. Association of Non-Pharmaceutical Interventions to Reduce the Spread of SARS-CoV-2 With Anxiety and Depressive Symptoms: A Multi-National Study of 43 Countries. *Int. J. Public Health* 2022, *67*, 1604430, doi:10.3389/ijph.2022.1604430.
89. Baccolini, V.; Rosso, A.; Di Paolo, C.; Isonne, C.; Salerno, C.; Migliara, G.; Prencipe, G.P.; Massimi, A.; Marzuillo, C.; De Vito, C.; et al. What Is the Prevalence of Low Health Literacy in European Union Member States? A Systematic Review and Meta-Analysis. *J. Gen. Intern. Med.* 2021, *36*, 753–761, doi:10.1007/s11606-020-06407-8.
90. Atek, S.; Pesaresi, C.; Eugeni, M.; De Vito, C.; Cardinale, V.; Mecella, M.; Rescio, A.; Petronzio, L.; Vincenzi, A.; Pistillo, P.; et al. A Geospatial Artificial Intelligence and Satellite-Based Earth Observation Cognitive System in Response to COVID-19. *Acta Astronaut.* 2022, *197*, 323–335, doi:10.1016/j.actaastro.2022.05.013.

91. Raghupathi, W.; Raghupathi, V. Big Data Analytics in Healthcare: Promise and Potential. *Health Inf. Sci. Syst.* 2014, 2, 3, doi:10.1186/2047-2501-2-3.
92. Benjamins, S.; Dhunoo, P.; Meskó, B. The State of Artificial Intelligence-Based FDA-Approved Medical Devices and Algorithms: An Online Database. *NPJ Digit. Med.* 2020, 3, 118, doi:10.1038/s41746-020-00324-0.
93. Park, Y.; Jackson, G.P.; Foreman, M.A.; Gruen, D.; Hu, J.; Das, A.K. Evaluating Artificial Intelligence in Medicine: Phases of Clinical Research. *JAMIA Open* 2020, 3, 326–331, doi:10.1093/jamiaopen/ooaa033.
94. Verma, S.; Rubin, J. Fairness Definitions Explained. *2018 IEEEACM Int. Workshop Softw. Fairness FairWare* 2018, 1–7, doi:10.1145/3194770.3194776.
95. O’Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*; Broadway Books, 2017; ISBN 978-0-553-41883-5.
96. Suresh, H.; Guttag, J.V. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. *Equity Access Algorithms Mech. Optim.* 2021, 1–9, doi:10.1145/3465416.3483305.
97. Schulam, P.; Saria, S. Reliable Decision Support Using Counterfactual Models. 2018, doi:10.48550/arXiv.1703.10651.