



**Università
degli Studi
di Ferrara**

**DOTTORATO DI RICERCA IN
MEDICINA MOLECOLARE**

CICLO XXXVII

COORDINATORE Prof. Paolo Pinton

**APPLICAZIONE DI TECNICHE DI ANALISI DI INTELLIGENZA
ARTIFICIALE NELLO STUDIO DI STATO DI SALUTE, EPIDEMIOLOGIA
ED ESPOSIZIONE PER L'IDENTIFICAZIONE DEI FATTORI DI RISCHIO
INDIVIDUALI NEL COVID-19 E NELLA RISPOSTA AI VACCINI**

Settore Scientifico Disciplinare MEDS-05/A

Dottoranda

Dott.ssa Francesca Remelli

Supervisor

Prof. Stefano Volpato

Co-Supervisor

Prof.ssa Elena Bellodi

Anni 2022/2025



SOMMARIO

1. INTRODUZIONE

- 1.1. Covid-19 nelle categorie a rischio.....
 - 1.1.1. Epidemiologia.....
 - 1.1.2. Patogenesi: teoria dell'*inflammaging*.....
 - 1.1.3. Clinica e prognosi.....
- 1.2. Evidenze sull'utilizzo dell'Intelligenza Artificiale in epidemiologia.....

2. STUDIO 1: Prognosi e risposta anticorpale al vaccino anti-SARS-CoV-2 negli anziani residenti in strutture di Lungodegenza Post-Acuzie: GeroCovid Vax Study

- 2.1. Scopo dello Studio
- 2.2. Materiali e Metodi
 - 2.2.1. Popolazione.....
 - 2.2.2. Raccolta dati.....
 - 2.2.3. Analisi statistica.....
- 2.3. Discussione
- 2.4. Conclusioni

3. STUDIO 2: Influenza dei fattori genetici sulla severità della malattia acuta da Covid-19 in adulti ospedalizzati: studio prospettico osservazionale

- 3.1. Scopo dello Studio
- 3.2. Materiali e Metodi
 - 3.2.1. Popolazione.....
 - 3.2.2. Raccolta dati.....
 - 3.2.3. Analisi statistica.....
- 3.3. Discussione
- 3.4. Conclusioni

4. CONCLUSIONI GENERALI

5. BIBLIOGRAFIA

INTRODUZIONE

1.1. Covid-19 nelle categorie a rischio

1.1.1. Epidemiologia

Identificata inizialmente in Cina nel dicembre 2019, l'infezione da SARS-CoV-2 (Covid-19) si è rapidamente diffusa su scala globale, portando l'Organizzazione Mondiale della Sanità (OMS) a dichiarare la pandemia nell'aprile 2020.¹ I dati di gennaio 2021 documentavano già un carico globale di oltre 102 milioni di infezioni e 2.2 milioni di decessi attribuibili alla malattia.²

L'impatto del Covid-19 sulla popolazione anziana e fragile è stato particolarmente significativo, riportando tassi estremamente più elevati in termini di ospedalizzazione e mortalità rispetto ad altre fasce d'età. Nelle fasi iniziali della pandemia, il tasso di letalità nei pazienti con più di 80 anni è stato stimato circa del 20%, valore che superava, durante la prima ondata, di quattro volte quello riscontrato nella popolazione generale e di otto durante la seconda.³ Nonostante l'età avanzata sia un fattore fortemente associato alla mortalità da Covid-19, è emerso come gli individui anziani possano presentare un'alta eterogeneità sia negli esiti clinici durante l'infezione acuta sia nella risposta umorale al vaccino anti-SARS-CoV-2.⁴ Tali differenze appaiono strettamente collegate a fattori individuali, quali presenza di comorbidità, disabilità e, più in generale, fragilità. A prescindere dall'età, la maggior parte dei pazienti ricoverati durante la pandemia presentava infatti molteplici patologie croniche preesistenti (fra queste, diabete, ipertensione arteriosa, cardiopatia ischemica cronica e demenza), riportando una maggiore probabilità di riportare un esito fatale di Covid-19.⁵

1.1.2. Patogenesi: teoria dell'inflammaging

Il SARS-CoV-2, virus a singolo filamento di RNA, appartiene alla classe dei *Betacoronavirus* e si trasmette via *droplets* sia fra animali che fra esseri umani.⁶ Essendo infatti in grado di legarsi tramite la proteina *spike* del pericapside al recettore di membrana Enzima Convertente l'Angiotensina 2 (ACE2), il SARS-CoV-2 è in grado di penetrare nella cellula e replicarsi al suo interno, con successiva citolisi della cellula ospite e rilascio di nuovi virioni.⁷ L'espressione sulla membrana cellulare di ACE2 sembra determinare il

tropismo tissutale del SARS-CoV-2, in particolare a livello polmonare, intestinale, cardiaco e renale, e questo sostiene dal punto di vista fisiopatologico molti dei sintomi della malattia.⁸ Da menzionare inoltre come, nei casi di malattia acuta severa, sia stata descritta una risposta immunitaria anomala - anche detta “tempesta citochinica” - caratterizzata da un massiccio rilascio sistemico di citochine pro-infiammatorie (come IL-6 e TNF α), e frequente sviluppo di Sindrome da Distress Respiratorio Acuto (ARDS) e insufficienza multi-organo (MOFS).⁹ Come già menzionato, l’età avanzata e la fragilità risultano associate a un maggior rischio di sviluppare forme cliniche gravi di infezione da SARS-CoV-2. Nel determinare tale fenomeno, tra le diverse ipotesi patofisiologiche proposte, un ruolo chiave sembra essere ricoperto da un’alterazione dei meccanismi anti-infiammatori.¹⁰ Infatti, nei soggetti anziani, e in misura ancor più marcata in quelli affetti da rilevante comorbidità, è stata documentata una risposta anti-infiammatoria insufficiente, incapace quindi di controllare in maniera efficace la risposta infiammatoria scatenata dal virus, favorendo così l’insorgenza della tempesta citochinica responsabile delle forme più severe di malattia.¹¹

A questo quadro si collega anche il concetto di *inflammaging*, secondo cui l’invecchiamento è caratterizzato da uno stato di infiammazione cronica di basso grado.¹² Tale condizione è alimentata da molteplici processi pro-infiammatori tipici dell’età avanzata, tra cui il rilascio persistente di mediatori infiammatori (come IL-1, IL-6 e TNF- α) da parte delle cellule senescenti e la presenza di infezioni croniche, oltre al microbiota intestinale alterato in senso pro-infiammatorio. La combinazione di questi stimoli infiammatori di bassa intensità, ma costanti, unitamente al declino dei meccanismi anti-infiammatori, determina un effetto complessivamente deleterio e particolarmente rilevante nella popolazione anziana fragile.¹³ In aggiunta al fenomeno dell'*inflammaging*, l’invecchiamento risulta anche caratterizzato da una graduale compromissione della risposta immunitaria adattativa, definita immunosenescenza. L’immunosenescenza, caratterizzata da una progressiva riduzione del numero e della funzionalità dei linfociti T e linfociti B maturi, è responsabile di una maggiore suscettibilità alle infezioni, di una ridotta tolleranza allo stress e di un elevato rischio di sviluppare forme cliniche severe in corso di malattie acute, inclusa l’infezione da SARS-CoV-2.¹⁴ Il fenomeno dell’immunosenescenza include anche una differenziazione ridotta dei linfociti T con successiva ridotta sensibilità allo stimolo dell’interferone gamma (INF- γ) - con minor produzione di IL-2 - e una riduzione nel numero delle cellule dendritiche

produttrici di INF- γ .¹⁵ La ridotta produzione di INF- γ , peptide fisiologicamente rilasciato a livello locale dalle cellule infette allo scopo di limitare la diffusione dell'agente antimicrobico, insieme alla ridotta sensibilità allo stesso, causano a loro volta una minor risposta della immunità adattativa al virus.¹⁶ Inoltre, è stato osservato come il SARS-CoV-2, estremamente sensibile all'azione del INF- γ , sia in grado a sua volta di ridurre la produzione di tale peptide e quindi di sfuggire più facilmente al sistema immunitario.^{11,17} Perciò, nel soggetto in età avanzata, la coesistenza di una ridotta produzione di INF- γ causata, da una parte, dalla immunosenescenza e, dall'altra, dal virus incrementa la capacità infettante di SARS-CoV-2 in questo tipo di soggetto. In quest'ottica, è noto come in corso di infezione da SARS-CoV-2, la comparsa di linfocitopenia agli esami bioumorali, presente in quasi l'80% dei pazienti ricoverati in ambiente intensivistico, si associ a una maggiore incidenza di malattia acuta severa.¹⁸

In sintesi, nei soggetti anziani fragili, l'infezione da SARS-CoV-2 tende a manifestarsi con maggiore frequenza e severità rispetto a soggetti adulti robusti, quindi con più elevata probabilità di sviluppare complicanze a medio e lungo termine.

1.1.3. Clinica e prognosi

L'estrema eterogeneità clinica della malattia da Covid-19, in particolare nel paziente anziano, ne complica la precisione e la tempestività della diagnosi.¹⁹ Tale variabilità è stata imputata sia a caratteristiche peculiari del virus, ma anche individuali dell'ospite, come, oltre all'età e al sesso, anche, per esempio, alla presenza di fragilità e comorbilità.²⁰ Fra i pazienti con infezione da Covid-19, i sintomi più frequentemente riportati sono febbre e sintomi respiratori (rinite, faringodinia, tosse e dispnea); altri sintomi comuni descritti, seppur con meno frequenza, sono cefalea, astenia, mialgie, anosmia e ageusia.^{15,21} Inoltre, da sottolineare anche l'alta prevalenza di pazienti asintomatici e paucisintomatici, che ha causato nelle prime fasi di pandemia serie difficoltà nell'applicazione di misure mirante al contenimento del virus.¹

Come noto, anche nel caso dell'infezione da Covid-19, nell'anziano le manifestazioni cliniche di esordio di malattia risultano spesso atipiche, complicandone ulteriormente la

diagnosi. Il *GeroCovid Observational Study*, studio multicentrico svolto su 981 pazienti ultra-sessantenni con Covid-19, riporta come in tali pazienti l'esordio clinico di malattia possa essere caratterizzato non da manifestazioni respiratorie, bensì da delirium, sintomi gastrointestinali (nausea, vomito, diarrea), anoressia e cadute.²² Tale risultato è concorde con quanto riportato da un altro studio multicentrico su 579 pazienti anziani ricoverati per infezione da SARS-CoV-2, in cui la coorte di pazienti ultra-settantacinquenni presentava più comunemente iporessia e apiressia all'esordio di rispetto alle classi più giovani.²³ Questi studi su popolazioni anziane con infezione da Covid-19 confermano quindi quanto è noto da tempo in letteratura in ambito geriatrico sulle presentazioni atipiche di malattia. A questo proposito, è stato osservato come circa un terzo dei pazienti anziani che accedono in Pronto Soccorso presenti sintomi atipici d'esordio della patologia sottostante, e, in caso di infezioni, tra le manifestazioni atipiche più frequenti venga riportata soprattutto l'apiressia.²⁴

Resta ancora incerto se le particolari caratteristiche cliniche del Covid-19 negli anziani dipendano esclusivamente dall'età avanzata oppure, più verosimilmente, da condizioni di salute caratteristiche della popolazione geriatrica, come la presenza di comorbidità e fragilità. Come riportato da una revisione pubblicata nel 2022 sull'argomento,¹⁹ diversi studi hanno infatti mostrato quadri clinici e livelli di gravità molto differenti in popolazioni anziane con Covid-19 anche all'interno della stessa fascia di età. L'ipotesi maggiormente condivisa dal mondo scientifico è che la presenza di sindromi cliniche, tipiche in età avanzata (come comorbidità e fragilità), possa influenzare in maniera determinante la presentazione clinica e la prognosi dell'infezione da SARS-CoV-2. A tale proposito, come descritto da uno studio multicentrico, i pazienti fragili, indipendentemente dall'età, riportavano più frequentemente come manifestazioni iniziali di malattia sintomi extrapolmonari, come delirium e disidratazione.²⁵ Il delirium, in particolar modo, sembra un caratteristico esordio clinico di malattia da Covid-19 nei pazienti fragili, come anche confermato dalla ricerca di Zazzara et al. svolta su due coorti cliniche (ospedaliera e territoriale).²⁶ Questo riscontro è in parte sostenuto dal fatto che una componente importante della fragilità, intesa come l'espressione di uno stato di vulnerabilità del soggetto ad eventi avversi,²⁷ sia determinata dalla presenza di disturbo neurocognitivo maggiore, fattore rischio più forte per l'insorgenza di delirium. Bianchetti et al, in un *position paper* pubblicato ad inizio della pandemia, riportava come il 67% degli anziani con disturbo neurocognitivo maggiore sviluppasse delirium come unico sintomo di esordio di infezione da SARS-CoV-2.²⁸

Relativamente alla prognosi, uno studio multicentrico riporta come circa il 20% dei pazienti anziani arruolati con Covid-19 sviluppi complicanze durante il ricovero, quali ARDS, shock settico, insufficienza renale acuta e scompenso cardiaco acuto.²⁹ Nonostante l'età avanzata e il sesso maschile³⁰ siano stati individuati come fattori di rischio indipendenti per severità di malattia acuta da SARS-CoV-2, come menzionato, la presenza di altre condizioni cliniche, più frequenti nell'anziano, appare influenzare l'andamento dell'infezione da Covid-19, una volta contratta. Fra queste, la presenza di fragilità sembra essere uno degli indicatori prognostici più attendibili per progressione di malattia.³¹ In ambito di medicina geriatrica è infatti noto come gli individui fragili siano esposti ad una maggiore incidenza di *outcome* sfavorevoli.³² Piuttosto che la sola età anagrafica, la valutazione della fragilità, concetto legato ma non coincidente con comorbidità,³³ è quindi di ausilio al clinico per stratificare il rischio dei pazienti in base alla riserva biologica individuale.³⁴ Diversi studi, effettuati in epoca pre-vaccinale su soggetti con infezione da SARS-CoV-2, supportano la validità del ruolo della fragilità anche in questo tipo di popolazione. Uno studio multicentrico pubblicato su *The Lancet* a giugno 2020, svolto un campione di 1564 pazienti ricoverati per Covid-19 (età media: 74 anni) riportava un'associazione significativa fra fragilità e decesso, indipendente dall'età.³⁵ In quest'ottica, Marengoni *et al.* riscontrano come pazienti anziani ricoverati per Covid-19 e deceduti durante la degenza ospedaliera fossero nella maggioranza già fragili prima dell'infezione;³⁷ inoltre, ad ogni aumento unitario del punteggio della *Clinical Frailty Scale* (CFS)³⁶ si osservava un aumento del rischio di morte intraricovero del 30%, indipendentemente dall'età.³⁷

Relativamente al ruolo della comorbidità come fattore prognostico nei pazienti con Covid-19, uno studio svolto su soggetti ricoverati con infezione acuta riporta come la maggior parte di questi presentasse due o più patologie croniche ed avesse un rischio più elevato di sviluppare forme gravi di malattia.³⁸ Anche nel lavoro svolto da Lian *et al.*, la classe più anziana di individui riportava un numero maggiore di comorbidità rispetto a quella dei più giovani ed un rischio incrementato di evoluzione severa della malattia infettiva.³⁹ Tali risultati sono stati confermati anche da uno studio su 319 ultra-sessantenni con Covid-19, in cui le patologie croniche cardiovascolari e cerebrovascolari e la broncopneumopatia cronica ostruttiva si associavano a una maggiore incidenza di forma severa di malattia acuta.⁴⁰ Oltre che su campioni *population-based* e ospedalieri, dati simili sono emersi anche da uno studio multicentrico svolto su soggetti residenti in strutture socio-assistenziali. In particolare, tra

soggetti con accertata infezione da SARS-CoV-2, quasi il 90% riportava tre o più malattie croniche, con un significativo impatto sulla prognosi, sia a breve (ospedalizzazione, decesso) che a lungo termine (demenza, disabilità fisica).⁴¹

Fra i fattori determinanti la variabilità individuale nella suscettibilità all'infezione da SARS-CoV-2 e nel rischio di esiti clinici avversi, diversi studi hanno evidenziato il ruolo di specifici polimorfismi a singolo nucleotide (*single-nucleotide polymorphisms* - SNPs).⁴² Fra questi, gli SNPs localizzati nei geni coinvolti nella risposta immunitaria innata (come TLR7, IFNAR2, OAS1) e nella regolazione dell'infiammazione (ad esempio IL6, IL1B e TNF) sono stati particolarmente indagati e appaiono associati ad un'aumentata incidenza di forme cliniche più gravi di malattia da Covid-19. Tali risultati sono probabilmente giustificati da un'alterazione delle vie INF-dipendenti e da risposta citochinica iper-infiammatoria.⁴³ Altri SNPs nei geni responsabili dell'ingresso virale nelle cellule ospiti (come ACE2 e TMPRSS2) sembrano modulare l'efficienza del legame con il virus e, di conseguenza, influenzare la carica virale iniziale e l'evoluzione clinica dell'infezione.⁴⁴ Nonostante ciò, il contributo dei singoli SNPs rimane modesto e si ipotizza che spesso dipenda da complesse interazioni con fattori ambientali, comorbidità e stato immunitario; tuttavia, la loro identificazione potrebbe rappresentare un passo rilevante nell'ottica della medicina personalizzata.

Nella prevenzione della malattia, i vaccini a RNA messaggero (mRNA) anti-SARS-CoV-2, approvati in Italia in dicembre 2020, hanno rappresentato un momento cruciale per il contenimento dell'infezione e la riduzione delle ospedalizzazioni da Covid-19. I primi commercializzati, BNT162b2 (*Pfizer-BioNTech*) e mRNA-1273 (*Moderna*), si basano sulla somministrazione di un frammento di mRNA codificante per la proteina *spike* del virus.⁴⁵ L'mRNA è incapsulato in nanoparticelle lipidiche, che, oltre a proteggere la molecola di mRNA dalla degradazione, ne facilitano l'ingresso all'interno delle cellule presentanti l'antigene. Una volta nel citoplasma, l'mRNA sfrutta i ribosomi cellulari per essere tradotto nella proteina virale *spike*, che viene quindi esposta sulla membrana cellulare o rilasciata in forma solubile. La proteina *spike* sintetica viene riconosciuta dal sistema immunitario dell'ospite, attivando due componenti fondamentali della risposta immunitaria adattativa: *i*) la risposta anticorpale, con produzione di anticorpi diretti contro la *spike*, capaci di impedire l'ingresso del virus nelle cellule; *ii*) la risposta cellulare, mediata dai linfociti T CD4+ e CD8+, che potenziano la memoria immunologica e contribuiscono all'eliminazione delle

cellule infettate. Il processo non interferisce con il DNA della cellula, poiché l'mRNA non entra nel nucleo e viene rapidamente degradato dopo la traduzione, lasciando come traccia solo la risposta immunitaria indotta dalla produzione della proteina *spike*.⁴² I vaccini a mRNA sono risultati rapidamente adattabili alle varianti emergenti, dimostrando un'elevata efficacia e con un profilo di sicurezza favorevole, caratteristiche risultate fondamentali durante l'emergenza pandemica.

Come già evidenziato, i soggetti di età avanzata sono stati tra i più colpiti durante la pandemia, sia per il maggior rischio di contagio sia per la maggiore probabilità di sviluppare esiti avversi, inclusa l'ospedalizzazione, per forme gravi di malattia.¹ Per questo motivo, nelle fasi iniziali della campagna vaccinale, fra i gruppi ad alto rischio ad essere per primi vaccinati sono stati gli anziani fragili. L'efficacia del vaccino anti-SARS-CoV-2, valutata sia in termini di risposta anticorpale sia di riduzione delle infezioni incidenti, pur risultando complessivamente buona, è apparsa variabile anche tra individui con stessa età anagrafica. Fra le ragioni di tale eterogeneità, come già descritto, si trovano i processi di immunosenescenza, caratteristici dell'invecchiamento. Tale fenomeno, infatti, comportando una riduzione della funzionalità del sistema immunitario innato e adattativo - con ridotta produzione di anticorpi e risposta dei linfociti T meno vigorosa - porterebbe a una minor risposta immunitaria secondaria a vaccino con un rapido declino della protezione nel tempo.^{14,15} Oltre all'immunosenescenza, diversi sono i fattori che sembrano ulteriormente influenzare la risposta vaccinale negli anziani. Le comorbidità giocano un ruolo centrale, ed in particolare: il diabete, le malattie cardiovascolari, l'insufficienza renale cronica e la broncopneumopatia ostruttiva cronica sono risultate associate a una minore risposta anticorpale a vaccino, soprattutto se presenti simultaneamente.⁴⁶ Anche la fragilità emerge come uno dei principali determinanti: secondo quanto riportato da diversi studi, i soggetti fragili sembrano sviluppare risposte anticorpali più deboli e risultano così più suscettibili alle future infezioni incidenti da Covid-19.⁴⁷ Allo stesso modo, la disabilità, la malnutrizione, la sarcopenia, la demenza e lo stato infiammatorio cronico (*inflammaging*) sembrano modulare negativamente la capacità dell'organismo di rispondere efficacemente alla vaccinazione.⁴⁸ Altri fattori che possono modificare la risposta vaccinale includono l'uso di farmaci immunosoppressori, quali corticosteroidi o terapie oncologiche, in grado di comprometterne l'efficacia. Il paziente anziano, frequentemente affetto da comorbidità, assume inoltre spesso una polifarmacoterapia, definita come l'assunzione cronica di cinque

o più farmaci,⁴⁹ con un'alta prevalenza di molecole capaci di interferire con la risposta immunitaria del soggetto. Anche la residenza in strutture sanitarie, dato l'elevato rischio di esposizione al virus (spazi comuni), sembra influenzare significativamente sia la protezione data dal vaccino sia il decorso clinico di un'eventuale infezione.⁴⁴

Nonostante ciò, la letteratura internazionale mostra che, seppur sia stata descritta una ridotta efficacia, la vaccinazione anti-SARS-CoV-2 appare diminuire significativamente il rischio di malattia grave, nonché di mortalità, nelle persone ultra-sessantacinquenni.⁵⁰ I primi dati raccolti durante le campagne vaccinali hanno confermato che due dosi di vaccino a mRNA erano sufficienti a generare livelli anticorpali protettivi sia negli adulti che negli anziani. Per quanto riguarda la durata della protezione nel tempo, invece, le classi di età più avanzata mostravano una riduzione più rapida dei livelli anticorpali rispetto ai soggetti più giovani, rendendo necessari richiami periodici, seppur su questo punto le evidenze disponibili siano ancora contrastanti.⁵¹ Durante la diffusione delle varianti Delta e Omicron, i *booster* hanno dimostrato di poter incrementare efficacemente la risposta immunitaria anche negli ultraottantenni e nei pazienti fragili, confermando il ruolo cruciale delle dosi aggiuntive in questa fascia di popolazione.⁴⁷

In conclusione, la letteratura attualmente disponibile evidenzia come, anche nel contesto dell'infezione da Covid-19, la sola età anagrafica appaia insufficiente per una valutazione accurata degli esiti clinici, inclusa la risposta immunitaria a vaccino. Ne deriva la necessità di integrare nella pratica clinica una valutazione multidimensionale, comprendente elementi quali comorbilità, disabilità fisica, deficit cognitivi e grado di fragilità, al fine di ottenere una stratificazione del rischio più corretta e affidabile in base alle caratteristiche specifiche del paziente.

1.2. Evidenze sull'utilizzo dell'Intelligenza Artificiale in epidemiologia

L'applicazione dell'Intelligenza Artificiale (*Artificial Intelligence* - AI) sta profondamente ridefinendo il paradigma dell'epidemiologia nello studio e nella gestione dei fattori di rischio individuali e di popolazione. Rispetto agli approcci di biostatistica tradizionale, l'uso dell'AI consente infatti di analizzare simultaneamente un maggior numero di variabili (*big data*), che a priori potrebbero influenzare il determinarsi dell'evento di interesse, consentendo di costruire modelli predittivi potenzialmente più robusti.⁵² In particolare, la possibilità di integrare dati provenienti da fonti estremamente eterogenee - come dispositivi *wearable*, sensori ambientali, immagini biomediche, dati amministrativi, referti testuali e profili genetici - consente di adottare un approccio realmente *data-driven*, in cui l'individuazione dei pattern di rischio non è vincolata alle ipotesi preliminari del ricercatore (approccio tradizionale *hypothesis-driven*), ma emerge in maniera induttiva dall'esplorazione diretta dei dati.

Un ambito di grande espansione dell'uso di modelli di AI riguarda i sistemi di sorveglianza epidemiologica in tempo reale: l'utilizzo di reti neurali e algoritmi di AI sembra supportare l'identificazione precoce di focolai epidemici, consentendo analisi predittive basate su dati a continuo aggiornamento provenienti da sistemi informativi sanitari, *mobility data*, piattaforme digitali e perfino indicatori di sintomi *self-reported*.⁵³ Tale approccio, già sperimentato durante la pandemia di Covid-19, potrebbe rivelarsi particolarmente utile nella gestione di agenti patogeni emergenti e nella previsione di scenari epidemiologici complessi, contribuendo a un processo decisionale più tempestivo da parte delle autorità sanitarie.

Potenzialmente non meno rilevante potrebbe essere il ruolo dell'AI nello studio dei fattori di rischio in caso di eventi clinici multifattoriali, risultanti quindi dall'interazione dinamica tra fattori biologici, comportamentali, ambientali e socio-economici, estremamente comuni in ambito di ricerca medica e, in particolare, geriatrica. La geriatria, con la sua intrinseca complessità clinica, rappresenta un campo di applicazione ideale per approcci computazionali avanzati capaci di modellare comorbidità multiple, fragilità, polifarmacoterapia e variazioni funzionali e cognitive nel tempo.⁵⁴ Modelli predittivi, costruiti su algoritmi di AI, potrebbero fornire un supporto importante per il clinico nella valutazione della prognosi del paziente, stratificando precocemente il rischio per *outcome* avversi, come declino funzionale, demenza, cadute, istituzionalizzazione e decesso,

permettendo così di orientare interventi personalizzati e ottimizzare le risorse sanitarie disponibili. Una revisione pubblicata a gennaio 2025 evidenzia come la flessibilità dei modelli di AI, applicati nello studio delle malattie croniche, sembra avere un potenziale vantaggio rispetto ai modelli definiti più rigidi della biostatistica standard, in quanto capaci di gestire dataset complessi, non lineari e ad alta dimensionalità, oltre che le loro potenziali interazioni ed eventuali pattern latenti.⁵⁵ Infatti, l'analisi dei dati medici, intrinsecamente complicata dall'incertezza derivante da vari fattori (quali differenze biologiche individuali, diversità nella progressione delle patologie, errori di misurazione e incompletezza dei *record* clinici) spesso impedisce ai metodi analitici tradizionali di generare modelli in grado di catturare completamente le relazioni probabilistiche presenti nei dati. Nonostante ciò, la qualità e la completezza delle informazioni raccolte condizionano fortemente anche le performance dei modelli di AI e il rischio di *bias* sistematici. Nel contesto epidemiologico, *bias* di selezione, misclassificazione e *missing* nella raccolta dei dati possono portare a modelli scarsamente predittivi, soprattutto quando le popolazioni che sviluppano l'evento di interesse sono sottorappresentate.⁵⁶ Per questi motivi, diventa essenziale implementare strategie di *data curation*, con controlli rigorosi di validazione esterna, al fine di garantire che i modelli predittivi risultino affidabili e applicabili nei differenti contesti clinici.

Fra gli strumenti di AI che si stanno affermando nell'analisi dei dati epidemiologici e biomedici, si ricordino il *Machine Learning* (ML) e il *Deep Learning* (DL), capaci di integrare flussi di dati eterogenei e multimodali (ad esempio, dati clinici da registri elettronici sanitari e informazioni ricavate da *social media*).

Il ML è una tecnica di AI che comprende una serie di algoritmi in grado di “imparare” automaticamente dai dati, utilizzando tecniche di apprendimento supervisionato - nel caso di diagnosi e previsione di *outcome* specifici - e non supervisionato - per l'identificazione di sottogruppi di pazienti (*clustering*), e migliorare così le proprie prestazioni.⁵⁷ In ambito epidemiologico, metodi supervisionati di ML, come le Reti Bayesiane, *Random Forest* (RF) e *Support Vector Machines* (SVM), sono stati applicati nello studio della stratificazione di rischio individuale, basato per esempio sulla presenza e il numero di malattie croniche, per specifici esiti clinici, ma anche in ambito di previsione della diffusione di malattie infettive.⁵⁸ Il ML si distingue quindi per la sua capacità di individuare interazioni non lineari complesse tra una vasta gamma di variabili (cliniche, genetiche, ambientali, socio-economiche) che i modelli di statistica tradizionale potrebbero non riuscire a cogliere.

I modelli di DL, come *Convolutional Neural Networks* (CNNs) e *Recurrent Neural Networks* (RNNs), rappresentano un'ulteriore evoluzione nell'estrazione ed analisi di *feature* da dati medici ad alta dimensionalità, quali profili genomici, ma anche registri clinici longitudinali.⁵⁹ Mentre le tecniche di CNNs sono attualmente molto usate nelle diagnosi in ambito sia oncologico che nello studio di malattie croniche, i modelli di RNNs, capaci di modellare i dati a serie temporale, consentono una previsione in tempo reale dell'*outcome* di interesse (come dei livelli di glicemia in base ai valori derivanti dal monitoraggio ematico continuo).⁶⁰ Nello studio del Covid-19, modelli di DL hanno dimostrato un'elevata efficacia, ad esempio, nell'analisi delle immagini derivanti da RX torace per la diagnosi precoce di polmonite e nella previsione della gravità clinica di malattia, ma anche nella previsione della diffusione a livello territoriale del contagio, integrando dati sanitari e socio-ambientali.⁶¹ Tali approcci permetterebbero quindi di derivare modelli altamente predittivi e di fornire supporto decisionale sia al clinico sia per interventi di sanità pubblica, come la pianificazione vaccinale o la gestione delle risorse ospedaliere.

Nonostante la potenzialità di questi approcci, un aspetto cruciale nell'uso dell'AI in epidemiologia riguarda la trasparenza e interpretabilità dei modelli. Mentre le tecniche di ML classiche consentono una certa tracciabilità del processo decisionale, i modelli di DL - spesso definiti *black box* - presentano difficoltà nell'interpretazione dei risultati, che potenzialmente limita la loro adozione nella pratica clinica.⁶² Per tale motivo, si stanno sviluppando approcci di *Explainable Artificial Intelligence* (XAI), volti a rendere interpretabili le decisioni dei modelli e a consentire una maggiore fiducia da parte di ricercatori e operatori sanitari.⁶³ Inoltre, la presenza di una stretta collaborazione fra clinici, epidemiologi ed ingegneri, sia nel disegno di studio che nella selezione delle variabili d'interesse, risulta essenziale per indirizzare correttamente l'utilizzo dei modelli di AI e garantire l'efficacia predittiva e la solidità delle conclusioni in ambito medico.

Diversi studi riportano come modelli basati su ML e DL sembrano migliorare l'accuratezza predittiva degli esiti clinici rispetto ai metodi tradizionali, nonché gestire più efficacemente complessi pattern non lineari nei dati epidemiologici.⁶⁴ Una meta-analisi pubblicata nel 2024, che confrontava tecniche di AI e algoritmi tradizionali nella predizione del rischio cardiovascolare a medio-lungo termine, evidenzia come modelli di ML applicati a dati clinici elettronici (EHR) mostrino performance superiori rispetto ai punteggi di rischio convenzionali.⁶⁵ In particolare, in ambito di medicina geriatrica, uno studio svolto su

4647 pazienti anziani con scompenso cardiaco riporta come il Random Forest sia risultato estremamente efficace nel predire la mortalità intra-ospedaliera, identificando simultaneamente più fattori associati a tale *outcome*.⁶⁶ In quest'ottica, Loutati et al dimostrano come l'utilizzo di modelli multimodali di ML risulti accurato anche nella predizione del rischio di riammissione ospedaliera a 30 giorni in pazienti ultra-settantacinquenni.⁶⁷ Studi che hanno applicato modelli di ML alla genetica della malattia di Alzheimer dimostrano altresì come algoritmi di *Gradient Boosting* e reti neurali siano in grado di replicare varianti *genome-wide* note e identificare nuovi loci non rilevati dai GWAS tradizionali, offrendo nuove prospettive in epidemiologia genetica geriatrica.⁶⁸

Anche per quanto riguarda la ricerca sul Covid-19, l'uso dell'AI ha fornito risultati promettenti. Una *scoping review* condotta da Burnazovic et al. (2023)⁶⁹ ha analizzato le applicazioni di AI nella stratificazione della prognosi di pazienti anziani con Covid-19, includendo 10 studi svolti su soggetti ospedalizzati. Fra le ricerche analizzate, l'età anagrafica e le comorbidità sono risultate fra i predittori più forti per *outcome* avversi, incluso il decesso.^{70,71,72} La revisione ha evidenziato come modelli di ML e DL sembrino essere strumenti efficaci nella predizione di esiti clinici e nell'identificazione di sottogruppi di pazienti ad alto rischio, suggerendo un potenziale concreto per l'integrazione di AI nella pratica clinica ospedaliera anche durante la pandemia da Covid-19. Relativamente alle popolazioni residenti in regime di lungodegenza, solamente due studi sono al momento disponibili su AI e Covid, entrambi svolti da Lee et al.^{73,74} Il primo, condotto su 80784 LTC soggetti, analizzava tramite modelli di ML 484 predittori di rischio di infezione da SARS-CoV-2, evidenziando buone performance metriche; caratteristiche relative alla zona geografica, ma anche comorbidità ed esami ematici, sono risultati associati all'*outcome* indagato.⁷³ Il secondo studio invece, svolto su 64733 partecipanti con infezione da SARS-CoV-2, ha analizzato 304 potenziali predittori per mortalità attraverso ML (Random Forest). I modelli ottenuti si sono dimostrati altamente predittivi ed, in particolare, lo stato funzionale, le comorbidità e gli esami di laboratorio si sono associati ad un maggior rischio di morte dopo infezione dal virus.⁷⁴

In sintesi, le evidenze disponibili supportano l'AI come un'opportunità significativa di avanzamento in ambito epidemiologico rispetto ad approcci inferenziali tradizionali, ma ulteriori studi sono necessari per valutare il suo utilizzo sistematico nel contesto della ricerca e pratica clinica.

STUDIO 1

**Prognosi e risposta anticorpale al vaccino anti-SARS-CoV-2
negli anziani residenti in strutture di lungodegenza:
GeroCovid Vax Study**

2.1. Scopo dello Studio

Indagare tramite modelli di AI i fattori associati ad *outcome* avversi (infezione da COVID-19, necessità di supporto con ossigenoterapia e decesso) e una minor risposta anticorpale a vaccino in anziani vaccinati con anti-SARS-CoV-2 residenti in regime di Lungodegenza Post-Acuzie (LPA).

2.2. Materiali e Metodi

Studio multicentrico longitudinale (*GeroCovid Vax Study*) svolto su pazienti anziani vaccinati con almeno una dose di anti-SARS-CoV-2 e residenti in LPA fra marzo 2021 e giugno 2022. Lo studio è stato promosso dalla Società Italiana di Geriatria e Gerontologia (SIGG) e dall'Istituto Superiore di Sanità (ISS) e sponsorizzato dall'Agenzia Italiana del Farmaco (AIFA). Lo studio è stato approvato dal Comitato Etico locale dell'Istituto Spallanzani, Roma (numero protocollo: 264_2021).

2.2.1. Popolazione

I criteri di inclusione sono stati:

- 1) età ≥ 60 anni;
- 2) aspettativa di vita ≥ 3 mesi;
- 3) residenza in LPA da almeno 3 mesi;
- 4) anamnesi positiva per almeno una dose di vaccinazione anti-SARS-CoV-2.

Per l'arruolamento nello studio, per ogni paziente è stato ottenuto il consenso informato.

Sono stati esclusi pazienti con età inferiore ai 60 anni, aspettativa di vita minore di 3 mesi, coloro con una infezione in corso da Covid-19 e con una dimissione programmata dalla LPA entro 3 mesi.

I pazienti sono stati valutati al baseline, quindi a 2, 6 e 12 mesi dopo la prima dose vaccinale.

2.2.2. Raccolta dati

Al baseline, per ogni paziente sono state raccolte: caratteristiche sociodemografiche, stile di vita (abitudine al fumo, consumo di alcol settimanale), *Basic and Instrumental Activities of Daily Living* (BADL, IADL), performance cognitive (*Mini Mental State Examination* - MMSE), mobilità (deambulazione autonoma con o senza ausili vs. non autonoma o allettamento), comorbilità (numero e tipo di malattie croniche, quali cardiopatie croniche, fibrillazione atriale, pneumopatie croniche, diabete, insufficienza renale cronica, neoplasia,

depressione, disturbo neurocognitivo maggiore o minore), fragilità, terapia farmacologica in corso (numero di farmaci ed assunzione di terapia immunosoppressiva), precedente infezione da Covid-19, tipo di vaccino anti-SARS-CoV-2 e numero di dosi ricevute ed analisi di laboratorio di routine.

Sono inoltre raccolti i dati relativi agli esiti clinici entro 12 mesi, fra cui infezione incidente da Covid-19, necessità di supporto con ossigenoterapia e decesso.

Il follow-up di studio, previsto di 12 mesi, includeva la rivalutazione ad intervalli programmati di tutti i parametri clinici e biologici raccolti al baseline. In particolare, i dati sono stati raccolti 7 giorni prima della prima dose di vaccino (baseline, t0), a 2 (t1), 6 (t2) e 12 mesi (t3) dopo la prima dose (**Figura 1**).

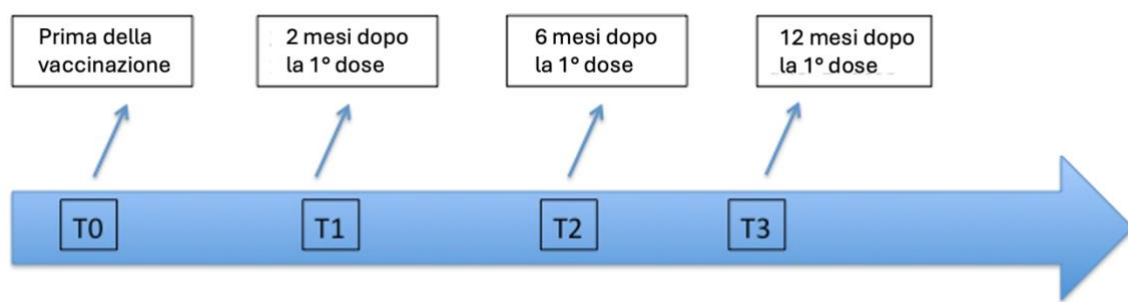


Figura 1. Timing di follow-up durante il GeroCovid Vax Study

In un sottocampione di pazienti, oltre alla raccolta dei dati clinici, sono stati prelevati dei campioni di sangue venoso al baseline e a tutti i timing del follow-up per analizzare la risposta immunitaria umorale. Tali campioni sono stati analizzati in un laboratorio centralizzato del Dipartimento di Malattie Infettive dell'ISS (Roma, Italia). Per la quantificazione dei livelli di IgG anti-Spike (S) è stato usato il test CLIA commerciale (Liaison SARS-CoV-2 trimeric S IgG assay), data la sua alta concordanza con i titoli degli anticorpi neutralizzanti. La presenza o l'assenza di IgG anti-Nucleocapside (N), ricercata ad ogni timing di follow-up, è stata valutata per verificare un'eventuale esposizione a SARS-CoV-2 al momento della raccolta del siero.

Tutti i dati raccolti sono stati inseriti su piattaforma online, fornita dalla *Bluecompanion company* (Francia).⁷⁵

2.2.3. *Analisi statistica*

L'analisi descrittiva delle variabili è stata espressa per le quantitative attraverso media e deviazione standard (DS), mentre per le categoriche attraverso frequenze assolute o relative. Il confronto tra gruppi in base al sesso è stato effettuato tramite test *t-Student* (variabili continue) e test chi-quadrato (variabili dicotomiche).

Per l'analisi degli *outcome* di interesse (infezione da Covid-19, necessità di ossigenoterapia, decesso e alta risposta anticorpale), sono stati sviluppati modelli di ML, includendo i seguenti predittori: età, sesso, abitudine al fumo, consumo di alcol, livello di mobilità, numero complessivo di farmaci assunti, uso di corticosteroidi/altri immunosoppressori, numero di comorbidità e presenza di specifiche patologie croniche (quali insufficienza renale cronica, malnutrizione, neoplasie, epatopatie, diabete, cardiopatie, fibrillazione atriale, pneumopatie croniche, depressione, disturbo neurocognitivo maggiore o minore), precedente infezione da Covid-19, tipologia di vaccino anti-SARS-CoV-2 e numero di dosi vaccinali ricevute. Le analisi sono state condotte in ambiente Python 3.9.21, utilizzando le librerie *numpy*, *pandas*, *matplotlib*, *seaborn*, *sklearn*, *xgboost* e *imblearn*. In base all'*outcome* esplorato, sono stati implementati i seguenti modelli: Decision Tree (DT), Random Forest (RF), XGBoost (XGB) e XGB-RandomForest (XGBRF), TabPfn, rete neurale *fully-connected costruita ad-hoc*, Light Gradient Boosting Machine (LGBM) e Gradient Boosting. Dopo la fase di *preprocessing* per la codifica delle variabili ed imputazione dei valori mancanti, vista l'elevata asimmetria dei dati, i dataset sono stati bilanciati con *undersampling* (Random Undersampling) o *oversampling* (Random Oversampling, SMOTE, ADASYN). La scelta del metodo è stata svolta in funzione della capacità di ciascun approccio di produrre performance metriche del modello più robuste relativamente all'*outcome* d'interesse. Per l'interpretazione dei modelli sono state utilizzate due classi di metriche: *i) impurity-based feature importance*, con quantificazione del contributo delle variabili alla riduzione dell'impurità negli alberi decisionali, e *ii) SHapley Additive exPlanations* (SHAP) *values*, utilizzati principalmente tramite rappresentazioni *beeswarm*, per valutare direzione ed entità dell'effetto delle singole variabili sulle predizioni. Relativamente al processo di addestramento e valutazione degli algoritmi di ML, prima dell'applicazione delle tecniche di bilanciamento, è stato creato un *test set* con la

distribuzione originale del target e, da questo, sono stati estratti i dati per un *validation set* per l'ottimizzazione degli iperparametri. Una volta identificata la configurazione ottimale per ciascun modello e dataset, il modello finale è stato addestrato sul *training set* completo e valutato sul *test set*. In base *outcome* esplorato, il modello con migliori performance (in termini di metriche *accuracy*, *precision*, *recall*, e *F1-score*) è stato scelto per le analisi. Per indagare la risposta umorale a vaccino sul sottocampione con dato disponibile, ad ogni timing di follow-up è stata calcolata la mediana dei valori anticorpali. I valori sono stati quindi classificati come “bassi” se inferiori o pari alla mediana, e come “alti” se superiori.

2.3. Risultati

L'età media dei 3276 partecipanti è stata di 83.4 anni (SD 9.1) e il 71.0% era di sesso femminile. Il 37.6% era capace di camminare autonomamente o con ausilio (bastone/walker); la mediana di farmaci assunti in terapia è stata di 6 (IQR 4-8). La sindrome di fragilità è stata riportata nel 14.3% dei casi; l'88.6% presentava una diagnosi di disturbo neurocognitivo.

Durante i 12 mesi di osservazione, il 9.3% (N=303) dei partecipanti ha sviluppato un'infezione incidente da Covid-19, l'1.4% (N=42) ha richiesto supporto con ossigenoterapia e il 20.9% (N=503) è deceduto.

In **Tabella 1** sono presentate le caratteristiche del campione in base al sesso. In particolare, le femmine erano più anziane dei maschi, riportavano una minor frequenza di abitudine al fumo, attiva o pregressa, e di consumo di alcol settimanale. Relativamente alle performance funzionali e cognitive, mentre non sono state evidenziate differenze significative fra i due sessi nella media delle BADL, le femmine presentavano punteggio medio inferiore sia di IADL che di MMSE. Inoltre, tale gruppo riportava una maggior frequenza di fragilità e disabilità al cammino con necessità assistenza, fino all'allettamento. Come mostrato in tabella, non sono state evidenziate differenze riguardo alla terapia cronica in corso, né per numero di farmaci né per frequenza di immunosoppressori assunti. Le donne riportavano una prevalenza superiore di pregressa infezione da Covid-19 rispetto agli uomini, mentre non sono state riscontrate differenze in merito al tipo di vaccino anti-SARS-CoV-2 né al numero di dosi ricevute. Infine, relativamente alle comorbidità, le femmine presentavano una frequenza maggiore di disturbo neurocognitivo e minore di pneumopatie croniche, diabete, insufficienza renale cronica e depressione rispetto ai maschi. Non sono state riscontrate differenze significative nel numero medio di comorbidità.

Tabella 1. Caratteristiche socio-demografiche e cliniche del campione in base al sesso.

	Maschi (N=948)	Femmine (N=2327)	p
Età, media (DS)	79.5 (9.7)	84.9 (8.4)	<0.001
Stile di vita			
Abitudine al fumo (attiva o pregressa)	159 (32.7)	105 (9.9)	<0.001
Consumo di alcol durante la settimana	156 (32.6)	102 (9.8)	<0.001
Performance funzionali e cognitive			
BADL, media (DS)	2.38 (3.35)	2.14 (6.37)	0.581
IADL, media (DS)	1.45 (2.34)	0.87 (1.59)	<0.001
Incapace di camminare in autonomia	277 (31.8)	811 (38.3)	0.001
Fragilità	119 (21.7)	348 (27.8)	0.008
MMSE, media (DS)	14.83 (9.92)	12.49 (9.58)	<0.001
Terapia domiciliare			
No. farmaci totale, media (DS)	6.2 (3.4)	6.1 (3.5)	0.531
Steroidi	12 (1.3)	26 (1.1)	0.857
Altri immunodepressori	18 (1.9)	35 (1.5)	0.510
Covid-19			
Precedente infezione da Covid-19	234 (27.9)	637 (32.3)	0.025
Tipo di vaccino anti-SARS-CoV-2 ricevuto			
<i>Pfizer-BioNTech</i>	883 (93.7)	2160 (93.1)	0.219
<i>Moderna</i>	58 (6.2)	160 (6.9)	
<i>Altro</i>	1 (0.1)	0 (0.0)	
Numero di dosi vaccinali ricevute, media (DS)	2.4 (0.6)	2.4 (0.6)	0.117
Comorbidità			
No. malattie croniche, media (DS)	4.7 (2.5)	4.6 (2.4)	0.287
Cardiopatie croniche	513 (54.1)	1312 (56.4)	0.252
Fibrillazione atriale	98 (10.3)	285 (12.2)	0.138
Pneumopatie croniche	225 (23.7)	386 (16.6)	<0.001
Diabete	244 (25.7)	465 (20.0)	<0.001
Insufficienza renale cronica	138 (16.0)	231 (10.9)	<0.001
Neoplasia	131 (15.2)	267 (12.5)	0.064
Malnutrizione	93 (10.8)	234 (11.0)	0.908
Depressione	407 (42.9)	778 (33.4)	<0.001
Disturbo neurocognitivo maggiore o minore	541 (57.1)	1549 (66.6)	<0.001

Legenda: BADL = Basic Activities of Daily Living; IADL = Instrumental Activities of Daily Living; MMSE = Mini Mental State Examination; DS = Deviazione Standard
Quando non specificato, in Tabella sono riportati valori assoluti e frequenze, N (%)

Relativamente al sottocampione di partecipanti con dosaggio anticorpale disponibile, la **Figura 2** mostra di livelli ematici di IgG anti-S al baseline ($p=0.925$), a 2 ($p=0.076$) e 6 mesi ($p=0.476$) nei due sessi; non è stato possibile esplorare i livelli anticorpali a 12 mesi per una frequenza di *missing data* in circa il 50%.

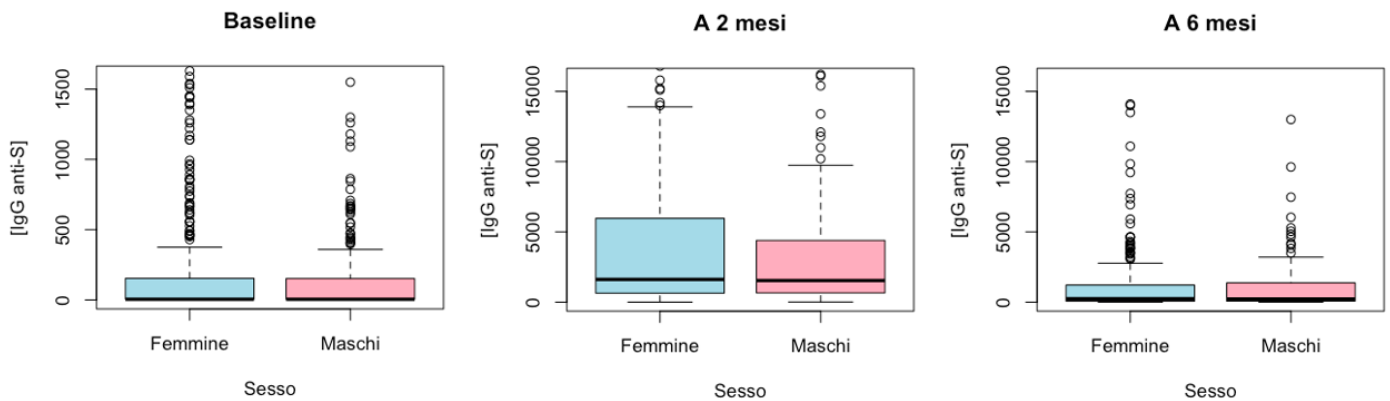


Figura 2. Box plot dei livelli ematici di IgG anti-Spike durante il follow-up in base al sesso

La **Figura 3** mostra la matrice di correlazione tra il tempo di insorgenza dell'infezione da Covid-19 e l'insieme dei predittori sociodemografici e clinici di interesse. L'intensità del colore rappresentato in figura riflette la forza delle correlazioni, riportate numericamente come coefficienti di Pearson. Nel complesso, l'*outcome* indagato non appare linearmente correlato a nessuno dei singoli predittori analizzati (range di valori più frequenti fra -0.10 e $+0.10$). Infatti, le correlazioni con valori più elevati riguardano quelle riscontrate fra i predittori stessi (ad esempio, fra fumo e cammino), senza coinvolgere direttamente l'*outcome*.

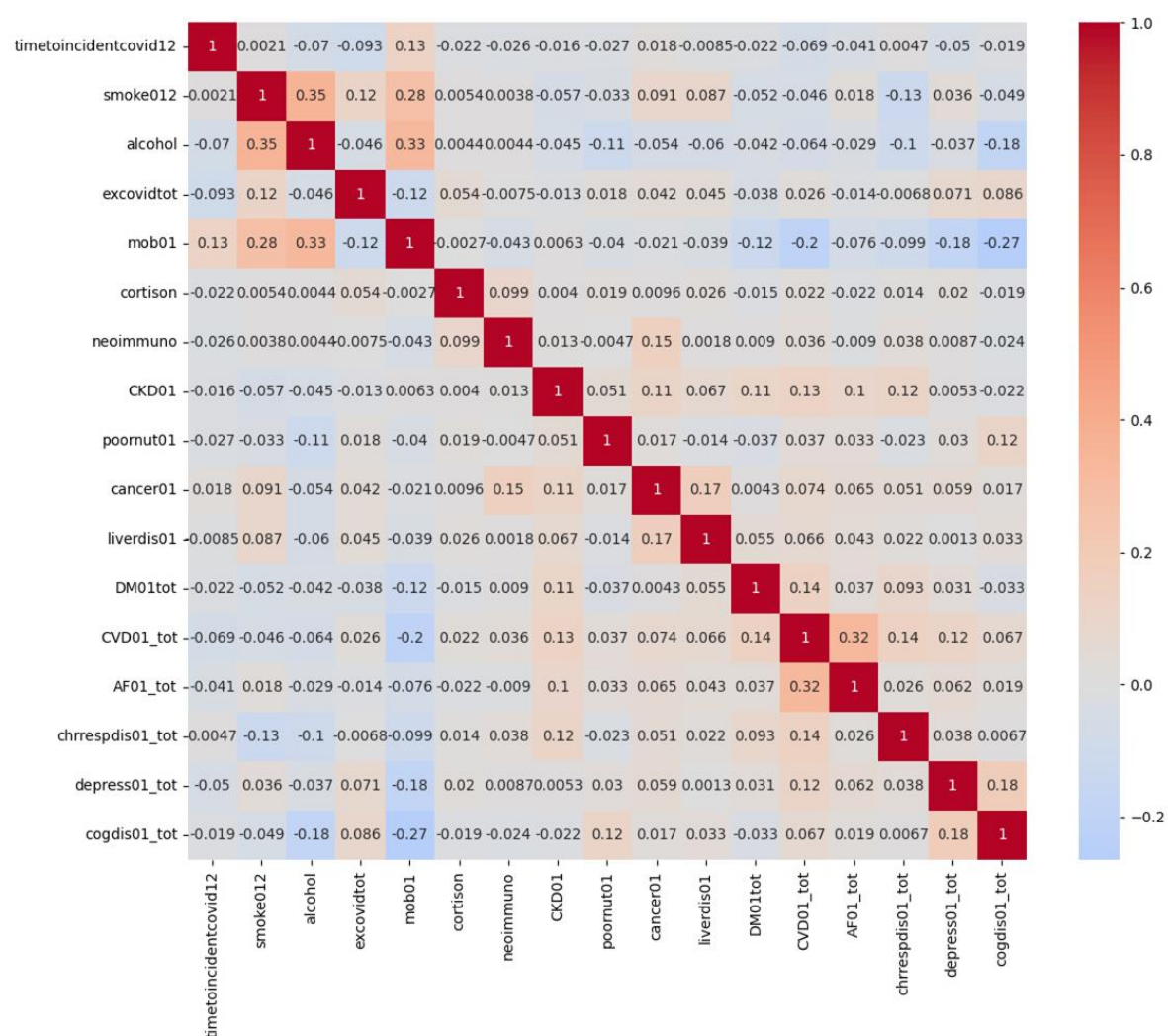


Figura 3. Matrice di correlazione fra i predittori analizzati nell'intero campione (N=3276)

Avendo escluso quindi la presenza di una correlazione lineare semplice fra predittori analizzati ed *outcome*, sono stati costruiti modelli di AI per l'analisi e la previsione sui dati. In relazione alle metriche ottenute, il modello predittivo risultato migliore nello studio dell'infezione incidente da Covid-19 è risultato il Random Forest, addestrato sul dataset con variabili numeriche convertite in categoriche e bilanciato tramite *undersampling* (*precision*=0.71, *recall*=0.88, *F1-score*=0.78 e *accuracy*=0.77); si veda **Tabella 2**.

Tabella 2. Metriche di performance nei modelli di AI testati sul dataset nella predizione dell'infezione incidente da Covid-19.

Modello	Bilanciamento	Precision	Recall	F1-Score	Accuracy
DT	<i>Undersampling</i>	0.68	0.81	0.74	0.70
RF	<i>Undersampling</i>	0.71	0.88	0.78	0.77
XGB	<i>Undersampling</i>	0.64	0.88	0.74	0.71
XGBRF	<i>Undersampling</i>	0.76	0.77	0.77	0.76

Legenda: DT = Decision Tree; RF = Random Forest; XGB = eXtreme Gradient Boosting; XGBRF = XGB Random Forests

L'analisi del modello ottenuto tramite Random Forest ha mostrato che la *feature* più importante era la pregressa infezione da parte del virus, che riduceva la probabilità di contrarre nuovamente l'infezione nei 12 mesi. Altre *features* associate all'*outcome*, sempre negativamente, sono state il basso consumo di alcol e un basso numero di farmaci assunti in terapia. Infine, la quarta *feature* per importanza era il numero di dosi vaccinali ricevute; in particolare, aver ricevuto due dosi di vaccino riduceva la probabilità di infezione incidente. In riferimento, si veda la **Figura 4**.

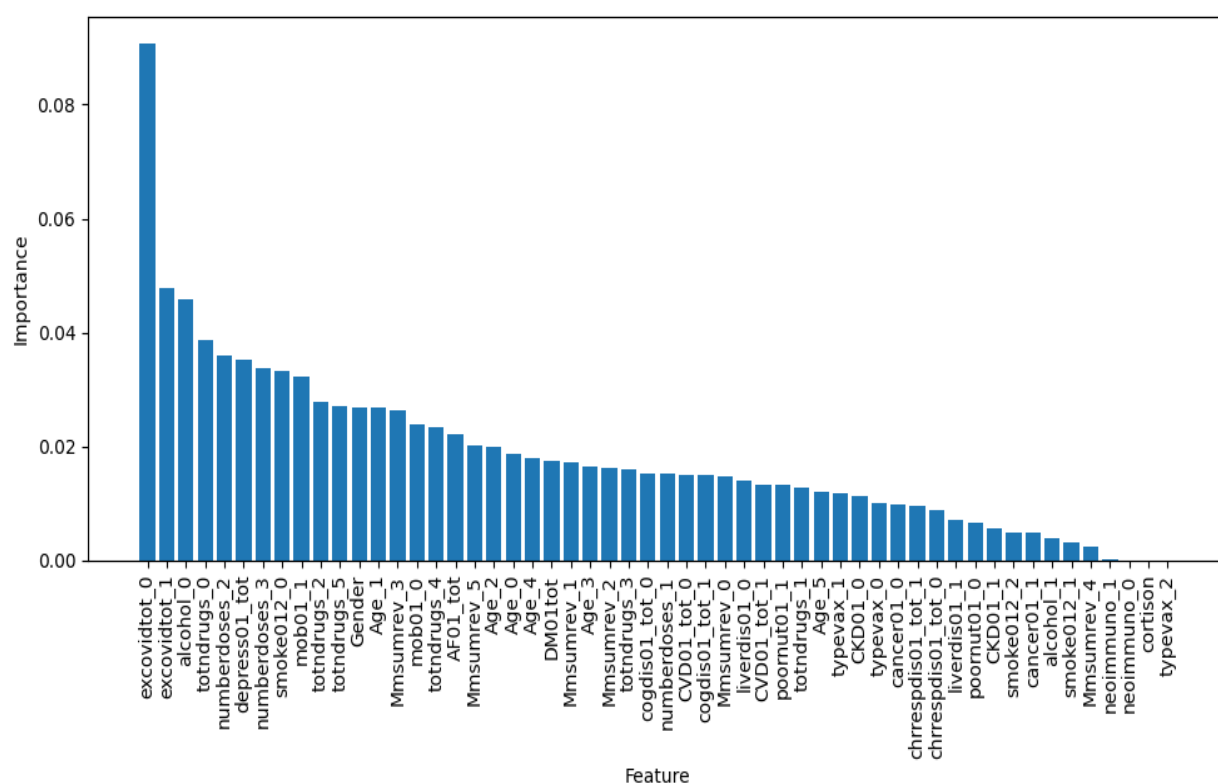


Figura 4. Importanza delle *features* nella predizione dell'infezione incidente da Covid-19

I risultati mostrati nella precedente figura sono stati coerenti con i valori SHAP riportati in **Figura 5**, che mostrano visivamente la direzione e l'intensità dell'influenza delle *feature* sul risultato predetto.

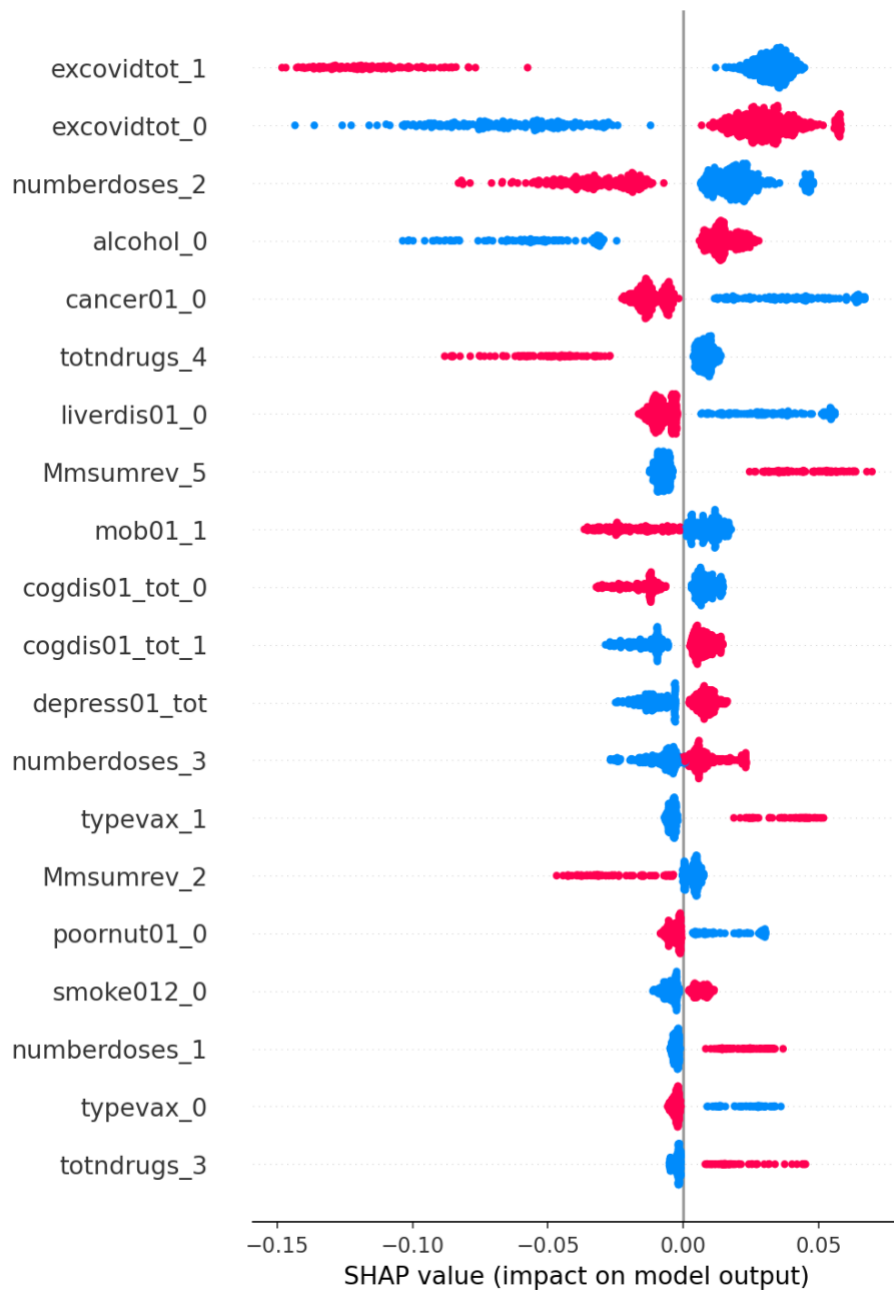


Figura 5. SHAP *values* dei predittori per infezione incidente da Covid-19

Riguardo al supporto con ossigenoterapia e al decesso a 12 mesi, i modelli utilizzati - sia di ML sia di DL che rete neurale - hanno mostrato metriche di performance eccessivamente basse per poter procedere ad esplorare tali *outcome* (Tabella 3 e 4).

Tabella 3. Metriche di performance nei modelli di AI testati sul dataset nella predizione della necessità di supporto con ossigenoterapia.

Model	Bilanciamento	Precision	Recall	F1-Score
XGB	Oversampling	0.55	0.29	0.38
TabPFN	Oversampling	0.55	0.29	0.38
CustomDL	Oversampling	0.12	0.43	0.19

Legenda: XGB = eXtreme Gradient Boosting; TabPFN = Tabular Prior-Data Fitted Network; CustomDL = Custom Deep Learning

Tabella 4. Metriche di performance nei modelli di AI testati sul dataset nella predizione del decesso.

Model	Bilanciamento	Precision	Recall	F1-Score
XGB	SMOTE-NC	0.67	0.61	0.64
XGB	ADASYN	0.72	0.52	0.61
LGBM	ADASYN	0.69	0.53	0.60

Legenda: XGB = eXtreme Gradient Boosting; LGBM = Light Gradient Boosting Machine

Allo scopo di valutare la risposta anticorpale a 2 e 6 mesi, è stata svolta un'analisi del sottogruppo di partecipanti con dati disponibili relativi ai livelli ematici di IgG anti-S. A 2 mesi, la popolazione analizzata di 472 pazienti è stata suddivisa in due classi di risposta anticorpale alta vs. bassa in base alla mediana di IgG anti-S, ottenuta di 1735 AU/mL (range: 5-42400 AU/mL). Dall'analisi della matrice di correlazione (**Figura 6**), l'*outcome* indagato non appare linearmente correlato a nessuno dei singoli predittori analizzati. Contrariamente, le due *feature* maggiormente correlate fra loro in maniera negativa sono state la precedente infezione da Covid-19 e il numero di dosi vaccinali ricevute, con un coefficiente di correlazione pari a -0.61. La correlazione positiva più alta è stata invece riscontrata tra le variabili relative al consumo settimanale di alcol e alla dipendenza nel cammino, con un valore del coefficiente di 0.66.

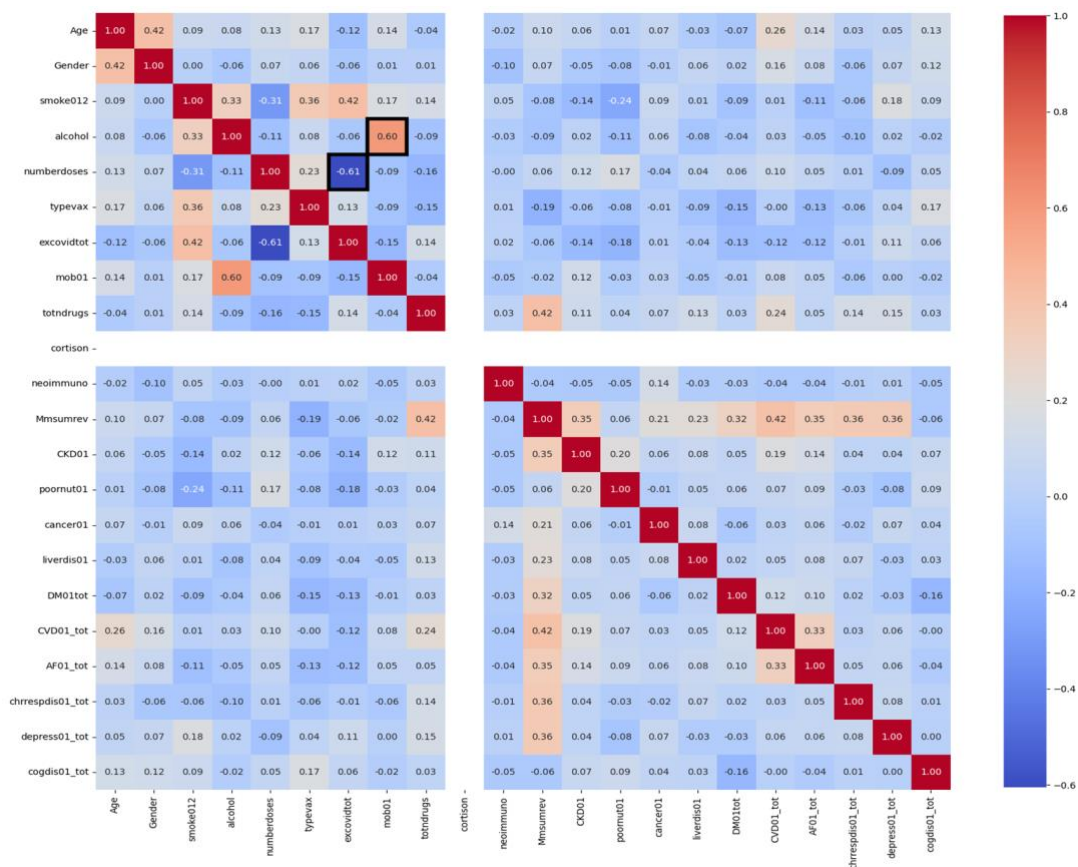


Figura 6. Matrice di correlazione fra i predittori analizzati nel sottogruppo con dati disponibili relativamente ai livelli ematici di IgG anti-Spike a 2 mesi (N=472)

Avendo escluso quindi la presenza di una correlazione lineare semplice fra predittori analizzati ed *outcome*, sono stati costruiti modelli di AI. In relazione alle metriche ottenute, il modello predittivo risultato migliore nello studio della risposta anticorpale alta vs. bassa a 2 mesi è risultato il Random Forest (*precision*=0.88, *recall*=0.78, *F1-score*=0.83 e *accuracy*=0.83); si veda **Tabella 5**.

Tabella 5. Metriche di performance nei modelli di AI testati sul dataset nella predizione della risposta anticorpale a 2 mesi (alta vs. bassa) a vaccino anti-SARS-CoV-2.

Modello	Bilanciamento	Precision	Recall	F1-Score	Accuracy
XGB*	-	0.75	0.73	0.74	0.75
XGB**	-	0.83	0.79	0.81	0.82
RF	-	0.88	0.78	0.83	0.83

Legenda: XGB = *eXtreme Gradient Boosting*; RF = *Random Forest*

**gamma*=0

***gamma*=3

Come mostrato in **Figura 7 (a)**, i primi quattro predittori per risposta anticorpale alta, in ordine di importanza, sono risultati: 1) pregressa infezione da Covid-19, 2) *Moderna* come tipo di vaccino anti-SARS-CoV-2 ricevuto, 3) abitudine al fumo (attiva o pregressa) e 4) presenza di disturbo neurocognitivo. Tali risultati si ritrovano anche nel grafico SHAP (**Figura 7, b**).

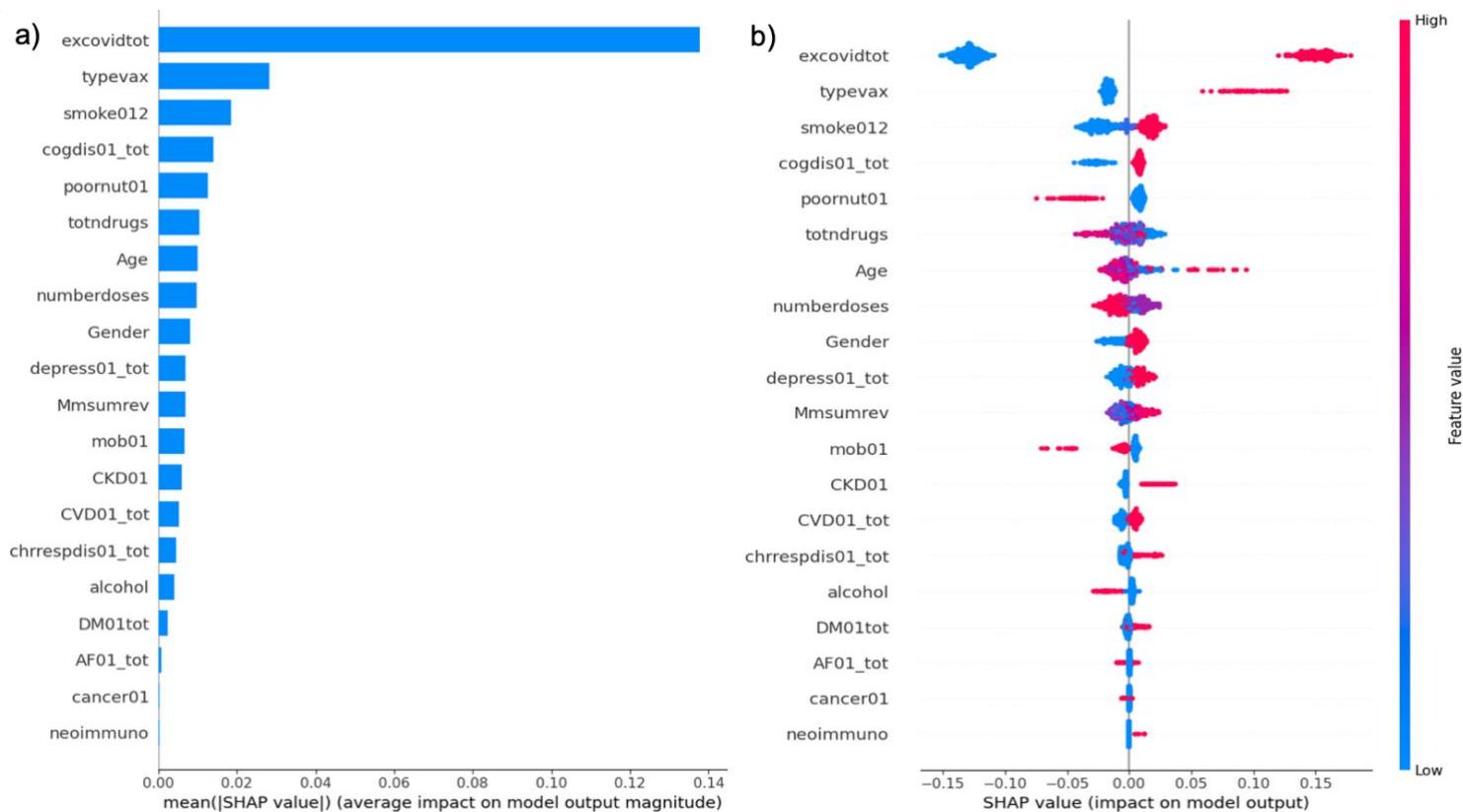


Figura 7. Importanza (a) e SHAP values (b) dei predittori per risposta anticorpale alta a 2 mesi

A 6 mesi, la popolazione analizzata di 381 pazienti è stata suddivisa in due classi di risposta anticorpale alta vs. bassa in base alla mediana di IgG anti-S, uguale a 280 AU/mL (range: 4.81-34200 AU/mL). In relazione alle metriche ottenute, il modello predittivo risultato migliore nello studio della risposta anticorpale alta vs. bassa a 6 mesi è risultato il Random Forest (*precision*=0.88, *recall*=0.78, *F1-score*=0.83 e *accuracy*=0.83); si veda **Tabella 6**.

Tabella 5. Metriche di performance nei modelli di AI testati sul dataset nella predizione della risposta anticorpale a 6 mesi (alta vs. bassa) a vaccino anti-SARS-CoV-2.

Modello	Bilanciamento	Precision	Recall	F1-Score	Accuracy
XGB	-	0.75	0.85	0.80	0.80
RF	-	0.89	0.78	0.83	0.83

Legenda: XGB = eXtreme Gradient Boosting; RF = Random Forest

Come mostrato **Figura 8 (a)**, i primi quattro predittori per risposta anticorpale alta, in ordine di importanza, sono risultati: 1) pregressa infezione da Covid-19, 2) basso numero di dosi vaccinali anti-SARS-CoV-2 ricevute, 3) abitudine al fumo (attiva o pregressa) e 4) *Moderna* come tipo di vaccino anti-SARS-CoV-2 ricevuto. Tali risultati si riscontrano anche nel grafico SHAP (**Figura 8, b**).

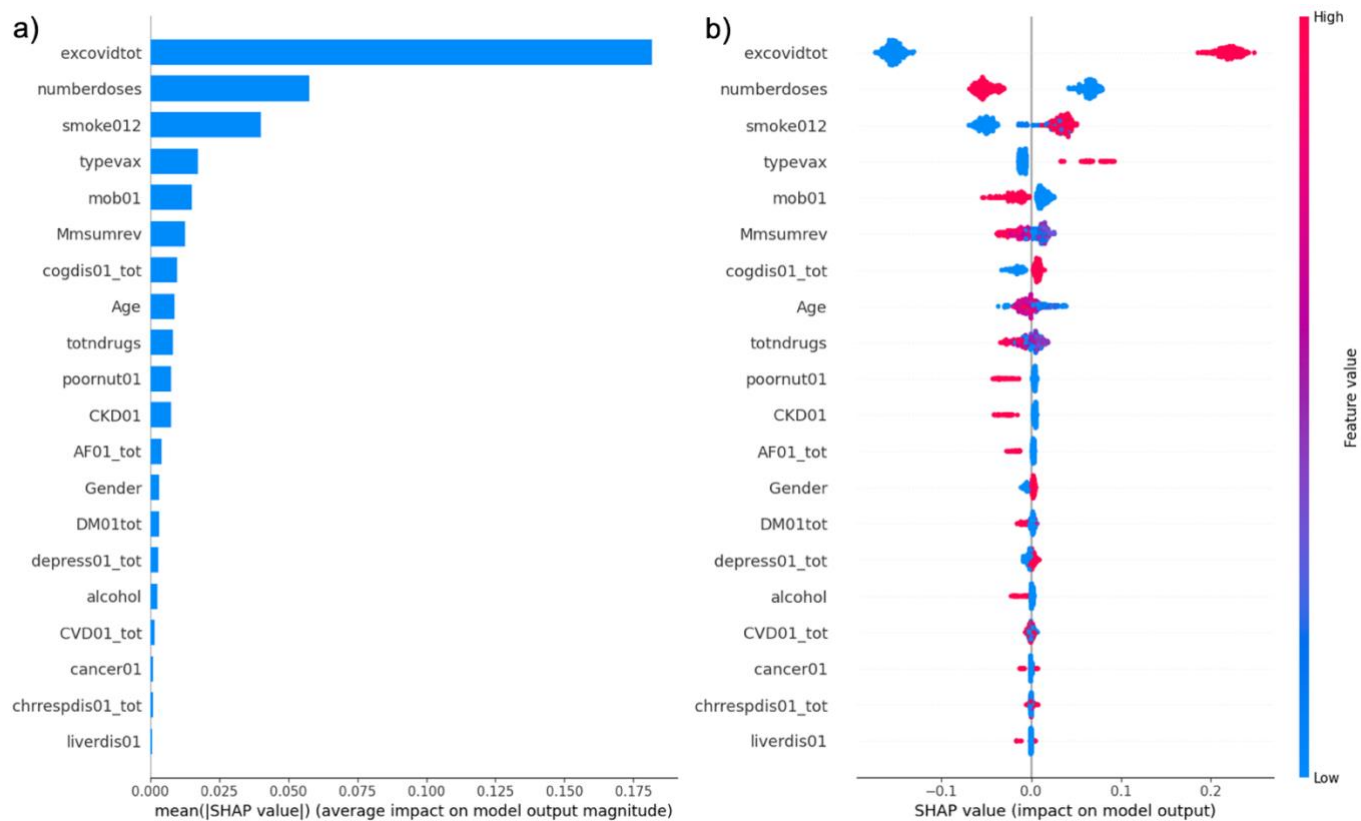


Figura 8. Importanza (a) e SHAP values (b) dei predittori per risposta anticorpale alta a 6 mesi

2.4. Discussione

In base ai risultati ottenuti mediante i modelli di AI, la presenza di una pregressa infezione da Covid-19 è stata la *feature* maggiormente associata sia a minor tasso di infezione da Covid-19 a 12 mesi che, nel sottocampione con dati disponibili, a più alta risposta umorale anti-S. Le altre caratteristiche dei partecipanti, in particolare il numero di dosi vaccinali ricevute, il numero di farmaci in terapia e specifiche malattie croniche, si sono associate in modo discontinuo agli *outcome* esplorati.

La coorte analizzata, che ha incluso 3276 persone ultra-sessantenni residenti in struttura, riportava un'età avanzata, una maggiore prevalenza di sesso femminile ed un'elevata complessità assistenziale, supportata dall'alta frequenza di disabilità funzionale e comorbidità; tali risultati sono in linea con quanto osservato in analoghi contesti di *long-term care*.⁷⁶ Inoltre, la vulnerabilità di questa specifica popolazione, diffusamente riportata in letteratura,⁷⁷ è anche testimoniata dall'alta prevalenza di disturbo neurocognitivo, maggiore o minore, attestatasi a quasi il 90% del campione raccolto. Le differenze in base al sesso osservate relativamente alle caratteristiche sociodemografiche e cliniche confermano trend consolidati: le donne, seppur più longeve, presentavano un profilo funzionale e cognitivo più compromesso, con minori punteggi di IADL e MMSE, maggior prevalenza di disabilità motoria e di disturbo neurocognitivo.⁷⁸ Parallelamente, la minore esposizione a fattori di rischio comportamentali (fumo e alcol) e la ridotta prevalenza di specifiche comorbidità cardiometaboliche e respiratorie nel sesso femminile rispecchiano pattern epidemiologici largamente descritti, anche nella popolazione geriatrica.⁷⁹

Durante il periodo di osservazione di 12 mesi, nel nostro campione l'incidenza di infezione da SARS-CoV-2 si è attestata di circa il 9%. Tale riscontro supporta indirettamente l'efficacia della campagna vaccinale anche in questa specifica popolazione, i cui dati relativi ad infezione ed ospedalizzazione da Covid-19 apparivano notevolmente superiori ad inizio della pandemia.⁸⁰ Infatti, durante la prima ondata, coloro residenti in LPA sono stati il gruppo maggiormente colpito dal Covid-19 e questo per diverse ragioni: oltre alla vulnerabilità intrinseca del soggetto in struttura, spesso anziano e comorbido, anche l'ambiente stesso, che prevede attività di gruppo e camere occupate da più persone, ha favorito la diffusione dell'infezione da SARS-CoV-2.⁴¹ E' noto, tuttavia, come i pazienti anziani e fragili possano presentare inoltre una ridotta risposta immunitaria a vaccino rispetto

a soggetti più giovani,⁸¹ ma pochi sono i dati attualmente disponibili in letteratura su questo tipo di popolazione. Uno studio pubblicato nel 2022, su 312 individui vaccinati con *Pfizer-BioNTech*,⁸² mostra come il gruppo di individui con età ≥ 60 anni riporti una risposta anticorpale di minor intensità, sia in termini di incremento dei livelli di anti-S dopo la vaccinazione che di picco assoluto raggiunto. Nonostante ciò, a 6 mesi dopo la somministrazione della prima dose vaccinale, nel gruppo degli anziani i livelli anticorpali presentavano una diminuzione meno rapida rispetto a quello dei più giovani. Anche relativamente alle cellule B della memoria, che negli over 60 risultavano inferiori nelle prime 3 settimane post-vaccinazione, a 3 e a 6 mesi raggiungevano livelli simili alla classe dei più giovani. Contrariamente a tali riscontri, due studi svolti su ospiti di *long term care*, rispettivamente 149⁸³ e 46 soggetti,⁸⁴ riportavano fra questi una risposta umorale subottimale rispetto ai controlli. L'alta prevalenza di *low responder* a vaccino fra individui anziani e fragili, almeno nelle prime settimane, potrebbe essere legata sia a fenomeni secondari all'invecchiamento, come l'immunosenescenza,^{14,15} sia alla presenza di molteplici malattie croniche e di disabilità, frequenti in questa popolazione.⁶⁵ Nonostante ciò, come mostrato da un'analisi di coorte pubblicata nel 2021, la vaccinazione appare ridurre significativamente il rischio di infezione da SARS-CoV-2 anche fra gli anziani fragili ospiti di struttura.⁸⁵ In questa direzione, una maggior protezione all'infezione, in seguito alla seconda dose vaccinale, è anche riportata in un ulteriore studio svolto su individui in *long term care*, a supporto della necessità di una campagna vaccinale che preveda almeno due dosi negli anziani fragili.⁸⁶

In base alle analisi di correlazione svolte, non sono state identificate relazioni lineari fra singoli predittori e infezione incidente da Covid-19 entro 12 mesi, evidenziando la complessità del fenomeno e giustificando l'impiego di approcci basati su AI. I modelli di AI, infatti, sono capaci di cogliere possibili effetti combinati o interazioni tra variabili, gestendo simultaneamente potenziali confondenti, presenza di dati mancanti e profili clinici eterogenei. In quest'ottica potrebbero quindi avere un potenziale applicativo diretto in epidemiologia, poiché implementabili nella pratica clinica per supportare la stratificazione del rischio e la personalizzazione delle strategie vaccinali o preventive, previa adeguata validazione esterna.^{53,55} Nel nostro dataset, il modello di Random Forest si è dimostrato quello con metriche di performance migliori nell'identificazione dei predittori di infezione incidente, evidenziando come la precedente esposizione al virus rappresenti l'elemento

protettivo più rilevante; tale riscontro è risultato in linea con i dati disponibili in letteratura. Fra questi, in una revisione sistematica con meta-analisi su 65 studi longitudinali, pubblicata su *The Lancet* (2023), la protezione complessiva (*pooled Odds Ratio*) data da una pregressa infezione da Covid-19 verso una reinfezione è stata stimata del 45.3% (Intervallo di Confidenza al 95%: 17.3-76.1).⁸⁷

In base ai nostri risultati, un basso numero di farmaci assunti in totale in terapia e il ridotto consumo di alcol settimanale si sono associati a una minore probabilità di infezione da Covid-19 nei 12 mesi. Questi risultati sono coerenti con la letteratura presente, che riporta come coloro in polifarmacoterapia presentino un rischio di ospedalizzazione incrementato, in particolar modo per cause infettive.⁸⁸ Inoltre, come sostenuto da una ricerca sul Covid-19 pubblicata nel 2023, il consumo cronico di alcol sembra ridurre la risposta antivirale al SARS-CoV-2, sia a livello polmonare che sistemico.⁸⁹

Infine, anche nel nostro studio e come già precedentemente discusso, le due dosi vaccinali sono apparse predittive per una minor infezione incidente da Covid-19, a supporto dell'instaurarsi di una protezione immunitaria efficace conferita dal ciclo vaccinale completo anche in questa popolazione anziana fragile.⁷⁰⁻⁷⁴

L'unico studio disponibile in letteratura sull'impiego di tecniche di AI per predire l'infezione da SARS-CoV-2 in residenti in regime di lungodegenza è quello svolto da Lee et al.⁷³ In particolare, gli autori riportano elevate metriche di performance nei modelli di ML sviluppati (XGBOOST). Seppur i predittori analizzati siano differenti da quelli inclusi nel nostro studio, i risultati appaiono coerenti con quanto da noi riscontrato: una maggiore complessità clinica, supportata da un maggior carico di comorbidità, è infatti emerso come fattore predittivo rilevante per infezione da parte del virus.

Nel sottogruppo con dati umorali disponibili, gli alti (vs. bassi) livelli di anticorpi anti-S a 2 e a 6 mesi non si sono correlati in modo significativo ai predittori analizzati. Testando sui dati i modelli di AI, il Random Forest, scelto per le migliori metriche di performance, ha evidenziato come la pregressa infezione da Covid-19 fosse il principale determinante di una risposta anticorpale elevata sia a 2 che a 6 mesi. Tale risultato è coerente con quanto già discusso precedentemente in relazione all'efficacia della protezione secondaria al vaccino verso infezione incidente da SARS-CoV-2 riportata in letteratura.⁷⁵

L'associazione positiva, osservata sia a 2 sia a 6 mesi, fra il vaccino *Moderna* e gli alti livelli di anti-S conferma precedenti studi, che riscontravano titoli anticorpali più alti in risposta a

tale vaccino rispetto agli altri a mRNA.⁹⁰ Tale differenza è stata principalmente attribuita alla maggiore quantità di mRNA per dose di vaccino e alla specifica formulazione lipidica del *Moderna*. Questi elementi determinerebbero un'espressione più robusta della proteina Spike sulle cellule dell'ospite e, di conseguenza, una risposta immunitaria più intensa e duratura scatenata da tale vaccino.⁹¹

Più complessa è l'interpretazione del ruolo dell'abitudine al fumo, della presenza di disturbo neurocognitivo e del numero maggiore di dosi vaccinali ricevute, risultati infatti come predittori per alta risposta anticorpale. Riguardo al fumo, i fumatori - attivi o pregressi - erano probabilmente soggetti con migliore riserva funzionale e cognitiva rispetto ai coetanei non fumatori, più disabili e fragili (*survivor bias*). Relativamente al disturbo neurocognitivo, la sua prevalenza prossima al 90% nella nostra coorte, potrebbe aver limitato la variabilità dell'esposizione e quindi influito sull'associazione ottenuta, risultata paradossa. Infine, l'associazione fra un numero maggiore di dosi vaccinali ricevute e una minor risposta anticorpale a 6 mesi potrebbe derivare da un effetto di *reverse causation*, per cui i soggetti più fragili e comorbidi, e verosimile minor livelli ematici di anti-S, siano stati candidati più frequentemente a ricevere dosi aggiuntive.

Gli *outcome* relativi al supporto con ossigeno e decesso a 12 mesi non hanno prodotto modelli predittivi con metriche di performance adeguate, evidenziando la necessità di disporre di dati clinici più accurati e multimodali, oltre ad una gestione più efficace durante la raccolta dei dati mancanti di follow-up. Tali limitazioni sottolineano anche il potenziale applicativo futuro dei modelli di AI: con dataset più completi e validazione esterna, essi potrebbero supportare decisioni cliniche, identificare pazienti ad alto rischio e ottimizzare interventi preventivi in contesti epidemiologici reali.

Fra i punti di forza del nostro studio, oltre al suo design multicentrico longitudinale e all'elevata numerosità del campione, è da riportare il tipo di popolazione analizzata, composta da residenti in LPA, generalmente esclusi dai trial clinici, nonostante siano la classe di soggetti su cui la pandemia ha avuto maggiore impatto.

Nell'interpretare i nostri risultati, alcuni limiti sono altresì da discutere. Nonostante sia stata effettuata un'estesa valutazione delle comorbidità dei soggetti, la durata e severità delle malattie non è stata raccolta e questo potrebbe aver ridotto la forza delle associazioni riscontrate. Inoltre, l'immunità cellulo-mediata indotta dal vaccino anti-SARS-CoV-2, essenziale nella protezione dal virus, non è stata quantificata e non esplorata nel nostro studio

come *outcome* come indicatore di efficacia della risposta immunitaria. In aggiunta, l'utilizzo della mediana del titolo anticorpale per la classificazione della risposta vaccinale, dipendendo dalla distribuzione dei dati raccolti e non seguendo un cut-off standardizzato, può aver ridotto la generalizzabilità dei risultati ottenuti a coorti esterne. Infine, l'alto numero di dati mancanti per alcune delle variabili registrate e lo scarso bilanciamento del dataset relativamente ad alcuni degli *outcome* indagati possono aver influenzato le capacità predittive dei modelli di AI utilizzati, riducendone le metriche di performance.

2.5. Conclusioni

In base ai modelli di AI, la presenza di una pregressa infezione da SARS-CoV-2 appare la caratteristica maggiormente predittiva per minor rischio di infezione a 12 mesi in pazienti anziani residenti in regime di LPA. Inoltre, nel sottocampione con dosaggio anticorpale disponibile, tale *feature* risulta quella maggiormente associata ad alti livelli di IgG anti-S a 2 e 6 mesi.

La nostra ricerca evidenzia come l'introduzione della AI nell'analisi dei dati epidemiologici possa offrire un contributo rilevante nella comprensione del rischio infettivo e della risposta immunitaria in anziani istituzionalizzati, popolazione caratterizzata da estrema eterogeneità clinica. Nonostante i modelli di AI utilizzati abbiano identificato predittori consistenti e clinicamente plausibili, la predizione degli esiti più severi (necessità di ossigenoterapia e decesso) ha mostrato basse metriche di performance, suggerendo la necessità di dati clinici più dettagliati e integrati da più fonti. Nel complesso, i risultati del nostro studio supportano l'utilità dell'AI come strumento complementare alle analisi statistiche classiche, con potenziale impatto nella prevenzione e nella personalizzazione degli interventi medici anche in contesti di *long-term care*.

STUDIO 2

Influenza dei fattori genetici sulla severità della malattia acuta da Covid-19

in adulti ospedalizzati:

studio prospettico osservazionale

3.1. Scopo dello Studio

Valutare tramite modelli di AI la presenza di un'associazione fra specifici loci genetici e decesso in adulti ospedalizzati per malattia acuta da Covid-19. Secondariamente, indagare se specifici loci genetici risultano associati alla necessità di intensificazione delle cure e all'insorgenza di delirium intraospedaliero durante la malattia acuta.

3.2. Materiali e Metodi

Studio monocentrico longitudinale svolto nella U.O. di Medicina Interna Universitaria dell'Azienda Ospedaliero-Universitaria Sant'Anna di Ferrara (Italia) su adulti ospedalizzati con malattia acuta da Covid-19 fra l'1 marzo 2020 e l'1 marzo 2021. Lo studio è stato approvato dal Comitato Etico locale Area Vasta Emilia Centro (CE-AVEC, numero di protocollo: 540/2020/Oss/AOUFe).

3.2.1. Popolazione

I criteri di inclusione sono stati:

- 1) presenza di infezione acuta da SARS-CoV-2;
- 2) ricovero presso U.O. di Medicina Interna.

Per l'arruolamento nello studio, per ogni paziente è stato ottenuto il consenso informato.

Sono stati esclusi coloro senza un'infezione in corso da SARS-CoV-2 all'ingresso in reparto.

I pazienti sono stati valutati al baseline e a 6 mesi dal reclutamento.

3.2.2. Raccolta dati

Al baseline, per ogni paziente sono state raccolte le seguenti informazioni: caratteristiche sociodemografiche, stile di vita (fumo), setting di provenienza pre-ricovero (domicilio, struttura e trasferimento ospedaliero), dipendenza funzionale (autonomo vs. cammino con ausilio vs. mobilità passiva letto-poltrona/allettamento), malattie croniche (*Charlson Comorbidity Index* – CCI – e tipo di malattie croniche, quali ipertensione arteriosa, cardiopatia ischemica cronica, scompenso cardiaco cronico, pregresso ictus ischemico, arteriopatia obliterante periferica, pneumopatie croniche, diabete, insufficienza renale cronica, epatopatia, neoplasia, demenza), terapia farmacologica domiciliare (numero di farmaci ed assunzione di anticoagulanti, ACE-inibitori, sartani, metformina, Inibitori di Pompa Protonica - PPI) ed esami bioumorali di routine.

Inoltre, all'ingresso in reparto, ad ogni paziente è stato effettuato un prelievo di sangue venoso per l'analisi del DNA. In particolare, sono stati analizzati i seguenti loci genetici: aplotipo APOE, ACE1, ACE2, HLA-A, TP53, OAS3, LZTFL1, gruppo sanguigno AB0 e PPARGC1A. La raccolta dei campioni di sangue venoso è stata svolta attraverso provette BD Vacutainer con processamento del plasma entro un'ora dal prelievo, quindi congelato a -80°C ed analizzato in cieco. Il DNA è stato estratto dal sangue intero congelato attraverso BioRobot EZ1 system, Qiagen (Hilden, Germania), strumento automatizzato per l'estrazione e la purificazione del DNA. Per la rilevazione degli SNPs selezionati è stata svolta la *Polymerase Chain Reaction* (PCR), in particolare: APOE (R158C, rs7412; C>T e C112R, rs429358; T>C), ACE2 (rs2285666; G>A), HLA-A (rs2571381; T>C e rs2499; T>G), TP53 (P72R, rs1042522; G>C), OAS3 (rs10735079; G>A), LZTFL1 (rs35044562; A>G), PPARGC1A (rs8192678; G>A) e gruppo sanguigno AB0 (rs657152; A>C). L'elettroforesi su gel di agarosio è stata invece utilizzata per identificare i polimorfismi di ACE1 (ins/del, rs4646994; alleli I/D: 490 bp / 190 bp). Come controllo interno di qualità, circa il 15% dei campioni per ciascun genotipo è stato rianalizzato in modo casuale.⁹²

A 6 mesi, tramite database ospedaliero, è stata valutata la sopravvivenza. Durante la degenza ospedaliera, il delirium è stato diagnosticato in base ai criteri diagnostici del Manuale Diagnostico e Statistico dei Disturbi Mentali, V Edizione (DSM-V),⁹³ mentre per intensificazione delle cure è stata considerata la necessità di trasferimento urgente presso l'U.O. di Pneumologia o di Terapia Intensiva.

3.2.3. Analisi statistica

L'analisi descrittiva delle variabili è stata espressa per le quantitative attraverso media e deviazione standard (DS), mentre per le categoriche attraverso frequenze assolute o relative. Il confronto tra gruppi in base al sesso è stato effettuato tramite test *t-Student* (variabili continue) e test chi-quadrato (variabili dicotomiche).

Relativamente all'analisi predittiva, l'associazione fra ogni locus genetico ed *outcome* è stata indagata attraverso diversi modelli di ML. I predittori inclusi nei modelli con aggiustamento di base (n. 1-10) sono stati età e sesso, mentre nei modelli completamente aggiustati (n. 11-20) includevano anche setting di provenienza, dipendenza funzionale, abitudine al fumo,

ipertensione arteriosa, cardiopatia ischemica cronica, scompenso cardiaco cronico, pregresso ictus, arteriopatia obliterante periferica, pneumopatie croniche, insufficienza renale cronica, epatopatia, neoplasia, diabete, demenza, CCI e numero di farmaci assunti in terapia domiciliare. Per esplorare il decesso a 6 mesi (*main outcome*), considerato come *time-to-event*, è stato utilizzato un Random Survival Forest (RSF),⁹⁴ valutato mediante *Concordance Index* (C-index). L'intensificazione delle cure e l'insorgenza di delirium intraricovero (*secondary outcomes*) sono stati indagati mediante Random Forest Classifier (RF) e le prestazioni di tali modelli sono state valutate attraverso *accuracy*, *precision*, *recall*, *F1-score*. Poiché i dataset relativi a questi ai due *outcome* secondari risultavano fortemente sbilanciati, sono state applicate tecniche di bilanciamento, tra cui StratifiedKFold, *class weight* e *oversampling* (Synthetic Minority Over-sampling Technique - SMOTE) per migliorare la stabilità e la robustezza delle predizioni. Per l'interpretazione dei modelli è stata utilizzata la *impurity-based feature importance*, che quantifica il contributo delle variabili alla riduzione dell'impurità negli alberi decisionali.

3.3. Risultati

L'età media dei 360 partecipanti è stata di 71.2 anni (SD 15.7) e il 47.6% era di sesso femminile. Il 18.1% dei partecipanti era fumatore e il 65.1% autonomo nel pre-ricovero. L'ipertensione arteriosa (62.2%), il diabete (19.2%), la demenza (18.3%) e la cardiopatia ischemica (15.0%) sono state le malattie croniche più frequenti; il punteggio medio di CCI è stato di 1.9 (SD 2.2). Il numero medio di farmaci assunti in terapia domiciliare era di 5.9 (SD 4.1).

La **Figura 1** mostra la prevalenza dei singoli SNPs dei loci genetici isolati nella popolazione

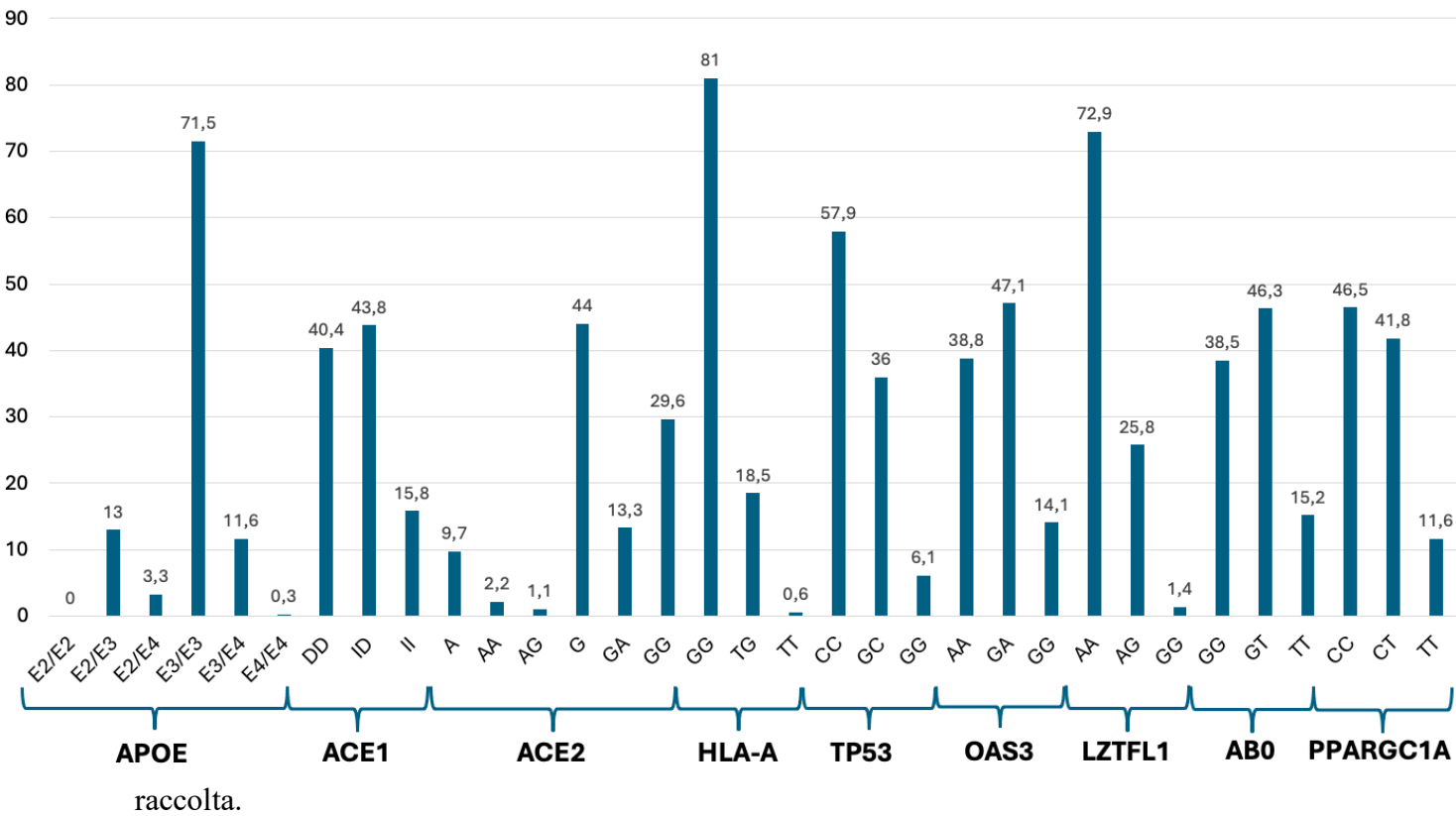


Figura 1. Prevalenza dei singoli SNPs dei loci genetici analizzati

La mortalità a 6 mesi è stata del 28.0%. Il 23.3% dei pazienti ha richiesto un'intensificazione delle cure, mentre l'8.3% ha avuto delirium durante il ricovero.

In **Tabella 1**, le caratteristiche del campione sono presentate in base al sesso. In particolare, le donne erano più anziane e riportavano con maggior frequenza disabilità al cammino nel

pre-ricovero rispetto agli uomini. Inoltre, relativamente alle comorbidità indagate, presentavano una maggior prevalenza di demenza, mentre una minore di insufficienza renale cronica. Non sono state riscontrate ulteriori differenze significative fra i due gruppi.

Tabella 1. Caratteristiche socio-demografiche e cliniche del campione in base al sesso.

	Maschi (N=188)	Femmine (N=172)	p
Età, media (DS)	68.2 (14.6)	74.5 (16.1)	<0.001
Stile di vita			
Abitudine al fumo (attiva o pregressa)	40 (21.3)	25 (14.5)	0.128
Setting di provenienza			
<i>Domicilio</i>	122 (64.9)	100 (58.1)	0.654
<i>Struttura</i>	22 (11.7)	29 (16.9)	
<i>Trasferimento ospedaliero</i>	33 (17.6)	32 (18.6)	
Performance funzionali			
Cammino			
<i>Autonomo</i>	140 (74.5)	94 (54.7)	0.002
<i>Utilizzo di ausilio</i>	17 (9.1)	37 (21.5)	
<i>Completamente dipendente</i>	31 (16.5)	41 (23.8)	
Terapia domiciliare			
No. farmaci totale, media (DS)	5.7 (4.3)	6.1 (3.9)	0.337
Anticoagulanti*	18 (9.6)	11 (6.4)	0.361
ACE Inibitori	40 (21.5)	44 (25.6)	0.433
Sartani	19 (10.2)	24 (14.0)	0.345
Metformina	25 (13.4)	13 (7.6)	0.110
Inibitori di Pompa Protonica (PPI)	63 (33.9)	61 (35.7)	0.599
Comorbidità			
<i>Charlson Comorbidity Index</i> , media (DS)	3.0 (2.3)	2.7 (2.0)	0.331
Ipertensione arteriosa	114 (60.6)	110 (64.0)	0.590
Cardiopatía ischemica cronica	35 (18.5)	19 (11.0)	0.063
Scompenso cardiaco cronico	21 (11.2)	21 (12.2)	0.887
Pregresso ictus ischemico	24 (12.8)	21 (12.2)	1.000
Arteriopatía obliterante periferica	15 (8.0)	8 (4.7)	0.277
Pneumopatie croniche	19 (10.1)	16 (9.3)	0.937
Diabete	41 (21.7)	28 (16.3)	0.241
Insufficienza renale cronica	37 (19.7)	18 (10.5)	0.023
Epatopatía	9 (4.8)	7 (4.1)	1.000

Neoplasia	30 (16.0)	38 (22.1)	0.306
Demenza	27 (14.4)	39 (22.7)	0.057
Gruppo sanguigno AB0			
<i>A</i>	78 (41.5)	64 (37.2)	0.473
<i>B</i>	21 (11.2)	18 (10.5)	
<i>AB</i>	17 (9.0)	11 (6.4)	
<i>O</i>	72 (38.3)	79 (45.9)	
Esami ematici all'ingresso in reparto			
Globuli bianchi (cells/mcl), media (DS)	7590.2 (4509.6)	8635.4 (9026.2)	0.160
Linfociti (cells/mcl), media (DS)	1041.1 (546.9)	1334.8 (2461.6)	0.112
Creatinina (mg/dl), media (DS)	1.39 (1.40)	1.03 (0.85)	0.004
Proteina C reattiva (mg/l), media (DS)	8.07 (6.64)	8.13 (7.21)	0.941
Procalcitonina (ng/ml), media (DS)	2.47 (13.34)	0.96 (3.35)	0.189

Legenda: DS = Deviazione Standard.

Quando non specificato, in Tabella sono riportati valori assoluti e frequenze, N (%)

**Gli anticoagulati includono il warfarin ed anticoagulanti orali ad azione diretta*

Relativamente all'analisi della sopravvivenza, il modello predittivo migliore è risultato il Random Survival Forest (C-index=0.812); **Figura 2.**

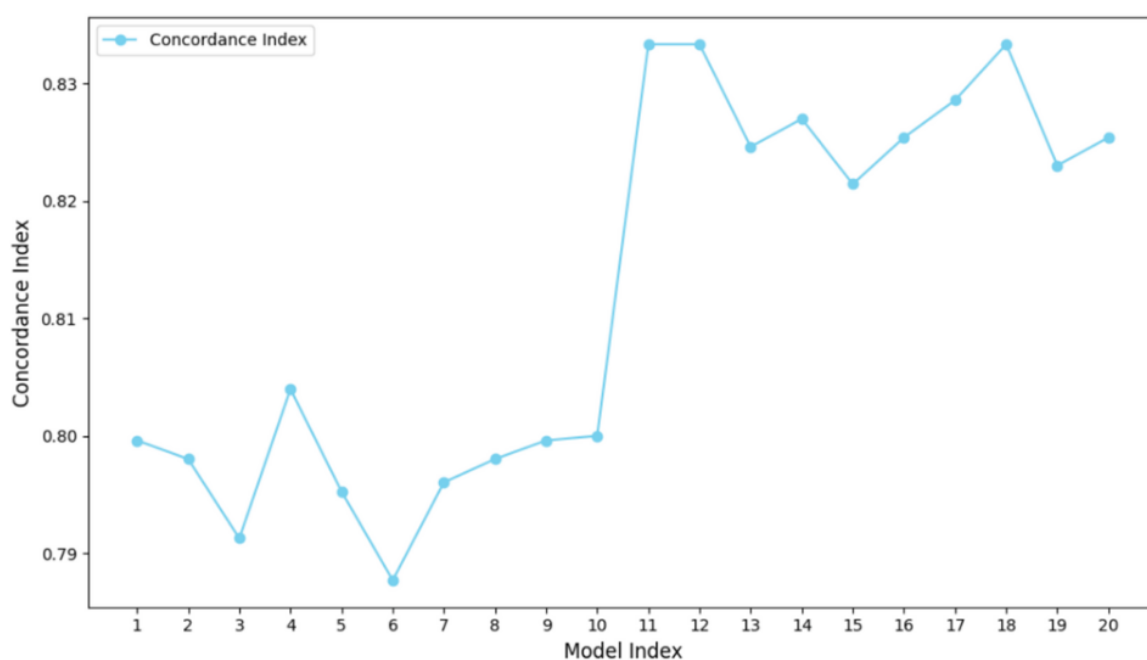


Figura 2. Concordance Index relativo ai differenti modelli Random Survival Forest

In base all'analisi delle *feature importance*, nei modelli con aggiustamento di base, la variabile con maggior peso nella predizione del decesso a 6 mesi è stata l'età (**Figura 3**). Per tutti i modelli esplorati, la seconda variabile per importanza è stata sempre il sesso, eccetto per il modello che includeva APOE.

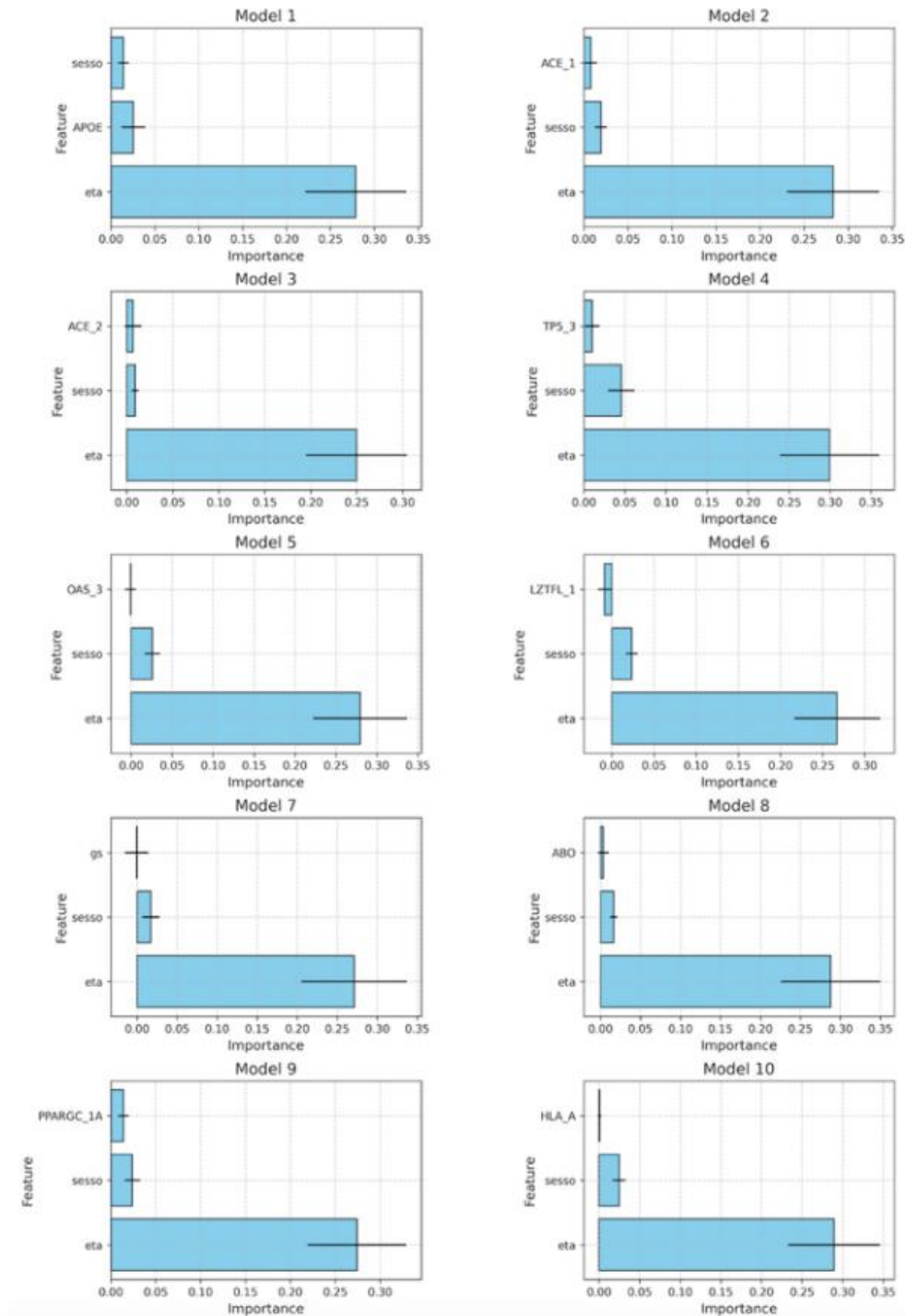
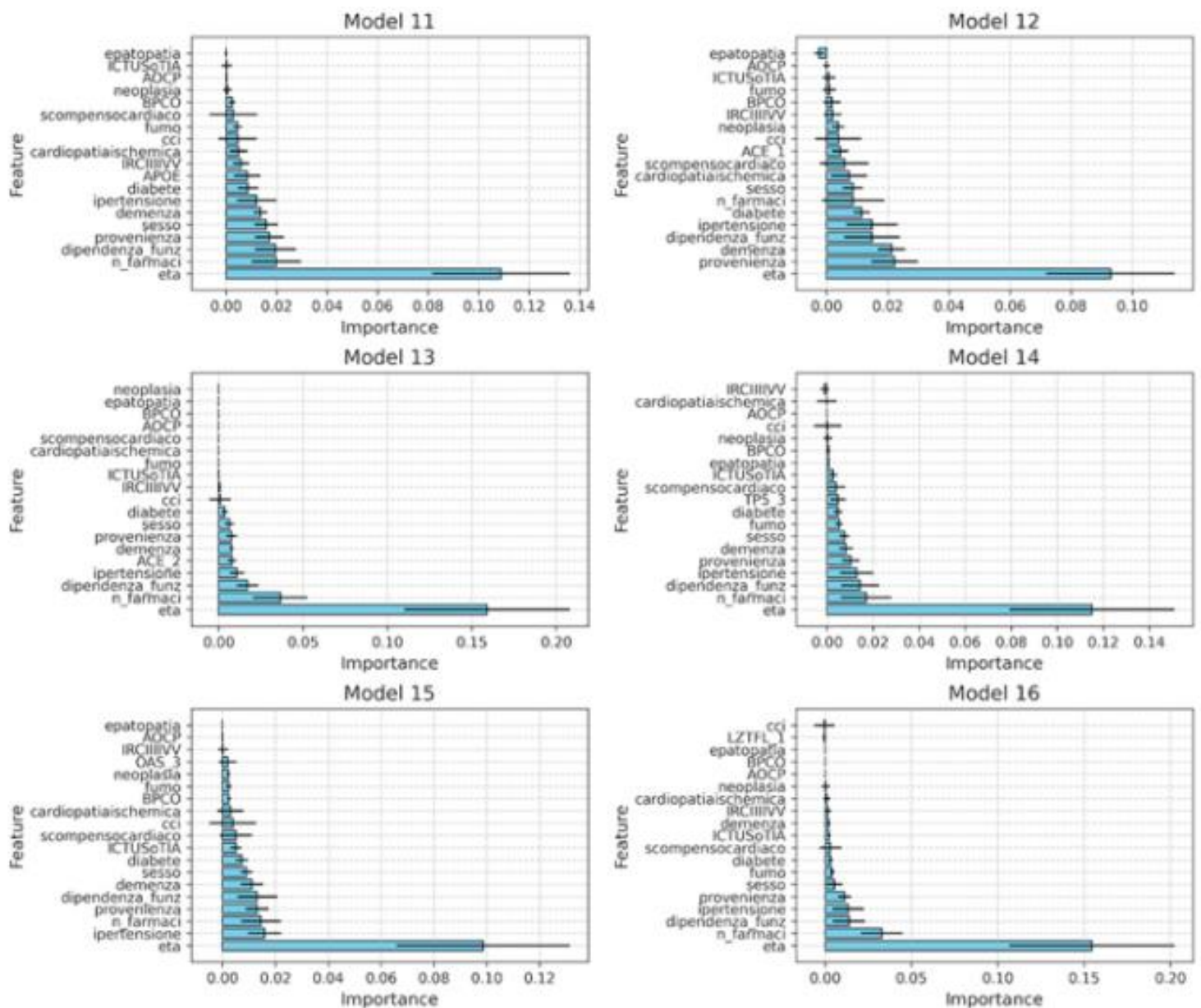


Figura 3. Importanza delle *features* nella predizione del decesso a 6 mesi (*basic adjusted models*)

Anche nei modelli completamente aggiustati, l'età è stato il predittore più importante per decesso a 6 mesi (**Figura 4**). La seconda *feature* più frequente per importanza è stata il numero di farmaci assunti in terapia domiciliare, in particolare nei modelli 11, 13, 14, 16, 17 e 18. Tra le cinque variabili con maggiore importanza predittiva, oltre all'età e alla terapia, emergono il setting di provenienza, la dipendenza funzionale e la presenza di ipertensione arteriosa o di demenza, con rilevanza diversa a seconda del modello. Inoltre, solo nei modelli che includevano ACE2 (modello 13) e PPARGC1A (modello 19), il locus genetico appariva come quinta *feature importance*.



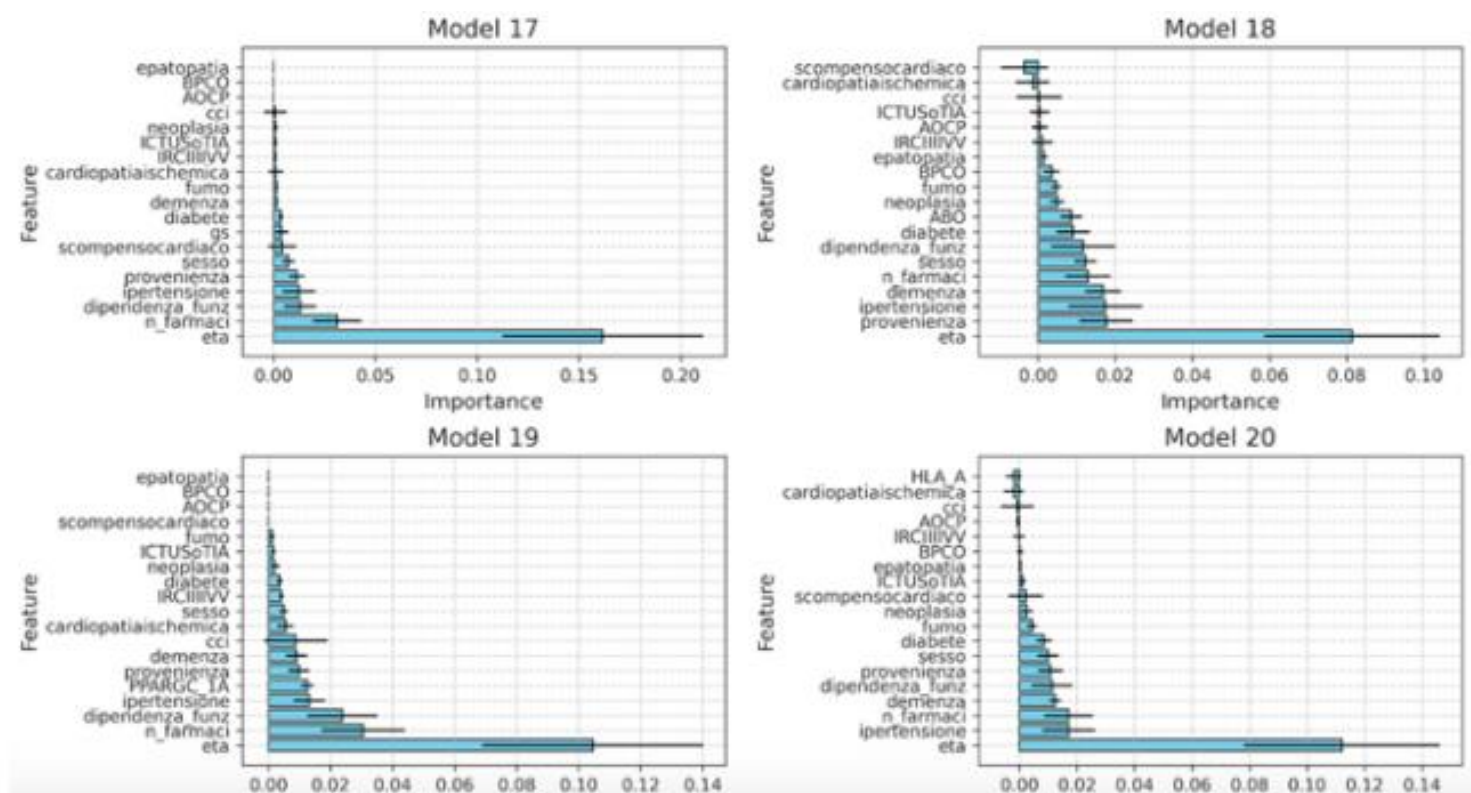


Figura 4. Importanza delle *features* nella predizione del decesso a 6 mesi (*fully adjusted models*)

Relativamente alla necessità di intensificazione delle cure e al delirium intraricovero, l'analisi delle *feature* non è stata svolta, in quanto i modelli predittivi addestrati sul dataset hanno fornito basse metriche di performance. In **Figura 5** sono riportate le matrici di confusione, utilizzate per valutare le prestazioni di modelli, che rappresentano la distribuzione delle predizioni rispetto alle classi reali.

a)

Confusion Matrix for Model 1 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	34	19
IntensCure	7	9

Confusion Matrix for Model 2 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	38	15
IntensCure	8	8

Confusion Matrix for Model 11 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	43	10
IntensCure	7	9

Confusion Matrix for Model 12 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	47	6
IntensCure	10	6

Confusion Matrix for Model 3 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	36	17
IntensCure	11	5

Confusion Matrix for Model 4 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	38	15
IntensCure	9	7

Confusion Matrix for Model 13 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	52	1
IntensCure	10	6

Confusion Matrix for Model 14 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	45	8
IntensCure	10	6

Confusion Matrix for Model 5 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	35	18
IntensCure	9	7

Confusion Matrix for Model 6 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	33	20
IntensCure	8	8

Confusion Matrix for Model 15 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	44	9
IntensCure	7	9

Confusion Matrix for Model 16 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	46	7
IntensCure	12	4

Confusion Matrix for Model 7 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	35	18
IntensCure	7	9

Confusion Matrix for Model 8 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	39	14
IntensCure	10	6

Confusion Matrix for Model 17 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	45	8
IntensCure	7	9

Confusion Matrix for Model 18 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	49	4
IntensCure	12	4

Confusion Matrix for Model 9 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	35	18
IntensCure	8	8

Confusion Matrix for Model 10 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	36	17
IntensCure	9	7

Confusion Matrix for Model 19 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	44	9
IntensCure	8	8

Confusion Matrix for Model 20 (intensificazione)

True label	Predicted label	
	No intensCure	IntensCure
No intensCure	44	9
IntensCure	6	10

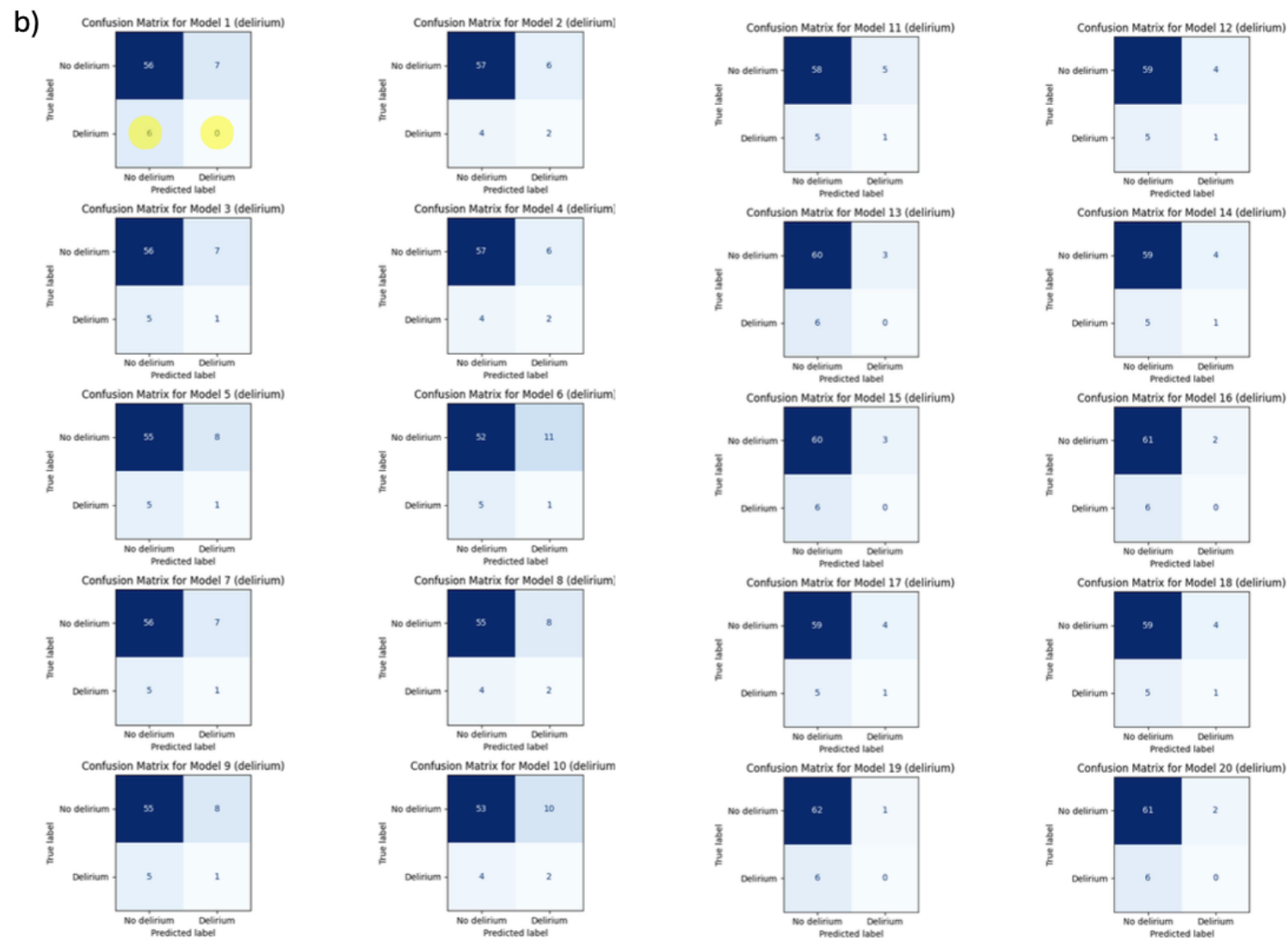


Figura 5. Matrici di confusione dei modelli per intensificazione delle cure e delirium

3.4. Discussione

Fra caratteristiche maggiormente predittive per decesso in adulti ospedalizzati per Covid-19, il nostro studio individua l'età, il numero di farmaci assunti in terapia domiciliare, il setting pre-ricovero, la dipendenza funzionale e la presenza di ipertensione arteriosa o demenza in anamnesi patologica remota. Seppur con un'importanza inferiore, alcuni loci genetici, quali ACE2 e PPARGC1A, hanno mostrato un rilevante potere predittivo.

La popolazione arruolata era caratterizzata da età avanzata e da un elevato carico di comorbidità. Questo dato riflette quanto riportato in letteratura, che identifica i soggetti ricoverati per Covid-19 come un gruppo di soggetti particolarmente vulnerabile e fragile. Nel nostro campione, infatti, lo sviluppo degli esiti avversi più severi della malattia da Covid-19 è risultato estremamente frequente, con un tasso di mortalità maggiore del 25% a 6 mesi, così come l'alta necessità di ricorso a cure intensive e di incidenza di delirium durante la degenza. Come mostrato in una meta-analisi su 26652 pazienti ospedalizzati per Covid-19, la prevalenza di fragilità variava dal 9.1% all'87.2%, con una stima complessiva del 51.4% (Intervallo di Confidenza al 95%: 39.9-62.9). Inoltre, fra i 21 studi inclusi, gli autori riscontravano la presenza di un'associazione significativa fra fragilità all'ingresso e mortalità entro 2 mesi (*polled Odds Ratio*: 3.0; Intervallo di Confidenza al 95%: 2.3–3.9), indipendentemente dall'età del soggetto.⁹⁵ Tale risultato viene confermato anche da uno lavoro svolto durante la pandemia su pazienti ultra-ottantenni ricoverati per infezione da SARS-CoV-2, di cui più del 75% era fragile al baseline: un maggiore grado di fragilità si associava in maniera significativa ad un aumentato rischio di morte entro 4 mesi.⁹⁶ Perciò, sia secondo i nostri risultati che in base alla letteratura disponibile, i soggetti ricoverati per Covid-19 appaiono più frequentemente fragili e con un maggior rischio di eventi avversi.

Relativamente all'analisi descrittiva svolta in base al sesso, le donne risultavano mediamente più anziane e più frequentemente affette da disabilità fisica e demenza rispetto agli uomini. Tale riscontro è coerente con quanto riportato in letteratura: il sesso femminile, pur presentando una maggiore aspettativa di vita, tende a mostrare una più elevata prevalenza di disabilità rispetto al sesso maschile, che si presume proprio secondaria alla maggiore longevità.⁹⁷ È altresì noto come le demenze, tra cui si ricordi la Malattia di Alzheimer come sottotipo più frequente dopo i 65 anni, presentino una predilezione per il sesso femminile, risultando in linea con quanto anche da noi riscontrato.⁹⁸ Relativamente alle altre malattie

croniche, nel nostro campione gli uomini mostravano una maggiore prevalenza di insufficienza renale cronica rispetto alle donne. Tale riscontro risulta consistente con quanto riportato in letteratura, secondo cui questa differenza sarebbe dovuta agli ormoni sessuali, con un effetto di rischio associato al testosterone e un ruolo protettivo attribuibile agli estrogeni.⁹⁹

Tra i modelli di ML esplorati, il Random Survival Forest ha mostrato le migliori performance predittive per decesso a 6 mesi. L'età è risultata il predittore dominante in tutti i modelli esplorati, sia in quelli con aggiustamento di base che completamente aggiustati, confermandone il ruolo centrale nella prognosi a lungo termine del Covid-19. Com'è noto da tempo in medicina geriatrica, nonostante l'età anagrafica sia spesso individuata negli studi epidemiologici come uno dei principali fattori di rischio per esiti avversi in corso di malattie infettive acute,¹⁰⁰ da sola non appare in grado di giustificare in maniera indipendente tale associazione.¹⁰¹ L'età anagrafica, infatti, risulta intrinsecamente associata ad un più frequente elevato carico di comorbidità, disabilità e fragilità,³³ veri determinanti di un maggior rischio di prognosi avversa (*effect modifier*).¹⁰² In ambito di Covid-19, una revisione sistematica, svolta ad inizio della pandemia su soggetti infetti, mostra come per ogni aumento unitario di anno di età, vi sia un incremento del tasso di mortalità del 7.4%.¹⁰³ Tale riscontro è supportato anche da uno studio osservazionale svolto in Italia su 1985 pazienti ospedalizzati con SARS-CoV-2, in cui l'incidenza maggiore di decesso viene riscontrata nelle classi di età più avanzate.¹⁰⁴ Come menzionato, tali risultati, incluso quello del nostro studio, possono essere supportati da un maggior carico di comorbidità e disabilità simultaneamente presenti nei soggetti di età avanzata. In questa direzione, uno studio condotto su 6512 pazienti ospedalizzati per Covid-19 riporta un tasso di mortalità più elevato nel gruppo con comorbidità, la cui età media risulta superiore rispetto ai soggetti non multimorbidi.¹⁰⁵ Coerentemente con quanto discusso, la *scoping review* svolta da Burnazovic et al. nel 2023⁶⁹ e relativa alle applicazioni di AI in ambito di Covid-19, riporta come l'età anagrafica e le comorbidità siano fra i predittori più rilevanti per *outcome* avversi, incluso il decesso, in pazienti anziani ospedalizzati con infezione acuta.

Nei modelli completamente aggiustati, il numero di farmaci assunti a domicilio emerge frequentemente come secondo predittore più importante per decesso. Tale riscontro è coerente con quanto noto in letteratura, in quanto la polifarmacoterapia appare un indicatore indiretto di comorbidità e fragilità, riflettendo una maggiore complessità clinica del

soggetto.¹⁰⁶ Inoltre, l'assunzione contemporanea di più farmaci, aumentando il rischio di interazioni farmacologiche, effetti collaterali e difficoltà nella gestione posologica, può contribuire ad incrementare il rischio di eventi avversi.¹⁰⁷ In questa direzione, uno studio retrospettivo svolto su 348 residenti in struttura con infezione da Covid-19 mostra come il numero di farmaci assunti *die* si associ in maniera indipendente ad una maggiore probabilità di decesso a 30 giorni (*Odds Ratio*: 1.09, Intervallo di Confidenza al 95%: 1.00-1.20), anche dopo aggiustamento per età, sesso, CCI, *Body Mass Index* e stato vaccinale anti-SARS-CoV-2.¹⁰⁸

Secondo i modelli di ML utilizzati, oltre ad età e terapia domiciliare, il setting pre-ricovero, la dipendenza funzionale, l'ipertensione arteriosa e la demenza appaiono predittori importanti per decesso. Come precedentemente discusso, tutti questi fattori, indicatori dello stato di salute del soggetto, confermano la vulnerabilità biologica e funzionale come determinante prognostico fondamentale anche nella popolazione anziana con Covid-19.^{33,91} Coerentemente, una revisione sistematica con meta-analisi svolta su 56 studi longitudinali, riporta che nei campioni *population-based* e in quelli provenienti da contesti ospedalieri, la presenza di disabilità si associa a un maggior rischio di decesso correlato a Covid-19, rispettivamente di 3.3 e 2.1 volte; tale risultato non è apparso confermato invece nelle popolazioni residenti in struttura, probabilmente per l'alta prevalenza di disabilità fra tali soggetti.¹⁰⁹ Relativamente alla demenza, noto fattore prognostico negativo, anche in soggetti con infezione da Covid-19 tale malattia appare associata ad un aumentato rischio di morte e di ricovero ospedaliero prolungato.¹¹⁰

Oltre a identificare le complesse relazioni fra i principali predittori, i modelli di AI potrebbero avere un'applicazione diretta nella pratica clinica: previa validazione esterna, il loro impiego in contesti ospedalieri potrebbe supportare i medici nella stratificazione del rischio di decesso, ottimizzare la gestione dei pazienti e guidare decisioni terapeutiche personalizzate.

L'analisi dei loci genetici mostra una prevalenza eterogenea dei diversi SNPs nella popolazione arruolata; tuttavia, la loro rilevanza nella predizione del rischio di decesso è risultata complessivamente marginale. Solo nei modelli che includevano ACE2 e PPARGC1A, il locus genetico appare tra le prime cinque variabili per importanza predittiva, suggerendo come, nel nostro campione, l'impatto dei fattori genetici sulla prognosi risulti nettamente inferiore rispetto ai determinanti sociodemografici e clinici. In letteratura, ACE2,

recettore primario di ingresso di SARS-CoV-2 nelle cellule umane, nonché enzima implicato nel sistema renina-angiotensina (RAS),¹¹¹ è stato individuato come principale locus genetico associato ad esiti avversi in corso di malattia da Covid-19.¹¹² Uno studio svolto su 176 adulti ricoverati per Covid-19 riporta come specifici polimorfismi di ACE2 appaiano fortemente predittivi per un andamento severo della malattia acuta.¹¹³ Il motivo fisiopatologico sembra ritrovarsi in uno sbilanciamento verso meccanismi pro-infiammatori, secondario all'interazione del virus con ACE2, determinante una *down-regulation* dell'attività di tale recettore.¹¹⁴ Anche il risultato riscontrato relativamente al locus PPARGC1A (*Peroxisome Proliferator-Activated Receptor Gamma Coactivator 1-Alpha*), importante regolatore del metabolismo mitocondriale,¹¹⁵ come predittore di decesso nei pazienti con Covid-19 appare fisiopatologicamente plausibile, seppur il suo ruolo sia meno indagato in letteratura rispetto ad ACE2. Infatti, una ridotta attività di PPARGC1A presente in caso di specifici polimorfismi potrebbe portare, in corso di infezioni acute come da Covid-19, ad una disfunzione mitocondriale con aumentato stress ossidativo ed infiammazione sistemica, con successivo viraggio a forme cliniche più severe di malattia.¹¹⁶ Tuttavia, il ruolo dei loci genetici negli esiti di infezione da SARS-CoV-2 necessita di ulteriori studi, che includano necessariamente la valutazione di fattori sia sociodemografici che clinici.

Per quanto riguarda gli *outcome* secondari, i modelli predittivi non hanno raggiunto performance adeguate per interpretare i dati raccolti, con matrici di confusione che mostrano una bassa accuratezza nella classificazione dei soggetti che hanno richiesto intensificazione delle cure o che hanno sviluppato delirium. Tale riscontro è probabilmente dovuto allo sbilanciamento dei dataset, che ha limitato la capacità del modello di apprendere correttamente le classi minoritarie, oltre che alla natura intrinsecamente multifattoriale di suddetti esiti clinici, meno strettamente associati a variabili standardizzabili e facilmente misurabili durante la raccolta dati. Ne consegue una ridotta capacità dei modelli di AI di cogliere la complessità dei processi fisiopatologici sottesi, suggerendo la necessità di campioni più ampi, dati multimodali e una caratterizzazione clinica più precisa per migliorare la predittività rispetto a tali *outcome*.

La natura prospettica dello studio, nonché l'inclusione fra i predittori di un vasto spettro di loci genetici, ne rappresentano importanti punti di forza. Fra i limiti, si ricordino il design monocentrico e la bassa numerosità del campione, che riducono la possibilità di generalizzazione dei risultati a coorti esterne. Inoltre, la durata e severità delle malattie

croniche, seppur raccolte in maniera estesa, non è stata registrata e questo può aver sottostimato le associazioni riscontrate. Da aggiungere, la mancata indagine di ulteriori eventi acuti o ospedalizzazioni nei 6 mesi successivi all'arruolamento può avere introdotto un *misclassification bias* degli esiti clinici intermedi, con conseguente sovrastima del tasso di decesso dovuto a Covid-19 nella popolazione indagata. Infine, lo scarso bilanciamento del dataset relativamente negli *outcome* secondari ha impedito l'analisi dei dati tramite modelli di AI utilizzati, a causa delle basse metriche di performance.

3.4. Conclusioni

In base ai modelli di AI utilizzati, in soggetti adulti ospedalizzati per Covid-19, i loci genetici maggiormente predittivi per decesso sono risultati ACE2 e PPARGC1A, anche se in misura minore rispetto ad altre *feature* sociodemografiche e cliniche, quali età, numero di farmaci, setting pre-ricovero, dipendenza funzionale, ipertensione arteriosa e demenza. Questi risultati suggeriscono che, pur avendo un ruolo nella prognosi dell'individuo, i dati genetici da soli non costituiscano un indicatore forte quanto i determinanti clinici tradizionali. Tuttavia, in base a tali risultati, consistenti con quanto noto in letteratura, l'impiego dell'IA in epidemiologia si conferma uno strumento promettente per integrare informazioni complesse ed eterogenee (cliniche, funzionali e genetiche), nell'ottica di approcci più personalizzati e accurati nella stratificazione del rischio.

4. CONCLUSIONI GENERALI

I dati raccolti durante il primo studio ed interpretati mediante algoritmi di ML mostrano come in soggetti anziani residenti in LPA, la presenza di una pregressa infezione da SARS-CoV-2 sia risultata la *feature* maggiormente predittiva per minor tasso di infezione da Covid-19 a 12 mesi e per più alta risposta umorale a vaccino anti-SARS-CoV-2. Le altre caratteristiche dei partecipanti, quali numero di dosi vaccinali ricevute, numero di farmaci in terapia domiciliare e comorbidità, sono associate in modo discontinuo agli *outcome* esplorati. Nel secondo studio, svolto su pazienti adulti ricoverati per Covid-19 in fase acuta, i loci genetici ACE2 e PPARGC1A sono risultati predittori per decesso a 6 mesi, seppur meno importanti rispetto a specifiche variabili sociodemografiche (età, setting pre-ricovero) e cliniche (numero di farmaci, dipendenza funzionale, ipertensione arteriosa, demenza).

In base ai risultati ottenuti, l'utilizzo della AI appare uno strumento promettente nel contesto degli studi di popolazione. Tuttavia, l'adozione di tale approccio necessita di un team multidisciplinare, basato sulla stretta collaborazione fra clinici, ingegneri e statistici, essenziale per una corretta creazione degli algoritmi e applicabilità dei risultati nella pratica clinica.

Bibliografia

1. World Health Organization. Coronavirus disease 2019 (COVID-19) situation report-97. World Health Organization, Geneva (2020).
2. Dixon, B. E. et al. Symptoms and symptom clusters associated with SARS-CoV-2 infection in community based populations: Results from a statewide epidemiological study. PLoS ONE 16, e0241875 (2021).
3. Onder, G., et al. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA 323, 1775-1776 (2020).
4. Nickel, C.H. Age, comorbidity, frailty status: effects on disposition and resource allocation during the COVID-19 pandemic. Swiss Med Wkly 150, 20169 (2020).
5. Wu, Z. et al. Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases from the Chinese Center for Disease Control and Prevention. JAMA 323, 1239–1242 (2020).
6. Jackson, C.B. et al. Mechanisms of SARS-CoV-2 entry into cells. Nature Reviews 23, 3-20 (2022).
7. Shang, J. et al. Structural basis of receptor recognition by SARS-CoV-2. Nature 581, 221–224 (2020).
8. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220 (2020).
9. Fajgenbaum, D.C. et al. Cytokine Storm. NEJM 383, 2255-2273 (2020).
10. Onder, G. et al. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA 323, 1775–1776 (2020).
11. Blanco-Melo, D. et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. Cell 181, 1036-1045.e9 (2020).
12. Franceschi, C. et al. Chronic inflammation (Inflammaging) and its potential contribution to age-associated diseases. Journals of Gerontology - Series A Biological Sciences and Medical Sciences 69, S4–S9 (2014).
13. Franceschi, C. et al. Inflammaging: a new immune–metabolic viewpoint for age-related diseases. Nature Reviews Endocrinology 14, 576–590 (2018).
14. Fulop, T. et al. Immunosenescence and inflammaging as two sides of the same coin: Friends or Foes? Frontiers in Immunology 8, 1960 (2018).

-
15. Goronzy, J. J. et al. Mechanisms underlying T cell ageing. *Nature Reviews Immunology* 19, 573–583 (2019).
 16. Agrawal, A. Mechanisms and implications of age-associated impaired innate interferon secretion by dendritic cells: A mini-review. *Gerontology* 59, 421–426 (2013).
 17. Chen, G. et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *Journal of Clinical Investigation* 130, 2620–2629 (2020).
 18. Yang, X. et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *The Lancet Respiratory Medicine* 8, 475–481 (2020).
 19. Remelli, F. et al. Clinical Features of SARS- CoV-2 Infection in Older Adults. *Clin Geriatr Med* 38(3), 483-500 (2022).
 20. Richardson, S. et al. Presenting Characteristics, Comorbidities, and Outcomes among 5700 Patients Hospitalized with COVID-19 in the New York City Area. *JAMA* 323, 2052–2059 (2020).
 21. Yang, J. et al. Prevalence of comorbidities and its effects in coronavirus disease 2019 patients: A systematic review and meta-analysis. *International Journal of Infectious Diseases* 94, 91–95 (2020).
 22. Trevisan, C. et al. COVID-19 as a Paradigmatic Model of the Heterogeneous Disease Presentation in Older People: Data from the GeroCovid Observational Study. *Rejuvenation Research* 25(3), 129-139 (2022).
 23. Guo, T. et al. Clinical Characteristics of Elderly Patients with COVID-19 in Hunan Province, China: A Multicenter, Retrospective Study. *Gerontology* 66, 467–475 (2020).
 24. Limpawattana, P. et al. Atypical presentations of older adults at the emergency department and associated factors. *Arch Gerontol Geriatr* 62, 97–102 (2016).
 25. Bavaro, D.F. et al. Peculiar clinical presentation of COVID-19 and predictors of mortality in the elderly: a multicentre retrospective cohort study. *Int J Infect Dis*, 105, 709–15 (2021).
 26. Zazzara, M.B. et al. Probable delirium is a presenting symptom of COVID-19 in frail, older adults: a cohort study of 322 hospitalised and 535 community-based older adults. *Age and Ageing* 50(1), 40–8 (2021).
 27. Rockwood, K. et al. A global clinical measure of fitness and frailty in elderly people. *CMAJ* 173, 489–495 (2005).
 28. Bianchetti, A. et al. Improving the care of older patients during the COVID-19 pandemic. *Aging Clin Exp Res* 32(9), 1883–8 (2020).

-
29. Guo, T. et al. Clinical Characteristics of Elderly Patients with COVID-19 in Hunan Province, China: A Multicenter, Retrospective Study. *Gerontology* 66(5), 467-475 (2020).
30. Gebhard, C. et al. Impact of sex and gender on COVID-19 outcomes in Europe. *Biol Sex Differ* 11(1), 29 (2020).
31. Kojima, G. et al. Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age Ageing* 47(2), 193-200 (2018).
32. Zeng, A. et al. Mortality in relation to frailty in patients admitted to a specialized geriatric intensive care unit. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 70, 1586–1594 (2015).
33. Vetrano, D. L. et al. Frailty and multimorbidity: A systematic review and meta-analysis. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 74, 659–666 (2019).
34. Dent, E. et al. Management of frailty: opportunities, challenges, and future directions. *Lancet* 394, 1376–1386 (2019).
35. Hewitt, J. et al. The effect of frailty on survival in patients with COVID-19 (COPE): a multicentre, European, observational cohort study. *The Lancet Public Health* 5, e444–e451 (2020).
36. Rockwood, K. et al. Using the clinical frailty scale in allocating scarce health care resources. *Canadian Geriatrics Journal* 23, 254–259 (2020).
37. Marengoni, A. et al. Beyond chronological age: Frailty and multimorbidity predict in-hospital mortality in patients with coronavirus disease 2019. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 76, E38–E45 (2021).
38. Zhou, F. et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 395(10229), 1054–62 (2020).
39. Lian, J et al. Analysis of epidemiological and clinical features in older patients with coronavirus disease 2019 (COVID-19) outside Wuhan. *Clin Infect Dis* 71(15), 740–7 (2022).
40. Wang, L. et al. Coronavirus disease 2019 in elderly patients: characteristics and prognostic factors based on 4-week follow-up. *J Infect* 80(6), 639–45 (2020).
41. Malara, A. et al. Clinical features of SARS-CoV-2 infection in Italian long-term care facilities: GeroCovid LTCFs observational study. *J Am Med Dir Assoc* 23, 15–8 (2022).
42. Abrahão Silva, M.J. et al. Functional and structural characterization of COVID-19 risk-associated exonic SNPs and identification of novel therapeutic sites: An in silico analysis. *Human Gene* 45, 2773-0441 (2025).

-
43. Neira-Goulart M, et al. Inflammasome genes polymorphisms are associated with progression to mechanical ventilation and death in a cohort of hospitalized COVID-19 patients in a reference hospital in Rio de Janeiro, Brazil. *Gene* 865, 147325 (2023).
44. Gavin, Y. et al. Oudit, Angiotensin-converting enzyme 2—at the heart of the COVID-19 pandemic. *Cell* 186(5), 906-922 (2023).
45. Bettini. E. et al. SARS-CoV-2 mRNA Vaccines: Immunological Mechanism and Beyond. *Vaccines* 9(2), 147 (2021).
46. Camacho-Moll, M.E. et al. The impact of comorbidity status in COVID-19 vaccines effectiveness before and after SARS-CoV-2 omicron variant in northeastern Mexico: a retrospective multi-hospital study. *Front Public Health* 12, 1402527 (2024).
47. Lin, T.Y. et al. Frailty and Humoral Immune Responses Following COVID-19 Vaccination among Patients Undergoing Hemodialysis. *The Journal of nutrition, health and aging* 27(11), 980-986 (2023).
48. Rattay, K. et al. Lessons learned: COVID-19 vaccinations and people with disabilities. *Vaccine* 42, 0264-410X (2024).
49. Gnjjidic,D. et al. Polypharmacy cutoff and outcomes: five or more medicines were used to identify community-dwelling older men at risk of different adverse outcomes. *Journal of Clinical Epidemiology* 65(9), 989-995 (2012).
50. Zheng, H. et al. COVID-19 vaccines for the older population: safety and efficacy go hand in hand. *The Lancet Healthy Longevity* 3(7), e455-e456 (2022).
51. Prasad, V. et al. An Evidence-Based Approach to Covid-19 Vaccination. *NEJM* 392, 2484-2486 (2025).
52. Sahu, M.K. AI in Epidemiology: Enhancing Public Health Surveillance and Response through Machine Learning Models. *Journal of Machine Learning in Pharmaceutical Research* 3 (2023).
53. Chenruo, L.V. et al. Innovative applications of artificial intelligence during the COVID-19 pandemic. *Infectious Medicine* 3(1), 2772-431X (2024).
54. Abadir, P. et al. Artificial Intelligence in Geriatrics: Riding the Inevitable Tide of Promise, Challenges, and Considerations. *The Journals of Gerontology: Series A* 79(2), glad279 (2024).
55. Vargas-Santiago M. et al. A State-of-the-Art Review of Artificial Intelligence (AI) Applications in Healthcare: Advances in Diabetes, Cancer, Epidemiology, and Mortality Prediction. *Computers* 14, 143 (2025).

-
56. Mandreoli, F. et al. Real-world data mining meets clinical practice: Research challenges and perspective. *Frontiers in Big Data* 5 (2022).
57. Alloghani, M. et al. A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. *Unsupervised and Semi-Supervised Learning* (2019).
58. Chowdhury, M.A. et al. Recent machine learning guided material research - A review. *Computational Condensed Matter* 29, 2352-2143 (2021).
59. Miotto, R. et al. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics* 19(6), (2017).
60. Vespa, M. et al. Discovery of Logic-Probabilistic Rules from COVID-19 Vaccine Antibody Response in Older People: Results from the GeroCovid VAX Study. *AIME* (2025)
61. Comito, C. et al. Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review. *Artif Intell Med* 128, 102286 (2022).
62. Pfeifer, B. et al. Explaining and visualizing black-box models through counterfactual paths. *Pattern Anal Applic* 28, 154 (2025).
63. Sajid, A. et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion* 99, 1566-2535 (2023).
64. Prasad, S. et al. Methods comparison and practical considerations for generalizability and interpretability. *Ann Epidemiol*, 111, 186-192 (2025).
65. Liu, T. et al. Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis. *Eur Heart J Digit Health* 6(1), 7-22 (2024).
66. Liu, X. et al. Machine learning for predicting in-hospital mortality in elderly patients with heart failure combined with hypertension: a multicenter retrospective study. *Cardiovasc Diabetol* 23(1):407 (2024).
67. Loutati, R. et al. Multimodal Machine Learning for Prediction of 30-Day Readmission Risk in Elderly Population. *Am J Med* 137(7), 617-628 (2024).
68. Bracher-Smith, M. et al. Machine learning in Alzheimer's disease genetics. *Nat Commun* 16(1), 6726 (2025).
69. Burnazovic, E. et al. Application of Artificial intelligence in COVID-19-related geriatric care: A scoping review. *Arch Gerontol Geriatr* 116, 105129 (2024).

-
70. Chatterjee, A. et al. Can predicting COVID-19 mortality in a European cohort using only demographic and comorbidity data surpass age-based prediction: An externally validated study. *PLOS ONE* 16(4), e0249920 (2021).
71. Horobet, A. et al. Europe's war against COVID-19: A map of countries' disease vulnerability using mortality indicators. *International Journal of Environmental Research and Public Health* 17(18), 6565 (2020).
72. Tezza, F. et al. Predicting in-hospital mortality of patients with covid-19 using machine learning techniques. *Journal of Personalized Medicine* 11(5), 343 (2021).
73. Lee, D.S. et al. Factors associated with SARS-CoV-2 test positivity in long-term care homes: A population-based cohort analysis using machine learning. *Lancet Reg Health Am* 6, 100146 (2022).
74. Lee, D.S. et al. Predictors of mortality among long-term care residents with SARS-CoV-2 infection. *J Am Geriatr Soc* 69(12), 3377-3388 (2021).
75. Trevisan, C. et al. Assessing the impact of COVID-19 on the health of geriatric patients: The European GeroCovid Observational Study. *Eur J Intern Med* 87, 29-35 (2021).
76. Zazzara, M.B. et al. Comorbidity patterns in institutionalized older adults affected by dementia. *Alzheimers Dement (Amst)* 14(1), e12320 (2022).
77. Hakimjavadi, R. et al. Cognitive and Functional Decline Among Long-Term Care Residents. *JAMA Netw Open* 8(4), e255635 (2025).
78. Beam, C.R. et al. Differences Between Women and Men in Incidence Rates of Dementia and Alzheimer's Disease. *J Alzheimers Dis* 64(4), 1077-1083 (2018).
79. Villiers-Tuthill, A. et al. The relationship of tobacco and alcohol use with ageing self-perceptions in older people in Ireland. *BMC Public Health* 16, 627 (2016).
80. Resciniti, N.V. et al. COVID-19 Incidence and Mortality Among Long-Term Care Facility Residents and Staff in South Carolina. *J Am Med Dir Assoc* 22(10), 2026-2031 (2021).
81. Chen, L. et al. Impact of Immunosenescence on Vaccine Immune Responses and Countermeasures. *Vaccines* 12(11), 1289 (2024).
82. Renia, L. et al. Lower vaccine-acquired immunity in the elderly population following two-dose BNT162b2 vaccination is alleviated by a third vaccine dose. *Nat Commun* 13, 4615 (2022).
83. Canaday, D.H. et al. Reduced BNT162b2 mRNA vaccine response in SARS-CoV-2-naive nursing home residents. *Clin Infect Dis* 16, ciab447 (2021).

-
84. Van Praet, J.T. et al. Humoral and cellular immunogenicity of the BNT162b2 mRNA Covid-19 Vaccine in nursing home residents. *Clin Infect Dis* 7, ciab300 (2021).
85. Britton, A. et al. Effectiveness of the Pfizer-BioNTech COVID-19 Vaccine Among Residents of Two Skilled Nursing Facilities Experiencing COVID-19 Outbreaks - Connecticut, December 2020-February 2021. *MMWR Morb Mortal Wkly Rep* 70(11), 396-401 (2021).
86. Dagan, N. et al. BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting. *N Engl J Med* 384(15), 1412-1423 (2021).
87. Stein, C. et al. Past SARS-CoV-2 infection protection against re-infection: a systematic review and meta-analysis. *The Lancet* 401(10379), 833-842 (2023).
88. Bonaga, B et al. Frailty, polypharmacy, and health outcomes in older adults: the frailty and dependence in Albacete study. *J Am Med Dir Assoc* 19:46-52 (2018).
89. Lewis, S.A. et al. Chronic alcohol consumption dysregulates innate immune response to SARS-CoV-2 in the lung. *eBioMedicine* 97, 104812 (2023).
90. Akingbola, A. et al. Comparing Moderna's mRNA-1083 and Pfizer's dual-target mRNA vaccines for influenza and COVID-19. *npj Vaccines* 10, 105 (2025).
91. Serwanga, J. et al. Sustained S-IgG and S-IgA antibodies to Moderna's mRNA-1273 vaccine in a Sub-Saharan African cohort suggests need for booster timing reconsiderations. *Front. Immunol.* 15 (2024).
92. Zamboni, P. et al. Clinical implications of gene polymorphisms in venous leg ulcer: A model in tissue injury and reparative process. *Thromb. Haemost* 98 (1), 131–137 (2007).
93. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed., text rev.) (2022).
94. Moncada-Torres, A. et al. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Scientific Reports* 6968 (2021).
95. Zou, Y. et al. Predictive value of frailty in the mortality of hospitalized patients with COVID-19: a systematic review and meta-analysis. *Ann Transl Med* 10(4), 166 (2022).
96. Covino, M et al. Long-Term Effects of Hospitalization for COVID-19 on Frailty and Quality of Life in Older Adults ≥ 80 Years. *J Clin Med* 11, 5787 (2022).
97. Volpato, S. et al. Comorbidities and impairments explaining the association between diabetes and lower extremity disability: The Women's Health and Aging Study. *Diabetes Care* 25(4), 678-83 (2002).
98. Mielke, M.M. Sex and Gender Differences in Alzheimer's Disease Dementia. *Psychiatr Times* 35(11), 14-17 (2018).

-
99. Harris, R.C. et al. The role of gender disparities in kidney injury. *Ann Transl Med* 8(7), 514 (2020).
100. High, K.P. Infection as a cause of age-related morbidity and mortality. *Ageing Research Reviews* 3(1), 1-14 (2004).
101. Salih, A. et al. Conceptual Overview of Biological Age Estimation. *Aging Dis* 14(3), 583-588 (2023).
102. Willadsen, T.G. et al. Multimorbidity and mortality: A 15-year longitudinal registry-based nationwide Danish population study. *J Comorb* 8(1), 2235042X18804063 (2018).
103. Romero Starke, K. et al. The isolated effect of age on the risk of COVID-19 severe outcomes: a systematic review with meta-analysis. *BMJ Glob Health* 6(12), e006434 (2021).
104. Santella, B et al. Age Differences and Prevalence of Comorbidities for Death and Survival in Patients with COVID-19: A Single-Center Observational Study in a Region of Southern Italy. *Life (Basel)* 14(11), 1376 (2024).
105. Agrawal, U. et al. Association between multimorbidity and mortality in a cohort of patients admitted to hospital with COVID-19 in Scotland. *J R Soc Med* 115(1), 22-30 (2022).
106. Sinnott, C. et al. Multimorbidity or polypharmacy: two sides of the same coin? *J Comorb* 5, 29-31 (2015).
107. Kratz, T. et al. Psychopharmacological Treatment in Older People: Avoiding Drug Interactions and Polypharmacy. *Dtsch Arztebl Int* 116(29-30), 508-518 (2019).
108. Visser, A.G.R. et al. The impact of polypharmacy on 30-day COVID-related mortality in nursing home residents: a multicenter retrospective cohort study. *Eur Geriatr Med* 14(1), 51-57 (2023).
109. Kuper, H. et al. Are people with disabilities at higher risk of COVID-19-related mortality?: a systematic review and meta-analysis. *Public Health* 222, 115-124 (2023).
110. Johnson, A.L. et al. Covid-19 outcomes among patients with dementia and age-matched controls who were hospitalized in 21 US health-care systems. *Alzheimers Dement* 20(9), 6395-6406 (2024).
111. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260 (2020).
112. Argenziano, M.G. et al. Characterization and clinical course of 1000 patients with coronavirus disease 2019 in New York: retrospective case series. *BMJ* 369, m1996 (2020).
113. Fagyas, M. et al. Circulating ACE2 activity predicts mortality and disease severity in hospitalized COVID-19 patients. *Int J Infect Dis* 115, 8-16 (2022).

-
114. Beyerstedt, S. et al. COVID-19: angiotensin-converting enzyme 2 (ACE2) expression and tissue susceptibility to SARS-CoV-2 infection. *Eur J Clin Microbiol Infect Dis* 40(5), 905-919 (2021).
115. Nitz, I. et al. Analysis of PGC-1 α variants Gly482Ser and Thr612Met concerning their PPAR γ 2-coactivation function. *Biochem Biophys Res Commun* 353, 481–6 (2007).
116. Yenmis, G. et al. The Distribution of Sport Performance Gene Variations Through COVID-19 Disease Severity. *Diagnostics (Basel)* 15(6), 701 (2025).