DIGITAL FUTURE
OR4 - Soluzioni verticali

Analisi di dominio Healthcare
TR4.9

AUTORE: EXPRIVIA S.P.A

UNIONE EUROPEA
FONDO EUROPEO DI SVILUPPO REGIONALE

Ministero Sviluppo Economico

REGIONE PUGLIA
AREA POLITICHE PER LO SVILUPPO ECONOMICO
IL LAVORO E L'INNOVAZIONE

Regolamento regionale della Puglia per gli aiuti in esenzione n. 17 del 30/09/2014

Titolo II – Capo 1 "Aiuti ai programmi di investimento delle Grandi Imprese"

POR PUGLIA FESR 2014 – 2020

CONTRATTO DI PROGRAMMA

"Digital Future"

CODICE PROGETTO: CP12PA6

Diritti di autore e riservatezza
Questo documento è proprietà esclusiva della società Exprivia S.p.A. e non può essere riprodotto, anche in forma parziale, senza un'autorizzazione scritta della società stessa.

Indice dei contenuti

DIGITAL FUTURE **2**
INDICE DEI CONTENUTI **4**
INDICE DELLE FIGURE **6**
INDICE DELLE TABELLE **7**
1. INTRODUZIONE **8**
1.1. SCOPO E CAMPO DI APPLICAZIONE 8
1.2. ORGANIZZAZIONE DEL DOCUMENTO 8
2. PROBLEMATICHE CLINICHE CHE POSSONO TRARRE VANTAGGIO DALLA CORRELAZIONE DEI DATI **9**
2.1. PROBLEMATICHE CLINICHE 10
2.1.1. Diabete *10*
2.1.2. Obesità *11*
2.1.3. Patologie cardiovascolari *13*
2.2. TELEMEDICINA: APPROCCIO SMART ALLA GESTIONE DELLE PATOLOGIE CRONICHE 14
3. ANALISI DEI MODELLI PREDITTIVI DEL RISCHIO DI PATOLOGIE CRONICHE PRESENTI IN LETTERATURA **16**
3.1. MODELLI PREDITTIVI PER IL RISCHIO DEL DIABETE DI TIPO 2 17
3.1.1. Analisi di una selezione di modelli predittivi disponibili in letteratura *18*
3.1.2. Confronto fra i modelli predittivi disponibili *34*
3.1.3. Selezione del modello predittivo più utili *40*
3.2. MODELLI PREDITTIVI PER IL RISCHIO CARDIOVASCOLARE 41
3.2.1. Analisi di una selezione di modelli predittivi disponibili in letteratura *42*
3.2.2. Confronto fra i modelli predittivi disponibili *54*
3.2.3. Selezione del modello predittivo più utile *55*
3.3. ADATTABILITÀ DEI MODELLI AL CONTESTO ITALIANO 56
3.4. CONSIDERAZIONI SUI MODELLI PREDITTIVI ANALIZZATI 57
4. ANALISI DELLE FONTI DATI PUBBLICHE CONTENENTI INFORMAZIONI UTILI ALLA CORRELAZIONE **58**
4.1. DATABASE PUBBLICI INDIVIDUATI 59
4.1.1. THIN Database *59*
4.1.2. Qresearch *61*
4.1.3. MalaCards *62*
4.1.4. Diseases Database *64*
4.1.5. KEGG Disease Database *65*
4.1.6. DisEASES *66*
4.1.7. DisGeNET *67*
4.1.8. OMIM *69*
4.1.9. REACTOME diseases *69*
4.1.10. Comparative Toxicogenomics Database (CTD) *70*
4.1.11. Nutrition Evidence *73*
4.1.12. VMH database *73*
4.1.13. CHE Toxicant and Disease Database *74*
4.1.14. Agency for Toxic Substances and Disease Registry *76*
4.1.15. Eurostat Health *77*
4.1.16. Pesticide-Induced Diseases Database *78*
4.1.17. Gene-Environment interactions (GxE) Database *78*
4.1.18. ClinVar *80*
4.1.19. HUMSAVAR *81*
4.1.20. eDGAR *81*

4.1.21. GWAS Catalog *81*
4.1.22. HGV Database *82*
4.1.23. Ensembl Variation *83*
4.1.24. dbSNP *83*
4.1.25. Genecards *83*
4.1.26. DRUGBANK *84*
4.1.27. Toxin and Toxin Target Database (T3DB) *84*
4.1.28. DGIdb (Drug Gene Interaction Database) *84*
4.1.29. PharmGKB *85*
4.1.30. KEGG Drug Database *85*
4.1.31. COSMIC (Catalog Of Somatic Mutations In Cancer) *86*
4.1.32. DiseaseEnhancer Database *86*
4.1.33. LncRNADisease database *87*
4.1.34. DECIPHER *88*
4.1.35. Registro tumori *88*
4.2. RIEPILOGO TIPOLOGIE DATABASE INDIVIDUATE 89
4.3. CONSIDERAZIONI SULL'ANALISI DEI DATABASE 91
5. ANALISI DELLE FONTI DATI AZIENDALI UTILIZZABILI PER I MODELLI PREDITTIVI 93
5.1. EREPOSITORY 93
5.2. EWARD 94
5.3. EVISIT 95
5.4. HMONITOR 96
5.5. HASSISTANT 97
5.6. ELIFECARE 98
6. BIBLIOGRAFIA 100

Indice delle figure

Figura 1 Cambridge Diabetes Risk Score [10] 20
Figura 2 Grafici SCORE per il rischio di malattie cardiovascolari per popolazioni europee ad alto e basso rischio basati sui valori del colesterolo totale 49
Figura 3 Rappresentazione in MalaCards delle patologie tra loro correlate 63
Figura 4 Patologie come stati perturbati 65
Figura 5 Categorie di dati tra cui sono previste relazioni all'interno del database 70
Figura 6 Relazioni inferite tra gene e patologia 71
Figura 7 Relazioni inferite tra sostanza chimica e patologia 72

Indice delle tabelle

Tabella 1 Modelli predittivi di base per il rischio di diabete di tipo II 37
Tabella 2 Modelli predittivi estesi per il rischio di diabete di tipo II 38
Tabella 3 Confronto tra i database pubblici contenenti informazioni utili per la correlazione di informazioni e la definizione di modelli predittivi 90

1. Introduzione

1.1. Scopo e campo di applicazione

Per il presente Technical Report ci si pone l'obiettivo di analizzare il dominio HealthCare allo scopo di evidenziare come le applicazioni cliniche di diagnostica e cura possano trarre vantaggio dall'utilizzo di una piattaforma come Digital Future in grado di inferire conoscenza attraverso l'analisi di grandi quantità di dati che tra loro sono generalmente poco integrati.

Il filo conduttore del documento sarà quello di individuare le principali problematiche cliniche che possono trarre beneficio, in termini di prevenzione, diagnosi e cura, dall'integrazione e correlazione delle differenti tipologie di informazioni. Successivamente saranno prese in considerazione le fonti informative contenenti dati utili alla creazione di modelli predittivi: saranno analizzate sia fonti pubbliche curate dalla comunità scientifica, sia fonti private prodotte ad esempio nel corso dei processi di cura (cartelle cliniche, sistemi di acquisizione e monitoraggio dei parametri vitali, etc.).

1.2. Organizzazione del documento

Il documento offrirà un'analisi del dominio Healthcare concentrandosi sulle seguenti tematiche:

- Problematiche cliniche che possono trarre vantaggio dalla correlazione dei dati
- Analisi dei modelli predittivi del rischio di patologie croniche presenti in letteratura
- Analisi delle fonti dati pubbliche contenenti informazioni utili alla correlazione
- Analisi delle fonti dati aziendali utilizzabili per i modelli predittivi

Tutto il documento sarà suddiviso in capitoli, ciascuno dei quali coprirà la singola tematica.

2. Problematiche cliniche che possono trarre vantaggio dalla correlazione dei dati

L'invecchiamento della popolazione e il cambiamento degli stili di vita stanno determinando lo sviluppo di nuove cronicità e l'esigenza di cure a più lungo termine. Il SSN si trova quindi in una fase strutturale di inadeguatezza economica in quanto da un lato deve limitare la spesa sanitaria e dall'altro vede aumentare i costi per prevenzione, diagnosi e cura a causa della necessità di macchinari più evoluti, farmaci più specifici e metodiche analitiche sempre più raffinate. In uno scenario di questo tipo è sempre più marcata la necessità della deospedalizzazione e dello spostamento degli sforzi della cura verso il territorio e verso una maggiore centralità del malato. Tale tendenza è facilitata dall'evoluzione della Società verso una maggiore connessione e disintermediazione che porterà i pazienti a informarsi e curarsi in autonomia e a gestire a distanza i rapporti con il sistema sanitario. Si sta affermando, inoltre, una nuova sensibilità con cui ciascun individuo si approccia al proprio benessere; tale sensibilità prevede una maggior valorizzazione della prevenzione e della cura del proprio stato fisico, attraverso l'adozione di stili di vita salutari, ed una maggior responsabilizzazione al monitoraggio della propria salute. I termini chiave del nuovo approccio sono quindi proattività ed empowerment; la maggiore diffusione di dispositivi digitali quali smartphone o misuratori di parametri fisiologici connessi al paziente (ed in grado di trasmettere le informazioni ai fornitori della cura) stanno facilitando questa fase. L'individuo diventa quindi partner dei sanitari che lo curano e protagonista nella pianificazione delle attività volte a prevenire malattie o evitare il peggioramento di quelle già manifeste. Queste nuove pratiche fanno parte del più ampio mondo della Telemedicina che quindi diventa una strada obbligata che le aziende Sanitarie devono tenere in considerazione per ripensare i propri processi e le proprie organizzazioni.

Un ruolo più attivo del paziente nella gestione della cura permette di ampliare il ventaglio di miglioramenti nei processi di prevenzione, diagnosi e cura. Uno dei miglioramenti è la possibilità di sfruttare la peculiarità delle informazioni associate e/o raccolte dal paziente (ambiente in cui vive, parametri fisiologici misurati, stili di vita, etc.).

L'ambiente che ci circonda e i fattori sociali quali lavoro, condizioni economiche, etc. possono influire sulla qualità della salute e sulla possibilità di insorgenza di molte malattie. Sarebbe ideale, quindi, ridisegnare le modalità di cura dei pazienti in maniera tale da tenere in considerazione tutti (o almeno i principali) fattori in grado di condizionare lo stato di salute. Tale obiettivo può essere raggiunto attraverso la profilazione del paziente a partire dai dati disponibili ossia:

- dati clinici (anamnesi, esame obiettivo, etc.);
- dati genomici (varianti genomiche, profili di espressione genica, etc.)

- dati rilevati attraverso questionari somministrati al paziente;
- dati di misurazioni effettuate quotidianamente mediante dispositivi indossabili e collegati alla rete;
- dati sui farmaci somministrati
- dati socio-ambientali
- altre evidenze cliniche (es. immagini)

La profilazione si prospetta complessa sia per la varietà dei dati da considerare, sia perché richiederà il coinvolgimento di strutture sanitarie in rete, di nuove tecnologie e dell'adozione di tecniche di intelligenza artificiale.

La necessità di profilare del paziente richiederà la messa a punto di modelli predittivi sia per la prevenzione che per la cura; il modello predittivo, attraverso la stratificazione dei pazienti in classi di rischio diverse, potrà essere un supporto per decisioni di tipo clinico, chirurgico e terapeutico.

2.1. Problematiche cliniche

Le problematiche cliniche più adatte a trarre beneficio (in termini di prevenzione, diagnosi e cura) dall'integrazione e dalla correlazione delle informazioni, sono quelle che sono in grado di generare tipologie di dati diversificate prodotte in maniera costante e protratta nel tempo. Patologie croniche come il diabete, l'obesità e le patologie cardiovascolari rispondono a tali requisiti sia perché, vista la loro prevalenza, garantiscono una quantità sufficiente di informazioni utilizzabili, sia perché gli elementi che concorrono alla loro insorgenza e al loro decorso sono diversificati.

Le patologie croniche rappresentano, a livello mondiale, la maggior percentuale di cause di morte. L'OMS ha calcolato che nel 2015 circa il 75% dei decessi [*Organizzazione Mondiale della Sanità - Prevenire le malattie croniche: un investimento vitale: rapporto globale dell'OMS*] è stato imputabile a malattiche croniche che, per la loro natura, hanno impatti importanti sui sistemi sociali e sanitari. La situazione è destinata a peggiorare poiché le aspettative di vita legate ad un maggior invecchiamento della popolazione aumentano e con esse i costi sociali ed economici di tali malattie.

2.1.1. *Diabete*

Il diabete oggi rappresenta una delle maggiori sfide sociali tanto per le economie in via di sviluppo quanto per quelle più industrializzate e caratterizzate da maggior benessere. Si tratta di quello che l'ONU ha definito come un grave problema di sanità pubblica per il pianeta, destinato a crescere. Secondo l'OMS, oggi, infatti, circa 50 milioni di persone in Europa soffrono di diabete. In Europa, a causa del fenomeno sempre più progressivo di invecchiamento della popolazione, si prevede un possibile aumento di circa il

40% entro il 2030 della popolazione adulta e conseguentemente degli individui affetti da iperglicemia cronica da diabete. Si tratta di un aumento in parte dovuto all'invecchiamento generale della popolazione, ma principalmente alla diffusione di condizioni a rischio come sovrappeso e obesità, mancanza di una corretta educazione alimentare, sedentarietà e disuguaglianze economiche. In Italia il diabete, nel 2009, è stato riconosciuto come causa principale di morte in oltre 20 mila casi. Oggi conta oltre 2 milioni di diabetici con un tasso di crescita di circa 100 mila nuovi casi all'anno a causa dell'invecchiamento della popolazione e di una conseguente maggiore diffusione della malattia. A crescere, tuttavia, non sono solamente le stime relative alla diffusione di tale patologia, ma anche gli oneri socio-economici che ne conseguono per i governi, necessari per poter fronteggiare al meglio tale sfida sociale. A gravare maggiormente, non è tanto il trattamento del diabete in sé, quanto quello delle sue complicanze; in Italia, infatti, solo il 7% della spesa riguarda i farmaci, mentre il 25% è legato alle terapie per le complicanze e le patologie concomitanti e il 68% è relativo al ricovero ospedaliero e alle cure ambulatoriali, con costi stimati nell'ordine di circa 4.5 miliardi di euro. Nel contesto delle complicanze del diabete, risulta importante il monitoraggio dello stato di salute degli arti inferiori ed in particolare del piede del paziente diabetico che rappresenta una delle parti maggiormente colpite dalla patologia. Circa il 50% di tutte le amputazioni degli arti inferiori riguarda infatti i diabetici. Fra queste, l'84% ha come causa ulcere del piede aggravatesi nel tempo. Il piede del diabetico necessita di particolare cura poiché anche la più insignificante lesione, se non propriamente curata, potrebbe condurre a successive complicazioni.

2.1.2. *Obesità*

Nel mondo l'obesità interessa almeno 400 milioni di adulti, secondo i dati dell'O.M.S. riferiti al 2005. Le proiezioni indicano che nel 2015 gli adulti obesi saranno più di 700 milioni. L'Italia, secondo dati ISTAT (Noi Italia 2013 facenti riferimento al 2011), si colloca tra i paesi europei con percentuale più bassa di obesità (11.4%) e la regione Puglia con prevalenza tra le più alte (11.5%), considerando la popolazione adulta con età di circa 18 anni. Nel complesso, l'obesità aumenta dopo i 35 anni, sia per gli uomini sia per le donne, con differenze di genere a svantaggio degli uomini che si annullano tra le persone anziane.

L'obesità è in genere correlata ad un'alimentazione scorretta e ad uno stile di vita sedentario. Un'assunzione di eccessive quantità di zuccheri e di grassi, soprattutto se associata a scarsa attività motoria, può accrescere il rischio di sviluppare l'obesità, in particolare tra le persone che già presentano una predisposizione. L'eccesso di massa grassa è il risultato di un insieme complesso e multifattoriale di processi che si sviluppano dall'interazione tra genotipo ed ambiente e coinvolgono aspetti metabolici, fisiologici, sociali, comportamentali e culturali.

Il controllo dell'obesità può avvenire andando a monitorare l'introito calorico e il dispendio energetico a partire dal livello di metabolismo basale, ossia dalla quantità di energia che l'organismo consuma a riposo.

Un mantenimento del peso corporeo si ottiene bilanciando l'introito calorico ottenuto con l'alimentazione e il dispendio energetico (derivante dalla somma dell'energia consumata per le normali attività giornaliere e dell'attività fisica praticata); un calo ponderale, invece, si ottiene aumentando il dispendio energetico e/o riducendo l'introito calorico attraverso una dieta ipocalorica.

I soggetti in sovrappeso od obesi sottoposti a monitoraggio generalmente ricevono dal medico delle indicazioni sotto forma di programmi individuali riportanti gli alimenti da assumere e l'attività fisica da praticare. Tali programmi, però, sono statici e aggiornati dal dietologo solo al momento della visita periodica e non tengono conto delle variazioni del metabolismo basale del paziente che, ad esempio, rallenta in condizioni di digiuno, con l'età, con il clima o aumenta e rimane persistentemente più elevato se si pratica attività fisica o se si acquisisce massa magra. Una misura frequente del metabolismo basale consente di avere a disposizione un'informazione più aggiornata che permette di adattare dinamicamente il programma individuale alle reali esigenze metaboliche del paziente e soprattutto permette di ottenere una valutazione oggettiva dell'andamento e della qualità del calo ponderale (un calo ponderale di massa grassa è qualitativamente migliore di un calo ponderale di massa magra). Una delle componenti che influenza il metabolismo basale è la composizione corporea ed in particolare la variazione della percentuale di massa magra [http://ajcn.nutrition.org/content/82/5/941.full]. Misurare quindi la composizione corporea consente di ottenere una stima del valore del livello del metabolismo basale.

Componente fondamentale di un programma di riduzione dell'obesità è anche la quantità e l'intensità dell'attività fisica praticata. Non è facile ottenere una valutazione oggettiva dell'attività fisica praticata poiché è il paziente a dichiarare il tipo e la durata dell'attività. Strumenti come i contapassi sono estremamente semplici da utilizzare, ma non sempre forniscono una misura accurata. Raccogliere informazioni accurate sulla durata e sull'intensità dello sforzo fisico fornirebbe un ulteriore elemento affidabile su cui poter contare per valutare l'andamento e la qualità del calo ponderale.

L'esigenza di personalizzazione del programma di riduzione del peso corporeo e di miglioramento degli stili di vita può essere ulteriormente soddisfatta andando a sfruttare la variabilità di risposta di ciascun individuo dovuto al proprio corredo genetico e alla sua interazione con l'ambiente. La nutrigenetica e la nutrigenomica possono fornire un'arma in più per rendere più efficace il miglioramento della salute di pazienti obesi. Attraverso la nutrigenetica è teoricamente possibile selezionare, in base al profilo genetico di ciascun individuo, uno stile di vita alimentare personalizzato in grado di mantenere più a lungo lo stato di salute. La nutrigenomica, invece, si occupa di valutare gli effetti dei nutrienti sull'espressione del DNA; tenendo in considerazione questo aspetto è possibile individuare quegli alimenti o in genere quei comportamenti alimentari che influiscono negativamente sul funzionamento dell'intero organismo o di una parte di esso [http://www.ncbi.nlm.nih.gov/pubmed/21625170].

Sovrappeso e obesità rappresentano dei fattori di rischio ad alta incidenza per la salute poiché causa o fattori aggravanti di malattie cardiovascolari, ictus, diabete, alcuni tumori, etc. L'ictus e l'infarto sono due delle conseguenze dell'obesità con maggior impatto socio-sanitario e possono essere individuati con un certo anticipo grazie a dei sintomi premonitori tra cui l'emiplegia di un arto per l'ictus cerebrale e un forte dolore toracico per l'infarto del miocardio. La possibilità di identificare alcuni sintomi premonitori può essere di notevole aiuto nella prevenzione di tali eventi.

2.1.3. *Patologie cardiovascolari*

Le patologie cardiovascolari costituiscono la prima causa di morte nel mondo (circa 15 milioni di persone) e sono responsabili di un terzo della mortalità globale; l'ictus cerebrale, invece, rappresenta la principale causa d'invalidità. Il numero di persone decedute a causa di ictus nel 2015 ammonta a 6.7 milioni ed ogni anno si prevede che si verifichino oltre 200.000 nuovi casi di ictus (incidenza 2-3 pazienti/anno/1000 abitanti) e che i soggetti che si prevede sopravvivano ad un ictus, con esiti più o meno invalidanti, saranno circa 900.000.

Ben l'80% di queste morti potrebbero essere evitate se si prestasse più attenzione ai fattori di rischio come il tabacco, l'alimentazione scorretta e l'inattività fisica.

Le patologie del sistema circolatorio causano 224.482 decessi (97.952 uomini e 126.530 donne), pari al 38,8% del totale dei decessi. Per le malattie ischemiche del cuore (infarto del miocardio, altre forme acute e subacute di cardiopatia ischemica, infarto miocardico pregresso, angina pectoris e altre forme croniche di cardiopatia ischemica) si registrano 75.046 decessi (37.827 uomini e 37.219 donne), pari al 33% circa del totale delle morti per malattie del sistema circolatorio. Negli uomini la mortalità è trascurabile fino all'età dei 40 anni, emerge fra i 40 e i 50 anni e poi cresce in modo esponenziale con l'età. Nelle donne il fenomeno si manifesta a partire dai 50-60 anni e cresce rapidamente; lo svantaggio degli uomini rispetto alle donne è più accentuato nell'età riproduttiva e tende a ridursi con l'avanzare dell'età.

L'infarto miocardico è la causa più frequente di mortalità e morbilità nel mondo; nel 98% dei casi è conseguenza dell'aterosclerosi coronarica. È più frequente nell'uomo che nella donna, almeno sino ai 40 anni, ed è strettamente legato a condizioni quali abuso di tabacco, ipercolesterolemia, obesità, vita sedentaria. A questi elementi, che possono essere considerati predisponenti, vanno aggiunte alcune situazioni che talora possono risultare scatenanti quali sforzi, emozioni, interventi chirurgici e gravi emorragie.

2.2. Telemedicina: approccio smart alla gestione delle patologie croniche

Gli ultimi decenni sono stati caratterizzati da un significativo miglioramento delle condizioni di salute nei Paesi OCSE dovuti a migliori condizioni socio-economiche, a maggiore attenzione agli stili di vita, alla riduzione di alcuni fattori di rischio e ai progressi nel campo della medicina che hanno contribuito ad aumentare la longevità. Se da un lato sono migliorate le condizioni di salute, dall'altro è mutato anche il bisogno di salute delle persone. Per far fronte all'aumento della domanda di assistenza, i sistemi sanitari sono cresciuti in complessità ed estensione con livelli di spesa progressivamente maggiori dovuti, in parte, all'introduzione di nuovi farmaci, protocolli e macchinari.

Il fenomeno dell'invecchiamento della popolazione (in Italia si prevede che nel 2045-2050 l'aspettativa di vita arriverà a 87,3 anni) pone importanti questioni tra le quali quelle che allarmano di più sono le ripercussioni economiche sui sistemi di assistenza. L'aumento delle patologie legate all'invecchiamento, infatti, determinerà un incremento di costi legati alle fasi di prevenzione, diagnosi e cura.

È indispensabile, quindi, che le organizzazioni sanitarie si riorganizzino per attivare la prevenzione e la diagnosi precoce delle malattie allo scopo di ridurre e/o ritardare le complicanze. Tale strategia, unita ad un costante monitoraggio dei parametri vitali del paziente, permetterà di aumentare l'efficacia delle azioni legate al piano di cura e di rivolgersi all'ospedale solo nei casi di urgenza o di approfondimento specialistico, aumentando così l'efficienza del percorso (si impegnano le strutture sanitarie solo quando realmente necessario) e aumentando contestualmente l'efficacia degli interventi (la cura a casa migliora la qualità di vita del paziente).

Questa riorganizzazione dei servizi territoriali non può che essere sostenuta da infrastrutture tecnologiche adeguate: dal monitoraggio remoto, alla telemedicina e al teleconsulto; dai sistemi di prenotazione virtuali, integrati con i sistemi di refertazione e archiviazione alla disponibilità digitale della storia clinica del paziente mediante dossier e fascicoli elettronici. L'obiettivo è quello di avere tutto integrato, sempre disponibile, fruibile da qualsiasi dispositivo in modo che le interazioni con le strutture sanitarie siano minimizzate alle reali necessità e che i percorsi di cura rappresentino i processi primari che guidano o meglio orchestrano i differenti attori che interagiscono col paziente.

Uno scenario ipotetico vede il domicilio del paziente come un'estensione fisica di un reparto di ospedale dove sono disponibili strumenti non invasivi di monitoraggio basati su tecnologie IoT. Il paziente, attraverso l'uso di tecnologie indossabili (wearable), viene costantemente monitorato e le informazioni, integrate con il monitoraggio di altri parametri ambientali, vengono inviate ad un centro di raccolta e rese disponibili al medico di medicina generale e allo specialista che ha in cura il paziente. Lo specialista, intervenendo sulle sole situazioni critiche determinate da alcuni parametri fuori range (ad es. usando le scale EWS – Early Warning Score), può decidere di cambiare il percorso di cura sospendendo l'assunzione di un farmaco,

somministrando un farmaco diverso, prescrivendo un esame specialistico (mirato ad una indagine più approfondita) o addirittura inviando soccorsi immediati.

La tendenza che si delinea, quindi, è quella di un passaggio graduale dall'Ospedale al Territorio e dal Territorio al domicilio del paziente. Tale scenario è arricchito dalla costituzione di centri servizi territoriali con capacità di gestire l'emergenza/urgenza ma anche la continuità di cura in prossimità dei pazienti, fino alla diffusione di servizi territoriali mediante tecnologie di teleconsulto e televisita.

I sistemi ICT di supporto alla governance sanitaria dovranno essere sempre più integrati con le tecnologie IoT ed avere la capacità di lavorare con altri sistemi informativi nello scambio di informazioni agendo in maniera coordinata. L'uso della telemedicina, associato all'uso delle cartelle cliniche elettroniche, del Dossier Sanitario e del Fascicolo Elettronico del paziente, contribuirà a dematerializzare e centralizzare i dati, razionalizzare la spesa legata alla diagnostica e a condividere la storia del paziente tra più strutture sanitarie.

Questa sfida sarà vinta se si provvederà ad aggregare interessi diversi intorno ai bisogni del paziente realizzando una Filiera della Sanità di servizi integrati per il trattamento delle malattie croniche nell'ottica di una più razionale gestione della spesa e dell'assicurazione di una migliore qualità di vita dei pazienti e contestualmente efficacia della cura.

3. Analisi dei modelli predittivi del rischio di patologie croniche presenti in letteratura

L'utilizzo di modelli predittivi in ambito sanitario può fornire un contributo non indifferente al miglioramento della salute. In genere ci si aspetta che tutti i medici siano esperti e che dietro le loro decisioni ci sia sempre una ricerca approfondita; ciò, tuttavia, non può essere sempre vero. I medici si sottopongono a continui aggiornamenti professionali, ma non sempre possono avere a mente tutta la conoscenza di cui hanno bisogno in ogni situazione. Anche se avessero accesso completo alla grande quantità di dati necessaria per trattare tutte le patologie che incontrano, avrebbero ancora necessità di tempo ed esperienza per analizzare l'informazione che hanno a disposizione e per applicarla al profilo medico del paziente. Per questo motivo, sempre più medici e compagnie assicurative stanno iniziando a tenere in considerazione le analisi predittive a sostegno del proprio lavoro.

L'analisi predittiva in ambito sanitario adopera tecnologie e metodi statistici per analizzare una grande quantità di informazioni allo scopo di predire esiti clinici per i singoli pazienti. Tra le informazioni analizzate ci sono gli esiti di trattamenti terapeutici effettuati in passato (dati storici), ma anche le evidenze emerse in ricerche mediche pubblicate in articoli scientifici. I metodi utilizzati per la predizione sono chiamati "learning models" perché possono migliorare in precisione quando alla base di conoscenza si aggiungono nuovi casi. L'analisi predittiva differisce dalla statistica tradizionale e dalla Evidence-Based Medicine per il fatto che nell'analisi predittiva le predizioni sono fatte per gli individui e non per i gruppi e perché l'analisi predittiva non si basa su una distribuzione normale (gaussiana) [1].

I principali benefici dell'adozione dell'analisi predittiva per la medicina sono [2]:

1. Incremento dell'accuratezza delle diagnosi: la predizione, in ogni caso, non rimpiazzerà il giudizio del medico, ma fungerà da supporto.
2. Miglioramento della medicina preventiva: l'analisi predittiva, specie in ambito genomico, permetterà: ai medici curanti di identificare ed intervenire precocemente sui pazienti a rischio; ai pazienti di variare in tempo i propri stili di vita per prevenire l'insorgenza di patologie croniche.
3. Personalizzazione della diagnosi e della cura: un trattamento che funziona per la maggioranza di una popolazione può non funzionare al meglio per uno specifico paziente. L'analisi predittiva può fornire ai medici indicazioni specifiche sul trattamento più opportuno per ogni specifico paziente, evitando trattamenti non necessari e potenzialmente dannosi.
4. Possibilità di prevedere in anticipo l'andamento dei costi dell'assistenza sanitaria

5. Possibilità di indirizzare la produzione dei farmaci in base alle esigenze: le aziende farmaceutiche possono adottare l'analisi predittiva per prevedere le necessità di medicinali per il pubblico, concentrando la produzione ad es. su alcuni farmaci.
6. Miglioramento dell'efficacia delle cure e della qualità della vita: l'analisi predittiva permette di semplificare la personalizzazione del trattamento farmacologico dei pazienti evitando la prescrizione di farmaci solo perché funzionano per la maggioranza delle persone. La qualità della vita del paziente migliorerà poiché sarà più informato e consapevole dei rischi.

L'efficacia dei modelli predittivi applicati in ambito sanitario aumenta se tali modelli sono utilizzati come punto di partenza per intraprendere delle azioni e se i medici sono direttamente coinvolti nel processo.
Accanto ai vantaggi, tuttavia, ci sono molte insidie che possono verificarsi nell'adozione di analisi predittive. Le principali sono [1]:

1. Avere a disposizione più dati non vuol dire necessariamente essere in grado di poter fare più deduzioni: può essere difficile, infatti, estrarre conclusioni clinicamente rilevanti anche da grandi quantità di dati
2. Deduzioni e valore non sono la stessa cosa: mentre molti risultati scientifici possono essere interessanti, possono esserlo poco per migliorare significativamente i risultati clinici
3. La capacità di interpretare i dati varia sulla base dei dati stessi: alcune volte anche i dati migliori possono permettere solo deduzioni limitate di rilevanza clinica.

Per la predisposizione di modelli predittivi del rischio di patologie croniche è necessario disporre di una certa quantità e tipologia di informazioni che possano contribuire a rendere il modello utile per una corretta interpretazione del rischio. In questo capitolo saranno analizzati alcuni modelli per la predizione rispettivamente del rischio di diabete di tipo 2 e di patologie cardiovascolari. Per ciascuno dei due rischi valutati saranno confrontati tra loro i principali modelli predittivi disponibili in letteratura e sarà individuato quello che è più adeguato da adottare per gli obiettivi del progetto.

3.1. Modelli predittivi per il rischio del diabete di tipo 2

Il diabete di tipo 2 è un grave problema sanitario a livello mondiale. Gli studi sul cambiamento dello stile di vita e l'intervento farmacologico in individui con ridotta tolleranza al glucosio hanno dimostrato in modo convincente che l'adozione di tali misure possono prevenire il diabete. L'identificazione precoce dei soggetti ad alto rischio di insorgenza di diabete è quindi importante per adottare strategie di prevenzione mirate ed è necessaria per concentrare gli sforzi per la prevenzione nell'alto numero di individui a rischio elevato,

evitando al tempo stesso l'onere della prevenzione e del trattamento farmacologico per il numero persino maggiore di persone a basso rischio [6].

Poiché questi studi si sono concentrati principalmente su individui con ridotta tolleranza al glucosio, per identificare quelli che necessitano un intervento è stato adottato un test orale di tolleranza al glucosio (OGTT). I disagi e i costi associati a questo test hanno indotto allo sviluppo di regole semplici che prevedono l'uso di informazioni cliniche prontamente disponibili in grado di prevedere il diabete con proprietà diagnostiche uguali o migliori rispetto al test orale di intolleranza al glucosio [7].

Le linee guida europee per la prevenzione del diabete di tipo 2 e la Federazione internazionale per il diabete raccomandano l'uso di un sistema affidabile, semplice e pratico di valutazione del rischio o di un questionario per identificare le persone ad alto rischio di diabete. Molti modelli di previsione sono stati sviluppati nel recente passato. Su questo argomento sono disponibili diverse review che descrivono i modelli previsionali esistenti e il valore predittivo di fattori di rischio specifici (come la sindrome metabolica) su un'ampia gamma di popolazioni. Poiché le prestazioni di un modello predittivo sono generalmente sopravvalutate nella popolazione in cui è stato sviluppato, la validazione esterna di tali modelli in una popolazione indipendente, idealmente da parte di ricercatori non coinvolti nello sviluppo dei modelli, è essenziale per valutare in generale le prestazioni e quindi l'utilità potenziale di tali modelli in diverse popolazioni e contesti. Di conseguenza, non è possibile raccomandare alcuni modelli di previsione per identificare i soggetti ad alto rischio di diabete se non si conosce la validità esterna dei modelli disponibili.

In questo capitolo saranno analizzati e confrontati tra loro i principali modelli predittivi del rischio di diabete di tipo 2 disponibili in letteratura. A seguito del confronto effettuato sarà identificato quello che è più conveniente usare per le esigenze del progetto.

3.1.1. *Analisi di una selezione di modelli predittivi disponibili in letteratura*

I modelli predittivi disponibili possono essere suddivisi in modelli che contengono solo predittori non invasivi ("modelli base") e modelli che comprendono anche biomarker convenzionali, come glucosio, HbA1c, lipidi, acido urico o γ-glutammiltransferasi ("modelli estesi") [6].

L'analisi fatta nelle review mostra che i modelli predittivi funzionano bene per identificare gli individui ad alto rischio di diabete futuro, tuttavia, i modelli dovrebbero essere adattati al contesto locale e allo scopo del modello e almeno dovrebbero essere corretti sulla base dell'incidenza del diabete nella popolazione in cui deve essere applicato. La principale utilità dei modelli previsionali è quella di identificare correttamente le persone ad alto rischio, evitando l'onere del trattamento per le persone a basso rischio.

Nella pratica della sanità pubblica, sarebbe preferibile utilizzare un modello che includa solo un numero limitato di predittori basati ad esempio su variabili non invasive con prestazioni più elevate (ad es. età, sesso, obesità, ipertensione, stili di vita, familiarità di diabete e tratti metabolici), favorendo in questo modo l'uso di un modello di base. Noble et al. [8] hanno suggerito sette modelli come i più promettenti per l'uso nella pratica clinica o della sanità pubblica; di questi tre sono modelli estesi (ARIC, Framingham e San Antonio) e quattro sono modelli di base (AUSDRISK, QDScore, FINDRISC e Cambridge Risk Score). La maggior parte dei modelli predittivi (come QDScore e AUSDRISK) è composto da sette o dieci predittori. È interessante notare che modelli di base che includevano solo sei predittori (come il KORA), hanno funzionato in modo simile ai modelli più estesi. I modelli estesi, che comprendono i biomarcatori, potrebbero essere adottati solo per coloro che sono stati individuati come ad alto rischio usando un modello predittivo di base. Infine, un modello sviluppato in un contesto (come quello dei dati della sanità pubblica) o in un determinato paese non deve necessariamente essere utile in un altro contesto (ad esempio le cure secondarie) o in un altro paese.

La maggior parte dei modelli di previsione di base hanno funzionato bene per identificare gli individui ad alto rischio di sviluppare il diabete di tipo 2 in una popolazione indipendente. I modelli estesi con dati aggiuntivi sui biomarcatori convenzionali avevano prestazioni leggermente migliori. Altri biomarcatori utili come potenziali predittori del diabete di tipo 2 includono varianti genetiche, trascritti di RNA, peptidi e proteine, lipidi e piccoli metaboliti. Alcuni di questi biomarcatori portano ad un aumento statisticamente significativo dell'accuratezza del modello, ma il loro valore incrementale sembra attualmente troppo piccolo per un uso clinico di routine [9]. La maggior parte dei modelli, tuttavia, sovrastima il rischio reale di diabete.

3.1.1.1. *Modelli predittivi di base*

I modelli predittivi di base sono quelli che prevedono l'utilizzo di variabili di predizione facilmente reperibili anche ad opera del paziente stesso o dal medico (ad esempio la cartella clinica personale del paziente) e che non richiedono la necessità di aver effettuato degli esami invasivi (es. analisi del sangue). Tra le variabili non invasive utilizzate come fattori di rischio ed utilizzate nei modelli predittivi di base ci sono: età, sesso, sovrappeso, circonferenza vita, stile di vita, familiarità di diabete, fumo, uso di farmaci corticosteroidi o antipertensivi, presenza di patologie cardiache e livello di deprivazione economico-sociale.

3.1.1.1.1. Cambridge Risk Score

Il Cambridge Diabetes Risk Score include solo informazioni sui fattori di rischio non biochimici; tali fattori stanno diventando sempre più abitualmente disponibili per il medico di famiglia. Il punteggio di rischio è stato sviluppato e convalidato in modo indipendente in studi trasversali di popolazione ed ha una sensibilità

del 77% ed una specificità del 72% nell'individuazione del diabete non diagnosticato rispetto al test orale di tolleranza al glucosio in una popolazione soggetta solo a cure primarie. Lo score è stato utilizzato per valutare in che misura lo score potrebbe prevedere la futura incidenza clinica del diabete in un'ampia coorte di popolazione nell'ambito dello studio European Prospective Investigation of Cancer (EPIC)-Norfolk [10].

Il Cambridge Diabetes Risk Score è stato calcolato per ogni partecipante allo studio EPIC-Norfolk utilizzando i coefficienti di regressione pubblicati per l'età, il sesso, la storia familiare di diabete, il fumo, la prescrizione di farmaci steroidei o antiipertensivi e l'indice di massa corporea (BMI) (Figura 1).

$$\frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)}}$$

α	-6.322	Constant
β_1	-0.879	Female
β_2	1.222	Prescribed antihypertensive medication
β_3	2.191	Prescribed steroids
β_4	0.063	x age in years
β_5	0	Body Mass Index <25 kg/m^2
	0.699	Body Mass Index 25-27.49 kg/m^2
	1.970	Body Mass Index 27.5-29.99 kg/m^2
	2.518	Body Mass Index ≥ 30 kg/m^2
β_6	0	No first degree relative with diabetes
	0.728	Parent or sibling with diabetes
	0.753	Parent and sibling with diabetes
β_7	0	Non-smoker
	-0.218	Ex-smoker
	0.855	Current smoker

Figura 1 Cambridge Diabetes Risk Score [10]

Lo score di rischio è stato poi ricodificato in quintili ed è stato confrontato il rischio di sviluppare il diabete in ogni quintile con la categoria di rischio più bassa (gruppo di riferimento). Per testare la performance del punteggio nell'identificazione degli individui con diabete incidente, è stato calcolato il valore di sensibilità, specificità e likelihood ratio positivo per diverse soglie di rischio ed è stata tracciata una receiver operating characteristic (ROC[1]) curve per esaminare la capacità di discriminazione del punteggio.

[1] In statistica, una curva ROC è un grafico che illustra la capacità diagnostica di un sistema classificatore binario al variare della soglia di discriminazione. La curva ROC viene creata tracciando il True Positive Rate (TPR) rispetto al False Positive Rate (FPR) a varie impostazioni di soglia. Il TPR è noto anche come sensibilità. L'FPR è noto anche come fall-out o probabilità di falsi allarmi e può essere calcolato come (1 - specificità). L'analisi ROC fornisce strumenti per selezionare i modelli possibilmente ottimali e scartare quelli non ottimali indipendentemente dalla distribuzione della classe. In sostanza, quindi, la ROC misura la discriminazione, cioè la capacità del test di classificare correttamente le persone con e senza la malattia.

Il Cambridge Diabetes Risk Score riesce ad effettuare moderatamente bene la previsione del diabete di tipo 2 nella coorte EPIC-Norfolk, con un'area sotto il ROC del 75%. Gli individui nel quintile superiore del rischio avevano una probabilità 22 volte maggiore di sviluppare il diabete rispetto a quelli nel quintile inferiore. Il Cambridge score sembra quindi essere uno strumento semplice ed efficace per identificare le persone a rischio di sviluppare il diabete di tipo 2 utilizzando informazioni non biochimiche raccolte regolarmente dalle cartelle cliniche dei pazienti.

Lo score di Cambridge funziona bene anche senza informazioni sulla storia familiare e sul fumo ed ha un buon valore predittivo quando viene testato su una popolazione diversa da quella su cui è stato originariamente sviluppato [10].

3.1.1.1.2. FINDRISK (Finnish Diabetes Risk Score)

Il modello FINDRISK ha l'obiettivo di prevedere il rischio di futuro diabete trattato con farmaci utilizzando parametri che sono facili da misurare con metodi non invasivi, che sono noti per essere associati con il rischio di diabete di tipo 2, che sono facilmente comprensibili e che costituiscono fattori di rischio modificabili del diabete [11].

Per sviluppare e validare lo score FINDRISK (Finnish Diabetes Risk Score) nel 1987 è stato prelevato un campione casuale dal registro nazionale della popolazione e nel 1992 è stato prelevato un altro campione indipendente (studi FINRISK). I dati sono stati raccolti attraverso un collegamento informatico con il registro dei farmaci dell'Istituto nazionale di previdenza sociale. Questo registro comprende informazioni su tutte le persone finlandesi che sono state autorizzate a ricevere gratuitamente un trattamento farmacologico per alcune malattie croniche, tra cui il diabete.

Poiché l'obiettivo era quello di produrre un semplice calcolatore di rischio che potesse essere comodamente utilizzato nelle cure primarie e anche dagli individui stessi, sono stati inseriti nel modello solo parametri facilmente valutabili senza alcun test di laboratorio o altre misurazioni cliniche che richiedessero particolari competenze.

I parametri indipendenti statisticamente significativi predittori di futuro diabete trattato con farmaci utilizzati sono stati l'età, l'indice di massa corporea (BMI), la circonferenza vita, l'uso di farmaci antiipertensivi e la storia di alti livelli di glucosio nel sangue. Il modello conciso comprende solo queste variabili statisticamente significative. Il modello completo comprende anche l'attività fisica e il consumo

---

Un'area ROC di 0.8 non significa che l'80% delle persone che svilupperanno il diabete saranno effettivamente identificate, ma piuttosto che la probabilità è dell'80% che a un caso selezionato casualmente (cioè una persona che svilupperà il diabete) venga assegnato un rischio di diabete stimato più elevato rispetto a un caso non selezionato casualmente (cioè una persona che rimarrà senza diabete).

giornaliero di frutta o verdura. Anche se queste due variabili non influiscono molto sul potere predittivo del modello statistico, sono state incluse nel Diabetes Risk Score per sottolineare l'importanza dell'attività fisica e della dieta nella prevenzione del diabete. L'IMC compreso tra 25 e 30 kg/m2 non era un predittore statisticamente significativo, tuttavia, è stato incluso nel punteggio finale del rischio di diabete perché è un indice della fase intermedia tra peso normale e obesità, con un impatto ragionevolmente elevato sul rischio di diabete [11].

Nel modello multivariato, il sesso maschile era un predittore statisticamente significativo del rischio di diabete trattato con farmaci. Tuttavia, l'inclusione o l'esclusione del sesso nei modelli ha modificato di poco i coefficienti delle altre variabili indipendenti. Pertanto, il sesso nel modello multivariato finale e nel punteggio finale del Diabete Rischio non è stato incluso.

La validazione del modello è stata effettuata analizzando le aree sotto la receiving operating curve (AUC). Le curve ROC dimostrano che il Diabetes Risk Score basato sulla coorte del 1987 ha predetto molto bene il diabete trattato con farmaci (AUC = 0,85). La previsione era altrettanto buona anche nella coorte del 1992 (AUC = 0,87).

Grazie al Diabetes Risk Score, l'interpretazione del rischio di diabete dell'individuo è facile e può essere espressa come probabilità. Il diabete trattato con farmaci è molto improbabile che si sviluppi in soggetti con un basso punteggio di rischio di diabete; pertanto, questi soggetti possono essere esclusi da ulteriori procedure come il test del glucosio senza causare un problema di risultati falsi-negativi.

Se il Diabetes Risk Score viene utilizzato come versione computerizzata, la probabilità stimata (p) di diabete trattato con farmaci (entro i prossimi 10 anni) per qualsiasi combinazione di fattori di rischio può essere calcolata dai coefficienti con la formula che segue [11]:

$$p(diabetes) = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots)}}$$

dove β0 è l'intercetta e β1, β2, etc. rappresentano i coefficienti di regressione delle varie categorie dei fattori di rischio x1, x2, etc.

Il Diabetes Risk Score è stato progettato per essere uno strumento di screening per identificare i soggetti ad alto rischio nella popolazione e per aumentare la consapevolezza dei fattori di rischio modificabili e dello stile di vita sano.

È disponibile un calcolatore online dello score alla pagina https://qxmd.com/calculate/calculator_236/findrisc-diabetes-risk-calculator.

3.1.1.1.3. QDscore

Il QDScore è un algoritmo di previsione per stimare il rischio di acquisire il diabete di tipo 2 diagnosticato su un periodo di 10 anni in una popolazione etnicamente e socio-economicamente diversificata. L'algoritmo è

stato concepito in modo da basarsi su variabili prontamente disponibili nelle cartelle cliniche elettroniche dei pazienti o che i pazienti stessi potrebbero verosimilmente conoscere - ossia senza bisogno di test di laboratorio o misurazioni cliniche - e quindi consentirne un'applicazione immediata ed efficace sotto il profilo dei costi nella pratica clinica di routine e nelle iniziative nazionali di screening. Il QDScore potrebbe essere utilizzato per identificare i pazienti ad alto rischio di diabete che potrebbero beneficiare degli interventi per ridurre il loro rischio [12].

L'algoritmo è stato progettato a partire da uno studio prospettico di coorte in un'ampia popolazione di pazienti con cure primarie a partire dalla versione 19 del database QResearch (www.qresearch.org). Si tratta di un'ampia banca dati elettronica convalidata per le cure primarie che contiene le cartelle cliniche di 11 milioni di pazienti registrati presso 551 studi medici generici utilizzando il sistema informatico Egton Medical Information System (EMIS). Le pratiche e i pazienti contenuti nella banca dati sono rappresentativi a livello nazionale per l'Inghilterra e il Galles.

Le variabili predittive incluse nel modello note per la loro influenza sul rischio di sviluppare il diabete e generalmente registrate nelle cartelle cliniche elettroniche dei pazienti come parte della pratica clinica di routine sono [12]:

- sesso
- etnia auto-assegnata;
- età;
- indice di massa corporea;
- fumo;
- punteggio di deprivazione Townsend che va da -6 nel più ricco a 11 nei più poveri;
- storia familiare di diabete in un parente di primo grado;
- diagnosi di malattie cardiovascolari;
- ipertensione trattata con farmaci, ossia diagnosi di ipertensione e più di due prescrizioni di farmaci antiipertensivi;
- trattamento con corticosteroidi sistemici, ossia almeno due prescrizioni nei sei mesi precedenti.

Per il disegno dell'algoritmo, i valori di queste variabili erano limitati a quelli registrati nella cartella clinica elettronica della persona prima della diagnosi di diabete di tipo 2. La categoria "white or not recorded" comprendeva i britannici, gli irlandesi e altri soggetti categorizzabili come bianchi, nonché quelli la cui etnia non era registrata. Tale categoria è stata designata come categoria di riferimento.

I set di dati del Regno Unito ricavati da pratiche di medici di famiglia hanno il vantaggio di avere una popolazione numerosa e ampiamente rappresentativa, con una tracciabilità storica dei dati riferita a più di un decennio nella maggior parte delle pratiche. Questi database contengono anche dati su molte delle variabili chiave che notoriamente sono associate al rischio di diabete di tipo 2, come età, sesso, etnia, fumo,

indice di massa corporea, anamnesi familiare di diabete, ipertensione curata, uso corrente di corticosteroidi e deprivazione sociale. La deprivazione è fortemente associata ad una maggiore prevalenza di diabete e di fattori di rischio correlati al diabete come la dieta, l'obesità e il fumo.

Un particolare punto di forza dello studio condotto per definire il punteggio è l'utilizzo di una vasta popolazione rappresentativa a partire da una banca dati convalidata. Un altro punto di forza importante è che tutte le variabili utilizzate nell'algoritmo sono raccolte durante la pratica clinica di routine e sono registrate all'interno della cartella clinica primaria di un singolo paziente nei paesi economicamente più sviluppati. Questo significa che l'algoritmo può essere utilizzato dai pazienti per l'autovalutazione in un calcolatore web (www.qdscore.org). In alternativa, può essere implementato all'interno di sistemi informatici clinici utilizzati nelle cure primarie ed essere utilizzato per stratificare la popolazione (età 25-79 anni) per il rischio su base continuativa senza la necessità di inserire manualmente i dati. Sebbene non esistano soglie ampiamente condivise per la classificazione dei pazienti ad "alto rischio", il QDScore potrebbe fungere da base per un programma sistematico di identificazione dei pazienti ad alto rischio per l'intervento o per contribuire ad una diagnosi precoce.

Uno dei limiti dello studio è che il risultato principale è stato il diabete di tipo 2 diagnosticato da un clinico e registrato sul sistema informatico clinico. Il risultato non è stato formalmente convalidato e lo studio non ha utilizzato i risultati degli esami di laboratorio per confermare la diagnosi. Tuttavia, questa diagnosi difficilmente sarebbe stata registrata se il paziente non avesse avuto il diabete.

Nel modello finale è stata inclusa una misura sostitutiva della deprivazione materiale, il punteggio di Townsend. Il punteggio di Townsend si basa sul codice postale dell'area in cui vice il paziente (corrispondente a circa 150 famiglie) ed è un punteggio composito che comprende la mancanza di auto, il tasso di disoccupazione, il sovrappopolamento e il non possedere una casa di proprietà. Alcune persone che vivono all'interno di un'area non saranno della stessa tipologia degli altri residenti, con conseguente errore di classificazione. La deprivazione è probabilmente associata ad altri fattori noti per aumentare il rischio di diabete, come la cattiva alimentazione, la mancanza di esercizio fisico e l'aumento dell'assunzione di alcol, e quindi terrà conto almeno in parte di alcuni degli effetti di questi fattori. I coefficienti di Townsend sono stati espressi in unità di deviazione standard in modo che il coefficiente possa essere applicato in un contesto non britannico in cui potrebbero essere applicati altri indici di deprivazione.

Uno dei vantaggi del QDScore è l'inclusione sia della deprivazione che dell'etnia di appartenenza autoassegnata, che sono indipendentemente associate al rischio di diabete incidente; questo è probabilmente utile per i problemi riscontrati con il punteggio di rischio Cambridge nella sua performance in popolazioni etnicamente diverse.

Il QDScore è stato convalidato testando le prestazioni dell'algoritmo in un set di dati di validazione. È stata calcolata l'area under the receiving operator curve (AROC), dove valori più elevati indicano una migliore

discriminazione: l'AROC per il dataset originale era di 0,83 per gli uomini e 0,85 per le donne; l'AROC per il dataset di validazione era rispettivamente 0,80 e 0,81 [12].

Esiste un tool per il calcolo online del QDScore alla pagina http://www.qdiabetes.org/; inoltre, sempre presso la stessa pagina, è possibile scaricare il codice sorgente di un'applicazione che permette di calcolare lo score fornendo i parametri necessari nella linea di comando.

3.1.1.1.4. AUSDRISK

AUSDRISK è uno tool australiano per la valutazione del rischio di diabete di tipo 2 per la previsione del rischio a cinque anni di diabete basato su misure demografiche, di stile di vita e antropometriche semplici; in particolare, lo strumento si basa su nove fattori di rischio noti o di facile autovalutazione: età, sesso, etnia, storia familiare di diabete, anamnesi di glicemia elevata, uso di farmaci antiipertensivi, fumo, inattività fisica e circonferenza vita [13].

Il modello è stato progettato utilizzando i dati provenienti da 5 anni di follow-up dello studio australiano sul diabete, l'obesità e lo stile di vita (AusDiab). Lo studio di base riguardava un'indagine trasversale, nazionale e basata sulla popolazione, condotta su 11247 adulti di età pari o superiore a 25 anni provenienti da 42 distretti del censimento selezionati casualmente. I dati sulle caratteristiche demografiche, il consumo di alcol e di alimenti, il fumo, l'attività fisica, il tempo trascorso davanti alla televisione, la storia familiare di diabete e l'anamnesi medica sono stati raccolti tramite un questionario somministrato da un intervistatore. L'etnia è stata classificata in base al paese di nascita, e il retaggio aborigeno e delle isole dello Stretto di Torres è stato accertato mediante una domanda specifica. L'attività lavorativa auto-dichiarata è stata codificata in gruppi di occupazione utilizzando la classificazione australiana standard delle professioni. I partecipanti sono stati codificati come fisicamente inattivi se il tempo totale impegnato a camminare o in attività moderata o vigorosa era inferiore a 150 minuti alla settimana. Ai partecipanti è stato domandato se fosse mai stato segnalato loro di avere alti livelli di glucosio nel sangue (anche durante la gravidanza per le donne) [13].

L'analisi di validazione ha mostrato una buona capacità di discriminazione.

L'etnia è stata precedentemente riconosciuta come un predittore significativo per il diabete incidente ed è stata quindi inclusa nei modelli dello studio Atherosclerosis Risk in Community (ARIC) e dello studio San Antonio Heart Study. Nei criteri diagnostici della sindrome metabolica proposti dalla Federazione internazionale per il diabete, si raccomandano soglie specifiche per etnia e sesso per la circonferenza della vita per definire l'adiposità centrale. Questo approccio è stato adottato nell'analisi per riflettere il fatto che il rischio di disturbi metabolici in aborigeni e abitanti dell'isola Torres Strait o asiatici sono associati ad una circonferenza vita più piccola rispetto alle persone di origine Europea. Poiché l'obiettivo dello studio era quello di fornire un semplice punteggio di rischio di diabete, sono state escluse diverse variabili con

significatività statistica. L'attività lavorativa è stata eliminata nonostante la presenza di una categoria (caratteristiche domestiche) che mostra una relazione statisticamente significativa con il diabete incidente. L'istruzione è stata esclusa in quanto di importanza marginale e non ha contribuito in modo significativo al modello. È stata presa in considerazione anche la possibilità di eliminare l'IMC dal modello. In primo luogo, dato che la categoria di IMC relativa agli obesi (30-< 35 kg/m2) aveva solo significatività statistica borderline, e solo la categoria obesità morbosa con IMC (35 kg/m2) è chiaramente entrata nel modello, la sua inclusione è stata ritenuta un'inutile complicazione. In secondo luogo, l'inclusione di sole categorie di IMC che indicavano obesità o obesità morbosa avrebbe potuto inviare un messaggio non appropriato di salute pubblica, suggerendo che coloro che sono in sovrappeso (25-30 kg/m2) non sono a rischio elevato di diabete. Misurare la circonferenza della vita è sempre più pubblicizzato come uno strumento di salute pubblica, ma può essere difficile per le persone misurare accuratamente la propria circonferenza vita. Pertanto, nel modello si è provato a sostituire l'IMC al posto della circonferenza vita, ma il risultato è stato una perdita di potere discriminante. Inoltre, anche se il peso e l'altezza sembrano più semplici da misurare, le bilance con misure accurate non sono normalmente disponibili al grande pubblico, inoltre, molte persone probabilmente si affidano al ricordo di una misurazione precedente; per questi motivi, la conversione del peso e dell'altezza in IMC (anche con tabelle appropriate) è probabile che porti ad errori.

Oltre al suo valore predittivo, l'AUSDRISK ha ottenuto buoni risultati nel discriminare tra coloro che avevano e quelli che non avevano diagnosticato il diabete nello studio trasversale di AusDiab.

Questo studio presentava alcuni limiti. Poiché sono state incluse solo le persone di età superiore ai 25 anni, l'uso dell'AUSDRISK nelle fasce di età più giovani avrebbe probabilmente sopravvalutato il rischio di diabete. In secondo luogo, dato che sono stati seguiti troppo pochi individui di origine aborigena e delle isole dello Stretto di Torres e delle isole del Pacifico per determinare con precisione il rischio di diabete per questo gruppo, essi sono stati combinati con i cittadini dell'Europa meridionale e asiatici per generare un gruppo etnico ad alto rischio. Questo approccio potrebbe non essere sufficientemente accurato per cogliere il rischio eccessivo di diabete in queste due popolazioni. Infine, l'AUSDRISK è stato sviluppato per la previsione di diabete incidente e non per la ridotta tolleranza al glucosio o l'alterazione del glucosio a digiuno.

AUSDRISK è stato convertito in un questionario basato su punti e di facile utilizzo per il paziente (disponibile su http://www.bakeridi.edu/aus_diabetes_risk) e in uno strumento interattivo online di valutazione del rischio (disponibile su http://health.gov.au/internet/main/publishing.nsf/ Content/diabetesriskassessment tool). Queste versioni includono un fattore di rischio relativo al consumo di frutta e verdura, che non era un predittore significativo del diabete nel modello finale, ma è stato aggiunto per il suo valore di messaggio di salute pubblica.

L'AROC del modello AUSDRISK sul dataset originale è di 0,78. Il modello è stato inoltre convalidato su due coorti indipendenti e l'AROC in questi casi è stato rispettivamente pari a 0,66 e 0,79.

3.1.1.2. *Modelli predittivi estesi*

I modelli predittivi estesi sono quelli che prevedono l'utilizzo, accanto a variabili di predizione facilmente reperibili ad opera del paziente o del medico (ad esempio la cartella clinica personale del un paziente), anche di variabili di predizione che richiedono la necessità di aver effettuato degli esami invasivi (es. analisi del sangue). Tra le variabili utilizzate come fattori di rischio ed utilizzate nei modelli predittivi estesi ci sono i valori di glicemia, colesterolo HDL, uricemia, trigliceridemia, emoglobina glicata e glicemia a digiuno.

3.1.1.2.1. ARIC

Lo scopo del modello è quello di prevedere l'alto rischio di sviluppare il diabete negli adulti di mezza età, bianchi e afroamericani utilizzando informazioni cliniche prontamente disponibili.
Nel periodo 1987-1989, lo studio Atherosclerosis Risk in Communities (ARIC) ha reclutato una coorte di 15792 uomini e donne provenienti da quattro comunità statunitensi. Sono stati studiati 7915 partecipanti di età compresa tra i 45 e i 64 anni, esenti da diabete a livello basale e sono stati accertati 1.292 casi di diabete per diagnosi clinica o test di tolleranza orale al glucosio [14].
I partecipanti sono stati valutati per l'uso di farmaci per il diabete e l'ipertensione, fumo e storia familiare di diabete (in entrambi i genitori) mediante interviste. L'IMC è stato calcolato come peso/altezza2 (kg/m2) e l'obesità è stata definita come IMC ≥30 kg/m2. Alla fine del follow-up è stato definito il diabete incidente tramite un OGTT o come rapporto di diagnosi clinica o trattamento del diabete durante il periodo di follow-up.
I fattori di rischio considerati sono stati: sesso, etnia, storia familiare di diabete, uso di farmaci per l'ipertensione, altezza, età, varie misure di obesità (circonferenza vita, peso, IMC, rapporto vita-fianchi), pressione sanguigna sistolica, glicemia a digiuno, colesterolo HDL, trigliceridi e insulina a digiuno.
Sono stati definiti due modelli: uno comprendente solo elementi clinicamente rilevabili che non richiedono una valutazione di laboratorio e l'altro comprendente solo la glicemia a digiuno. Successivamente sono stati definiti altri due modelli. Il primo combina elementi dei due modelli iniziali e il secondo conteneva anche il colesterolo HDL e i trigliceridi. In questi modelli non sono stati inclusi né l'IMC né l'insulina a digiuno, poiché il contributo supplementare all'AUC, sebbene statisticamente significativo, era minimo per ciascuno di essi. Le proprietà diagnostiche delle funzioni di rischio sono state successivamente valutate nel campione di prova.

Rispetto al modello con la sola variabile clinica, il modello che combinava gli elementi clinici con la glicemia a digiuno era più predittivo (AUC 0,78), mentre quello che includeva i lipidi era il miglior predittore (AUC 0,80). Le funzioni di rischio, composte dalle variabili cliniche più la glicemia a digiuno (AUC 0,78) e, in aggiunta, i trigliceridi e il colesterolo HDL (AUC 0,80) hanno dato risultati migliori. Anche se le migliori proprietà diagnostiche sono state quelle derivate da una funzione di rischio, comprese le variabili lipidiche (AUC 0.80), il guadagno è piccolo rispetto a quelle derivate da una funzione di rischio senza lipidi (AUC 0.78) e, a seconda dell'impostazione, non può giustificare l'aumento delle risorse necessarie per le misure lipidiche. Una funzione di rischio basata esclusivamente sulle variabili cliniche, pur avendo una capacità predittiva complessiva inferiore (AUC 0,71), è stata in grado di ricavare regole con sensibilità che vanno dal 40% al 77% e specificità corrispondenti che vanno dall'84% al 55%. Pertanto, le regole basate solo su informazioni cliniche possono essere utili, ad esempio, come primo passo nelle strategie diagnostiche per la prevenzione primaria in ambito comunitario [14].

Le sensibilità e le specificità del modello per gli adulti di mezza età, bianchi e afroamericani possono non essere applicabili ai gruppi più giovani o anziani o a quelli in altri contesti, con conseguente diverso rischio sottostante di sviluppare il diabete.

In conclusione, le regole derivate da informazioni cliniche prontamente disponibili, da sole o combinate con semplici misure di laboratorio, possono caratterizzare gruppi di adulti di mezza età come soggetti a vari gradi di rischio di diabete. Questa categorizzazione permette di classificare l'intensità delle azioni preventive in base al grado di rischio di ciascun paziente.

Un calcolatore online del rischio ARIC è disponibile all'indirizzo http://aricnews.net/diabrisk/diabrc1.html.

3.1.1.2.2. KORA

Il modello KORA ha l'obiettivo di prevedere il diabete di tipo 2 per la popolazione anziana e di verificare in che misura l'aggiunta di misurazioni di glucosio a due ore (test di tolleranza del glucosio orale) e biomarcatori consente di migliorare il potere predittivo dei punteggi di rischio basati su parametri clinici convenzionali e non biochimici.

Test di tolleranza al glucosio orale sono stati effettuati su un campione di 1353 soggetti, di età compresa tra i 55 e i 74 anni nella Germania meridionale dal 1999 al 2001. La coorte è stata oggetto di nuove indagini nel 2006-2008. Sono stati progettati tre modelli. Età, sesso, IMC, familiarità per diabete, fumo e ipertensione sono stati selezionati come variabili per il modello 1. Il modello 2 comprendeva inoltre la glicemia a digiuno, l'emoglobina glicata e l'acido urico. Per il modello 3 sono state selezionate le stesse variabili ed in più la glicemia a 2 ore. L'area under the receiver operating characteristic curve (AROC) è aumentata significativamente da 0,763 (modello 1) a 0,844 (modello 2) e 0,886 (modello 3). I biomarkers come l'adiponectina e l'insulina non hanno migliorato le capacità predittive dei modelli 2 e 3 [15].

3.1.1.2.3. Framingham

L'obiettivo del punteggio di Framingham è quello di prevedere il diabete mellito incidente a 7 anni negli adulti di mezza età. Nel corso dello studio si sono verificati 160 casi di nuovo diabete di tipo 2; modelli di regressione sono stati utilizzati per prevedere nuovi casi di diabete di tipo 2, a partire da caratteristiche note al soggetto (modello personale, vale a dire, età, sesso, storia familiare di diabete e indice di massa corporea), aggiungendo semplici misurazioni cliniche che includevano caratteristiche di sindrome metabolica (modello clinico semplice) e, infine, valutando complessi modelli clinici che includevano (1)il test di tolleranza al glucosio a due ore; (2) l'indice di sensibilità dell'insulina Gutt; (3) il modello omeostatico di resistenza all'insulina e il modello omeostatico degli indici di sensibilità della resistenza all'insulina delle cellule beta. Il potere discriminante è stato valutato mediante il calcolo dell'AROC (area under the receiver operating characteristic curve) [16].

Le variabili del modello personale, ad eccezione del sesso, erano predittori statisticamente significativi del diabete di tipo 2 (AROC di 0,72). Nel modello clinico semplice, la storia familiare di diabete e di obesità sono predittori significativi, insieme con l'ipertensione, bassi livelli di colesterolo HDL, livelli elevati di trigliceridi e squilibri della glicemia a digiuno (AROC, 0,85). I modelli clinici complessi non hanno mostrato ulteriori miglioramenti nelle discriminazioni dei modelli (AROC di 0,850-0,854) e non sono stati di livello superiore rispetto al semplice modello clinico [16].

Il diabete familiare, l'obesità e le caratteristiche della sindrome metabolica prevedono efficacemente il rischio di diabete di tipo 2 in un campione di popolazione bianca di mezza età e sono stati utilizzati per sviluppare un semplice algoritmo di previsione del diabete di tipo 2 per stimare il rischio di nuovo diabete di tipo 2 durante un intervallo di follow-up di 7 anni [16].

È disponibile online un calcolatore dello score alla pagina http://www.framinghamheartstudy.org/risk-functions/diabetes/#.

3.1.1.2.4. San Antonio

Lo studio di San Antonio ha avuto lo scopo di determinare se i modelli multivariabili sono superiori ai test di tolleranza al glucosio per identificare le persone ad alto rischio di diabete mellito. Lo studio è stato condotto come studio prospettico di coorte a San Antonio in Texas. La coorte era costituita da 1791 messicani americani e da 1112 bianchi non ispanici non affetti da diabete, selezionati casualmente da censimenti [17].

Le variabili misurate sono state: anamnesi medica, indice di massa corporea, pressione sanguigna, livelli di glicemia a digiuno e dopo 2 ore, colesterolo LDL e HDL e livello di trigliceridi.

Per la previsione di un'incidenza a 7,5 anni di diabete di tipo 2, l'AROC (area under the receiver-operating characteristic) per un modello multivariabile che coinvolge variabili cliniche prontamente disponibili era significativamente maggiore dell'AROC relativa al parametro con il solo valore della glicemia dopo 2 ore. L'aggiunta della misurazione della glicemia dopo 2 ore al modello multivariabile ha aumentato l'AROC, ma solo dall'84,3% all'85,7% [17].

Anche se l'aggiunta della variabile della glicemia dopo 2 ore al modello migliora la previsione, il leggero miglioramento che ne risulta comporta maggiori costi e disagi.

3.1.1.3. *Modelli predittivi delle complicanze del diabete*

Con l'introduzione delle cartelle cliniche elettroniche vengono acquisite e memorizzate ulteriori informazioni sulle interazioni medico-paziente. I modelli predittivi hanno in questo modo il potenziale per fondere in un contesto clinico diversi tipi di informazioni sul paziente. Il diabete è una delle principali priorità in ambito medico-scientifico e nella gestione dell'assistenza sanitaria; pertanto sono disponibili numerosi dati e informazioni su questi pazienti. Il diabete, inoltre, è una patologia che può portare ad un gran numero di complicanze a lungo termine molto gravi come la cecità, l'amputazione e le malattie cardiache se non curate in tempo.

Il trattamento, specialmente con l'insulina, non è privo di effetti collaterali come il rischio di ipoglicemia e aumento di peso. I modelli predittivi potrebbero potenzialmente contribuire alla gestione di questi problemi legati al diabete. Negli ultimi decenni si è assistito ad una rapida presa di coscienza delle possibilità di utilizzare le informazioni disponibili per prevedere gli esiti del diabete.

I modelli predittivi spesso includono più predittori per stimare la probabilità o il rischio di un determinato esito o per definire che un determinato risultato è presente/assente (modello di previsione diagnostico) o accadrà in un determinato lasso di tempo (modello di previsione prognostico) in un individuo.

Sono stati compiuti molti sforzi per sviluppare modelli predittivi da utilizzare nella gestione del diabete e delle sue complicanze. Tuttavia, in generale, la maggior parte di questi modelli non sono stati attuati e l'impatto clinico non è stato esaminato. Sebbene manchino le evidenze dell'applicazione pratica, si ritiene che i modelli predittivi abbiano il potenziale di permettere ai fornitori di assistenza sanitaria di acquisire molte più conoscenze e di prendere decisioni più informate attingendo alla grande quantità di dati clinici memorizzati elettronicamente [18].

Nei paragrafi seguenti è riportata una breve descrizione di alcuni modelli di previsione delle complicanze a breve e lungo termine del diabete.

3.1.1.3.1. Modelli predittivi per le complicanze a lungo termine

Retinopatia

La retinopatia diabetica è una delle cause principali di cecità in tutto il mondo e questa grave complicazione del diabete è già presente al momento della diagnosi clinica del diabete di tipo 2 in alcuni pazienti. La prassi è quella di consigliare ai pazienti con diabete di tipo 2 di sottoporsi ad un primo esame oculistico completo da parte di un oftalmologo o optometrista poco dopo la diagnosi di diabete.

C'è un lungo periodo di latenza prima della perdita visiva e la progressione di questa malattia è in larga misura prevedibile e curabile. Diversi studi si sono concentrati sull'individualizzazione dell'intervallo di screening in base ai fattori di rischio per la progressione della retinopatia. Looker et al. [19] hanno utilizzato hidden Markov models per calcolare le probabilità di estendere l'intervallo per persone senza retinopatia visibile. Mehlsen et al. [20] hanno costruito un modello di regressione logistica multipla per regolare l'intervallo di screening nei pazienti a basso rischio. I predittori inclusi nel modello erano HbA1c, numero di emorragie retiniche ed essudati, maggior durata del diabete e pressione sanguigna. Cichosz et al. [18] hanno pubblicato un modello utilizzabile per selezionare un gruppo ad alto rischio tra i pazienti con diabete di recente diagnosi.

Lo stesso gruppo di ricercatori del QDScore ha realizzato anche un modello predittivo per il rischio di cecità in pazienti diabetici. Tale score è stato calcolato sulla base dei dati raccolti regolarmente da molte migliaia di medici di medicina generale in tutta la Gran Bretagna e strutturati nella banca dati QResearch [25]. Il modello predittivo per la cecità negli uomini e nelle donne include età, il rapporto colesterolo totale / colesterolo HDL, la pressione sistolica, l'HbA1c, la deprivazione sociale, la durata del diabete, il tipo di diabete, la malattia renale cronica e la retinopatia o la maculopatia rigenerativa esistente. L'indice di massa corporea e il fumo non erano significativamente associati al rischio. Valori crescenti di età, HbA1c e pressione sanguigna sistolica sono stati associati con un aumento del rischio di cecità. Anche valori crescenti del rapporto colesterolo totale/colesterolo HDL sono stati associati con un aumento del rischio di cecità. L'aumento della durata del diabete è stato associato ad un aumento del rischio nonostante l'adeguamento all'età e ad altri fattori di rischio. È stata riscontrata un'interazione significativa tra malattia renale ed età. La retinopatia o maculopatia preesistente è il fattore di rischio più forte, con un incremento di 2,7 volte superiore per le donne e un incremento di 2,9 volte superiore per gli uomini.

È stato messo a disposizione un calcolatore online del rischio alla pagina http://qdiabetes.org/amputation-blindness/index.php.

Neuropatia

Il 50% delle persone con diabete di tipo 2 sono affetti da neuropatia e pertanto hanno un più alto rischio di sviluppare un'ulcera da piede diabetico. L'ulcera al piede è una delle principali complicanze nei pazienti diabetici, con un rischio di amputazione del 15%. Con la diagnosi precoce e un trattamento multidisciplinare

appropriato, la frequenza di amputazione può essere ridotta fino al 60-85%. Nel 2006, Boyko et al[21] hanno seguito 1285 veterani diabetici e hanno pubblicato un modello di previsione basato su 7 variabili cliniche comunemente disponibili per lo sviluppo di ulcere al piede. Alla fine non è stato adottato alcun sistema prestabilito e l'applicazione dei modelli di convalida nella pratica clinica rimane limitata. Il potenziale dei modelli di previsione dell'ulcera al piede è grande, ma sono necessari ulteriori studi.

Come per la cecità, il gruppo di ricercatori del QDScore ha realizzato anche un modello predittivo per il rischio di amputazione in pazienti diabetici [25]. Il modello finale per l'amputazione degli arti inferiori nelle donne includeva l'età, la pressione sistolica, l'HbA1c, la deprivazione sociale, la durata del diabete, il fumo, l'etnia, la presenza di artrite reumatoide, l'insufficienza cardiaca, le malattie vascolari periferiche e le malattie renali croniche. Il modello finale negli uomini comprendeva anche il tipo di diabete e la fibrillazione atriale. L'indice di massa corporea e il rapporto colesterolo totale / colesterolo HDL non sono stati significativamente associati al rischio negli uomini e nelle donne. L'aumento della durata del diabete è stato associato ad un aumento del rischio di amputazione degli arti inferiori negli uomini e nelle donne. L'aumento dei livelli di fumo era associato ad un maggiore rischio di amputazione; l'associazione era più marcata nelle donne che negli uomini. Per i fumatori pesanti rispetto ai non fumatori, il rischio di amputazione è aumentato di 1,9 volte per le donne e di 1,3 volte per gli uomini. Una malattia vascolare periferica preesistente era associata a rischi più elevati (di quattro volte nelle donne e di tre volte negli uomini), seguita da malattie renali croniche (di 2,7 volte nelle donne e di 2,3 volte negli uomini).

Un calcolatore online del rischio è disponibile alla pagina http://qdiabetes.org/amputation-blindness/index.php.

Nefropatia

L'identificazione delle persone ad alto rischio di rapido declino della funzione renale è importante e sono stati dimostrati interventi basati su evidenze scientifiche per prevenire o rallentare lo sviluppo verso fasi avanzate della nefropatia. La maggior parte dei modelli sviluppati per prevedere la progressione della malattia renale sono stati testati in un contesto generale. I fattori più comunemente usati in questi modelli sono il sesso, l'età, l'IMC, lo stato del diabete, la pressione sanguigna, la creatinina serica, le proteine presenti nelle urine e il rapporto albumina serica / proteine totali.

Cardiopatia

Il diabete è un noto fattore di rischio per le cardiopatie coronariche. Il diabete aggiunge un rischio circa 2 volte superiore per una vasta gamma di malattie vascolari, indipendentemente da altri fattori di rischio convenzionali.

Uno dei migliori modelli è il punteggio Framingham, che è stato ampiamente accettato e include il diabete come predittore. Diversi punteggi sono stati sviluppati specificamente per prevedere le malattie cardiache nei pazienti con diabete. L'AUC di questi modelli varia da 0,59 a 0,80. I predittori più frequentemente inclusi sono sesso, età, pressione sanguigna sistolica, colesterolo e fumo. Nonostante gli sforzi profusi in questo campo, la maggior parte dei modelli deve ancora essere dimostrata valida nella cura quotidiana. Secondo l'International Diabetes Federation, questi modelli non sono adeguati o sono limitati perché non si sono dimostrati utili in popolazioni con più di 65 anni e perché sono stati applicati in persone in cui era già stato avviato un trattamento per prevenire le malattie cardiache. La ricerca futura dovrebbe concentrarsi sull'impatto dell'uso di modelli di previsione delle cardiopatie coronariche nella cura quotidiana dei pazienti affetti da diabete.

3.1.1.3.2. Modelli predittivi per le complicanze a lungo termine

Ipoglicemia

Le persone con diabete di tipo 1 spesso sperimentano episodi di ipoglicemia perché hanno bisogno di ridurre il livello di zucchero nel sangue utilizzando l'insulina. Anche i pazienti con diabete di tipo 2 possono avere episodi di ipoglicemia a causa dell'uso crescente di insulina in questo gruppo. La paura indotta dall'ipoglicemia è pronunciata e gli esiti clinici di questa condizione sono gravi. L'incidenza di ipoglicemia che richiede assistenza in emergenza raggiunge il 7,1% all'anno tra i pazienti con diabete e ben il 6% di tutti i decessi in pazienti con diabete di tipo 1 sono dovuti all'ipoglicemia.

L'introduzione del sistema di monitoraggio continuo della glicemia (CGM) ha permesso di misurare frequentemente la glicemia interstiziale e molti scienziati hanno analizzato le opportunità offerte da questa nuova tecnologia. Tuttavia, l'uso del CGM per la previsione dell'ipoglicemia comporta l'accettazione di una certa percentuale di falsi allarmi positivi. L'ipoglicemia colpisce l'intero sistema nervoso autonomo, compreso il cuore, il cervello e la traspirazione. Ciò ha portato allo sviluppo di sistemi di previsione che includono informazioni da EEG, misure di impedenza cutanea ed elettrocardiogrammi.

Guadagno di peso associate all'insulina

Nella maggior parte dei pazienti con diabete di tipo 2 sarà necessario iniziare il trattamento dell'insulina per raggiungere l'obiettivo terapeutico di HbA1c < 7 mmol/l. Il problema dell'aumento di peso indotto dall'insulina è stato documentato a lungo come un problema nel trattamento del diabete. L'aumento del peso può influenzare negativamente il profilo di rischio cardiovascolare e aumentare la morbilità e la mortalità quando il trattamento intensivo è rinviato a causa del timore del paziente di aumentare di peso. La previsione dell'aumento di peso associato all'insulina ha attirato poca attenzione in letteratura rispetto ad altre complicanze del diabete. È risaputo che il dosaggio dell'insulina è un forte fattore predittivo

dell'aumento di peso. Jansen et al. [22] hanno seguito 65 pazienti con diabete durante il trattamento con insulina e hanno proposto un modello di regressione per la "previsione" dell'aumento di peso. Tuttavia, il modello non è adatto all'uso prospettico in quanto richiede dati su 0-12 mesi di dosaggio dell'insulina e su eventuali cambiamenti nel dosaggio dell'insulina. Inoltre, in questo studio non sono riportate misure di performance. Balkau et al [23] hanno riportato dati sui fattori associati all'aumento di peso associato all'insulina in 2179 pazienti con diabete di tipo 2. Hanno anche proposto un modello che potrebbe spiegare parte dell'aumento di peso, ma il loro modello non era funzionale per l'uso prospettico in clinica. I fattori inclusi in questo modello sono stati HbA1c, IMC e informazioni sull'insulina. Cichosz et al. [24] hanno sviluppato un altro modello per prevedere se i pazienti nel trattamento dell'insulina sono in eccesso di aumento di peso a 1,5 anni dalla linea di base. Questo modello aveva una AROC di 0,80. Le informazioni di base sono state incorporate in un modello di regressione multilogistica combinato con l'aumento di peso nei primi 3 mesi. In futuro, sono necessari ulteriori studi in questo campo e il modello di Cichosz et al. [24] può essere utilizzato nella pratica clinica per identificare le persone con il rischio di grandi aumenti di peso, ma manca uno studio di validazione.

3.1.2. *Confronto fra i modelli predittivi disponibili*

Per effettuare un confronto tra i modelli predittivi presi in considerazione è possibile analizzare le tabelle che seguono. Nella Tabella 1 sono riportati i modelli predittivi di base per il rischio di diabete, mentre nella Tabella 2 sono riportati i modelli predittivi estesi. Per poter confrontare ciascun modello sono stati presi in considerazione i seguenti elementi:

- I fattori di rischio (variabili) non invasivi utilizzati dal modello
- I fattori di rischio (variabili) metabolici utilizzati dal modello
- Il range di età in cui è applicabile (basato principalmente sulla coorte utilizzata per determinare il modello)
- L'area AROC ("under the receiving operator curve) calcolata sul dataset originale
- L'area AROC calcolata sul dataset di validazione
- La disponibilità o meno di un calcolatore online dello score
- La disponibilità di un tool da linea di comando per il calcolo dello score
- L'adattabilità dello score ad altri contesti (es. altre nazioni)

Tra gli elementi selezionati per il confronto, l'AROC è stata selezionata in quanto rappresenta una misura di discriminazione, ossia quanto bene un modello predittivo di rischio identifica coloro che svilupperanno una malattia nel periodo di follow-up nell'ambito di uno studio di coorte. Dato che i punteggi di rischio

mostrano spesso un eccesso di overfit ai set di dati utilizzati per lo sviluppo del modello, è necessaria la validazione esterna di un modello predittivo prima del suo utilizzo diffuso. La convalida esterna consiste nell'applicare il modello di previsione a un insieme di dati con persone diverse e nel rivalutare le misure delle prestazioni del modello. La convalida interna non è pienamente equivalente a quella esterna, in quanto la convalida interna si basa ancora sui dati originali.

L'analisi dei vari modelli predittivi effettuata in diverse review [9] permette di identificare alcuni punti chiave per la valutazione dell'efficacia dei fattori di rischio sperimentati e quindi dei modelli predittivi stessi. Sebbene l'uso di misurazioni ripetute per la previsione del diabete sembra essere un approccio interessante, solo i fattori di rischio con traiettorie molto diverse sono suscettibili di migliorare la capacità predittiva di un determinato punteggio di rischio [9].

Si può sostenere che sia ovvio migliorare la previsione basata su misurazioni multiple rispetto alle singole misurazioni della glicemia, ma il fatto che i punteggi attuali non facciano uso di misurazioni multiple nella pratica clinica per migliorare la valutazione del rischio individuale sembra degno di nota. Una problematica importante per quanto riguarda le misurazioni ripetute dei fattori di rischio è che questo approccio potrebbe avere effetti negativi sulla prevenzione delle malattie, in quanto potrebbe ritardare l'avvio degli sforzi preventivi. Alcuni studi hanno confrontato l'accuratezza dei modelli di previsione senza e con i metaboliti e hanno riscontrato miglioramenti relativamente modesti negli AROC per i modelli che includevano la metabolomica oltre ai fattori di rischio accertati per il diabete di tipo 2.

È stata valutata anche l'influenza di varianti genomiche sul rischio di diabete; fattori di rischio basati sulla combinazione di 40 SNP hanno portato ad AROC di 0,55-0,63, sostanzialmente inferiori a quelle ottenute da età, sesso e IMC solamente. In alcuni studi, l'aggiunta di informazioni sul genotipo a modelli basati su fattori di rischio antropometrici e clinici accertati ha portato ad aumenti statisticamente significativi degli AROC, ma questi miglioramenti non erano generalmente superiori a 0,03.

Una perfetta previsione del diabete potrebbe non essere possibile per almeno cinque motivi [9]:

1. La diagnosi del diabete non è chiara come quella di altre malattie croniche (e.g. il cancro). Ad esempio, la diagnosi di una persona con diabete quando il livello di glicemia dopo 2 h è di 201 mg/dl, ma non quando il livello è di 199 mg/dl dipende dall'imprecisione della misurazione.
2. L'imprecisione della misurazione si applica anche a tutti i predittori utilizzati nei punteggi di rischio.
3. I punteggi di rischio non riescono a cogliere i cambiamenti dello stile di vita o dei farmaci in seguito alla valutazione del rischio individuale.
4. I casi di diabete incidente nello studio di coorte utilizzato per sviluppare un modello di previsione potrebbero non essere stati rilevati perché si sono verificati dopo la fine del periodo di follow-up; ciò contribuisce all'errore di misurazione nel risultato del diabete di tipo 2.

5. Molti nuovi biomarcatori descritti come fattori di rischio indipendenti per il diabete di tipo 2 sono correlati con i tradizionali fattori di rischio o altri biomarcatori. Pertanto, esse forniscono solo informazioni supplementari limitate e non contribuiscono a una migliore discriminazione.

Diventa necessario, quindi, individuare in quale condizione i nuovi marcatori possono avere un valore incrementale maggiore. Spesso i biomarcatori sono fortemente correlati ai fattori di rischio convenzionali, in modo da non fornire ulteriori informazioni predittive. I nuovi biomarcatori che saranno sperimentati miglioreranno la previsione del diabete solo se sono debolmente correlati con i fattori di rischio già accertati.

Tabella 1 Modelli predittivi di base per il rischio di diabete di tipo II

Modello predittivo	Fattori di rischio non invasivi	Range di età in cui è applicabile	AROC dataset originale	AROC dataset validazione	Calcolatore online disponibile?	Tool da linea di comando disponibile?	Adattabile ad altri contesti?
Cambridge Risk Score	• Età • Sesso • IMC • Fumo • Uso di farmaci corticosteroidi • Uso di farmaci antipertensivi • Storia familiare di diabete	?	0.74	0.72	?	No	Sì
FINDRISC	• Età • IMC • Circonferenza vita • Uso di farmaci antipertensivi • Storia di alti livelli di glicemia • Inattività fisica • Dieta (frutta, verdura, etc.)	> 18	0.85	0.77	Sì anche tramite questionario	No	Sì
QDScore	• Età • Sesso • Etnia • IMC • Fumo • Familiarità per diabete • Punteggio di deprivazione Townsend • Ipertensione trattata	25-84	0.83 (M) 0.85 (F)	0.80 (M) 0.81 (F)	Sì	Sì	Sì mediante adeguamento indice di deprivazione

	• Cardiopatia • Uso di corticosteroidi						
AUSDRISK	• Età • Sesso • Etnia • Familiarità per diabete • Storia di alti livelli di glicemia • Uso di farmaci antipertensivi • Fumo • Inattività fisica • Circonferenza vita	> 25 anni	0.78	0.79	Sì anche tramite questionario	No	Sì

Tabella 2 Modelli predittivi estesi per il rischio di diabete di tipo II

Modello predittivo	Fattori di rischio non invasivi	Fattori di rischio metabolici	Range di età in cui è applicabile	AROC dataset originale	AROC dataset validazione	Calcolatore online disponibile?	Tool da linea di comando disponibile?	Adattabile ad altri contesti?
ARIC	• Età • Etnia • Circonferenza vita • Altezza • Pressione sistolica • Familiarità per diabete	• Glicemia a digiuno • Trigliceridi • Colesterolo HDL	45-64	0.80	0.84	Sì	No	?
KORA	• Età • Sesso • IMC • Familiarità per diabete	• Glicemia a digiuno • Emoglobina glicata	55-74	0.84	0.81	?	No	?

	• Fumo • Ipertensione	• Acido urico						
Framingham	• IMC • Familiarità per diabete • Pressione arteriosa	• Glicemia a digiuno • Colesterolo HDL • Trigliceridi	45-64	0.85	0.78	Sì	No	?
San Antonio	• Età • Sesso • IMC • Etnia • Pressione sistolica • Storia familiare per diabete	• Glicemia a digiuno • Colesterolo HDL	?	0.84	0.83	?	No	?

3.1.3. *Selezione del modello predittivo più utili*

Alla luce dell'analisi e del confronto effettuati sui modelli predittivi presi in considerazione in questo capitolo è possibile individuare il modello predittivo che ha le caratteristiche più adatte per essere adottato all'interno del progetto.

Gli elementi presi in considerazione per determinare la scelta di un modello predittivo (e che in parte coincidono con gli elementi riportati nelle tabelle del paragrafo precedente) sono state:

1. Semplicità di reperimento delle informazioni necessarie
2. Applicabilità alla più ampia popolazione di individui per quanto riguarda età e tipologia di popolazione
3. Capacità di discriminazione
4. Fruibilità del modello da parte del pubblico

Semplicità di reperimento delle informazioni necessarie

I modelli predittivi, come descritto nei paragrafi precedenti, si distinguono in modelli di base e modelli estesi. I modelli di estesi sono quelli che necessitano, per il calcolo dello score, di fattori di rischio / variabili non sempre prontamente disponibili perché richiedono analisi del sangue (ad es. valori ematici di colesterolo HDL, acido urico, trigliceridi, etc.). I modelli di base, invece, sono stati costruiti con variabili che ogni paziente conosce su sé stesso o che il medico conosce del paziente e che non richiedono analisi ematiche. Tra le due categorie di modelli predittivi è preferibile usare quelli di base perché applicabili ad una più ampia popolazione di pazienti; per cui, nella scelta del modello, saranno presi in considerazione esclusivamente i modelli di base.

Applicabilità alla più ampia popolazione di individui per quanto riguarda età e tipologia di popolazione

I modelli predittivi sono stati disegnati a partire da coorti di individui selezionati nell'ambito di studi specifici o a partire da banche dati di pazienti. Tra i vari modelli analizzati esiste una leggera variazione per la validità del modello stesso in base all'età del paziente di cui si vuole calcolare il rischio. La scelta del modello più opportuno da usare deve quindi basarsi su quello che permette un'applicabilità al più ampio range possibile di età. Tutti i modelli predittivi di base coprono un ampio range di età e quindi sono ugualmente validi. Nello specifico, però il FINDRISC è utilizzabile per tutte le età al di sopra dei 18 anni. Il QDScore è l'unico che ha un limite superiore di validità di età corrispondente ad 84 anni.

Un altro elemento che determina l'applicabilità ad altri contesti è la nazionalità degli individui che hanno costituito la coorte. Gli autori che hanno definito i vari modelli hanno sperimentato i modelli in dataset di validazione ed hanno potuto verificare che il modello è applicabile anche per pazienti di altre nazionalità rispetto a quelle utilizzate per la definizione del modello. È interessante menzionare tra i vari modelli il QDScore che contiene un elemento in grado di personalizzare lo score in base alla nazione in cui viene applicato; tale elemento è l'indice di deprivazione che costituisce una misura dello svantaggio socioeconomico e che, nel caso del QDScore, è stato considerato come un fattore di rischio per l'insorgenza del diabete di tipo 2. L'indice di deprivazione è collegato al luogo in cui un paziente vive ed è possibile calcolarlo a partire dai dati di censimento disponibili in ciascuna nazione (nel capitolo 5 saranno forniti maggiori dettagli sulla possibilità di adattamento dell'indice di deprivazione al contesto italiano). Quindi, considerando i parametri dell'applicabilità, il QDScore sembra il modello più completo.

Capacità di discriminazione

La capacità di discriminazione dei vari modelli predittivi viene effettuata confrontando le AROC ossia l'"Area under the Receiving Operating Curve". I modelli predittivi che hanno una AROC superiore a 0.80 per il dataset di riferimento sono FINDRISC e QDScore. L'unico modello predittivo che ha una AROC maggiore di 0.80 anche nel dataset di validazione è il QDScore. Quindi tra i vari modelli, quello che appare più discriminate è il QDScore.

Fruibilità del modello da parte del pubblico

Dei quattro modelli predittivi di base presi in esame solo per il Cambridge Risk Score non è stato trovato alcun tool online che permette il calcolo dello score. Degli altri tre esiste per ciascuno di essi una pagina web all'interno della quale è possibile effettuare il calcolo dello score compilando i campi oppure rispondendo alle domande di un questionario. Solo per il QDScore è disponibile, oltre al calcolatore web, un'applicazione per windows e linux da linea di comando che permette il calcolo dello score. Di questo tool è anche disponibile il codice sorgente.

Alla luce delle considerazioni fin qui effettuate, il modello predittivo QDScore è quello che possiede le caratteristiche più appropriate che lo rendono adatto ad essere adottato all'interno del progetto.

3.2. Modelli predittivi per il rischio cardiovascolare

Le malattie cardiovascolari sono la principale causa di morte prematura e di invalidità. I pazienti asintomatici ritenuti ad alto rischio di malattie cardiovascolari devono essere identificati precocemente in

modo da poter ricevere consigli su cambiamenti dello stile di vita, come la cessazione del fumo, la pratica di attività fisica e la dieta, sul trattamento per abbassare la pressione sanguigna e per ridurre i livelli di colesterolo, etc. Molte linee guida raccomandano che il rischio di malattie cardiovascolari venga stimato combinando diversi fattori di rischio sotto forma di una stima numerica. Sono disponibili vari calcolatori del rischio, come grafici, tabelle, programmi informatici e strumenti basati sul web. Le equazioni derivate dallo studio di coorte americano di Framingham sono le più diffuse nel Regno Unito [26].

Diversi studi suggeriscono che i tassi di mortalità per le malattie cardiovascolari (CVD) delle persone con diabete di tipo 2 sono da due a quattro volte superiori a quelli della popolazione generale. È pertanto necessario sviluppare strumenti specifici per valutare e ridurre il rischio di CVD in questa popolazione di individui [27].

Le stime del rischio di CVD possono essere utili sia per i clinici che per i pazienti: per i medici è un'informazione prognostica che può supportarli nella scelta delle strategie terapeutiche e preventive; per i pazienti può essere uno strumento di motivazione ad adottare stili di vita sani e ad osservare le terapie prescritte che riducono i rischi.

In questo capitolo saranno analizzati i principali punteggi di rischio di CVD: dallo studio Framingham sviluppato negli anni' 60 fino agli ultimi modelli.

3.2.1. *Analisi di una selezione di modelli predittivi disponibili in letteratura*

Considerando che il diabete mellito comporta solitamente la coesistenza di diversi fattori di rischio cardiovascolare, sono stati sviluppati approcci multifattoriali per la valutazione del rischio di malattie cardiovascolari. I primi studi volti a sviluppare strumenti affidabili per la valutazione del rischio di malattie cardiovascolari basati su una combinazione di diversi fattori di rischio sono stati condotti negli Stati Uniti negli anni' 60 (Framingham score).

Alla fine degli anni' 90, uno studio della finlandese ha suggerito per la prima volta che le persone affette da diabete, ma senza storia di malattie cardiovascolari (CVD), presentavano un rischio di malattie cardiovascolari simile a quello delle persone senza diabete sopravvissute a un evento di CVD. In seguito a questa osservazione, diversi studi hanno sostenuto il concetto di diabete come "equivalente al rischio di malattia cardiovascolare", ossia si ritiene che la presenza del diabete mellito conferisca un rischio di CVD a 10 anni simile a quello di individui non affetti da diabete con una precedente storia di CVD. Questo approccio è stato confermato da algoritmi attualmente sviluppati sia da coorti americane che europee, come lo European Systematic COronary Risk Evaluation (SCORE) e il modello Prospective Cardiovascular Munster (PROCAM). In questi modelli, sia tutti i pazienti con diabete mellito che quelli con una patologia

cardiovascolare preesistente sono considerati persone ad alto rischio e vengono trattati come se avessero bisogno di una prevenzione secondaria per le patologie cardiovascolari [27].

3.2.1.1. *Score di Framingham*

Il Framingham Risk Score è un algoritmo utilizzato per stimare il rischio cardiovascolare a 10 anni di un individuo. Il punteggio è stato inizialmente sviluppato sulla base di uno studio di coorte a lungo termine della comunità ed è applicabile alla popolazione in generale. Uno dei principali limiti della versione iniziale di questo modello è che non considerava lo stato del diabete o qualsiasi altro indicatore di iperglicemia cronica. Negli anni successivi, lo stato del diabete è stato aggiunto al modello, ma solo come variabile di categoria e lo strumento è stato convalidato solo nella popolazione generale [27].
Il punteggio è valido per gli individui di età compresa tra i 30 e i 74 anni e senza malattie cardiovascolari all'esame di riferimento. Le variabili predittive utilizzate nell' ultima versione del punteggio sono:

- Sesso
- Età
- Diabete
- Fumo
- pressione sistolica trattata e non trattata
- Colesterolo totale
- Colesterolo HDL
- IMC (in sostituzione dei lipidi in una versione più semplice dello score) [28].

È disponibile online un calcolatore dello score alla pagina https://www.framinghamheartstudy.org/risk-functions/index.php.

3.2.1.2. *UK Prospective Diabetes Study (UKPDS) risk engine*

Il punteggio di rischio UK Prospective Diabetes Study (UKPDS) è stato inizialmente concepito per stimare separatamente il rischio di cardiopatie coronariche (CHD) e il rischio di ictus. È stato sviluppato su una coorte di 5102 pazienti con diabete di tipo 2 seguiti per circa 10 anni. Altri modelli per la valutazione del rischio di malattie cardiovascolari nella popolazione generale hanno utilizzato variabili "binarie" per la glicemia, ossia la presenza / assenza di diabete. L'approccio specifico per il diabete del modello UKPDS includeva per la prima volta l'emoglobina glicata (HbA1c) come variabile continua; ha anche sostituito l'età come fattore di rischio con due variabili specifiche per il diabete: l'età alla diagnosi di diabete e il tempo trascorso dalla diagnosi di diabete. I fattori di rischio inclusi nel primo modello UKPDS sono stati [27]:

- Età di diagnosi del diabete
- Sesso
- Gruppo etnico
- Fumo
- Emoblogina glicata HbA1c (%)
- Pressione Sistolica (mmHg)
- Rapporto Colesterolo totale / Colesterolo HDL

Nel 2007 è stato pubblicato un nuovo modello che stima direttamente il rischio di malattie cardiovascolari (definito come la prima insorgenza di infarto miocardico mortale o non fatale, morte cardiaca improvvisa, altre cardiopatie ischemiche, ictus mortale o non fatale o malattia vascolare periferica fatale). L'equazione di rischio sottostante è stata prima convalidata nella coorte Collaborative Atorvastatin Diabetes Study (CARDS), uno studio di prevenzione primaria comprendente 2838 pazienti con diabete, e poi nella European Prospective Investigation of Cancer (EPIC)-Norfolk Cohort, uno studio prospettico di coorte in cui i pazienti di età compresa tra i 40 e i 79 anni sono stati reclutati da medici generici nella regione di Norfolk nel Regno Unito. Un campione di 10137 pazienti, di cui 272 con diabete diagnosticato, è stato estratto da questa coorte per valutare le prestazioni dello score UKPDS rispetto alle equazioni di rischio di Framingham sia nella popolazione generale che in quella diabetica. Entrambe queste equazioni hanno funzionato ragionevolmente bene per identificare i pazienti con elevato rischio di malattie cardiovascolari. UKPDS ha ottenuto migliori risultati nella popolazione diabetica rispetto al punteggio di Framingham, anche se entrambe le equazioni hanno sovrastimato il rischio in questo gruppo di pazienti. Considerando che in studi precedenti il punteggio Framingham usato nella popolazione diabetica aveva sottostimato il rischio di malattie cardiovascolari, questi strumenti sembrano essere molto utili per supportare i medici nella gestione del trattamento, ma non del tutto adeguati per comunicare informazioni sui rischi ai pazienti.
E' possibile scaricare degli applicativi per Windows e MAC per il calcolo dello score UKPDS dalla seguente pagina web: https://www.dtu.ox.ac.uk/riskengine/download.php.

3.2.1.3. *Diabetes Audit and Research in Tayside, Scotland (DARTS)*

Per il modello Diabetes Audit and Research in Tayside, Scotland (DARTS) sono stati considerati dieci fattori di rischio utili per sviluppare un'equazione di rischio di malattie cardiovascolari; la principale misura di risultato per la sua convalida è stato il primo evento di cardiopatia coronarica importante (infarto miocardico acuto mortale o non fatale o morte da cardiopatia coronarica) [27].

Rispetto all'UKPDS, il modello scozzese non ha restrizioni di età, include l'indice di massa corporea (BMI), l'altezza, i trigliceridi e il trattamento antiipertensivo (variabili binarie: sì/no) e non tiene conto del background etnico.

3.2.1.4. *ADVANCE risk engine*

I modelli di rischio di malattie cardiovascolari Framingham e UKPDS sono stati convalidati su un ampio campione etnicamente diverso di pazienti affetti da diabete dallo studio di coorte "Action in Diabetes and Vascular disease: preterAx and diamicroN-MR Controlled Evaluation (ADVANCE)". La coorte era composta da 7168 persone con diabete senza storia di malattie cardiovascolari precedenti. Questo studio di convalida ha rivelato che il rischio assoluto a 4 anni di eventi cardiovascolari era sovrastimato da entrambi i modelli [27].

È stato quindi valutato un nuovo modello di previsione del rischio per migliorare le prestazioni in una coorte multietnica di pazienti. Lo score ADVANCE fornisce il rischio di avere un evento cardiovascolare in 4 anni; esso è stato progettato appositamente per le persone con diabete di tipo 2 ed è basato su fattori di rischio comunemente valutati nella pratica clinica di routine.

Nel modello di rischio ADVANCE sono stati inclusi dieci fattori di rischio: età alla diagnosi, sesso, durata del diabete, pressione cardiaca, retinopatia, fibrillazione atriale, emoglobina glicata HbA1c, log del rapporto albumina urinaria/creatinina, colesterolo e ipertensione trattata.

È disponibile online un calcolatore dello score alla pagina http://www.advanceriskengine.com/#page1_cv.

3.2.1.5. *Fremantle Australian risk score*

Nell'ambito del Fremantle Diabetes Study (FDS), in Australia, è stato condotto uno studio su una coorte di individui con diabete di tipo 2 non affetti da malattie cardiovascolari, per valutare le prestazioni degli score UKPDS e Framingham nella previsione delle patologie cardiovascolari a 5 anni. Entrambi questi algoritmi non hanno dato risultati soddisfacenti nella popolazione australiana oggetto di studio. Gli sperimentatori hanno quindi deciso di sviluppare e convalidare una funzione di rischio multi-variabile per la previsione del rischio cardiovascolare a 5 anni in questi pazienti. Il modello comprende diverse variabili normalmente disponibili nelle cure primarie: età, sesso, etnia, storia di malattie cardiovascolari, trattamento del diabete (dieta, agenti ipoglicemici orali, insulina), colesterolo HDL, emoglobina glicata (HbA1c), rapporto albumina urinaria / creatinina. A differenza del modello UKPDS, il modello australiano include le persone con diabete

con precedenti di malattie cardiovascolari, tiene conto del background etnico, non considera l'età alla diagnosi e la durata del diabete ma solo l'età attuale, ed esclude la pressione sanguigna sistolica o il trattamento antiipertensivo [27].

Presso il seguente link è possibile scaricare un foglio di calcolo Excel che permette il calcolo dello score: http://www.medpharm.uwa.edu.au/research/fremantle_diabetes_study/?a=2345213.

3.2.1.6. *New Zeland Diabetes cohort study*

Il modello neozelandese per la valutazione del rischio di CVD nelle persone con diabete è stato implementato dallo studio Diabetes Cohort Study (DCS), una coorte creata da un programma nazionale annuale di revisione delle cure primarie. Un totale di 36127 pazienti con diabete di tipo 2 senza precedenti malattie cardiovascolari è stato considerato per ricavare un'equazione per il rischio. I fattori di rischio presi in considerazione sono stati: età alla diagnosi, durata del diabete, sesso, pressione sanguigna sistolica, fumo, rapporto tra colesterolo totale e HDL, etnia, emoglobina glicata HbA1c e rapporto tra albumina e creatinina nelle urine. Una delle caratteristiche più importanti di questo modello è che è stato valutato su una coorte molto ampia di pazienti con diabete, grazie all'uso di dati raccolti di routine. Gli autori hanno sottolineato l'importanza di prendere in considerazione l'etnia come fattore di rischio, considerato un punto critico per ridurre le disuguaglianze sanitarie [27].

3.2.1.7. *The Swedish national diabetes register*

Il registro nazionale svedese del diabete è stato utilizzato per produrre un'equazione di previsione del rischio di malattie cardiovascolari a cinque anni nella popolazione diabetica. Lo studio si è basato su 11646 pazienti di sesso femminile e maschile, di età compresa tra 18 e 70 anni. I fattori di rischio considerati in questo studio longitudinale sono stati: età di insorgenza del diabete, durata del diabete, emoglobina glicata (HbA1c), IMC, pressione sanguigna sistolica, sesso, l'uso di farmaci antiipertensivi, di farmaci per la riduzione dei lipidi e l'abitudine al fumo. Un indicatore interessante utilizzato in questo score è il trattamento con i farmaci per abbassare il livello di lipidi, ma il limite principale è che non ha incluso la valutazione della funzione renale [27].

3.2.1.8. *QRISK score*

Nel 2007 è stato pubblicato il primo modello QRISK per stimare il rischio di malattie cardiovascolari a 10 anni. Lo studio QRISK, basato su dati raccolti dalla medicina generale del Regno Unito, mirava a sviluppare un nuovo algoritmo per stimare il rischio decennale di patologiche cardiovascolari (cardiache o cerebrovascolari) e per convalidare le sue prestazioni rispetto all'algoritmi di Framingham. La coorte di derivazione del QRISK consisteva in persone esenti da diabete e patologie cardiovascolari pregresse [26]. QRISK è stato seguito da un modello aggiornato (QRISK2) nel 2008, che includeva l'origine etnica e fattori di rischio aggiuntivi (diabete di tipo 2, artrite reumatoide, fibrillazione atriale e malattie renali croniche). Da allora, QRISK2 è stato aggiornato ogni anno e ricalibrato all'ultima versione del database QResearch, da cui i dati sono stati estratti.

Il QRISK2 si basava su una coorte aperta di 2,29 milioni di pazienti provenienti da studi nel Regno Unito e appartenenti a diversi gruppi etnici: la fascia d'età in cui si applica è stata estesa da 35-74 anni a 25-84 anni, il diabete di tipo 1 è stato incluso come variabile separata, il fumo è stato valutato a cinque livelli invece che a due e il punteggio di Townsend (per la deprivazione sociale) è stato aggiornato utilizzando i valori più recenti del censimento del 2011; ciò contribuisce a garantire che gli algoritmi riflettano i cambiamenti nelle caratteristiche della popolazione [29].

I fattori di rischio considerati nel QRISK2 sono:

Etnia autovalutata

- Età (25-84 anni)
- Sesso
- Codice postale (per determinare il punteggio di deprivazione)
- Fumo
- Stato del diabete
- Anamnesi familiare di cardiopatia coronarica in un parente di primo grado sotto i 60 anni
- Malattia renale cronica
- Fibrillazione atriale
- Pressione sanguigna sistolica
- Trattamento antipertensivo
- Artrite Reumatoide
- Rapporto Colesterolo totale / Colesterolo HDL
- Indice di massa corporea

Nel 2014 è stata pubblicata la linea guida NICE (National Institute for health and Clinical Excellence) sulla valutazione del rischio cardiovascolare. Questa linea guida ha evidenziato una serie di condizioni associate

ad un aumento del rischio cardiovascolare che potrebbero non essere completamente colte dal QRISK2, tra cui l'HIV/AIDS, le malattie renali di stadio 3, il lupus eritematoso sistemico, la concomitanza di gravi malattie mentali e l'uso di antipsicotici atipici o corticosteroidi. Queste condizioni non sono specificamente identificate all'interno del QRISK2 e ciò potrebbe portare a sottostimare il rischio nei gruppi di pazienti interessati. Inoltre una ricerca pubblicata di recente ha evidenziato un aumento del rischio cardiovascolare e una potenziale importanza prognostica della disfunzione erettile, dell'emicrania e della variabilità della pressione sanguigna. Per far fronte a queste sfide, è stata derivata e validata una nuova versione dell'algoritmo, il QRISK3, per determinare se questi fattori debbano essere incorporati nell'algoritmo per migliorare la stima del rischio cardiovascolare per questi pazienti [29].

QRISK3 include più fattori rispetto a QRISK2 per aiutare i medici a identificare quelli più a rischio di malattie cardiache e ictus. Questi sono:

- Malattia renale cronica
- Emicrania
- Uso di corticosteroidi
- Lupus eritematoso sistemico (SLE)
- Uso di antipsicotici atipici
- Malattia mentale grave
- Disfunzione erettile
- Misura della variabilità della pressione sanguigna sistolica

Un calcolatore web per il punteggio QRISK3 è disponibile alla seguente pagina web: https://qrisk.org/three/. E' anche possibile scaricare una versione a riga di comando dello stesso strumento. Il punteggio QRISK2 aggiornato alla versione 2017 è disponibile alla seguente pagina web: https://qrisk.org/2017/.

3.2.1.9. *SCORE charts (SCORE Project)*

La Società europea di cardiologia e la Second Joint Task Force hanno promosso lo sviluppo di un sistema di stima dei rischi basato su un ampio gruppo di insiemi di dati europei rappresentativi, in grado di rilevare la variazione regionale del rischio cardiovascolare. Ciò ha portato alla creazione del progetto SCORE (Systematic COronary Risk Evaluation) come progetto di azione concertata europea finanziato nell'ambito del programma BIOMED dell'Unione europea. L'obiettivo del progetto SCORE è quello di sviluppare un sistema di stima del rischio per la pratica clinica in Europa. Questo processo si articola in tre fasi: in primo luogo, l'elaborazione di semplici grafici di rischio cartacei per le popolazioni europee ad alto e basso rischio; in secondo luogo, lo sviluppo di metodi per la creazione di grafici di rischio nazionali o regionali basati su

dati pubblicati sulla mortalità; infine, l'integrazione della stima del rischio in un'applicazione informatica per la gestione dei fattori di rischio [30].

Il modello SCORE è applicabile nella prevenzione primaria e stima il rischio decennale di un evento cardiovascolare mortale. I fattori di rischio inclusi nel modello sono:

- Età (lo score è più dettagliato nel range di età 50-65 anni; i grafici vengono forniti però per il range 40 - 65)
- Sesso
- Pressione sistolica
- Colesterolo totale
- Colesterolo HDL
- Fumo

Per valutare il punteggio sono state approntate due coppie di grafici: una che tiene in considerazione il colesterolo totale come fattore di rischio e un'altra che tiene in considerazione il rapporto colesterolo totale /colesterolo HDL (la Figura 2 mostra solo i grafici che considerano il colesterolo totale come fattore di rischio). Il rischio viene letto arrotondando l'età della persona all'età più vicina indicata sul grafico, il colesterolo totale o il rapporto colesterolo totale/colesterolo HDL all'unità intera più vicina e la pressione sanguigna al multiplo di 20 mmHg più vicino.

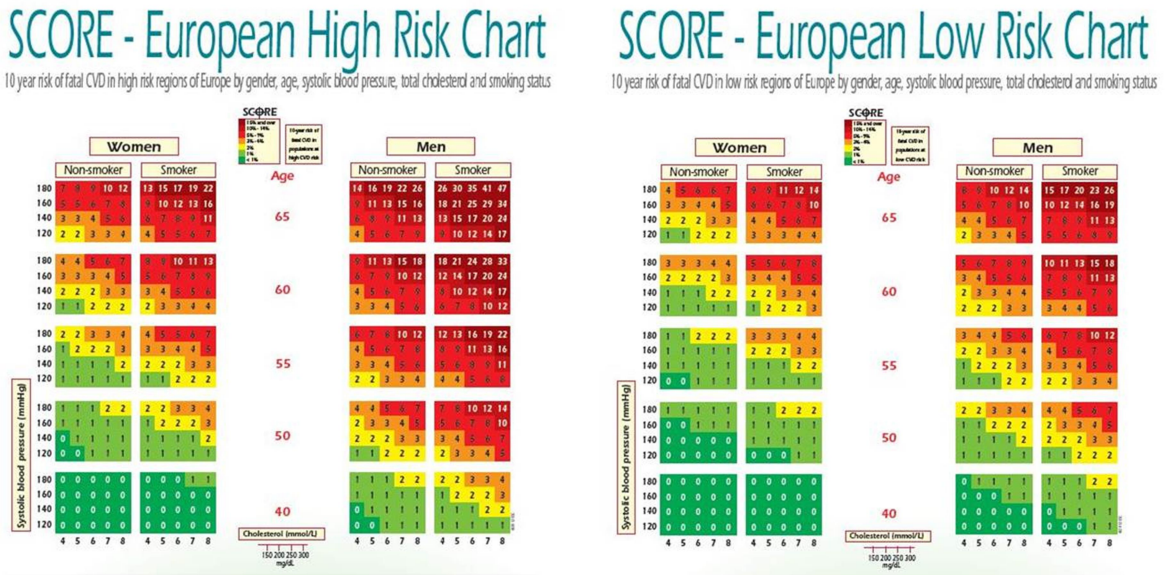

Figura 2 Grafici SCORE per il rischio di malattie cardiovascolari per popolazioni europee ad alto e basso rischio basati sui valori del colesterolo totale

(Fonte: https://www.escardio.org/static_file/Escardio/Subspecialty/EACPR/Documents/score-charts.pdf)

Poiché l'età è un fattore determinante del rischio coronarico e le fasce d' età delle coorti sono piuttosto eterogenee, i ricercatori hanno limitato il calcolo del modello per adattarlo alla fascia d'età da 45 a 64 anni. Le prestazioni dei grafici basati sul colesterolo e sul rapporto colesterolo totale / colesterolo HDL sono molto simili.

SCORE è finalizzato alla stima del rischio cardiovascolare totale piuttosto che del rischio di cardiopatia coronarica. Tale approccio è stato adottato poiché dovrebbe fornire una migliore stima del rischio per la persona ed anche una migliore riflessione sulle implicazioni per il servizio sanitario dei fattori di rischio cardiovascolare. Le malattie cardiovascolari non coronariche sono importanti perché rappresentano una proporzione maggiore di tutti i rischi cardiovascolari nelle regioni europee con bassi tassi di cardiopatie coronariche. Il metodo adottato calcola il rischio totale in due parti, la componente coronarica e la componente non coronarica, consentendo di effettuare calcoli sulle conseguenze del trattamento.

Il progetto SCORE si concentra sulla stima del rischio su eventi mortali di malattie cardiovascolari, anziché su eventi fatali e non mortali combinati. Sia i pazienti che i medici sono interessati sia agli eventi non fatali che alle malattie cardiovascolari mortali; inoltre la morbilità e l'invalidità causate da eventi di malattie cardiovascolari non fatali rappresentano il maggiore onere economico per il sistema sanitario e la società. Gli eventi di malattie cardiovascolari non fatali pongono tuttavia una serie di problemi per lo sviluppo dei sistemi di stima del rischio, in quanto dipendono in modo significativo dalle definizioni e dai metodi utilizzati per la loro determinazione.

Il diabete è notoriamente associato ad un marcato aumento del rischio di malattie cardiovascolari; tra le persone con diabete di tipo 2 l'aumento del rischio di malattie cardiovascolari è almeno 2 volte superiore negli uomini e addirittura 4 volte maggiore nelle donne. I principali fattori di rischio "convenzionali" sono noti per avere una relazione quasi simile con il rischio di malattie cardiovascolari in persone diabetiche e non diabetiche, ma ad ogni livello di fattore di rischio o con qualsiasi combinazione di fattori di rischio, il rischio assoluto è maggiore nelle persone diabetiche. Questo rischio superiore nelle persone diabetiche è in parte spiegato da fattori "correlati al diabete", come il tipo e la durata del diabete, il controllo glicemico e la presenza di retinopatia o microalbuminuria o proteinuria. Il gruppo United Kingdom Prospective Diabetes Study ha già sviluppato modelli matematici per la stima del rischio assoluto di cardiopatie coronariche e del rischio di ictus negli uomini e nelle donne con diabete di tipo 2 di nuova diagnosi. I grafici di rischio SCORE possono tuttavia essere utilizzati per una valutazione approssimativa del rischio cardiovascolare nelle persone diabetiche, perché il rapporto dei principali fattori di rischio "convenzionali" con le malattie cardiovascolari è quasi parallelo nei soggetti diabetici e non diabetici, anche se il rischio dei soggetti diabetici è molto più elevato.

Una versione elettronica, interattiva e migliorata di SCORE è disponibile per il download (previa registrazione) all'indirizzo http://www.heartscore.org/en_GB/download [27]. I grafici SCORE, invece, sono

disponibili alla seguente pagina web: https://www.escardio.org/Education/Practice-Tools/CVD-prevention-toolbox/SCORE-Risk-Charts.

3.2.1.10. *Progetto Cuore*

Il modello Cuore stima il rischio decennale di evento cardiovascolare mortale e non fatale nella prevenzione primaria.

La versione italiana delle Linee Guida per la Prevenzione delle malattie cardiovascolari ha indicato che per l'Italia è preferibile utilizzare gli strumenti del Progetto CUORE dell'Istituto Nazionale della Salute per stimare il rischio di tali malattie nella popolazione in generale. Diversi studi sulle coorti italiane hanno messo a confronto gli strumenti di SCORE e CUORE: uno studio pubblicato nel 2010 ha evidenziato che il primo riflette abbastanza bene la mortalità cardiovascolare italiana e, di conseguenza, le coorti italiane del Progetto CUORE sono abbastanza rappresentative dei paesi europei a basso rischio di mortalità cardiovascolare. Altri studi hanno riscontrato un moderato livello di convergenza tra i due strumenti, in particolare per quanto riguarda la discriminazione dei soggetti ad alto rischio.

Il punteggio di rischio cardiovascolare globale del Progetto CUORE è stato sviluppato utilizzando i dati di diverse coorti arruolate nel nord, centro e sud Italia tra gli anni '80 e '90, i cui fattori di rischio sono stati raccolti utilizzando procedure standardizzate. Il punteggio individuale (disponibile sul sito web www.cuore.iss.it) può essere calcolato in soggetti di età compresa tra i 35 e i 69 anni e prevede il rischio decennale di un primo evento cardiovascolare mortale e non fatale sulla base dei seguenti fattori di rischio:

- Età (i grafici vengono forniti per il range 40 – 69; il range di età valido per lo score individuale è 35-69)
- Sesso
- Pressione sistolica
- Trattamento antipertensivo (valido solo per il calcolo dello score individuale)
- Colesterolo totale
- Colesterolo HDL (valido solo per il calcolo dello score individuale)
- Fumo
- Diabete

Oltre al punteggio individuale, sono disponibili anche i grafici di rischio, che differiscono per il numero dei fattori di rischio utilizzati e per la fascia d' età in cui sono applicati e per l'accuratezza delle informazioni. Secondo questo strumento, gli individui sono considerati ad alto rischio se il loro rischio di malattie cardiovascolari decennale è pari o superiore al 20%; invece, sono a basso rischio se è <5%.

Diversamente dal Progetto CUORE, i grafici di rischio SCORE non includono una variabile binaria che tiene conto della presenza di diabete nella funzione di rischio, perché c'è stata una mancanza di uniformità nell'accertamento del diabete. Inoltre, le istruzioni per l'uso di questi grafici indicano che essi sottostimano il rischio nei pazienti diabetici e dovrebbero essere usati solo in pazienti con diabete di tipo 1 senza danni agli organi target. Va sottolineato che le linee guida ESC (European Society of Cardiology) sulla Prevenzione delle malattie cardiovascolari, anche nelle sue versioni per diabete, pre-diabete e malattie cardiovascolari, classifica i soggetti con diabete mellito (sia di tipo 1 che di tipo 2) a rischio elevato o molto elevato, a seconda che presentino uno o più fattori di rischio cardiovascolare e/o danni agli organi bersaglio.
Il "Piano Nazionale di Prevenzione" 2005-2007 e 2010-2012 prevedeva una valutazione del rischio cardiovascolare decennale utilizzando gli strumenti del Progetto CUORE e ciò ha permesso la realizzazione di un programma nazionale di formazione per i medici generici.
Le linee guida e gli studi su questi temi sottolineano la necessità di aggiornare i grafici e i punteggi per la valutazione dei rischi, dal momento che il profilo di rischio della popolazione e la mortalità per le malattie cardiovascolari cambiano nel tempo. È quindi necessario iscrivere nuove coorti o aggiornare quelle esistenti. In quest'ottica, il Progetto CUORE ha ampliato il database dei fattori di rischio cardiovascolare aggiungendo la coorte dell'Osservatorio Epidemiologico Cardiovascolare (OEC). Le analisi preliminari hanno confermato il ruolo predittivo dei principali fattori di rischio già inclusi nell'algoritmo, ma ulteriori studi dovrebbero essere condotti per valutare l'inclusione di nuovi fattori di rischio, principalmente la glicemia.
Nel recente passato, la valutazione del rischio cardiovascolare ottenuta attraverso gli strumenti del Progetto CUORE non è stata quasi mai applicata, nonostante le raccomandazioni del Ministero della Salute per la valutazione del rischio cardiovascolare per il rimborso delle statine nella prevenzione primaria [27].

3.2.1.11. *RISKARD score*

Un altro strumento per la valutazione del rischio cardiovascolare sviluppato in Italia dal Gruppo di Ricerca per la Valutazione del Rischio Cardiovascolare è lo score Riskard, sviluppato sia sotto forma di grafico dei rischi che sotto forma di software denominato Riskard 2002. Lo studio ha iscritto una coorte di oltre 9000 persone provenienti da tre studi italiani, di età compresa tra i 35 e i 74 anni. Negli anni successivi, il grafico e il software sono stati ulteriormente aggiornati (Riskard 2005), con l'obiettivo di offrire strumenti basati su popolazioni più ampie e diversificate, un maggior numero di individui esposti a rischi ed eventi, periodi di follow-up più lunghi e altri concetti innovativi. Il grafico, per sua natura, può incorporare solo un numero limitato di fattori di rischio, mentre molti altri possono essere inclusi nel software. I dati sono stati raccolti da nove studi sulla popolazione in otto regioni italiane, per un totale di 17153 soggetti. Il nuovo grafico che permette di stimare il rischio decennale di subire un primo evento cardiovascolare è stato prodotto per

uomini e donne di età compresa tra i 45 e i 74 anni, esenti da patologie cardiovascolari. Sono state elaborate stime per il rischio assoluto e per il rischio relativo per le misure compensative, quest'ultimo rispetto ai livelli previsti nella popolazione in generale [27].

Il software, invece, è stato realizzato per prevedere il rischio di eventi coronarici, cerebrovascolari e cardiovascolari di rilievo per il follow-up a 5, 10 o 15 anni, in uomini e donne di età compresa tra i 35 e i 74 anni non affetti da malattie cardiovascolari. I fattori di rischio presi in considerazione nel software Riskard 2005 sono stati sesso, età, IMC, pressione sanguigna media fisiologica, colesterolo HDL, colesterolo non-HDL, fumo di sigaretta, diabete e frequenza cardiaca. I risultati hanno prodotto diversi indicatori, quali il rischio assoluto, il rischio relativo, il rischio ideale (per un profilo di rischio molto favorevole) e l'età biologica del rischio.

3.2.1.12. *QuED score*

Negli ultimi anni, la pressione esercitata sui sistemi sanitari affinché fornissero cure di alta qualità e controllassero i costi è progressivamente aumentata. In questo contesto è stato sviluppato lo studio "Qualità ed Esito in Diabetologia" (QuED), il cui obiettivo principale è stato quello di sviluppare un punteggio per la valutazione della qualità dell'assistenza diabetologica. In questo studio è stata documentata la stretta relazione tra un punteggio di qualità della cura del diabete e i risultati a lungo termine anche sul rischio cardiovascolare. Lo studio QuED è stato condotto in Italia dal Consorzio Mario Negri Sud-Dipartimento di Farmacologia Clinica ed Epidemiologia, in collaborazione con il Centro per la Ricerca sulle Politiche Sanitarie - Università della California, Irvine. Complessivamente, 101 cliniche ambulatoriali e 103 medici generici di tutte le regioni d'Italia hanno partecipato allo studio, registrando rispettivamente 2448 e 785 pazienti con diabete di tipo 2 indipendentemente dall'età, dalla durata o dal trattamento. Il punteggio è stato calcolato utilizzando indicatori di processo e risultati intermedi prontamente disponibili (HbA1c, pressione sanguigna, colesterolo LDL, microalbuminuria). Nella costruzione del modello, è stato attribuito un punteggio discreto ad ogni variabile come segue [27]:

- 0 punti se il paziente non è stato trattato per la condizione specifica nonostante i valori elevati, o se il paziente ha mostrato valori insoddisfacenti nonostante il trattamento;
- punti se la misurazione di un parametro non è stata effettuata negli ultimi 12 mesi;
- 10 punti se sono stati raggiunti gli obiettivi desiderati.

Questo punteggio variava da 0 a 40, con un punteggio più alto che rispecchiava una migliore qualità delle cure. Questo punteggio è stato definito anche come Q-Score.

Lo studio mostra che il rischio di sviluppare un evento cardiovascolare era superiore dell'89% nei pazienti con un punteggio ≤10 e del 43% in quelli con un punteggio tra 10 e 20, rispetto ai pazienti con un punteggio >20. È anche importante sottolineare che è stata documentata una relazione lineare tra il punteggio QuED e l'incidenza di eventi cardiovascolari sia in pazienti con che senza precedenti eventi cardiovascolari, e questo è vero per diversi esiti, come CHD, eventi cerebrovascolari, malattie vascolari periferiche e mortalità cardiovascolare. Il punteggio QuED non è destinato a sostituire altri modelli di rischio utilizzati per prevedere il rischio individuale di eventi cardiovascolari, ma piuttosto un indicatore della qualità media dell'assistenza per il diabete fornita da una specifica impostazione nel sistema sanitario.

3.2.2. *Confronto fra i modelli predittivi disponibili*

Nel presente capitolo sono stati analizzati i principali punteggi di rischio per malattie cardiovascolari, a partire dal primo sviluppato negli anni' 60, lo studio Framingham, fino agli ultimi modelli specifici per il diabete. Negli ultimi anni, il diabete è incluso nella maggior parte dei modelli, anche se, nella maggior parte di essi, sotto forma di variabile binaria che valuta l'assenza o la presenza della malattia. Il livello di HbA1c o di glicemia sembra essere un indicatore importante per la valutazione del rischio di malattie cardiovascolari sia nella popolazione diabetica che in quella generale. In generale, gli studi più recenti concordano sull'importanza di utilizzare coorti specifiche di pazienti diabetici per valutare adeguatamente i modelli di rischio di malattie cardiovascolari nella popolazione. Inoltre, le strutture di cura del diabete e il gruppo etnico devono essere presi in considerazione nello sviluppo e nella valutazione del rischio di malattie cardiovascolari nei pazienti diabetici. I dati della letteratura evidenziano che i modelli di previsione del rischio cardiovascolare non funzionano in modo ottimale nelle persone con diabete di gruppi etnici diversi. Pertanto, può essere opportuno che ciascun paese confermi i modelli esistenti sulle proprie popolazioni e li adatti alle diverse situazioni, per migliorare la gestione mirata dei rischi [27].

Le ultime linee guida della European Society of Cardiology (ESC) sulla prevenzione delle malattie cardiovascolari nella pratica clinica raccomandano la stima del rischio totale utilizzando molteplici fattori di rischio, come il modello SCORE, per gli adulti asintomatici senza evidenze di malattie cardiovascolari. Sempre nel contesto europeo, i recenti aggiornamenti delle linee guida sulla alterazione dei lipidi e la valutazione del rischio cardiovascolare del National Institute for Health and Care Excellence (NICE) nel Regno Unito raccomandano l'utilizzo di un nuovo strumento per la valutazione del rischio delle malattie cardiovascolari nella prevenzione primaria fino all' età di 84 anni e anche per le persone con diabete di tipo 2: il QRISK2 [27].

In una recente review che ha messo a confronto le prestazioni dei modelli di rischio di malattie cardiovascolari sviluppate esclusivamente su persone con diabete e quelle sviluppate sulla popolazione generale, è stato inserito anche uno studio condotto in Italia in tre cliniche di Modena. Lo studio ha coinvolto 1532 pazienti e mirava a valutare l'accuratezza prognostica di quattro algoritmi utilizzati per stimare il rischio cardiovascolare: l'UKPDS (costituito da una popolazione di persone con diabete di tipo 2), il Framingham e due algoritmi italiani, il Riskard e il CUORE.

Il modello di Framingham mostra una sovrastima del rischio nella popolazione italiana; inoltre, la funzione proposta dallo studio UKPDS, quando applicata alla popolazione diabetica italiana, tende a sopravvalutare il rischio cardiovascolare, senza differenze tra i sessi.

Tenendo conto della situazione italiana, la funzione più ampiamente adottata, l'algoritmo CUORE, sottovaluta il rischio cardiovascolare, in particolare nelle donne. L' algoritmo di Riskard appare più coerente con la valutazione del rischio cardiovascolare nella coorte esaminata; tuttavia, suddividendo la popolazione per sesso, emerge una sovrastima degli eventi negli uomini e una grande sottovalutazione nelle donne. Quindi, questi algoritmi sono concordi nel sovrastimare il rischio negli uomini e nel sottovalutare il rischio nelle donne.

Le conclusioni sono che la stima del rischio cardiovascolare dipende dall'algoritmo utilizzato e dal rischio di base della coorte di riferimento, e per aumentare le prestazioni di tali funzioni è opportuno adottare funzioni concepite per una popolazione specifica, comprese le variabili di rischio proprie del diabete.

3.2.3. *Selezione del modello predittivo più utile*

Alla luce del confronto effettuato nel paragrafo precedente, il modello predittivo che appare più utile da utilizzare ai fini del progetto è il QRISK2. Le motivazioni che concorrono alla scelta di tale modello sono:

1. Il modello è stato suggerito nelle recenti linee guida del NICE (National Institute for Health and Care Excellence)
2. Le variabili utilizzate per il calcolo dello score sono di facile reperibilità sia da parte del medico che del paziente
3. Una delle sue variabili, l'indice di deprivazione, è adattabile ad un contesto regionale differente da quello in cui è stato sviluppato.
4. È disponibile un tool da linea di comando che permette l'integrazione del modello in applicazioni informatiche più complesse

3.3. Adattabilità dei modelli al contesto italiano

Uno dei parametri utilizzati negli score QDiabetes e QRISK è l'indice di deprivazione, un indicatore che permette di valutare la misura della deprivazione materiale e sociale dell'individuo soggetto ad indagine. Tale indice è composto da 5 indicatori di svantaggio [31]:

- x1: % di popolazione con istruzione pari o inferiore alla licenza elementare (ossia mancato raggiungimento dell'obbligo scolastico)
- x2: % di popolazione attiva disoccupata o in cerca di prima occupazione
- x3: % di abitazioni occupate in affitto (ossia % di famiglie che vivono in una casa non di proprietà)
- x4: % di famiglie monogenitoriali con figli dipendenti conviventi
- x5: densità abitativa (numero di occupanti per 100 m2 nelle abitazioni)

L'indice di deprivazione è la somma di ciascuno dei 5 indicatori di svantaggio standardizzato (Z-score).

La standardizzazione è un procedimento che riconduce una variabile aleatoria distribuita secondo una media μ e varianza σ^2, ad una variabile aleatoria con distribuzione "standard", ossia di media zero e varianza pari a 1.

Per ottenere la standardizzazione, il procedimento prevede di sottrarre alla variabile aleatoria la sua media e dividere il tutto per la deviazione standard, ovvero utilizzando la formula utile a trovare i punti zeta (Z-score o standard score):

$$Z = \frac{x - \mu}{\sigma}$$

Quindi se Z_{x_i} è lo Z-score dell'indicatore di svantaggio xi, l'indice di deprivazione I è dato dalla somma degli Z-score dei 5 indicatori di svantaggio, ossia:

$$I = Z_{X_1} + Z_{X_2} + Z_{X_3} + Z_{X_4} + Z_{X_5}$$

L'indice di deprivazione è strettamente legato al luogo in cui si trova l'individuo; per cui, per determinare l'indice di deprivazione deve essere fornita la località in cui vive l'individuo; tale località nel caso dell'indice di deprivazione calcolato nel Regno Unito corrisponde al codice postale.

L'indice di deprivazione è adattabile alle varie nazioni se per ciascuna nazione esistono i dati di censimento relativi agli indicatori di svantaggio riportati precedentemente.

Nel caso dell'Italia è possibile ricavare i 5 indicatori dal censimento effettuato nel 2011 i cui dati sono disponibili sul sito web dell'ISTAT (http://dati-censimentopopolazione.istat.it/Index.aspx).

I corrispondenti indicatori di svantaggio disponibili per il censimento ISTAT del 2011 sono:

- x1: indice di non conseguimento della scuola del primo ciclo

- x2: % di popolazione attiva disoccupata o in cerca di prima occupazione (dato dal rapporto tra numero di persone in cerca di occupazione e forza lavoro)
- x3: % di famiglie in affitto
- x4: % di famiglie mono-genitoriali con figli dipendenti conviventi (dato dal rapporto della somma di famiglie costituite da padre con figli e madre con figli sul numero di nuclei familiari)
- x5: densità abitativa (dato dalla proporzione della superficie a disposizione per occupante su 100 m2)

Ciascuno degli indicatori di svantaggio è disponibile per ciascuno dei circa 8000 comuni italiani (dove i dati sono disponibili). Nel caso in cui per l'individuo di cui si vuole calcolare l'indice di deprivazione non fosse noto il comune di residenza, è possibile calcolarlo in base alla provincia o regione di residenza. Nel caso in cui tali informazioni fossero mancanti, l'indice di deprivazione calcolato sarà quello medio italiano.
I Z-score dei 5 indicatori di svantaggio ottenuti dai dati del censimento ISTAT del 2011 sono stati calcolati usando 5 differenti fogli excel allegati a questo documento.

3.4. Considerazioni sui modelli predittivi analizzati

Nei paesi sviluppati l'analisi predittiva è la prossima grande opportunità di evoluzione per la medicina.
I pazienti dovranno essere meglio informati e assumersi maggiori responsabilità per la propria assistenza, se vogliono utilizzare le informazioni ottenute. Il ruolo del medico probabilmente cambierà e fungerà più da consulente che da decision maker, che consiglierà, avvertirà e aiuterà i singoli pazienti. I medici potranno provare più soddisfazione nella pratica man mano che i risultati positivi aumentano e gli esiti negativi diminuiscono. Il tempo di interazione con i singoli pazienti aumenterà e i medici potranno avere ancora una volta il tempo di creare relazioni positive e durature con i loro pazienti. Anche gli ospedali, le aziende farmaceutiche e i fornitori di assicurazioni vedranno cambiamenti. Per esempio, ci possono essere meno ricoveri non necessari, con conseguente iniziale diminuzione dei ricavi. Nel tempo, tuttavia, le ammissioni saranno più significative, il mercato si adeguerà e i risultati aumenteranno. Inizialmente, i ricavi possono anche essere persi dalle aziende farmaceutiche e dalle aziende che producono dispositivi, ma poi offerte più specializzate e individualizzate permetteranno di aumentare i profitti.
Questi cambiamenti possono letteralmente rivoluzionare il modo in cui la medicina viene praticata per migliorare la salute e ridurre le malattie [1].

4. Analisi delle fonti dati pubbliche contenenti informazioni utili alla correlazione

I gruppi di ricerca che si occupano di effettuare statistiche sullo stato di salute della popolazione o che si occupano di individuare modelli predittivi per il rischio, ad esempio, di patologie croniche, utilizzano i dati necessari per le loro analisi e ricerche a partire da due tipologie di fonti dati:

1. Studi condotti a livello nazionale o dai gruppi scientifici stessi
2. Banche dati sanitarie generalizzate

La prima tipologia è frequentemente utilizzata per scopi scientifici: ogni gruppo di ricerca definisce delle coorti (ossia "un insieme di individui, facenti parte di una popolazione, che sono caratterizzati dall'aver sperimentato uno stesso evento in un periodo predefinito") su cui effettuerà le verifiche delle ipotesi alla base della ricerca. Dato che tali informazioni sono utilizzate per rispondere a specifici obiettivi di ricerca, i dati disponibili variano parecchio in termini di tipologia e numero. Inoltre, il numero di individui che costituiscono le coorti è limitato e raramente supera alcune decina di migliaia di soggetti.

La seconda tipologia, invece, è caratterizzata da dati estratti generalmente dalle cartelle cliniche compilate all'interno delle strutture sanitarie. Le informazioni disponibili sono più omogenee rispetto a quelle disponibili per gli studi, ma il vantaggio principale è che essendo tali dati raccolti nel corso della normale attività di assistenza sanitaria svolta dalle strutture sanitarie, è possibile avere a disposizione un gran numero di individui coinvolti (dell'ordine di qualche decina di milioni di pazienti) e relativi ad un ampio intervallo temporale.

Per la costruzione di modelli predittivi sono utili, oltre alle informazioni più strettamente cliniche appena citate, anche quelle informazioni messe a disposizione da banche dati specializzate utilizzate in ambito di ricerca; esempi di tali banche dati sono quelle che riportano informazioni di correlazione tra varianti genetiche e patologie o tra patologie e stili di vita o esposizione a sostanze chimiche, etc.

All'interno del presente capitolo saranno descritti i database pubblici individuati e considerati come potenzialmente utili all'attività di definizione di modelli predittivi per l'insorgenza di patologie croniche e per l'individuazione di correlazioni tra la patologia ed altri fattori (ambientali, chimici, genetici, etc.).

4.1. Database pubblici individuati

4.1.1. *THIN Database*

"The Health Improvement Network" (THIN) database rappresenta una collaborazione tra due aziende: In Practice Systems (INPS), che ha sviluppato il software Vision utilizzato dai medici di medicina generale in Inghilterra per gestire i dati dei pazienti e IMS Health che ha poi fornito l'accesso ai dati per l'uso nella ricerca medica. I dati del database THIN sono raccolti durante la pratica medica di routine e sono periodicamente inviati al network. Gli studi di ricerca per la pubblicazione condotti utilizzando i dati di THIN sono approvati da un comitato etico accreditato a livello nazionale. Per l'utilizzo del database è necessaria una licenza a pagamento [3].

Caratteristiche

I dati di THIN attualmente contengono dati di EMR (Electronic Medical Record) di 11.1 milioni di pazienti (3.7 milioni di pazienti attivi) provenienti da 562 medici di medicina generale in Gran Bretagna, coprendo il 6.2% della popolazione. I pazienti sono rappresentativi della popolazione UK per età, sesso, condizioni mediche e livelli di mortalità adeguati in base alla demografia e alla deprivazione sociale. I dati raccolti sono completamente anonimizzati, sono costantemente aggiornati e possono essere seguiti durante la loro evoluzione nel tempo.

Struttura dei dati

I dati dei pazienti sono strutturati in file standardizzati, uno per ogni pratica. Le diagnosi sono codificate nella codifica READ in maniera tale da permettere la standardizzazione del modo con cui le informazioni sono registrate. I codici delle diagnosi sono raggruppati in capitoli tematici (e.g. cancro) e includono termini relative a sintomi, diagnosi, procedure e test di laboratorio. Le prescrizioni sono inserite usando i codici Multilex rilasciati da First Databank; tali codici possono essere facilmente collegati ai codici del British National Formulary (BNF). La maggior parte delle pratiche hanno anche dati di laboratorio automaticamente scaricate nell'EMR (Electronical Medical Record).

I principali file di dati contenuti in THIN sono:

1. PATIENT: dati demografici come età, sesso, data di registrazione della pratica medica, data di chiusura della pratica medica, data di morte.
2. MEDICAL: dati clinici come diagnosi mediche, rinvii a ospedali e specialisti, note a testo libero. Per ognuno di questi eventi è disponibile anche l'informazione sulla data e sul luogo (e.g., studio del medico, ospedale, etc.).
3. THERAPY: tutte le prescrizioni con la data di rilascio, la formulazione, la quantità e le istruzioni di dosaggio, l'indicazione per il trattamento di tutte le nuove prescrizioni (dedotte dal riferimento

trasversale agli eventi medici con la stessa data) e gli eventi che hanno portato al ritiro di un farmaco o di un trattamento.

4. ADDITIONAL HEALTH DATA (AHD): vaccinazioni e contraccettivi di prescrizione; varie informazioni quali fumo, altezza, peso, immunizzazioni, gravidanza, nascita, morte e risultati di laboratorio.
5. POSTCODE VARIABLE INDICATORS (PVI): il codice di avviamento postale è collegato agli indici socioeconomici, etnici e ambientali
6. CONSULTATION: data, ora e durata della consultazione
7. STAFF: sesso e ruolo del personale che ha inserito i dati

Ogni voce del file medico di un paziente può essere associata a un commento in quanto il sistema Vision consente l'immissione di testo libero o informazioni scansionate.

Può essere possibile ottenere informazioni ulteriori sul paziente attraverso il servizio Additional Information Service (AIS) che include:

- questionari anonimi completati dal paziente o dal medico curante
- copie della corrispondenza basata sul paziente
- un intervento specificato (ad esempio un test di laboratorio per confermare la diagnosi)
- certificati di morte

Il costo di utilizzo di AIS varia a seconda del tipo e della quantità di informazioni richieste e richiederà l'approvazione completa del MREC (Multicentre Research Ethics Committee).

Punti di forza e di debolezza

Il database THIN ha punti di forza e debolezza:

Punti di forza

- È un insieme di dati molto grande contenente record di circa il 5,7% della popolazione britannica.
- Fornisce una fonte informativa rappresentativa della popolazione britannica.
- Permette di studiare tutti i pazienti con una particolare malattia.
- Può selezionare i soggetti di controllo dalla stessa popolazione di origine
- I dati vengono raccolti in modo non interventistico e quindi riflettono la "vita reale"
- L'informazione viene continuamente aggiornata, permettendo di esaminare gli effetti di nuovi interventi / trattamenti.
- Non è necessario raccogliere dati, riducendo quindi i tempi e i costi dell'indagine
- È adattabile alla maggior parte dei disegni di studio epidemiologici (cioè coorte, case-control, case-series).
- Può essere utilizzato per studiare condizioni cliniche o outcomes relativamente rari.

- Un certo numero di pratiche sono state collegate ai dati HES (statistiche degli episodi ospedalieri).

Punti di debolezza

- L'utilizzo primario del software Vision è la gestione del paziente e non la ricerca medica e i dati riflettono solo gli eventi ritenuti rilevanti per la cura del paziente
- L'analisi richiederà l'intervento di un gestore dati esperto con accesso a software e hardware appropriati
- Anche un insieme di dati così grande può essere problematico quando l'esposizione e gli esiti sono entrambi rari.
- Le situazioni in cui i dati THIN potrebbero non essere appropriati includono:
    - Studi in cui l'etnia, l'occupazione e / o lo stato socio-economico dell'individuo sono variabili importanti (non disponibili come dati sul livello del paziente).
    - La mancata aderenza alle prescrizioni dei medicinali può essere un problema per le condizioni cliniche correlate all'assunzione di farmaci.
    - Studi dove i dati sono principalmente legati all'assistenza secondaria (studi sulla cura del cancro).
    - Studi che esaminano farmaci da banco, in quanto questi non vengono di solito registrati.
    - Studi che esaminano i risultati dei test di laboratorio prima dell'informatizzazione dovrebbero essere consapevoli del fatto che possono essere stati immessi solo valori anormali.

4.1.2. *Qresearch*

QRESEARCH è un database derivato dalle cartelle cliniche pseudonimizzate di oltre 24 milioni di pazienti. I dati attualmente arrivano da circa 1500 medici di medicina di base che utilizzano il sistema informatico clinico EMIS. I servizi di medicina di base sono diffusi in tutto il Regno Unito e includono i dati dei pazienti attualmente registrati nonché i dati storici dei pazienti che possono anche essere deceduti. I record storici si estendono fino all'inizio degli anni '90, rendendo Qresearch uno dei più grandi e ricchi database di medicina di base del mondo [5].

Le principali caratteristiche del database sono:

- Comprende i dettagli socioeconomici associati al codice postale di ciascun paziente
- I dati estratti non contengono identificatori dei pazienti poiché i dati vengono anonimizzati alla fonte
- I pazienti possono rinunciare a partecipare all'iniziativa

- Il database è aperto ai ricercatori accademici dietro approvazione del comitato etico
- Ai ricercatori non sarà fornito tutto il dataset di dati, bensì solo i dati anonimizzati necessari per rispondere al loro quesito di ricerca
- I costi di utilizzo dei dati sono attentamente valutati per consentire il finanziamento del sistema, ma al tempo stesso per consentire un accesso agevole ai ricercatori accademici
- Sono state effettuate analisi per dimostrare l'accuratezza e la completezza dei dati e sono disponibili per analisi di morbilità
- L'intero database è stato collegato a cause di dati di morte, tumori e dati ospedalieri a livello di singolo paziente con collegamenti che si estendono indietro fino al 1993.
- I dati collegati sono disponibili solo presso l'Università di Nottingham che è anche l'unico controllore dei dati per singoli set di dati (vale a dire morte, cancro, dati ospedalieri) e per il database risultante, inclusi i set di dati collegati.

4.1.3. *MalaCards*

MalaCards (Maladies Cards) è un database integrato di malattie e disordini umani arricchito con annotazioni e modellato sul database dei geni umani GeneCards. MalaCards è organizzato in "schede di malattia", ognuna delle quali contiene le informazioni note sulla patologia tra cui gli alias e le annotazioni; tra le annotazioni si trovano informazioni circa sintomi, farmaci, articoli di letteratura scientifica referenziati, geni coinvolti, studi clinici effettuati, malattie o disturbi correlati, etc. Un motore di ricerca automatico delle informazioni si occupa di popolare le schede di malattia, utilizzando dati remoti insieme alle informazioni raccolte utilizzando la piattaforma GeneCards.

Più in dettaglio, le sezioni di MalaCards sono popolate a partire da [32]:

1. risorse esterne contenenti informazioni sulle malattie per stabilire i nomi delle patologie, sinonimi, sommari, farmaci/terapie/trattamenti, caratteristiche cliniche, test genetici e contesto anatomico.
2. GeneCards per le pubblicazioni correlate e i geni associati.
3. GeneAnalytics per l'analisi di gruppi genici associati alla patologia allo scopo di produrre pathways
4. Riferimenti interni di MalaCards stesso per ulteriori malattie o disturbi correlati.

Le sezioni principali di una scheda di MalaCards sono:

1. ***Aliases & Classifications*** che contiene:
    - Sinonimi e alias per la malattia in questione
    - Caratteristiche tratte da OMIM, HPO, Orphanet e GeneReviews; per ciascuna patologia vengono anche riportate, se disponibili, la mortalità, l'età di insorgenza, l'età in cui si verifica la morte, l'ereditarietà e la penetranza.

- L'albero di classificazione della patologia
- *Cross references* ad identificativi di database esterni od ontologie.

2. ***Summaries***: contenente descrizioni e brevi definizioni della malattia, degli organi coinvolti, dell'eziologia e dei principali sintomi.
3. ***Related Diseases*** contenente le patologie correlate rappresentate mediante "network images". Ogni malattia correlata costituisce un nodo, mentre le linee rappresentano una connessione con la malattia bersaglio. Lo spessore delle linee è proporzionale agli elementi a supporto della connessione. Un'altra forma di connessione malattia-malattia mostrata in tale sezione è la comorbilità.

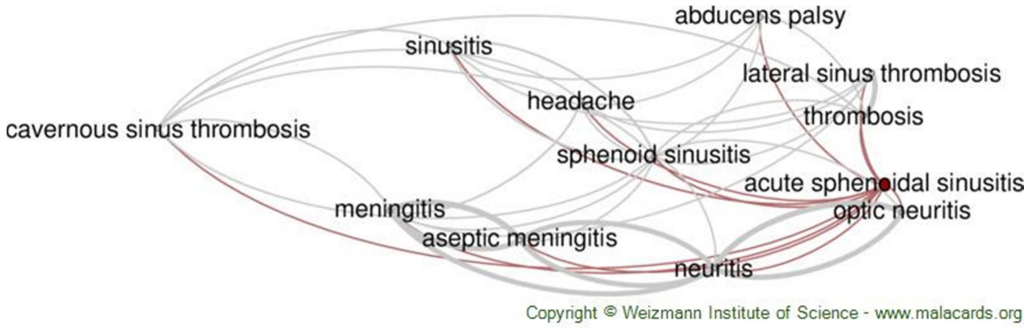

Figura 3 Rappresentazione in MalaCards delle patologie tra loro correlate

4. *Symptoms & Phenotypes*: contiene informazioni e link sui sintomi e altri attributi clinici della malattia.
5. *Drugs & Therapeutics* contiene informazioni riguardanti:
   - Farmaci approvati dall'FDA: i dati includono il nome del farmaco, il principio attivo, l'azienda farmaceutica e la data di approvazione
   - Farmaci per le patologie (da DrugBank, HMDB, DGIdb, PharmGKB, IUPHAR): patologie e farmaci sono mappati alle sperimentazioni cliniche presenti in ClinicalTrials.gov; in questa sezione sono mostrati i farmaci che condividono una sperimentazione clinica con la malattia.
   - Approcci terapeutici basati su cellule da Lifemap Discovery: vengono mostrati approcci terapeutici candidati a base di cellule staminali e cellule di colture embrionali e adulte.
6. *Genetic Tests* contiene descrizioni di test genetici e citogenetici specializzati e test biochimici per i disturbi ereditari.
7. *Anatomical Context* contiene descrizioni di cellule, comparti e organi rilevanti per la malattia.
8. *Publications* contiene la lista delle referenze alle pubblicazioni associate alla malattia, attualmente ottenute ricercando tutte le pubblicazioni presenti nel database di GeneCards.

9. *Genes* contiene l'elenco dei geni associati alla malattia. I geni "elite" di MalaCards (contrassegnati con il carattere *) sono quelli che possono essere associati alla causa della malattia, poiché le loro associazioni gene-malattia sono supportate da fonti curate manualmente e affidabili.
10. *Variations*: contiene le variazioni causali note per la malattia estratte da diverse fonti informative.
11. *Expression*: contiene dati sui geni espressi in modo differenziale nel tessuto malato rispetto al tessuto normale corrispondente e derivano da esperimenti pubblicati nel Gene Expression Omnibus (GEO) e/o dalla letteratura scientifica e da informazioni sull'espressione genica provenienti da set di dati pubblici online su larga scala.
12. *Pathways*: contiene informazioni sui pathways collegati alla patologia.
13. *GO Terms*: contiene i termini dell'ontologia di Gene Ontology relativi all'insieme di geni legati alla patologia.

Nella scheda, ogni malattia ha un simbolo unico composto dalla prima lettera del suo nome, seguito dalle due consonanti successive e da un numero di serie. Ad esempio, il simbolo generato per la sindrome di Rett è RTT001. Gli utenti registrati a GeneCards Plus possono scaricare i dati della scheda in un foglio excel. Il database è accessibile al seguente link: https://www.malacards.org/ .

4.1.4. *Diseases Database*

Il "Diseases Database" è un archivio di patologie umane, farmaci, sintomi, segni, risultati di indagini anomale, etc. Il database fornisce un sistema di ricerca che copre le seguenti aree [33]:

- Disturbi medici interni
- Sintomi e segni
- Disturbi congeniti ed ereditari
- Malattie ed organismi infettivi
- Farmaci e medicinali
- Anomalie comuni in esami ematologici e biochimici

Il Diseases Database, analogamente a molte ontologie, utilizza un modello entità-attributo-valore con regole sugli attributi (tipi di relazione) che le entità (oggetti) possono assumere.
Ogni voce è assegnata ad una o più categorie del Disease Database. Ad esempio, l'aspirina è assegnata alla categoria "farmaci, ormoni e mediatori", la distrofia muscolare di Duchenne a "condizioni autosomiche dominanti". Le categorie identificano le relazioni a cui un oggetto potrebbe plausibilmente appartenere, vale a dire che l'aspirina potrebbe "appartenere ad una famiglia di farmaci", mentre una "condizione

dominante autosomica" non potrebbe. La distrofia muscolare di Duchenne potrebbe "essere allelica con" altre condizioni mendeliane, mentre un "insetto" non potrebbe.

A differenza di molte ontologie, il Diseseas Database non utilizza, tuttavia, una logica descrittiva per organizzare i contenuti. La maggior parte delle relazioni presenti all'interno del database sono assertive piuttosto che definitive. Per chiarire con un esempio, l'infarto miocardico può causare dolore toracico ma non sempre lo fa e il dolore toracico non definisce la condizione. Alcune ontologie avanzate, tuttavia, potrebbero supportare la seguente serie di inferenze: il muscolo cardiaco è coinvolto nell'infarto; gli infarti sono dolorosi; il cuore è presente nel torace; quindi l'infarto miocardico può causare dolore toracico [34].

4.1.5. *KEGG Disease Database*

KEGG (Kyoto Enciclopedia of Genes and Genomes) è un insieme di database utilizzato per comprendere funzioni e utilità di alto livello sui sistemi biologici, come la cellula, l'organismo e l'ecosistema, a partire da informazioni di livello molecolare, come quelle generate dal sequenziamento.

In KEGG, le patologie sono viste come stati perturbati del sistema della rete molecolare. I fattori genetici ed ambientali delle malattie, così come i farmaci, sono considerati perturbanti di questo sistema. Diversi tipi di malattie, comprese quelle monogeniche, quelle multifattoriali e quelle infettive, sono tutte trattate in modo unificato riunendo tali perturbanti e le loro interazioni.

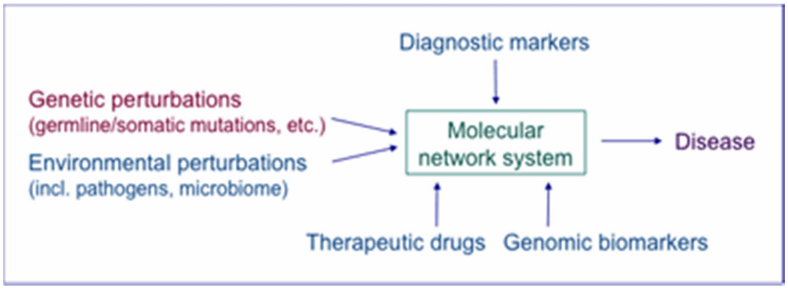

Figura 4 Patologie come stati perturbati

Nel database KEGG PATHWAY, le conoscenze note sulle reti molecolari perturbate sono rappresentate come mappe dei pathway della patologia. Sebbene i geni della malattia (perturbanti genetici) siano contrassegnati in rosso nelle mappe, i dettagli delle perturbazioni, come le mutazioni, non sono forniti perché tali informazioni sono riportate nel database KEGG NETWORK.

KEGG DISEASE è in pratica una raccolta di entry di malattie focalizzate solo sui perturbanti, poiché i dettagli delle reti molecolari sono sconosciuti per la maggior parte delle condizioni. Ogni voce è identificata dal codice H e contiene un elenco di fattori genetici noti (geni di malattia), fattori ambientali, agenti patogeni e farmaci terapeutici. Le entry relative alle malattie sono classificate in file gerarchici in formato BRITE.

La categoria Human Diseases del database KEGG PATHWAY è una raccolta di mappe di pathway della patologia. Contiene malattie multifattoriali come tumori, malattie del sistema immunitario, malattie neurodegenerative, malattie cardiovascolari e malattie metaboliche in cui i geni noti della malattia sono contrassegnati in rosso. Contiene anche malattie infettive in cui sono rappresentate le reti molecolari che interagiscono tra gli agenti patogeni e l'uomo.
Lo strumento di ricerca in KEGG Mapper permette all'utente di caricare un set di geni umani per verificare se esistono delle patologie che possono essere associate al set di dati [35].

4.1.6. *DisEASES*

DisEASES (https://diseases.jensenlab.org/Search) è una risorsa web aggiornata settimanalmente che integra le evidenze sulle associazioni geni - malattia estratte mediante text mining automatico dei testi, mediante analisi manuale della letteratura scientifica, dai dati sulle mutazioni tumorali, etc. Il sistema supporta l'evidenza assegnando punteggi di confidenza che facilitano il confronto tra i diversi tipi e fonti di evidenza.
Più in dettaglio, il sistema permette di estrarre le associazioni di geni della malattia dagli abstract biomedici e consiste in un tagger altamente efficiente basato su dizionari per il riconoscimento di geni e malattie umane, combinato con uno schema di punteggio che tiene conto delle co-occorrenze sia all'interno delle frasi che tra una frase e l'altra. L'approccio è in grado di estrarre la metà di tutte le associazioni curate manualmente con un tasso di falsi positivi di solo lo 0,16%. DISEASES integra i risultati del text mining con informazioni sulle associazioni di geni che causano una malattia curate manualmente, dati sulle mutazioni tumorali e studi di associazione genome-wide da database esistenti come Genetics Home Reference (GHR), UniProt Knowledgebase (UniProtKB), GWAS e dati sulle mutazioni dal Catalog of Somatic Mutations in Cancer (COSMIC). Per rendere i dati facili da utilizzare per analisi su larga scala, tutte le fonti di evidenza sono state mappate su identificatori comuni, sono assegnate ad esse dei punteggi di qualità comparabili e rese disponibili per il download.
Diverse banche dati esistenti si concentrano o contengono associazioni malattia-gene, ottenute principalmente attraverso l'analisi manuale della letteratura biomedica. Purtroppo, la maggior parte di queste utilizza un vocabolario interno controllato delle malattie e sono soggette a licenze restrittive, rendendo difficile così la loro integrazione sia da un punto di vista tecnico che giuridico. Il più famoso dei database è Online Mendelian Inheritance in Man (OMIM), a cui sono stati forniti collegamenti ipertestuali sia per i geni che per le malattie. Altri database collegati sono lo Human Gene Mutation Database (HGMD), il Comparative Toxicogenomics Database (CTD) e il Genetics Home Reference (GHR) [36].

4.1.7. *DisGeNET*

DisGeNET (http://www.disgenet.org/) è una piattaforma di ricerca che contiene una delle più grandi collezioni pubbliche di geni e varianti associate alle patologie umane. DisGeNET integra informazioni sulle associazioni geni-malattie umane (GDA) e associazioni di varianti-malattie (VDA) provenienti da vari archivi, comprese malattie mendeliane, complesse e ambientali. L'integrazione viene effettuata attraverso la mappatura del vocabolario genetico e delle malattie ed utilizzando l'ontologia per il tipo di associazione DisGeNET.

Le informazioni gene-malattia in DisGeNET sono organizzate in base ai tipi di database di origine nelle seguenti categorie:

- CURATED: associazioni geni-malattia (GDA) da UniProt, PsyGeNET, Orphanet, CGI, CTD (human data), ClinGen, etc.
- ANIMAL MODELS: GDA provenienti da modelli murini
- INFERRED: GDAs da Human Phenotype Ontology (HPO), e GDA inferite da VDA riportate in Clinvar, GWAS, etc.
- ALL: GDA dalle risorse citate in precedenza e da LHGDN e BeFree

I dati curati contengono le associazioni geni-malattia (GDA) fornite da risorse curate dagli esperti, quali:

- UNIPROT: contenente informazioni curate sulla sequenza, la struttura e la funzione delle proteine. I GDA sono ottenuti dal file "humsavar" prendendo in considerazione solo le associazioni contrassegnate come Disease.
- CTD (Comparative Toxicogenomics Database): contiene informazioni curate manualmente sulle relazioni gene-malattia con particolare attenzione alla comprensione degli effetti delle sostanze chimiche ambientali sulla salute umana. Dal CTD sono incluse solo le associazioni contrassegnate come marker/meccanismo o terapeutico.
- ORPHANET: un portale di riferimento per le informazioni sulle malattie rare e i farmaci orfani il cui scopo è quello di contribuire a migliorare la diagnosi, la cura e il trattamento dei pazienti con malattie rare; da Orphanet vengono mantenute tutte le GDA.
- CLINGEN (Clinical Genome Resource) è dedicato alla costruzione di una risorsa centrale che definisce la rilevanza clinica dei geni e delle varianti per l'uso in medicina e ricerca di precisione; da ClinGen sono incluse tutte le GDA ad eccezione di quelle etichettate come "refuted".
- Genomics England PanelApp è una base di conoscenza accessibile al pubblico che permette di creare, memorizzare e interrogare panel genetici virtuali relativi a disturbi umani. Da questa fonte sono conservate tutte le GDA.

- CGI (Cancer Genome Interpreter) è uno strumento che (i) identifica le alterazioni oncogene conosciute; (ii) predice i potenziali driver tra quelli di significato sconosciuto e (iii) identifica le alterazioni nel tumore che influenzano la risposta ai farmaci antitumorali. Da questo insieme di dati, vengono conservati solo dati convalidati.

Inferred Data si riferisce a GDA desunte da HPO e da VDA. Nel caso dell'HPO, le GDA sono dedotte da associazioni fenotipo-malattia e gene-malattia tramite "triangolazione". Nel caso delle VDA, viene creata una GDA per ogni gene annotato alla variante e la malattia annotata alla variante. Al fine di dedurre GDA da VDA, sono considerate solo le varianti che mappano le sequenze codificanti dei geni. La variante viene assegnata a uno o più geni utilizzando le mappature di dbSNP.
I dati inferiti contengono le associazioni geni-malattia (GDA) fornite da risorse quali:

- HPO (Human Phenotype Ontology) mira a fornire un vocabolario standardizzato delle anomalie fenotipiche riscontrate nelle malattie umane. Dall'HPO vengono importate le FREQUENT_FEATURES dal file genes_to_phenotype.
- ClinVar è un archivio pubblico e liberamente accessibile di report sulle relazioni tra varianti e fenotipi di rilevanza medica, con evidenze a sostegno. Da Clinvar sono importate solo le associazioni varianti-malattia con le etichette: Pathogenic, Likely pathogenic, Risk factor and Affects.
- GWAS CATALOG è una raccolta di tutti gli studi di associazione genome-wide pubblicati che analizzano almeno 100.000 SNPs e tutte le associazioni di SNP-trait. Dal catalogo GWAS vengono conservate le VDA con valori $p < 1,0 \times 10^{-6}$.

La versione 6.0 del database contiene quasi 630000 associazioni gene-malattia (GDA) e più di 210000 associazioni varianti-malattia.
DisGeNET può essere utilizzato per diversi scopi di ricerca, tra cui lo studio dei presupposti molecolari delle patologie umane e delle loro comorbidità, l'analisi delle proprietà dei geni malattia, la generazione di ipotesi sull'azione terapeutica dei farmaci e sugli effetti avversi dei farmaci e la validazione dei geni malattia predetti computazionalmente [37].
Il database DisGeNET è reso disponibile con una licenza che non permette di utilizzare il materiale per scopi commerciali.

4.1.8. *OMIM*

Online Mendelian Inheritance in Man (OMIM) (http://omim.org/) è un catalogo aggiornato di geni umani, malattie e tratti genetici focalizzato sulla relazione molecolare tra variazione genetica ed espressione fenotipica [38].

Le mutazioni sono catalogate in OMIM nella sezione Varianti Alleliche. Per la maggior parte dei geni sono incluse solo mutazioni selezionate; i criteri per l'inclusione includono la prima mutazione scoperta, la frequenza nella popolazione, il fenotipo distintivo, il significato storico, il meccanismo insolito di mutazione, il meccanismo patogenetico e l'eredità distintiva (ad esempio, dominante con alcune mutazioni, recessiva con altre mutazioni nello stesso gene). La maggior parte delle varianti alleliche rappresentano mutazioni che causano una malattia. Sono inclusi alcuni polimorfismi, molti dei quali mostrano una correlazione positiva con particolari disturbi comuni.

OMIM Gene Map e Morbid Map presentano rispettivamente le localizzazioni citogenetiche dei geni e dei disturbi descritti in OMIM.

L'uso di OMIM è libero per uso personale, per uso educativo o accademico o per scopi di ricerca attraverso il front end del database. Gli utenti che intendono fare un uso commerciale e che desiderano scaricare tutto o parte del database OMIM devono ottenere una licenza stipulando un contratto.

4.1.9. *REACTOME diseases*

REACTOME (https://reactome.org) è un database di pathway curato manualmente e sottoposto a revisione. L'obiettivo del database è quello di fornire strumenti innovativi per la visualizzazione, l'interpretazione e l'analisi della conoscenza dei pathway a supporto della ricerca di base e clinica, dell'analisi del genoma, della modellazione e della systems biology.

Alla base di Reactome c'è un database relazionale di molecole coinvolte nella regolazione e nel metabolismo e le loro relazioni organizzate in pathway e processi biologici. L'entità centrale del modello dati di Reactome è la reazione. Le entità (acidi nucleici, proteine, complessi, vaccini, terapie antitumorali e piccole molecole) che partecipano alle reazioni formano una rete di interazioni biologiche e sono raggruppate in pathway. Esempi di pathway biologici in Reactome includono il metabolismo intermedio, la regolazione trascrizionale, l'apoptosi e le patologie.

Il sito web Reactome è stato progettato per fornire all'utente una mappa grafica dei processi e dei pathway biologici conosciuti; esso fornisce anche un'interfaccia attraverso la quale l'utente può "cliccare" per ottenere informazioni curate e dettagliate sui componenti e le loro relazioni.

Reactome è utilizzato da clinici, genetisti, ricercatori di genomica e biologi molecolari per interpretare i risultati di studi sperimentali high-throughput, da bioinformatici che cercano di sviluppare nuovi algoritmi per estrarre conoscenza da studi genomici e da systems biologists che costruiscono modelli predittivi di pathway normali e varianti causati da una malattia.

Tutti i dati e il software sono liberamente disponibili per il download. I dati di interazione, reazione e pathway sono disponibili in formato flat file, Neo4j GraphDB, MySQL, BioPAX, BioPAX, SBML e PSI-MITAB e sono accessibili anche attraverso le API dei servizi web [39].

Reactome può annotare e visualizzare i pathway associati alla malattia. Le annotazioni di malattia presenti includonon quelle relative a cancro, malattie metaboliche e immunitarie e malattie infettive. Dove possibile, i pathway di malattia di Reactome includono anche l'interazione di farmaci terapeutici rilevanti [40].

4.1.10. *Comparative Toxicogenomics Database (CTD)*

Il CTD (http://ctdbase.org) è una base dati pubblica il cui scopo è quello di raccogliere informazioni utili ad evidenziare come l'esposizione a fattori ambientali influisce sulla salute umana. Il database fornisce informazioni curate manualmente su:

- interazioni tra composti chimici e geni/proteine
- malattie dovute a prodotti chimici
- relazioni tra geni e malattie genetiche

Questi dati sono integrati con dati funzionali e di pathway per aiutare a sviluppare ipotesi sui meccanismi alla base delle malattie influenzate dall'ambiente.

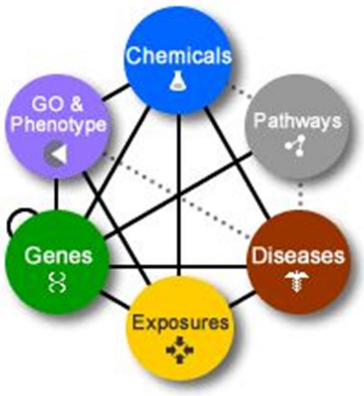

Figura 5 Categorie di dati tra cui sono previste relazioni all'interno del database

Ci sono diverse categorie di dati e interazioni ricercabili nel database:

- **Chemicals** riporta informazioni su sostanze chimiche, informazioni curate su geni e proteine che interagiscono con esse, relazioni curate e inferite sulle patologie, pathway e annotazioni funzionali.
- **Diseases**: il vocabolario delle malattie MEDIC del CTD è un sottoinsieme modificato di descrittori della categoria "Diseases" delle Medical Subject Headings (MeSH), combinato con le malattie genetiche del database OMIM. I biocuratori del CTD hanno mappato le malattie OMIM su termini all'interno del vocabolario gerarchico delle malattie MeSH per espandere la rappresentazione della malattia. Questo vocabolario combinato è usato per curare le associazioni gene-malattia e composti chimici-malattia.
- **Genes**: Il vocabolario genetico del CTD (simboli, nomi e sinonimi) deriva dal database Gene dell'NCBI.
- **Phenotypes**: il database CTD distingue tra patologie e fenotipi, laddove fenotipo si riferisce ad un evento biologico non riferito a patologia; per chiarire con degli esempi: arresto anomalo del ciclo cellulare (fenotipo) vs. cancro polmonare (malattia), aumento della differenziazione delle cellule adipose (fenotipo) vs. obesità (malattia), diminuzione della spermatogenesi (fenotipo) vs. infertilità maschile (malattia), etc. CTD usa Gene Ontology (GO) come fonte di termini relativi al fenotipo per gli effetti biologici.
- **Chemical–Gene/Protein Interactions**: Per migliorare la comprensione dei meccanismi delle azioni dei prodotti chimici, le informazioni sulle interazioni tra prodotti chimici e geni/proteine sono curate a partire dalla letteratura scientifica. Le interazioni sono curate utilizzando un vocabolario di tipo interattivo gerarchico che caratterizza le comuni interazioni fisiche, regolatorie e biochimiche tra sostanze chimiche e geni o proteine.
- **Gene–Disease Associations**: Le associazioni gene-malattia possono essere inferite a partire da associazioni curate tra prodotti chimici e geni e prodotti chimici e patologie. Il database CTD contiene associazioni curate da esperti ed inferite su associazioni geni - patologia. Le associazioni geni-patologie curate da esperti sono estratte dalla letteratura pubblicata dai biocuratori del CTD, o derivano dalla banca dati OMIM.

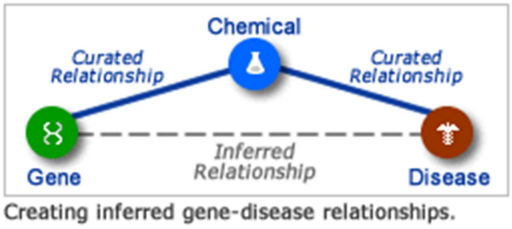

Figura 6 Relazioni inferite tra gene e patologia

Le associazioni inferite sono stabilite attraverso interazioni tra sostanze chimiche e geni curate da CTD (ad esempio, il gene A è associato alla malattia B perché esiste per il gene A un'interazione curata con la sostanza chimica C, e la sostanza chimica C presenta un'associazione curata con la malattia B). Le associazioni curate e dedotte sono identificate e aiutano gli utenti a sviluppare ipotesi sui meccanismi alla base delle malattie con cause ambientali.

- **Chemical–Disease Associations**: Le associazioni malattie – sostanze chimiche presenti nel database sono curate e desunte dalla letteratura scientifica pubblicata dai biocuratori del CTD. Le associazioni inferite sono stabilite tramite interazioni tra sostanze chimiche e geni (ad esempio, la sostanza chimica A è associata alla malattia B perché la sostanza chimica A ha un'interazione curata con il gene C, e il gene C ha un'associazione curata con la malattia B).

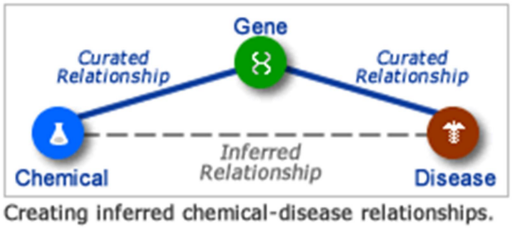

Figura 7 Relazioni inferite tra sostanza chimica e patologia

- **Chemical–Phenotype Interactions**: Un'affermazione di interazione tra prodotto chimico e fenotipo in CTD include 8 tipi di dati (C-Q-E-A-T-M-S-P) annotati utilizzando 8 vocabolari controllati, tra cui:
    - C, una sostanza chimica del vocabolario chimico del CTD;
    - Q, un qualificatore di azione del CTD che riflette la direzione dell'interazione ("aumenta" "diminuisce" o "influenza");
    - E, il fenotipo di entità come indicato in Gene Ontology;
    - A, un termine anatomico del ramo "Anatomia [A]" del MeSH;
    - T, un organismo della NCBI Taxonomy;
    - M, un codice del metodo CTD (in vivo, in vitro);
    - S, il codice che indica la sorgente delle informazioni CTD (abstract, testo completo);
    - P, l'identificativo dell'articolo (PMID) di PubMed.
- **Gene–Gene Interactions**: Il database contiene le interazioni gene-gene di BioGRID (che consistono in interazioni genetiche) e di proteine curate dalla letteratura primaria per tutti i principali organismi modello.

References: CTD contiene articoli di riferimento relativi a geni di vertebrati e invertebrati tossicologicamente significativi, malattie e prodotti chimici associati. I riferimenti sono stati identificati mediante metodi di ricerca di informazioni e comprendono un sottoinsieme di PubMed.

Pathways: I dati sui pathway di KEGG e REACTOME descrivono le reti note di interazione e reazione molecolare. Questi dati sono integrati con le sostanze chimiche, i geni e le malattie riportate nel database CTD per fornire informazioni sulle reti molecolari che possono essere compromesse da sostanze chimiche e sui possibili meccanismi alla base delle malattie ambientali.

Exposures: Il CTD sta lavorando per migliorare la capacità di identificare le connessioni ambiente-malattia sviluppando un'ontologia dell'esposizione (ExO) che sarà utilizzata per curare e presentare i dati sull'esposizione.

4.1.11. *Nutrition Evidence*

Nutrition Evidence (http://www.nutrition-evidence.com) è una raccolta online di studi scientifici di alta qualità nel campo della nutrizione e dello stile di vita, progettata per supportare le decisioni cliniche dei professionisti con l'obiettivo di fornire raccomandazioni efficaci e personalizzate basate su riscontri concreti. Le caratteristiche principali includono:

- Quasi 40.000 articoli di letteratura scientifica di alta qualità che riguardano tutti i professionisti nei settori della nutrizione funzionale, della dietetica e della medicina dello stile di vita.
- Articoli del BANT (British Association for Nutrition and Lifestyle Medicine) basati su un modello di medicina dello stile di vita che tiene conto non solo degli squilibri clinici, ma anche di altri aspetti chiave per la salute dell'individuo come lo stress, il sonno e l'attività fisica.
- Un'ampia selezione di filtri permette di trovare articoli mediante fattori e parole chiave. Ad esempio, è possibile cercare argomenti di carattere ambientale (ad es. dieta o inquinamento) o sull'integrità strutturale (ad es. salute muscoloscheletrica).

La banca dati è accessibile tramite il link http://www.nutrition-evidence.com/. Con un account registrato gli utenti possono accedere a tutte le funzioni del database [41].

4.1.12. *VMH database*

Il metabolismo svolge un ruolo cruciale nella salute umana e nelle malattie ed è modulato da fattori intrinseci (es. genetici) ed estrinseci (es. dieta e microbiota intestinale). Se considerati individualmente, questi fattori non spiegano a sufficienza lo sviluppo e la progressione di molte malattie complesse non trasmissibili, tra cui la sindrome metabolica e le malattie neurodegenerative. È necessario, pertanto, un

approccio sistemico per chiarire il contributo di ciascuno di questi fattori e per consentire lo sviluppo di strategie terapeutiche nuove ed efficaci.

Il database Virtual Metabolic Human (VMH) (https://www.vmh.life) racchiude le attuali conoscenze del metabolismo umano in cinque risorse interconnesse:

1. Metabolismo umano
2. Microbioma intestinale
3. Patologia
4. Nutrizione
5. ReconMaps

Nello specifico, il database VMH raccoglie informazioni sul metabolismo microbico umano e intestinale e collega queste informazioni a centinaia di malattie e dati nutrizionali. Alla sia base ci sono centinaia di ricostruzioni metaboliche su scala genomica curate manualmente, che sono state assemblate sulla base di dati genomici, biochimici e fisiologici. Il VMH facilita analisi e interpretazioni rapide di dati complessi derivanti da studi biomedici su larga scala, consentendo interrogazioni complesse del suo contenuto, fornendo una rappresentazione grafica dettagliata del metabolismo umano e distribuendo modelli computazionali per simulare il metabolismo umano e microbico.

Il VMH rappresenta un database interdisciplinare molto utile per l'interpretazione dei dati e la generazione di ipotesi per la comunità biomedica [42].

4.1.13. *CHE Toxicant and Disease Database*

Le malattie umane sono il risultato di complesse interazioni tra geni e ambiente. L'esposizione ambientale ad agenti chimici, fisici e biologici può causare o contribuire alla comparsa di malattie in individui sensibili. I fattori dello stile di vita, come la dieta, il fumo, il consumo di alcol, il livello di esercizio fisico e l'esposizione ai raggi UV, sono spesso al centro dell'attenzione quando si considerano le cause prevenibili di una malattia. L'esposizione a contaminanti chimici sul posto di lavoro, a casa, all'aperto e anche nell'utero, sono sempre più riconosciuti come fattori importanti e prevenibili che contribuiscono alle malattie umane.

Negli ultimi 50 anni sono state sviluppate, distribuite e smaltite nell'ambiente più di 80.000 sostanze chimiche. La maggior parte di esse non sono state testate per i potenziali effetti tossici sull'uomo o sugli animali. Alcune di queste sostanze chimiche si trovano comunemente nell'aria, nell'acqua, nel cibo, nelle case, nei luoghi di lavoro, etc. Mentre la tossicità di una sostanza chimica può non essere compresa completamente, la comprensione dell'effetto dell'esposizione a miscele di sostanze chimiche è ancora meno completa. Le sostanze chimiche possono avere effetti opposti, additivi o addirittura sinergici. Un

esempio di effetto sinergico è il fumo di tabacco abbinato all'esposizione all'amianto, che aumenta il rischio di cancro ai polmoni di 25 volte, un rischio molto più elevato di quello risultante dalla somma dei rischi dei singoli agenti.

Gli effetti tossici degli agenti chimici spesso non sono ben compresi o recepiti dagli operatori sanitari. Alcune sostanze chimiche, come l'amianto, il cloruro di vinile e il piombo, sono ben note come cause di malattie umane. Vi sono prove che l'aumento di alcuni tipi di cancro, asma e disturbi dello sviluppo può essere attribuito all'esposizione a sostanze chimiche, in particolare nei bambini.

Dato che la sperimentazione umana non è eticamente accettabile, gli effetti dell'esposizione a sostanze chimiche negli esseri umani sono difficili da studiare. Dati limitati sull'uomo sono disponibili per esposizioni accidentali, overdose o studi basati su dati di lavoratori esposti professionalmente.

Gli effetti dell'esposizione alle sostanze chimiche possono variare a seconda dell'età di esposizione (in utero, infanzia, adulto), della via di esposizione (ingestione, inalazione, contatto dermico), della quantità e della durata dell'esposizione, dell'esposizione a più sostanze chimiche contemporaneamente, e di altri fattori di suscettibilità personale, compresa la variabilità genetica [43].

Il CHE Tossicant and Disease Database riassume i collegamenti tra fattori di rischio chimico e altri fattori di rischio ambientale e più di 200 malattie o condizioni umane. Il database riflette lo stato attuale delle conoscenze sulle sostanze tossiche e sulle malattie umane, organizzate per categorie di malattie. Il database si concentra principalmente su studi epidemiologici sull'uomo.

Il database è interrogabile per malattia, categoria di malattia, sostanza tossica o parola chiave.

I principali sistemi interessati sono quello Cardiovascolare, Endocrino, Gastrointestinale, Genitourinario, Circolatorio, Immunitario, Epatico, Muscoloscheletrico, Neurologico, Respiratorio, Renale e Cutaneo.

Gli agenti tossici che sono stati collegati ad una condizione sono collocati in una delle tre categorie in base alla solidità delle prove per l'associazione.

1. La categoria "**strong evidence**" è riservata agli agenti tossici per i quali è stata verificata un'associazione causale con la malattia. La tossicità di queste sostanze tossiche è stata ben accettata dalle comunità di ricerca medica o scientifica. Le sostanze tossiche per le quali è stato stabilito che esistono prove sufficienti per provocare il cancro negli esseri umani sono incluse in questa categoria.
2. La categoria delle "**good evidence**" include le sostanze tossiche associate ad una malattia attraverso studi epidemiologici (trasversali, serie di casi o studi caso-controllo) o per le sostanze tossiche con qualche evidenza umana e forti evidenze animali che corroborano l'associazione. Sono inclusi in questa categoria anche gli agenti tossici che mostrano prove limitate come causativi del cancro negli esseri umani e prove sufficienti negli animali.

3. La categoria "**limited/conflicting evidence**" contiene sostanze tossiche debolmente associate a malattie umane attraverso casi di pochi individui esposti (case reports), da studi epidemiologici umani in conflitto che hanno dato risultati misti o equivoci, o, in alcuni casi, da rapporti che dimostrano chiaramente la tossicità in animali in cui non esistono dati umani.

La maggior parte delle sostanze tossiche del database rientra nella categoria delle prove "limitate / conflicting". Questo perché gli studi epidemiologici sull'uomo sono molto complessi, difficili da progettare e interpretare e non possono essere facilmente ripetuti.
Questo database ha limiti significativi che sono importanti da tenere in considerazione [44]:

- Le sostanze chimiche elencate sono una rappresentazione di sostanze tossiche che contribuiscono alle malattie e ai disturbi umani.
- Il database non prende in considerazione i tempi, la durata o la quantità di esposizione richiesta per provocare una particolare condizione. Alcune sostanze chimiche possono essere tossiche solo se inalate, mentre altre devono essere ingerite per essere tossiche. Alcune malattie derivano solo da esposizioni a dosi elevate, mentre esposizioni di basso livello possono essere meno importanti.
- La banca dati non fa alcun tentativo di quantificare la percentuale di malattie causate o a cui contribuiscono fattori ambientali specifici. Ad esempio, il mesotelioma, una forma rara di cancro, è quasi interamente dovuto all'esposizione all'amianto. Al contrario, la percentuale di casi di cancro ai polmoni causato dall'esposizione all'amianto è relativamente bassa rispetto al numero di casi causati dal fumo di tabacco o dal radon.

4.1.14. *Agency for Toxic Substances and Disease Registry*

L'Agency for Toxic Substances and Disease Registry (ATSDR) (https://www.atsdr.cdc.gov) è un'agenzia federale degli Stati Uniti che rende disponibile un portale attraverso il quale è possibile ricercare informazioni su sostanze tossiche e su come queste possono influenzare la salute.
ATSDR mette a disposizione una sezione chiamata "Toxicological Profiles" che contiene una raccolta unica di informazioni tossicologiche sulle sostanze pericolose. Ogni profilo tossicologico sottoposto a valutazione incrociata riflette una valutazione completa ed estesa, una sintesi e un'interpretazione delle informazioni tossicologiche ed epidemiologiche disponibili su una sostanza [45].
Le sostanze tossiche possono essere ricercate selezionandole da una lista in ordine alfabetico oppure in base:

- all'effetto sulla salute (es. sul peso corporeo, sul cancro, sul sistema cardiovascolare, sullo sviluppo, sul sistema endocrino (ghiandole ed ormoni))
- alla classe chimica (es. ftalati, diossine, pesticidi, etc.)

4.1.15. *Eurostat Health*

Eurostat Health [46] è una risorsa che raccoglie le statistiche sanitarie che misurano sia gli aspetti oggettivi che soggettivi della salute delle persone. Esse riguardano diversi tipi di aspetti connessi alla salute, come gli indicatori chiave sul funzionamento dei sistemi sanitari e sulla salute e la sicurezza sul luogo di lavoro. I dati disponibili riguardano i seguenti argomenti principali:

- Salute pubblica
    - Stato di salute: salute e morbilità, limitazioni funzionali e di attività, lesioni;
    - Fattori determinanti della salute: sovrappeso e obesità, attività fisica, comportamenti a rischio come il consumo di tabacco e alcol;
    - Sanità: spesa sanitaria, risorse (personale e strutture) e attività (servizi ospedalieri e ambulatoriali);
    - Disabilità: prevalenza della disabilità, occupazione, ostacoli all'integrazione sociale delle persone disabili;
    - Cause di morte: mortalità nazionale e regionale per cause di morte.
- Salute e sicurezza sul lavoro
    - Infortuni sul lavoro: numero di infortuni, cause e circostanze degli infortuni.

Le statistiche sullo stato di salute e sui determinanti sanitari si concentrano su vari aspetti dello stato di salute della popolazione e dei suoi determinanti non medici, degli stili di vita e dei comportamenti in materia di salute e sono quelle più utili per l'analisi di correlazione. I dati consentono di analizzare le problematiche di sanità pubblica, nonché i modelli demografici e socioeconomici e le disparità di stato di salute e dei suoi determinanti e costituiscono uno strumento di monitoraggio degli effetti delle politiche sanitarie. I dati sono strutturati come segue:

- Stato di salute
    - Anni di vita sana
    - Salute e benessere autopercepito
    - Limitazioni funzionali e di attività
    - Morbilità cronica auto-dichiarata
    - Lesioni da incidenti
    - Assenza dal lavoro per problemi di salute

- Fattori determinanti della salute
    - Indice di massa corporea
    - Attività fisica
    - Consumo di frutta e verdura
    - Consumo di tabacco
    - Consumo di alcool
    - Ambiente sociale

4.1.16. *Pesticide-Induced Diseases Database*

Il legame di patologie come asma, autismo, difficoltà di apprendimento, difetti congeniti, disfunzioni riproduttive, diabete, morbo di Parkinson, Alzheimer, cancro, etc. con l'esposizione ai pesticidi continua a rafforzarsi nonostante gli sforzi per limitare l'esposizione ad agenti chimici.
La banca dati sulle malattie indotte dai pesticidi facilita l'accesso a studi epidemiologici e di laboratorio basati su scenari reali di esposizione nel mondo reale che collegano gli effetti sulla salute pubblica ai pesticidi. La letteratura scientifica documenta elevati tassi di malattie croniche tra le persone esposte ai pesticidi, con un numero crescente di studi associati sia a malattie specifiche che a una serie di malattie.
Il database è accessibile attraverso una serie di link sul sito web di Beyond Pesticides. Le malattie segnalate nelle banche dati sono [47]:

- Asma
- Difetti alla nascita
- Cancro
- Diabete
- Disfunzioni del sistema endocrino
- Ritardi nell'apprendimento/sviluppo
- Malattia di Parkinson
- Disfunzioni sessuali e riproduttive

4.1.17. *Gene-Environment interactions (GxE) Database*

L'ambiente (dieta, esercizio fisico, consumo di alcool e tabacco) ha un ruolo nel modificare il rapporto tra varianti genetiche e misure cliniche della malattia. Il database GxE, che si occupa di tenere traccia delle interazioni rilevanti tra alimentazione, malattie cardiovascolari e diabete di tipo 2, è stato compilato a

partire dalla letteratura con il risultato di oltre 550 interazioni incorporate nel database, insieme a oltre 1430 casi in cui è stata riscontrata una mancanza di significatività statistica. Il database è una risorsa utile per i ricercatori in genetica e nutrizione al fine di comprendere quali punti del genoma umano sono sensibili alle variazioni della dieta, dell'attività fisica e dell'uso di alcol, tra le altre scelte di stile di vita.

L'insorgenza di malattie come le malattie cardiovascolari, le malattie coronariche, l'ipertensione e il diabete di tipo 2 possono essere prevenute o ritardate con adattamenti dello stile di vita. I fattori dello stile di vita presi in considerazione includono la dieta, l'attività fisica, il consumo di alcool e tabacco, il sonno, la latitudine e l'altitudine di residenza. Numerosi geni coinvolti nell'omeostasi dei biomarcatori clinici hanno alleli che si associano con il biomarcatore al baseline o hanno alleli che mostrano interazioni gene-ambiente (GxE). Un'interazione GxE è esemplificata da una variazione puntiforme di sequenza nel genoma dove una versione o un allele si associa ad un rischio per la salute solo quando un fattore ambientale supera una determinata soglia.

La catalogazione delle interazioni GxE fornisce le informazioni di base su quali fattori di stile di vita un individuo potrebbe adottare prima di assumere un farmaco terapeutico, che a sua volta può avere effetti collaterali dannosi.

Al fine di identificare le pubblicazioni che presentano interazioni statisticamente significative tra una variante genetica (ad es. SNP) e un fattore ambientale, viene consultata la letteratura scientifica con alcune parole chiave. Tra le strategie di ricerca, quelle più utilizzate prevedono l'uso di una parola chiave relativa ad una variante genetica (es. polimorfismo, variante o SNP) e una parola chiave relativa ad una "interazione" o ad un fattore ambientale (dieta, esercizio fisico, alcool, tabacco, sigarette, etc.). Un'altra strategia di ricerca prevede l'uso dei termini "gene" e "interazione" abbinati ad uno specifico fattore ambientale (ad esempio, alcool o esercizio fisico). Gli abstract dei documenti identificati sono stati poi collegati a quelli relativi alle malattie metaboliche e all'obesità [48].

I dati sono estratti manualmente leggendo gli articoli raccolti. Per l'inserimento di questi dati sono stati stabiliti diversi campi: Gene, accession number dell'SNP, alias o nomi comuni dell'SNP, allele di rischio, fenotipo, fattore ambientale, condizione del fattore ambientale, sesso, popolazione, etc. Il campo "condizione del fattore ambientale" descrive il valore del fattore ambientale al di sopra o al di sotto del quale è stata osservata l'interazione GxE. Solo quelle interazioni GxE descritte come statisticamente significative, generalmente con un valore $P \leq 0.05$, sono state selezionate dall'analisi della letteratura.

Va tenuto comunque conto che il database ha alcune limitazioni:

1) Sebbene l'interesse primario della ricerca attuale si sia concentrato sulle interazioni statisticamente significative, è necessario un maggiore sforzo sulle interazioni biologicamente significative.

2) I dati inclusi provengono da studi genetici candidati. Ad es., nello studio di associazione tra varianti del gene CYP1A2, assunzione di caffeina e infarto del miocardio, l'associazione tra questi elementi non è stata osservata fino a quando non è stata considerata la doppia stratificazione del genotipo e dell'assunzione di caffeina; ciò denota una potenziale carenza di tali studi, ma suggerisce anche che molte altre interazioni GxE possono essere descritte con un'attenta integrazione di dati accurati sulle esposizioni ambientali.
3) I metodi di valutazione delle esposizioni e dei risultati ambientali sono spesso diversi in ogni studio e quindi alcuni dati possono includere una combinazione di dati con diversi livelli di confidenza. Campioni di piccole dimensioni possono contribuire a falsi positivi di interazioni gene-ambiente.

Nonostante i punti deboli, questo database fornisce indicazioni sugli alleli di rischio in base ai fattori ambientali e permette di studiare popolazioni campionate e persino esempi di assenza di interazione. Queste informazioni sono importanti per i ricercatori per estrarre prove e generare nuove ipotesi e possono corroborare affermazioni asserite dai test genetici. Questo database dovrebbe essere utilizzato come strumento per l'analisi dei dati relativi all'associazione dei genotipi. I ricercatori saranno in grado di interrogare questo database prima dell'analisi statistica come mezzo per concentrare tali analisi su un particolare gene, fenotipo o fattore ambientale.
Il file di dati scaricabile per questo database è un file Excel accessibile al seguente indirizzo: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3275815/bin/NIHMS285264-supplement-Supplementary_Tables.xls

4.1.18. *ClinVar*

ClinVar (https://www.ncbi.nlm.nih.gov/clinvar) è un archivio pubblico e liberamente accessibile dell'NCBI che contiene informazioni sulle relazioni tra le varianti genomiche e i fenotipi. ClinVar elabora le segnalazioni di varianti trovate nei campioni di pazienti, le ipotesi sul loro significato clinico, le informazioni su chi ha sottomesso l'informazione e altri dati a supporto. Gli alleli descritti nelle osservazioni sono mappati su sequenze di riferimento e riportati secondo lo standard HGVS (Human Genome Variation Society).
ClinVar supporta submission di diversi livelli di complessità. La sottomissione può essere semplice come una rappresentazione di un allele e della sua interpretazione (a volte definita una presentazione a livello di variante), o dettagliata come fornire molteplici tipi di osservazioni strutturate (a livello di caso) o prove sperimentali sull'effetto della variante sul fenotipo.

4.1.19. *HUMSAVAR*

HUMSAVAR (https://www.uniprot.org/docs/humsavar) è un indice di polimorfismi umani e mutazioni che causano malattie. Nella versione 2015_08 del luglio 2015, il numero totale di singole varianti di aminoacidi annotate in UniProtKB/Swiss-Prot è stato di 71.795 unità. HUMSVAR indica anche se è stata segnalata una singola variante di aminoacidi associata alla malattia, un polimorfismo neutro o una variante non classificata.

4.1.20. *eDGAR*

eDGAR (http://edgar.biocomp.unibo.it/gene_disease_db/) è un database che raccoglie e organizza i dati sulle associazioni gene-malattia derivate da OMIM, Humsavar e ClinVar.
Per ogni gene associato alla malattia, eDGAR raccoglie informazioni su [49]:

1) Interazioni recuperate da PDB, BIOGRID e STRING;
2) annotazioni di Gene Ontology condivise;
3) Pathway condivisi di KEGG e REACTOME;
4) interazioni regolatorie derivate da TRRUST;
5) localizzazione su cromosomi e/o co-localizzazione in loci limitrofi.

L'attuale versione di eDGAR comprende 2672 malattie, relative a 3658 geni diversi, per un totale di 5729 associazioni gene-malattia. Il 71% dei geni è legato a 621 malattie multigeniche ed eDGAR evidenzia i termini GO comuni, i pathways KEGG/REACTOME, le interazioni fisiche e regolatorie. Il database è liberamente disponibile all'indirizzo http://edgar.biocomp.unibo.it .

4.1.21. *GWAS Catalog*

Il catalogo Genome-Wide Association Studies (GWAS) (https://www.ebi.ac.uk/gwas) è un catalogo di studi di associazione genome-wide. Questi studi offrono l'opportunità di indagare l'impatto di varianti comuni su malattie complesse; identificare i GWAS pubblicati, tuttavia, può essere difficile poiché la grande ricchezza di dati contenuti in queste pubblicazioni è di fatto inaccessibile ai ricercatori senza una catalogazione sistematica e una sintesi delle associazioni osservate [50]. Il sito web e l'infrastruttura del catalogo è ospitato dall'EMBL-EBI. I file sono disponibili per il download da https://www.ebi.ac.uk/gwas/downloads

4.1.22. *HGV Database*

La banca dati HGV (Human Genome Variation) (disponibile all'indirizzo https://hgv.figshare.com) è una banca dati online liberamente consultabile delle variazioni del genoma. È possibile effettuare ricerche e filtrare per diverse variabili, tra cui malattie specifiche, geni, popolazioni o regioni. Un insieme di informazioni standard su ogni variazione viene raccolto da articoli pubblicati in letteratura scientifica e aggiunti al database dopo verifica da parte dei revisori.

I Data Report sono brevi rapporti sulla variazione del genoma umano e sulla variabilità che descrivono le varianti genomiche che causano la malattia e/o le loro frequenze. Il Database HGV raccoglie le informazioni rilevanti riportate nei Data Reports pubblicati. Tutti i manoscritti sottoposti a Human Genome Variation ricevono una rigorosa revisione e controlli di qualità da parte di revisori e redattori prima della pubblicazione. Gli autori del Data Report controllano le informazioni sulla descrizione della mutazione con il mutalyzer name checker (https://mutalyzer.nl) o con il relativo sistema di controllo della descrizione per assicurarsi che la descrizione segua la nomenclatura HGVS prima della presentazione.

I dati raccolti per la banca HGV comprendono [51]:

- Nomi dei geni
- Codice OMIM
- Tipo di mutazione
- Patologia / Fenotipo
- GenBank accession number
- Mutazione (formato HGVS)
- Alterazioni a livello di proteina
- Variazione del Codone / Base azotata
- Cromosoma
- Posizione genomica
- Mutazione de novo / ereditata
- Regione
- Etnicità/ Popolazione
- Referenza

4.1.23. *Ensembl Variation*

Il database Ensembl Variation [52] memorizza le aree del genoma che differiscono tra i singoli genomi ("varianti") e, dove disponibili, le informazioni sulle malattie associate e sul fenotipo.
Esistono diversi tipi di varianti per diverse specie:

- Single Nucleotide Polymorphisms (SNPs)
- Short Nucleotide Insertions e/o deletions
- Longer variants classificate come varianti strutturali (comprese le CNVs)

4.1.24. *dbSNP*

Le varianti di sequenza sono responsabili delle caratteristiche fenotipiche individuali, compresa la propensione di una persona verso disturbi complessi come le malattie cardiache e il cancro.
Il database Single Nucleotide Polymorphism (dbSNP) (https://www.ncbi.nlm.nih.gov/snp) è un archivio di pubblico dominio per una vasta collezione di polimorfismi genetici semplici. Questa raccolta di polimorfismi comprende sostituzioni di nucleotidi a base singola (noti anche come polimorfismi a nucleotide singolo o SNPs), delezioni o inserzioni multibase su piccola scala (chiamate anche polimorfismi di inserimento di delezione o DIPs), e inserzioni di elementi trasponibili e microsatelliti (chiamate anche brevi ripetizioni in tandem o STRs). Ogni voce dbSNP include il contesto di sequenza del polimorfismo (cioè, la sequenza fiancheggiante), la frequenza di occorrenza del polimorfismo (per popolazione o individuo), e i metodi sperimentali, i protocolli e le condizioni utilizzate per testare la variazione.

4.1.25. *Genecards*

GeneCards (https://www.genecards.org) raccoglie informazioni complete e di facile utilizzo su tutti i geni umani annotati e predetti. Le informazioni (tra cui informazioni genomiche, trascrittomiche, proteomiche, genetiche, cliniche e funzionali) sono estratte automaticamente e integrate da una miriade di fonti di dati, dando luogo ad una scheda dettagliata per ciascuna delle decine di migliaia di entry di geni umani [53].
Il database consente ai ricercatori di navigare e interagire efficacemente con l'ampio universo di geni umani, malattie, varianti, proteine, cellule e pathway biologici. Al suo interno sono riportati anche collegamenti gene-malattia tramite MalaCards e pathway biologici combinati tramite PathCards, così come informazioni sui farmaci e l'espressione del proteoma.

4.1.26. *DRUGBANK*

La banca dati DrugBank (https://www.drugbank.ca) è una banca dati online completa, liberamente accessibile, contenente informazioni sui farmaci e sui loro target. Come risorsa bioinformatica e cheminformatica, DrugBank combina dati dettagliati sui farmaci (cioè chimici, farmacologici e farmaceutici) con informazioni complete sui target dei farmaci (sequenza, struttura e pathway). La vasta gamma di dati sui farmaci e sui loro target ha permesso la scoperta e la riutilizzazione di un certo numero di farmaci esistenti per il trattamento di malattie rare e recentemente identificate.
Ogni voce della DrugCard contiene più di 200 campi di dati, metà delle informazioni è dedicata ai dati farmacologici/chimici e l'altra metà ai dati relativi al target di farmaci o alle proteine [54].
DrugBank è offerta al pubblico come risorsa liberamente disponibile; tuttavia l'uso e la ridistribuzione dei dati, in tutto o in parte, per scopi commerciali (compreso l'uso interno) richiede una licenza.

4.1.27. *Toxin and Toxin Target Database (T3DB)*

Il Toxin and Toxin Target Database (T3DB), o Toxic Exposome Database (http://www.t3db.ca/), è una risorsa bioinformatica che combina dati dettagliati sulle tossine con informazioni complete sui bersagli tossici. Ogni record di tossine (ToxCard) contiene oltre 90 campi di dati e contiene informazioni quali proprietà chimiche e descrittori, valori di tossicità, interazioni molecolari e cellulari e informazioni mediche. Queste informazioni sono state estratte da oltre 18000 fonti, che comprendono altre banche dati, documenti governativi, libri e letteratura scientifica.
L'obiettivo del T3DB è quello di fornire meccanismi di tossicità e proteine target per ciascuna tossina. Questa duplice natura del T3DB, in cui i record dei bersagli delle tossine e delle tossine stesse sono collegati interattivamente in entrambe le direzioni, lo rende unico tra i database esistenti. È interrogabile e supporta ricerche estese su testi, sequenze, struttura chimica e interrogazioni relazionali. Le potenziali applicazioni del T3DB includono la previsione del metabolismo delle tossine e la previsione dell'interazione tossina/farmaco rendendolo applicabile a vari campi.

4.1.28. *DGIdb (Drug Gene Interaction Database)*

Il Drug-Gene Interaction database (DGIdb) (http://www.dgidb.org/) estrae informazioni da risorse esistenti (documenti, database e risorse web) che generano ipotesi su come i geni mutati potrebbero essere mirati terapeuticamente o prioritizzati per lo sviluppo di farmaci. Il riscontro di un'interazione farmaco-gene o di una categoria potenzialmente farmacologicamente accettabile non indica necessariamente l'efficacia (o la

sua assenza) di alcun farmaco o regime terapeutico. Un ritrovamento di nessuna interazione o di nessuna categoria potenzialmente farmacologicamente attiva non indica necessariamente la mancanza di efficacia di qualsiasi regime terapeutico o di trattamento.

DGIdb standardizza i contenuti provenienti da 30 fonti diverse e consente una navigazione avanzata, una ricerca e un filtering per facilitare l'accesso attraverso un'interfaccia utente web intuitiva o tramite interfaccia API [55].

DGIdb è un database con un'interfaccia web (www.dgidb.org) open source disponibile su GitHub (https://github.com/griffithlab/dgi-db). È possibile effettuare il dump di dati SQL del database dal seguente link http://dgidb.org/downloads.

4.1.29. *PharmGKB*

La Pharmacogenomics Knowledgebase (PharmGKB) (https://www.pharmgkb.org/) è una risorsa che raccoglie, cura e diffonde informazioni sull'impatto delle varianti genetiche umane sulle risposte ai farmaci. Fornisce informazioni clinicamente rilevanti, comprese le linee guida sul dosaggio, le etichette annotate dei farmaci e le associazioni gene-farmaco potenzialmente attivabili e le relazioni genotipo-fenotipo. I curatori assegnano i livelli di evidenza alle associazioni variante-farmaco utilizzando criteri ben definiti basati su un'attenta revisione della letteratura. PharmGKB è un'utile fonte di informazioni di alta qualità a supporto di progetti di implementazione di farmaci personalizzati [56].

4.1.30. *KEGG Drug Database*

KEGG DRUG (https://www.genome.jp/kegg/drug) è una risorsa di informazioni sui farmaci approvati in Giappone, Stati Uniti ed Europa, unificata sulla base della struttura chimica e/o della componente chimica, e associata a target terapeutico, enzima metabolizzante e altre informazioni della rete di interazione molecolare.

Ogni entry KEGG DRUG è identificata dal numero D che distingue la struttura chimica delle sostanze chimiche o il componente chimico delle ricette. È associata alle seguenti informazioni (ad esempio, D01441):

- Nomi generici (associati alle strutture chimiche)
- Nomi commerciali rappresentativi
- Link alle informazioni sulle etichette dei farmaci approvati dalla FDA in DailyMed
- Struttura chimica, componente chimica, sequenza peptidica

- Classe farmacologica definita da KEGG DGROUP
- Categoria terapeutica, codice ATC, gruppo chimico definito da KEGG DGROUP
- Attività, efficacia e malattie indicate
- Le molecole bersaglio nel contesto delle mappe di percorso KEGG
- Enzimi metabolizzanti e trasportatori di farmaci
- Altre molecole interagenti come l'inibizione e l'induzione dell'enzima/trasportatore
- Dati sulle interazioni avverse farmaco-farmaco (raccolti dai foglietti illustrativi giapponesi)
- Storia dello sviluppo dei farmaci rappresentata come mappa della struttura di KEGG DRUG
- Informazioni sulla classificazione dei farmaci nella gerarchia BRITE
- Collegamenti a banche dati esterne

4.1.31. *COSMIC (Catalog Of Somatic Mutations In Cancer)*

COSMIC (Catalogue of Somatic Mutations in Cancer) (https://cancer.sanger.ac.uk/cosmic) è una risorsa dettagliata e completa utile per esplorare l'effetto delle mutazioni somatiche nel cancro umano.
COSMIC include quasi 6 milioni di mutazioni codificanti in 1,4 milioni di campioni di tumore, tratte da oltre 26.000 pubblicazioni. Oltre alle mutazioni codificanti, COSMIC copre tutti i meccanismi genetici attraverso i quali le mutazioni somatiche promuovono il cancro, comprese le mutazioni non codificanti, le fusioni geniche, le varianti di numero di copia e le mutazioni di resistenza ai farmaci. COSMIC è principalmente curata a mano, garantendo qualità, accuratezza e acquisizione di dati descrittivi [57].
Tutti i dati COSMIC sono disponibili come file di download. I file scaricabili sono gratuiti per uso accademico; l'uso commerciale richiede una licenza.

4.1.32. *DiseaseEnhancer Database*

Studi di sequenziamento su larga scala hanno evidenziato l'esistenza di varianti genetiche che si verificano all'interno degli enhancers, elementi in grado di regolare l'espressione dei geni attraverso interazioni a lungo raggio sulla cromatina. Tali varianti potrebbero influire sulla regolazione degli enhancer modificando i legami dei fattori di trascrizione o il dirottamento degli enhancer e, a loro volta, dare un contributo essenziale alla progressione della malattia.
Il database DiseaseEnhancer fornisce una mappa completa degli enhancers associati alla malattia. Le informazioni sui geni target disegregolati da enhancer sono raccolti manualmente insieme ai tipi di varianti associate (compresa la variante a singolo nucleotide, le mutazioni somatiche, l'alterazione del numero di indel e del numero di copie) e i legami dei fattori di trascrizione interessati. Ulteriori dati sul genoma sono

stati integrati nel database DiseaseEnhancer per aiutare a caratterizzare gli enhancers associati alla malattia [58].

Le caratteristiche del database includono informazioni di base sull'enhancer (ad es. localizzazione genomica e geni bersaglio), i tipi di malattia associati, le varianti associate sull'enhancer e i loro fenotipi mediati (ad es. guadagno/perdita di enhancer e le alterazioni dei legami del fattore di trascrizione). Il database è interrogabile ed è possibile esportare i risultati della ricerca in un file o scaricare il database completo. DiseaseEnhancer fornisce una strada promettente per i ricercatori per facilitare la comprensione di una migliore regolazione della patogenesi delle malattie e identificare nuovi biomarcatori per la diagnosi e la terapia della malattia [59].

4.1.33. *LncRNADisease database*

I long non coding RNA (lncRNA) rappresentano gli RNA che hanno una dimensione >200 nucleotidi e che non hanno la capacità di codificare proteine. Sebbene gli lncRNAs siano tra le molecole meno conosciute, studi emergenti hanno rivelato che gli lncRNAs giocano un ruolo critico in una vasta gamma di processi biologici e sono associati ad una serie di malattie, come il cancro, le malattie cardiovascolari e le malattie neurodegenerative. Il database LncRNADisease (http://www.lncrnadb.org) non è solo una risorsa che ha curato i dati sperimentalmente supportati dall'associazione lncRNA-malattia, ma anche una piattaforma che ha integrato uno o più strumenti per la predizione di nuove associazioni malattia - lncRNA. Inoltre, il database LncRNADisease ha anche curato le interazioni con lncRNA a vari livelli, tra cui proteine, RNA, miRNA e DNA [60].

Attualmente gli utenti possono:

- navigare e scaricare i dati sperimentalmente supportati dall'associazione lncRNA-malattia;
- prevedere potenziali malattie associate per un nuovo lncRNA in base al suo contesto genomico;
- scaricare le sequenze di lncRNA;
- inviare nuove voci al database di lncRNADisease.

Ogni voce del database contiene informazioni di riferimento sull'RNA, incluse sequenze, informazioni strutturali, contesto genomico, espressione, localizzazione subcellulare, conservazione, prove funzionali e altre informazioni rilevanti. lncRNAdb può essere ricercato interrogando i nomi e gli alias, le sequenze, le specie e i geni codificanti associati, così come i termini contenuti nelle annotazioni, come i tessuti in cui sono espresse le trascrizioni e le malattie associate. Inoltre, lncRNAdb è collegato all'UCSC Genome Browser per la visualizzazione e al Noncoding RNA Expression Database (NRED) per la visualizzazione di informazioni provenienti da una varietà di fonti. lncRNAdb fornisce una piattaforma per la raccolta della letteratura

relativa a lncRNAs e la loro associazione con altri elementi genomici [61]. lncRNAdb è disponibile all'indirizzo: http://www.lncrnadb.org/

4.1.34. *DECIPHER*

Molti pazienti affetti da Malattie Rare ospitano varianti genomiche (varianti di sequenza o varianti del numero di copie) che, interrompendo la normale espressione genica, portano alla malattia. Tuttavia, molte varianti sono nuove o estremamente rare, rendendo problematica l'interpretazione clinica e incerte le correlazioni genotipo-fenotipo-fenotipo. L'identificazione di pazienti che condividono varianti in un dato gene e che hanno caratteristiche fenotipiche in comune porta ad una maggiore certezza nella natura patogena del gene e consente di definire il ruolo dei nuovi geni nello sviluppo e nella malattia. Inoltre, l'analisi del tipo di variante genomica e delle sue conseguenze (es. perdita di funzione o guadagno di funzione) permette di comprendere il meccanismo della malattia e i potenziali bersagli terapeutici.
DECIPHER (Database of genomiC varIation and Phenotype in Humans using Ensembl Resources) (http://www.sanger.ac.uk/PostGenomics/decipher) è un database interattivo che incorpora una suite di strumenti progettati per facilitare l'interpretazione delle varianti genomiche. DECIPHER migliora la diagnosi clinica recuperando informazioni da una varietà di risorse bioinformatiche rilevanti per la variante trovata nel paziente. La variante del paziente viene visualizzata nel contesto sia della variazione normale che della variazione patogena riportata in quel locus, facilitando così l'interpretazione.

4.1.35. *Registro tumori*

I Registri Tumori (RT) sono strutture deputate alla raccolta e registrazione di tutti i tumori incidenti in un determinato territorio. Il ruolo primario di un Registro Tumori è quello di costituire e gestire nel tempo un archivio di tutti i nuovi casi di tumore diagnosticati ed assicurare che la registrazione dei dati avvenga in modo rigoroso e con carattere continuativo e sistematico.
La maggior parte dei registri italiani sono registri di popolazione ovvero, raccolgono i dati relativi alle malattie tumorali di tutti i residenti di un determinato territorio (può essere una singola città o un'intera regione, una provincia o il territorio di una ASL).
Accanto ai Registri generali di popolazione, vi sono i Registri specializzati, che raccolgono informazioni su un singolo tipo di tumore o su specifiche fasce di età o su tumori professionali [62].

La banca dati AIRTUM è l'archivio nazionale, ospitato presso l'Istituto per lo Studio e la Prevenzione e la Rete Oncologica di Firenze, in cui confluiscono tutti i dati sui tumori raccolti dai singoli registri accreditati [63]. Attualmente vi sono depositate informazioni su:

- oltre 3 milioni di malati di tumore (3.521.574 tumori maligni diagnosticati nel periodo 1967-2015)
- oltre un milione di deceduti per tumore (1.832.445 decessi avvenuti nel periodo 1967-2015).

4.2. Riepilogo tipologie database individuate

La tabella che segue riepiloga i risultati dell'analisi delle banche dati individuate riportando, per ciascun database, la principale tipologia di dati in essi contenuti. Le categorie di informazioni individuate sono state:

- Dati clinici: riferiti ad informazioni su pazienti recuperate da cartelle cliniche o dati statistico-epidemiologici
- Dati genetici: contenenti informazioni circa i geni e/o le relative varianti coinvolti in una determinata patologia
- Dati su patologie: contenenti informazioni utili a caratterizzare, sotto diversi aspetti, le patologie
- Dati su farmaci: ossia farmaci utilizzati per la cura di patologie specifiche o riportanti interazioni con altri farmaci
- Dati su sostanze chimiche o tossiche: utili per determinare se l'esposizione a specifiche sostanze chimiche sia in responsabile o meno dell'insorgenza di patologie
- Dati su fattori ambientali: utili per determinare l'associazione con l'insorgenza di patologie
- Dati relativi a pathway biologici: utili per dettagliare e comprendere meglio i meccanismi molecolari di insorgenza delle patologie o di azione dei farmaci
- Dati su stili di vita e/o alimentazione: utili per determinare il contributo nell'insorgenza di patologie o nell'efficacia dell'azione farmacologica
- Dati utili per la medicina personalizzata: per determinare gli elementi necessari in grado di ottimizzare una cura in base alle specifiche caratteristiche di un individuo
- Dati utili all'individuazione di correlazioni: contenuti in database che riportano esplicitamente gli elementi in grado di corroborare la correlazione, ad es., tra una patologia e lo stile di vita, l'alimentazione, l'esposizione a sostanze tossiche, la presenza di specifiche varianti genetiche, etc.

Quasi tutti i database individuati sono popolati con dati provenienti da più di una delle categorie suddette.

Tabella 3 Confronto tra i database pubblici contenenti informazioni utili per la correlazione di informazioni e la definizione di modelli predittivi

Nome database / risorsa	Tipologia dati									
	Clinici	Genetici	Patologie	Farmaci	Sostanze chimiche	Ambientali	Pathway	Stile di vita / alimentaz.	Medicina personalizzata	Correlazioni
THIN DATABASE	X									
QRESEARCH	X									
MALACARDS		X	X	X			X			X
DISEASES DATABASE			X	X						X
KEGG DISEASE DATABASE		X	X	X						X
DISEASES		X	X							
DISGENET		X	X							
OMIM		X	X							
REACTOME DISEASES			X				X			X
COMPARATIVE TOXICOGENOMICS DATABASE (CTD)		X	X		X	X	X			X
NUTRITION EVIDENCE								X		
VMH DATABASE			X					X		X
CHE TOXICANT AND DISEASE DATABASE			X		X					X
AGENCY FOR TOXIC SUBSTANCES AND DISEASE REGISTRY			X		X					X
EUROSTAT HEALTH	X							X		X
PESTICIDE-INDUCED DISEASES DATABASE			X		X					X
GENE-ENVIRONMENT INTERACTIONS (GXE) DATABASE		X	X			X		X		X
CLINVAR		X	X							
HUMSAVAR		X	X							
EDGAR		X	X							
GWAS		X	X							

CATALOG										
HGV DATABASE		X	X							
ENSEMBL VARIATION		X	X							
DBSNP		X	X							
GENECARDS	X	X	X							
DRUGBANK				X						
TOXIN AND TOXIN TARGET DATABASE (T3DB)					X					
DGIDB (DRUG GENE INTERACTION DATABASE)		X		X					X	X
PHARMGKB		X		X					X	X
KEGG DRUG DATABASE				X					X	
COSMIC (CATALOG OF SOMATIC MUTATIONS IN CANCER)		X								
DISEASEENHANCER DATABASE		X	X							X
LNCRNADISEASE DATABASE		X	X							X
DECIPHER		X								
REGISTRO TUMORI	X									

4.3. Considerazioni sull'analisi dei database

Dall'analisi della Tabella 3 è possibile notare che la maggior parte dei database pubblici che possono fornire informazioni di correlazione tra patologie ed altri fattori sono quelli che riportano anche informazioni di tipo genetico. Il fatto che vi siano così tanti database che riportano contemporaneamente informazioni su patologie e profilo genetico è legato strettamente al fatto che gran parte delle patologie o predisposizioni sono causate da varianti genetiche. Di conseguenza, le fonti dati che possono essere più informative per l'individuazione di correlazioni o per la definizione di modelli predittivi sono quelle che riportano anche altre tipologie di informazioni, ma soprattutto quelle che contengono informazioni curate su correlazioni note tra i vari dati.

Tra i database pubblici contenenti informazioni cliniche sul paziente, quelli considerati più interessanti per via della dimensione del campione di pazienti coinvolto sono stati THIN Database (contenente dati di cartelle cliniche elettroniche di circa 11 milioni di pazienti provenienti da più di 500 medici di medicina generale) e Qresearch (derivato dalle cartelle cliniche pseudonimizzate di oltre 24 milioni di pazienti). Non è possibile, tuttavia, adottare tali database poiché le informazioni raccolte possono essere utilizzate solo con determinate restrizioni da enti di ricerca.

Altri database pubblici analizzati sono stati quelli che riportano correlazioni tra patologie, profilo genomico e di espressione, esposizione a sostanze chimiche, nutrizione, farmaci, etc. Tra i database analizzati quelli che sono in grado di fornire dati utili alla costruzione dei modelli predittivi sono:

- MalaCards: riporta varie informazioni sulle patologie come sintomi, farmaci, geni coinvolti, trial clinici condotti, patologie correlate, etc.
- KEGG Disease: fornisce informazioni sulle patologie focalizzandosi sull'elemento che "perturba" il pathway molecolare coinvolto
- DISEASES: riporta le evidenze sulle associazioni geni-malattia individuate mediante text-mining degli abstract di letteratura scientifica
- Comparative Toxicogenomics Database (CTD): fornisce informazioni accurate sulle interazioni tra sostanze chimiche e geni o patologie

5. Analisi delle fonti dati aziendali utilizzabili per i modelli predittivi

Le fonti aziendali potenzialmente utilizzabili per la definizione di modelli predittivi di insorgenza di patologie croniche o per la correlazione di informazioni sono quelle che stanno alla base delle soluzioni a catalogo dell'offerta di Exprivia per la sanità racchiuse nella suite e4Cure.

e4Cure è la suite delle soluzioni di Exprivia per la sanità che permette la gestione del paziente dal suo ingresso nella struttura sanitaria fino alle dimissioni ed include la CCE (Cartella Clinica Elettronica) che digitalizza e rende disponibile l'intero percorso sanitario del paziente, inclusa la diagnostica di laboratorio attraverso dati e immagini. Tale sistema informativo assicura la continuità assistenziale semplificandola grazie alla condivisione dei dati tra strutture diverse, consentendo di non replicare inutilmente indagini recenti a vantaggio dello stesso utente, oppure di confrontarle in caso di rapida evoluzione dello stato clinico a tutela del paziente, del medico e della struttura.

L'ottimizzazione del sistema sanitario attraverso e4cure consiste nell'adozione di diverse soluzioni flessibili, integrate nel sistema generale e suscettibili di adeguamenti in ragione delle nuove necessità strutturali, socio-sanitarie o emergenziali.

Tra le soluzioni della suite, quelle che possono fornire le informazioni adeguate allo scopo di definire modelli predittivi sono:

- eRepository
- eWard
- eVisit
- hMonitor
- hAssistant
- eLifeCare

5.1. eRepository

eRepository è la soluzione per l'archiviazione e la condivisione dei documenti clinici elettronici prodotti nei contatti tra i pazienti e le strutture ospedaliere e sanitarie durante il percorso diagnostico terapeutico.

eRepository consente la visibilità completa di tutte le tipologie di documenti prodotti dalle applicazioni utilizzate dalla struttura sanitaria. Le informazioni raccolte - dati clinici, immagini radiografiche, tracciati

biometrici e referti dei servizi diagnostici (laboratorio, radiologia, anatomia patologica) - entrano a far parte del repository che sarà poi consultabile con semplicità.

eRepository contiene la storia sanitaria del paziente, rendendo disponibili risultati, immagini, documenti e ogni altra informazione sull'attività clinica, utile per la cura della persona.

Il repository multimediale punta a ottimizzare la ricerca delle informazioni, per garantire la continuità della cura dentro e fuori l'ospedale, consentendo ai reparti e ai servizi radiologici di gestire e distribuire elettronicamente immagini e referti, con notevole risparmio di tempo e di costi.

La sua utilità è data dal fatto che, potendo il paziente usufruire indifferentemente delle diverse strutture sanitarie sul territorio, diventa importante la consultazione on line della sua storia clinica da parte delle diverse strutture sanitarie.

eRepository è predisposto per essere utilizzato come piattaforma operante a livello intra-ospedaliero, oppure in una configurazione inter-ospedaliera, consentendo di raccogliere i dati relativi al paziente indipendentemente dalla struttura sanitaria in cui il cittadino si reca, e arricchendo il suo profilo clinico di tutte le informazioni che lo riguardano.

Per quanto riguarda gli operatori, la disponibilità immediata delle immagini radiografiche e dei documenti clinici in reparto crea nuovo valore per il personale sanitario e consente maggiore tempestività nelle decisioni cliniche.

5.2. eWard

eWard è la cartella clinica elettronica a supporto del personale medico ed infermieristico per la gestione dei processi di cura del paziente ricoverato. Nasce dall'esigenza di fornire uno strumento per la gestione organica e strutturata dei dati relativi al ricovero del paziente, garantendo il supporto dei processi clinici e assistenziali e, al tempo stesso, ottimizzando la cura del paziente mediante la condivisione ed il recupero dei dati clinici.

La soluzione fornisce tutti gli strumenti per garantire la gestione dell'intero processo diagnostico-terapeutico–assistenziale del paziente in reparto che, di norma, comprende informazioni relative all'assessment clinico (anamnesi) e infermieristico (rilevazione dei fabbisogni infermieristici), esame obiettivo, diario clinico integrato (medico e infermieristico), referti di prestazioni ambulatoriali e di esami diagnostico-specialistici, (ad es. laboratorio, anatomia patologica, radiologia, etc.), gestione delle attività di nursing e gestione della lettera di dimissione e dei vari documenti clinico/amministrativi correlati al ricovero.

Man mano che vengono svolte le attività previste durante il ricovero, i ritorni informativi (referti, risultati, ecc.) sono acquisiti come parti dell'evento e contribuiscono all'aggiornamento dinamico del patient record.

eWard è stato progettato con l'obiettivo di fornire una piattaforma unica a livello di azienda sanitaria, in grado di soddisfare i requisiti comuni a tutti i reparti ospedalieri, e di configurarsi e specializzarsi in base alle specifiche esigenze dei reparti.

La soluzione è costituita, quindi, da una serie di moduli con forte caratterizzazione "aziendale" (definizione piano terapeutico, gestione somministrazioni, diario infermieristico, ecc.) e da un insieme di moduli che possono adattarsi alla disciplina nella quale viene utilizzata.

Per questa specifica peculiarità, la soluzione è stata sviluppata attraverso il componente Form Manager, uno strumento interattivo per la realizzazione di specializzazioni cliniche che consente di disegnare documentazione clinica contestuale al ricovero (esame obiettivo, cartella infermieristica e lettera di dimissione), con sofisticati livelli di specializzazione e con elevata flessibilità.

5.3. eVisit

eVisit è la soluzione Exprivia per la refertazione e la documentazione clinica orientata all'informatizzazione di tutti gli ambulatori e servizi diagnostici. Il prodotto è progettato con l'obiettivo di fornire una piattaforma unica per l'intera azienda sanitaria in grado di soddisfare i requisiti di tutta la specialistica ambulatoriale e strumentale attraverso l'impiego di tecnologie e soluzioni all'avanguardia. Le principali caratteristiche di eVisit sono:

- possibilità di realizzare specializzazioni cliniche mediante il componente Form Manager, attraverso il quale è possibile disegnare in modalità interattiva schede di refertazione e documentazione clinica multimediali;
- un ambiente evoluto di integrazione con le apparecchiature diagnostiche che permette di estendere il sistema PACS al trattamento di segnali biometrici (quali ECG, Holter etc.), all'acquisizione da diagnostiche non DICOM (quali endoscopi, laparoscopi, etc.), fino all'importazione automatica di misure calcolate dalle apparecchiature;
- una gestione dell'anamnesi strutturata, generale e specialistica, integrata con gli altri moduli Exprivia della suite e4cure per la gestione del reparto;

eVisit dispone di numerosi moduli specialistici, tra i quali:

- Ecocardiografia;
- Elettrofisiologia e Aritmologia;
- Chirurgia Vascolare
- Emodinamica;
- Ambulatorio di Cardiologia ed ECG;

- Screening Mammografico;
- Gastroenterologia;
- Ostetricia e Ginecologia.

5.4. hMonitor

hMonitor è la soluzione che si pone l'obiettivo di fornire un supporto informatico utile ai fini di una gestione del processo di monitoraggio costante e protratto nel tempo dello stato di salute di pazienti affetti da malattie del metabolismo, quali l'obesità o il diabete, nella forma anche pre-patologica. Altro obiettivo rilevante della soluzione è la costruzione di un'ampia ed articolata base dati per l'analisi su vasta scala delle malattie del metabolismo e la valutazione delle caratteristiche utili alla caratterizzazione clinica dei fenomeni e all'individuazione precoce dei soggetti a rischio. hMonitor consente la remotizzazione e il monitoraggio continuo per periodi più o meno prolungati dello stato di salute del paziente; tale remotizzazione è fondata sul connubio di diverse tecnologie ICT e permette di ridurre enormemente i costi dovuti all'ospedalizzazione di tali pazienti che oggi è necessaria in assenza di queste tecnologie abilitanti.
L'utilizzo di nuovi pattern per il processo di monitoraggio delle malattie è uno degli aspetti innovativi della soluzione hMonitor, possibile grazie alla flessibilità offerta dalle tecnologie abilitanti adottate quali smartphone, qrcode, wireless networking, sensoristica non intrusiva, MEMS (Micro Electro Mechanical Systems).
Il sistema di monitoraggio remoto di pazienti obesi/diabetici previsto da hMonitor è costituito da un'APP per smartphone collegata ad un sistema centrale di raccolta dati.
La soluzione, attraverso l'APP, permette di:

- indicare gli alimenti, i farmaci assunti e l'attività fisica praticata
- rilevare peso, pressione, pulsazioni, glicemia, mediante dispositivi di misura la cui rilevazione può essere letta dallo smartphone in modalità wireless anche senza digitazione manuale dei valori;
- visualizzare graficamente l'andamento del proprio stile di vita;
- incentivare i pazienti a seguire corretti stili di vita, grazie anche al controllo costante da parte del care giver;
- comunicare con i care giver e inviare in tempo reale al sistema centrale le informazioni rilevate;
- ricevere in tempo reale e in mobilità variazioni del programma individuale (alimentazione, attività fisica e terapia farmacologica) definite dal care giver;
- ricevere avvisi nel caso in cui i parametri rilevati siano fuori norma.

Attraverso l'applicazione di back office, hMonitor consente al Care Giver di:

- gestire l'anagrafica, l'anamnesi e il programma individuale dell'assistito;

- visualizzare grafici che descrivono l'andamento (aggiornato in tempo reale) dello stile di vita dell'assistito;
- ricevere messaggi di attenzione sullo stato di salute dell'assistito;
- comunicare con l'assistito e correggere in qualsiasi momento l'alimentazione, l'attività fisica consigliata o la terapia farmacologica prescritta;
- esportare i dati raccolti per analisi statistico-epidemiologiche.

5.5. hAssistant

La soluzione Mobile Assistance for Care (hAssistant) rientra nell'ambito delle evoluzioni tecnologiche finalizzate ad organizzare i piani d'intervento di assistenza sanitaria domiciliare agli anziani e ad automatizzare l'acquisizione delle informazioni circa i parametri vitali dell'assistito con l'uso combinato di sensori di parametri vitali e smartphone Android.
hAssistant consente la registrazione degli accessi domiciliari/residenziali effettuati dall'equipe domiciliare attraverso l'utilizzo di strumenti mobile e il monitoraggio continuo e per periodi più o meno prolungati dello stato di salute del paziente; il controllo da remoto si basa sulla gestione pianificata degli interventi in base alle disponibilità dei soggetti erogatori.
La soluzione hAssistant si propone come strumento utile e funzionale nell'ambito dell'Assistenza Domiciliare Integrata con tecnologia mobile attraverso la quale:

- gli erogatori di assistenza visualizzeranno sullo smartphone le prestazioni da erogare presso il domicilio del paziente e le relative informazioni cliniche;
- una centrale operativa potrà gestire la pianificazione delle prestazioni da erogare attraverso un'applicazione di back office e i responsabili potranno controllare e valutare l'operato degli erogatori;
- gli erogatori possono comunicare con la centrale operativa e ricevere in tempo reale variazioni del Piano Assistenziale Individuale definito dall'UVM;

Attraverso l'applicazione di back office, la Centrale Operativa potrà:

- gestire l'anagrafica, i Piani Assistenziali Individuali e i parametri rilevati dall'erogatore sull'assistito;
- inviare notifiche e localizzare sulla mappa gli erogatori, gestire le agende di disponibilità, consuntivare e effettuare statistiche sulle erogazioni con export dati;
- definire le equipe domiciliari;
- pianificare le attività a domicilio in base alle agende degli erogatori e visualizzare le pianificazioni programmate;

- acquisire le informazioni sugli assistiti ed erogatori da sistemi informativi esterni.

La soluzione hAssistant può essere utilizzata anche in contesti differenti da quello domiciliare; la capacità di acquisire informazioni da strumentazione per la misura di parametri vitali lo rende adatto anche in contesti ospedalieri nei casi di dimissioni protette.

5.6. eLifeCare

eLifeCare è la piattaforma di Telemedicina che contiene l'insieme di soluzioni e servizi forniti da Exprivia per Teleconsulto, Telerefertazione, Telepresenza e Telemonitoraggio a supporto di tutti gli operatori coinvolti nella cura e nel monitoraggio del paziente in cura a domicilio.

La piattaforma eLifeCare rivoluziona l'approccio alla cura domiciliare del paziente in quanto fornisce le infrastrutture tecnologiche e tutti i servizi necessari alla gestione completa e integrata di tutti i processi e i servizi di assistenza, disponibili e fruibili da qualsiasi dispositivo:

- Monitoraggio da remoto
- Telemedicina e Teleconsulto
- Monitoraggio approvvigionamento farmaci
- Sistemi di refertazione e archiviazione
- Storia clinica del paziente grazie a dossier sanitari e fascicoli elettronici

eLifeCare raccoglie la sfida della nuova Sanità Digitale che mette al centro il malato garantendo servizi continuativi che migliorano la qualità della vita e contestualmente aiutano a contenere i costi delle aziende sanitarie e degli ospedali mediante la deospedalizzazione del paziente.

La piattaforma si basa su una applicazione web attraverso la quale gli operatori di una Control Room hanno la possibilità di monitorare e gestire in tempo reale tutti i pazienti facendo da intermediario operativo tra lo stesso paziente e l'equipe o il medico specialista che ha in cura il malato.

La stessa applicazione è a disposizione, mediante smartphone o tablet, anche degli specialisti medici che possono accedere alla piattaforma per consultare i dati di cartella inerenti la terapia, i parametri vitali, etc. ed eventualmente modificare le terapie stesse condividendo specifici protocolli terapeutici oltre a consultare anche eventuali referti a supporto della terapia.

La piattaforma eLifeCare dispone di specifiche applicazioni mobile ad uso e utilizzo del paziente che permettono la partecipazione attiva dell'assistito, del care giver o dell'equipe domiciliare e che consentono la gestione da remoto di:

- Erogazione prestazioni a domicilio del paziente in base alla pianificazione dettata dai Piani Assistenziali.

- Rilevazione parametri vitali.
- Sessioni di video consulto.
- Geolocalizzazione paziente.
- Somministrazione terapia farmacologica.
- Richiesta approvvigionamento farmaci.

6. Bibliografia

[1] Linda A. Winters-Miner, "Seven ways predictive analytics can improve healthcare", https://www.elsevier.com/connect/seven-ways-predictive-analytics-can-improve-healthcare
[2] David Crockett, "4 Essential Lessons for Adopting Predictive Analytics in Healthcare", https://www.healthcatalyst.com/predictive-analytics-healthcare-lessons/
[3] "THIN Database", https://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/database
[4] "THIN data: Strengths & Limitations", https://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/database/pros-cons
[5] "Qresearch", http://www.qresearch.org/SitePages/What%20Is%20QResearch.aspx
[6] Abbasi et al., "Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study", BMJ 2012;345:e5900
[7] Schmidt MI et al., "Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study.", Diabetes Care. 2005 Aug;28(8):2013-8.
[8] Noble et al., "Risk models and scores for type 2 diabetes: systematic review", BMJ 2011;343:d7163
[9] Herder C, Kowall B, Tabak AG, Rathmann W., "The potential of novel biomarkers to improve risk prediction of type 2 diabetes", Diabetologia. 2014 Jan;57(1):16-29.
[10] Rahman et al., "A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study", Fam Pract. 2008 Jun;25(3):191-6. doi: 10.1093/fampra/cmn024. Epub 2008 May 30.
[11] Lindström J, Tuomilehto J., "The diabetes risk score: a practical tool to predict type 2 diabetes risk", Diabetes Care. 2003 Mar;26(3):725-31.
[12] Julia Hippisley-Cox, "Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore", BMJ 2009;338:b880
[13] Chen L et al., "AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures.", Med J Aust. 2010 Feb 15;192(4):197-202.
[14] Schmidt et al., "Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study", Diabetes Care. 2005 Aug;28(8):2013-8
[15] Rathmann et al., "Prediction models for incident type 2 diabetes mellitus in the older population: KORA S4/F4 cohort study.", Diabet Med. 2010 Oct;27(10):1116-23. doi: 10.1111/j.1464-5491.2010.03065.x.
[16] Wilson et al., "Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study.", Arch Intern Med. 2007 May 28;167(10):1068-74.
[17] Stern et al., "Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test?", Ann Intern Med. 2002 Apr 16;136(8):575-81.
[18] Cichosz et al., "Toward Big Data Analytics: Review of Predictive Models in Management of Diabetes and Its Complications.", J Diabetes Sci Technol. 2015 Oct 14;10(1):27-34. doi: 10.1177/1932296815611680.
[19] Looker HC, Nyangoma SO, Cromie DT, . Predicted impact of extending the screening interval for diabetic retinopathy: the Scottish Diabetic Retinopathy Screening programme. Diabetologia. 2013;56:1716-1725.

[20] Mehlsen J, Erlandsen M, Poulsen PL, Bek T. Individualized optimization of the screening interval for diabetic retinopathy: a new model. Acta Ophthalmol. 2012;90:109-114.
[21] Boyko EJ, Ahroni JH, Cohen V, Nelson KM, Heagerty PJ. Prediction of diabetic foot ulcer occurrence using commonly available clinical information: the Seattle Diabetic Foot Study. Diabetes Care. 2006;29:1202-1207
[22] Jansen HJ, Vervoort GMM, de Haan AFJ, Netten PM, de Grauw WJ, Tack CJ. Diabetes-related distress, insulin dose, and age contribute to insulin-associated weight gain in patients with type 2 diabetes mellitus: results of a prospective study. Diabetes Care. 2014; 37(10):2710-7.
[23] Balkau B, Home PD, Vincent M, Marre M, Freemantle N. Factors associated with weight gain in people with type 2 diabetes starting on insulin. Diabetes Care. 2014; 37(8):2108-13
[24] Cichosz SL, Christensen LL, Dencker M, Tarnow L, Almdal TP, Hejlesen O. Prediction of excessive weight gain in insulin treated patients with type 2 diabetes. 2015; Unpublished manuscript.
[25] Julia Hippisley-Cox et al., "Development and validation of risk prediction equations to estimate future risk of blindness and lower limb amputation in patients with diabetes: cohort study", BMJ 2015;351:h5441
[26] Julia Hippisley-Cox et al., "Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study", BMJ 2007;335:136
[27] Borsari et al., "Cardiovascular Risk Assessment in People Affected with Diabetes in Primary Care", https://www.intechopen.com/books/howtoreference/primary-care-in-practice-integration-is-needed/cardiovascular-risk-assessment-in-people-affected-with-diabetes-in-primary-care
[28] https://www.framinghamheartstudy.org/risk-functions/cardiovascular-disease/10-year-risk.php#
[29] Julia Hippisley-Cox et al., "Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study", BMJ 2017;357:j2099
[30] Conroy et al., "Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project.", Eur Heart J. 2003 Jun;24(11):987-1003.
[31] Caranci et al., "L'indice di deprivazione italiano a livello di sezione di censimento: definizione, descrizione e associazione con la mortalità.", Epidemiol Prev 2010; 34 (4): 167-176.
[32] Noa Rappaport et al., "MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search", Nucleic Acids Research, Volume 45, Issue D1, January 2017, Pages D877–D887, https://doi.org/10.1093/nar/gkw1012
[33] http://www.diseasesdatabase.com/content.asp
[34] http://www.diseasesdatabase.com/item_relationship_constraints.asp
[35] https://www.genome.jp/kegg/disease/
[36] Pletscher-Frankild S. et al., "DISEASES: Text mining and data integration of disease–gene associations", Methods. 2015 Mar;74:83-9. doi: 10.1016/j.ymeth.2014.11.020. Epub 2014 Dec 5.
[37] http://www.disgenet.org/dbinfo
[38] https://www.omim.org/help/faq
[39] https://reactome.org/what-is-reactome
[40] https://reactome.org/userguide/diseases
[41] https://bant.org.uk/centre-of-excellence/evidence-based-databases/nutrition-evidence-database/
[42] Alberto Noronha et al., "The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease", Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D614–D624, https://doi.org/10.1093/nar/gky992
[43] https://www.healthandenvironment.org/our-work/toxicant-and-disease-database/

[44] https://www.healthandenvironment.org/our-work/toxicant-and-disease-database/about-the-toxicant-and-disease-database
[45] https://www.atsdr.cdc.gov/toxprofiledocs/index.html
[46] https://ec.europa.eu/eurostat/web/health/overview
[47] https://www.beyondpesticides.org/resources/pesticide-induced-diseases-database/overview
[48] Smith R., "Beyond conflict of interest. Transparency is the key". BMJ. 1998;317(7154):291–292. doi:10.1136/bmj.317.7154.291
[49] Babbi, Giulia et al. "eDGAR: a database of Disease-Gene Associations with annotated Relationships among genes." BMC genomics vol. 18,Suppl 5 554. 11 Aug. 2017, doi:10.1186/s12864-017-3911-3
[50] https://www.ebi.ac.uk/gwas/docs/about
[51] https://www.nature.com/hgv/db
[52] https://www.ensembl.org/info/genome/variation/index.html
[53] Stelzer G. et al., "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses.", Curr Protoc Bioinformatics. 2016 Jun 20;54:1.30.1-1.30.33. doi: 10.1002/cpbi.5.
[54] https://www.drugbank.ca/about
[55] Kelsy C Cotto et al., "DGIdb 3.0: a redesign and expansion of the drug–gene interaction database", Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D1068–D1073, https://doi.org/10.1093/nar/gkx1143
[56] Whirl-Carrillo M. et al., "Pharmacogenomics knowledge for personalized medicine.", Clin Pharmacol Ther. 2012 Oct;92(4):414-7. doi: 10.1038/clpt.2012.96.
[57] John G Tate et al., "COSMIC: the Catalogue Of Somatic Mutations In Cancer", Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D941–D947, https://doi.org/10.1093/nar/gky1015
[58] http://biocc.hrbmu.edu.cn/DiseaseEnhancer/
[59] Guanxiong Zhang et al., "DiseaseEnhancer: a resource of human disease-associated enhancer catalog", Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D78–D84, https://doi.org/10.1093/nar/gkx920
[60] http://www.cuilab.cn/lncrnadisease
[61] Paulo P. Amaral et al., "lncRNAdb: a reference database for long noncoding RNAs", Nucleic Acids Research, Volume 39, Issue suppl_1, 1 January 2011, Pages D146–D151, https://doi.org/10.1093/nar/gkq1138
[62] https://www.registri-tumori.it/cms/pagine/cosa-sono-i-registri
[63] https://www.registri-tumori.it/cms/contenuto/la-banca-dati-airtum