



Special #3

Modelli matematici e intelligenza artificiale per la prevenzione del diabete di tipo 2

Alessia Paglialonga, Laura Azzimonti, Paolo Tieri

All'interno del progetto europeo **PRAESIIDIUM**, numerosi gruppi di ricerca uniscono le proprie competenze per sviluppare strategie innovative di prevenzione del diabete di tipo 2, combinando modelli matematici, simulazioni biologiche e strumenti di intelligenza artificiale spiegabile.

Lo speciale di questo numero, ancora una volta, è dedicato al racconto di tre esperti coinvolti nel progetto europeo **PRAESIIDIUM**:

Alessia Paglialonga Ricercatrice Senior al CNR - Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni (CNR-IEIIT) di Milano. Nel progetto **PRAESIIDIUM** il CNR-IEIIT contribuisce con due principali filoni di ricerca. Il primo riguarda la **spiegabilità degli algoritmi** (*AI explainability*), ovvero l'adozione di tecniche di spiegabilità sia intrinseche sia post-hoc per rendere interpretabili le decisioni dei modelli predittivi a clinici e pazienti. L'obiettivo è garantire un equilibrio fra **generalizzazione** (applicabilità a coorti ampie, come dataset di popolazioni caucasiche non a rischio) e



PROGETTO HORIZON EUROPE PRAESIIDIUM
Previsione in tempo reale del rischio di prediabete

SPECIAL #3

Laura Azzimonti
Docente e Ricercatrice Senior
SUPSI - IDSIA
Lugano, Svizzera

Paolo Tieri
Ricercatore Senior
CNR-IAC
Professore Bioinformatica e
Network Medicine
Università Sapienza Roma

Alessia Paglialonga
Ricercatrice Senior
CNR-IEIIT
Milano, Italia

SUPSI
IDSIA
Dalle Molle Institute for Artificial Intelligence
USI-SUPSI

CNR-IEIIT

personalizzazione (raccomandazioni specifiche basate su profili individuali), per rendere i modelli addestrati sui dati più trasparenti e comprensibili anche per i non esperti. Il secondo fronte riguarda lo sviluppo di **modelli matematici semplificati** per valutare l'impatto dell'attività fisica (durata, intensità, regolarità) sul rischio di sviluppare diabete nel lungo termine.

Laura Azzimonti Docente e Ricercatrice Senior presso **SUPSI-IDSIA**, Lugano Svizzera. All'interno del progetto sviluppa modelli di machine learning "physics-informed", che combinano conoscenza matematica del metabolismo con dati clinici reali. Questo approccio permette di costruire modelli più interpretabili e generalizzabili, fondamentali in ambito medico. In collaborazione con il CNR-IAC, SUPSI lavora a una versione più veloce del simulatore del diabete mellito tipo 2 (T2DM), capace di fornire previsioni personalizzate del rischio di diabete a basso costo computazionale, anche integrabili in un'applicazione.

Paolo Tieri Senior Researcher presso il **CNR-IAC – Istituto per le Applicazioni del Calcolo** e professore a contratto di **Bioinformatica e Network Medicine** all'Università Sapienza di Roma. Nel progetto **PRAESIIDIUM**, il CNR-IAC fornisce dati sintetici derivanti da un modello integrato del metabolismo e del sistema immunitario umano. Questo modello, calibrato su caratteristiche specifiche del paziente (dieta, attività fisica, parametri immunologici), consente di simulare l'evoluzione dello stato metabolico sia su periodi brevi (giorni) che su periodi prolungati (anni). I dati personalizzati così generati alimentano modelli di machine learning e deep learning sviluppati dai partner, rendendo possibili previsioni più precise e adattabili. Tali modelli puntano a descrivere l'organismo in un'ottica sistemica multi-livello, avvicinandosi progressivamente a simulazioni che includano scala cellulare, tessutale e d'organo.

Integrazione di modelli matematici e dati clinici per la personalizzazione del rischio di diabete

Il nostro compito come **SUPSI-IDSIA**, spiega **Laura Azzimonti**, è integrare i dati clinici e individuali dei pazienti con la conoscenza esperta racchiusa nel modello MT2D sviluppato dal **CNR** (coordinato da **Paolo Tieri**). Un esempio concreto è la valutazione dell'impatto dell'attività fisica sul rischio di sviluppare diabete di tipo 2.

I modelli che sviluppiamo sono progettati per descrivere l'evoluzione di diversi metaboliti su scale temporali sia brevi che lunghe, fino al dettaglio di variazioni indotte da un singolo pasto o da una sessione di esercizio fisico. Integrando questi modelli con i dati raccolti dal progetto, possiamo arricchire dataset limitati a specifiche popolazioni con informazioni aggiuntive, migliorando così la capacità di estrapolazione e di personalizzazione. In questo modo, i modelli non restano ancorati a un "paziente medio", ma vengono adattati alle caratteristiche del singolo individuo.

L'approccio di **PRAESIIDIUM** si fonda proprio su questa integrazione di fonti eterogenee: non un unico modello o metodo, ma una combinazione di strategie e tecniche di generazione ed elaborazione dei

dati. Per i modellisti, la sfida è duplice: da un lato, lavorare su modelli di popolazione; dall'altro, integrare le informazioni relative al singolo paziente (abitudini alimentari, livello di attività fisica, parametri clinici). Il percorso è graduale: si parte dai dati considerati essenziali dai clinici, per poi estendere e generalizzare quando possibile, oppure approfondire il dettaglio laddove necessario per descrivere meglio processi specifici.

L'obiettivo finale è sviluppare un sistema di valutazione e previsione **personalizzato, adattabile e sempre sotto la supervisione del medico**, che rimane il decisore ultimo nell'interpretazione e nell'applicazione clinica dei risultati.

I modelli di machine learning

In generale, spiega **Alessia Paglialonga**, i modelli di machine learning si sviluppano in due fasi principali. La prima riguarda **l'acquisizione e la selezione dei dati con cui allenare il modello**. Ad esempio, un modello di previsione del rischio viene addestrato su una coorte di pazienti e successivamente validato su una popolazione diversa, al fine di verificarne l'accuratezza predittiva.

La seconda fase consiste **nell'applicazione del modello ai dati di pazienti singoli**, in diversi momenti temporali, aggiornandone lo stato clinico. In questo processo le previsioni possono variare. I modelli più complessi, addestrati su grandi quantità di dati longitudinali, richiedono metriche robuste per monitorare costantemente l'accuratezza e una validazione continua su nuove popolazioni. In tal modo, il modello che evolve nel tempo può modificare le relazioni interne tra le variabili, restituendo per lo stesso paziente previsioni differenti rispetto a quelle fornite in passato.

Accanto a questi approcci, **modelli matematici** che descrivono l'evoluzione temporale di specifici biomarcatori collegati al rischio di diabete di tipo 2 permettono di studiare i cambiamenti nel tempo e di aggiornare in maniera dinamica lo stato del paziente.

Insieme a **Laura Azzimonti (SUPSI)** abbiamo condotto uno studio per stimare piani di attività fisica personalizzati con l'obiettivo di ridurre il rischio di diabete di tipo 2 in pazienti simulati. I piani venivano aggiornati progressivamente in base all'evoluzione clinica osservata nel tempo. Questo approccio ha mostrato come l'intervallo di prevenzione del rischio possa modificarsi dinamicamente in funzione dello stato del paziente e ha evidenziato come l'integrazione di diversi metodi migliori la precisione delle predizioni e l'adattabilità dei modelli.

La potenza di questi modelli, aggiunge **Tieri**, risiede nella capacità di simulare un'ampia gamma di traiettorie e scenari possibili, fornendo un repertorio di risposte teoriche che possono supportare il lavoro dei clinici. Il valore di **PRAESIIDIUM** sta proprio nella **flessibilità dei modelli, capaci di descrivere situazioni diverse e in continua evoluzione**.

Il valore dei modelli

L'accuratezza dei modelli varia a seconda dell'approccio. Le **equazioni differenziali** e, più recentemente, i modelli statistici di machine learning e deep learning possono raggiungere livelli teoricamente molto elevati: la possibilità di stimare numerosi parametri immunologici e metabolici consente di descrivere con dettaglio l'evoluzione dello stato del paziente. Più complessa è invece la validazione dei modelli ad agenti, che cercano di rappresentare sistemi molto ampi, a scala di organismo, e che richiedono ulteriori sviluppi.

Un livello di accuratezza superiore richiede simulazioni **multilivello**, capaci di descrivere processi dall'interno della cellula fino ai tessuti, agli organi e all'intero organismo, come fa il modello MT2D che integra un modello ad agenti del sistema immunitario con un modello a equazioni differenziali del metabolismo. Oggi questo obiettivo sembra raggiungibile, anche se è necessario ancora molto lavoro per il fine tuning e la validazione di un modello così complesso.

La validazione rimane cruciale, afferma **Paglialonga**, soprattutto quando si introducono nuovi dati, così come l'interazione con i medici, a cui è necessario chiarire anche i limiti del modello. Ogni previsione rappresenta un compromesso tra la complessità del sistema da descrivere, il numero di equazioni impiegate e i costi computazionali. Siamo ancora distanti da una validazione pienamente individuale, ma la verifica su trend di popolazione e con la letteratura scientifica fornisce un solido punto di riferimento. **In conclusione, non esiste un'unica metrica di accuratezza: dipende dal contesto e dall'obiettivo del modello.**

L'importanza dei dati e del dialogo con i decisori

La qualità, la quantità e la continuità dei dati sono essenziali, ma oggi mancano ancora raccolte sistematiche e digitalizzate in grado di integrarsi con la **cartella clinica elettronica**, segnala **Paglialonga**. Questo mette in evidenza l'urgenza del dialogo con decisori e stakeholder, affinché la **mappatura digitale su larga scala** diventi una priorità, non solo per la popolazione a rischio ma per la prevenzione di molte malattie croniche. Alcuni Paesi europei hanno già infrastrutture più avanzate, segnala **Tieri**, capaci di alimentare modelli predittivi e integrarli nei sistemi clinici come strumenti di supporto decisionale. Un progetto come **PRAESIIDIUM** può contribuire a sensibilizzare anche la politica sanitaria su questo fronte. Nel futuro, immaginiamo modelli in grado di adattarsi a ogni singolo dato presente nelle cartelle, sfruttando appieno questo patrimonio informativo.

Il rapporto con medici e pazienti

Oggi i clinici hanno familiarità con l'uso di algoritmi, mentre i pazienti mostrano atteggiamenti più eterogenei. Alcuni sono motivati dal sapere e intraprendono cambiamenti nello stile di vita; altri, invece, percepiscono le informazioni come fonte di ansia: *“Se non sono malato ora, perché preoccuparmi? Mi attiverò solo quando sarà necessario”*. Questo limite rimane anche in presenza di modelli predittivi molto accurati.

In realtà, la collaborazione **modello-clinico-paziente** è centrale. Il paziente stesso fornisce dati che alimentano il modello, e l'esperienza – come emerso con il COVID – mostra quanto sia cruciale la fiducia tra popolazione e sistema sanitario. Un modello funziona solo se i cittadini sono consapevoli dell'importanza di condividere dati, consentendo di allenare sistemi predittivi su un ventaglio più ampio di diversità. Ma questo richiede di considerare con attenzione lo stato della **fiducia pubblica** verso istituzioni mediche, sanitarie e scientifiche.

Open Science

I ricercatori non operano isolati nei laboratori: i modelli che sviluppiamo hanno come obiettivo il miglioramento della qualità della vita. La **trasparenza** della ricerca è per noi un principio cardine, perché la condivisione – tra laboratori, con i cittadini, con i decisori – è il motore dell'innovazione.

In **PRAESIIDIUM** questo approccio diventa concreto: condividere dati, feedback, risultati, impatti e aspettative non è solo buona pratica, ma la condizione necessaria per costruire soluzioni realmente utili e accettate. È questa l'essenza dell'**open science**: ricerca partecipata, co-creata e condivisa.

Contattaci:

Contattaci per maggiori informazioni



praesiidium.project@spindox.it



<https://praesiidium.spindoxlabs.com>



PRAESIIDIUM