

Università degli studi di Genova

Scuola superiore IANUA

Anno accademico 2022/2023

Concorso per Allievi dell'indirizzo IANUA in Scienze e
Tecnologie per la Società dell'Informazione

Andrei Apostu, Pietro Firpo, Sonia Spinelli



Non soltanto la cura, ma anche il benessere della persona, sono obiettivi a cui la medicina moderna è chiamata a rispondere in maniera sempre più interdisciplinare. In quest'ottica, l'Information and Communication Technology (ICT) gioca un ruolo fondamentale nel fornire strumenti di generazione, raccolta, conservazione, elaborazione e analisi di dati utili non solo a esigenze diagnostiche e prognostiche, ma anche nella programmazione di azioni personalizzate di prevenzione a lungo termine.

La nostra università, insieme ad altre istituzioni pubbliche e realtà industriali della Liguria, è chiamata ad un contributo importante in questa direzione, con la recente costituzione del primo "Centro Nazionale di Medicina Computazionale e Tecnologica" in Italia, che ha beneficiato di un ingente investimento di fondi PNRR. La Medicina Computazionale e Tecnologica congiunge infatti la ricerca biomedica e farmaceutica con gli ambiti propri della Scuola di Scienze Matematiche Fisiche e Naturali e della Scuola Politecnica.

In relazione alle esigenze della realtà locale introdotta e descritta nei riferimenti proposti, tenendo in considerazione le conoscenze e le competenze acquisite nell'ambito del proprio Corso di Studio e del percorso IANUA-STSI, si chiede di ragionare sulle sfide scientifiche e tecnologiche che dovranno essere affrontate dal Centro e a cui dovranno essere indirizzati opportuni investimenti.

Indice

1	Introduzione	3
2	La sfida tecnologica	3
2.1	I dati	3
2.2	La computazione	4
3	La sfida della privacy	5
3.1	L'importanza della privacy dei dati sanitari	5
3.2	Strategie di gestione della privacy dei dati	6
3.2.1	Consensi informati	6
3.2.2	Benefici per i partecipanti	6
3.2.3	Sicurezza dei dati	6
3.3	Promozione della partecipazione	6
3.3.1	Campagne di sensibilizzazione	6
3.3.2	Approccio "opt-in"	7
3.3.3	Trasparenza continua	7
4	La sfida comunicativa	7
4.1	Un team eterogeneo	8
4.2	Una piattaforma centralizzata	8
4.3	Una comunicazione efficace	8
5	In conclusione	9

1 Introduzione

Per porsi l'obiettivo di occuparsi del benessere della persona con un approccio olistico, il nuovo Centro Nazionale di Medicina Computazionale e Tecnologica di Genova sarà chiamato a rispondere a sfide di carattere tecnologico, informatico e comunicativo.

In primo luogo, si propone di usare tecnologie all'avanguardia nel campo della ricerca scientifica. Tecnologie simili possono richiedere grandi quantità di dati, che potrebbero non essere subito utilizzabili, e prevedono la necessità di svolgere grandi quantità di calcoli, che per essere svolti necessitano di una infrastruttura hardware adeguata.

In aggiunta, un centro di ricerca di tale portata richiede la disponibilità di numerosi dati sanitari, i quali costituiscono un tesoro di informazioni estremamente personali e sensibili. La protezione di queste preziose informazioni è fondamentale, poiché riflette il rispetto per la privacy e la dignità di ciascun individuo.

Infine, da non sottovalutare, il Centro deve affrontare l'eterogeneità del gruppo di ricercatori che andrà ad ospitare. È opportuno garantire un'efficace comunicazione e il corretto funzionamento della collaborazione tra professionisti con varie esperienze e competenze di base.

2 La sfida tecnologica

2.1 I dati

Uno degli strumenti tecnologici di cui il nascente Centro Nazionale di Medicina Computazionale e Tecnologica farà presumibilmente ampio uso è l'intelligenza artificiale, in varie sue forme. Una delle sue declinazioni che potrebbero trovare miglior spazio di applicazione nell'ambito sanitario è l'apprendimento automatico (o *machine learning*), la cui abilità principale è riconoscere pattern di vario tipo nei dati che ad esso vengono sottoposti: ad esempio, un modello di machine learning ben addestrato può essere in grado di capire che una determinata condizione clinica può comportare (o comunque facilitare) l'insorgere futuro di particolari patologie, permettendo di prendere misure preventive.

Qualsiasi siano i modelli utilizzati, l'apprendimento automatico necessita di grandi quantità di dati, tanto più grandi quanto più è complesso il modello; l'altro lato della medaglia consiste nel fatto che modelli più complessi, in linea di massima, sono in grado di riconoscere pattern più complessi. Tralasciando considerazioni riguardo l'argomento privacy, che analizzeremo in modo più sistematico nel seguito, tali dati sono già in presenti negli archivi del Sistema Sanitario Nazionale, ma solo parzialmente in forma digitale, in quanto la digitalizzazione della sanità italiana è un processo relativamente recente e ancora in fase di avanzamento. Pertanto, nel caso si usassero modelli particolarmente sofisticati, i soli dati digitali di cui ad oggi il Sistema Sanitario Nazionale è in possesso potrebbero non essere sufficienti allo scopo e si dovrebbe procedere alla

digitalizzazione di dati precedenti, archiviati in forma cartacea, e per la natura estremamente eterogenea dei documenti in questione il lavoro va svolto manualmente da operatori opportunamente addestrati: una scansione digitale non è sufficiente, in quanto gli algoritmi di apprendimento automatico necessitano di dati in formato machine readable; sarebbe pertanto necessaria la predisposizione di un'apposita piattaforma nella quale inserire i dati che formeranno i training set su cui allenare gli algoritmi. Un processo come quello appena descritto comporta una mole di lavoro non indifferente: la cartella clinica (cartacea) di un paziente che ha subito un'operazione di routine che prevede meno di una settimana di ricovero può arrivare ad una lunghezza di quasi cento pagine, il cui contenuto deve essere convertito in dati strutturati o semistrutturati; a partire da tale considerazione si può immaginare che onere sia ripetere simili procedure, ad esempio, per ogni paziente che è stato ricoverato in un ospedale di una grande città.

2.2 La computazione

Uno degli obiettivi che Centro Nazionale di Medicina Computazionale e Tecnologica si prefigge è lo studio di modelli che possano descrivere in modo quanto più accurato possibile il funzionamento di sistemi biologici complessi come organi o interi apparati, in modo da poterne creare delle vere e proprie copie virtuali (*digital twin*) che permettano di prevedere l'evoluzione del sistema stesso in differenti scenari.

Sia l'addestramento di algoritmi di machine learning che l'implementazione dei digital twin necessitano di grandi quantità di potenza computazionale. Al giorno d'oggi, la soluzione più usata per simili questioni è il *cloud computing*, un paradigma che prevede l'accesso su richiesta e da qualsiasi punto della rete a un insieme di risorse computazionali condivise che possono essere rese disponibili e rilasciate rapidamente con minimo sforzo di gestione e minima interazione con il fornitore del servizio; secondo un'indagine del Synergy Research Group, nel periodo aprile-giugno 2023, i primi tre fornitori di servizi cloud, ordinati per market share, sono stati Amazon Web Services, Microsoft Azure e Google Cloud. Una tale soluzione comporta grande flessibilità: ad esempio, non sempre è facile stimare la quantità di potenza di calcolo realmente necessaria e facendo uso di risorse cloud è estremamente semplice variare in corso d'opera la quantità di risorse della propria istanza, acquistandone di nuove se necessarie o rilasciandone se non utilizzate; facendo uso di servizi in cloud non è necessario occuparsi della manutenzione dell'infrastruttura fisica, che è a carico del fornitore: tra le conseguenze si ha che, a lungo termine, i costi potrebbero non essere particolarmente convenienti, in quanto non si paga solo l'uso dell'infrastruttura ma anche tutto ciò che intorno ad essa ruota. Un'altra conseguenza è che si ha un controllo limitato sull'infrastruttura stessa, ad esempio sulla sua ubicazione o sull'hardware utilizzato.

L'alternativa è la costruzione di un proprio centro di calcolo: in tal caso l'investimento iniziale è sicuramente maggiore, sia per l'acquisto dell'hardware sia per le competenze necessarie per progettare l'infrastruttura che meglio si

adatta alle proprie esigenze. Tuttavia, in un contesto come quello di gestione e uso dei dati sanitari precedentemente discussi, per motivi normativi legati alla privacy può essere molto importante poter scegliere liberamente dove collocare la propria infrastruttura. Secondo quanto dichiarato dalle istituzioni il Centro Nazionale di Medicina Computazionale e Tecnologica presso un nuovo ospedale situato presso il Great Campus di Erzelli, dove sono presenti importanti realtà aziendali e di ricerca in campo tecnologico come Siemens, Ericsson, Esaote e iit; una possibile opzione per ridurre i costi di un proprio data center potrebbe essere costruirlo nella sede stessa di Erzelli e concedere l'accesso alle sue risorse computazionali alle aziende che operano in loco, con tariffe agevolate rispetto ai cloud provider presenti sul mercato.

Per via dell'hardware utilizzato, i centri di calcolo ad alte prestazioni generano grandi quantità di calore che solitamente viene dissipato nell'aria. Un'ulteriore possibilità per ridurre l'onere economico previsto dalla costruzione di un proprio data center è quella di recuperare l'energia proveniente da esso sotto forma di calore ed utilizzarla per generare elettricità o riscaldare gli edifici circostanti; soluzioni simili negli ultimi anni, vista la crescente preoccupazione per le questioni di sostenibilità ambientale, stanno prendendo iniziando ad attirare l'attenzione di grandi player pubblici e privati; vari test, perlopiù guidati da aziende, sono attualmente in corso e i risultati sono incoraggianti.

3 La sfida della privacy

La gestione della privacy dei dati sanitari è una questione di fondamentale importanza in un settore come questo. I dati sanitari sono considerati altamente sensibili e la loro protezione è regolamentata da leggi rigorose. Questi dati comprendono informazioni personali, diagnosi mediche, trattamenti, e altro ancora, e garantire la loro sicurezza è essenziale per preservare la fiducia della popolazione e rispettare le normative vigenti.

3.1 L'importanza della privacy dei dati sanitari

La privacy dei dati sanitari riveste un'importanza cruciale per diverse ragioni:

- **Diritti individuali:** Ogni individuo ha il diritto fondamentale alla privacy delle proprie informazioni mediche. La condivisione non autorizzata o l'abuso di tali dati possono violare questi diritti.
- **Fiducia dei pazienti:** La fiducia dei pazienti è fondamentale per il successo del centro e la partecipazione attiva della popolazione alla ricerca medica. La promessa di proteggere la privacy dei dati è essenziale per instaurare questa fiducia.
- **Conformità legale:** Il centro deve aderire in modo rigoroso alle leggi sulla privacy dei dati sanitari, compreso il Regolamento Generale sulla Protezione dei Dati (GDPR) dell'Unione Europea e le leggi nazionali.

3.2 Strategie di gestione della privacy dei dati

Per affrontare la sfida della privacy dei dati sanitari, il centro dovrà implementare una serie di strategie.

3.2.1 Consensi informati

Una strategia fondamentale potrebbe essere l'introduzione del concetto di "consenso informato" per i pazienti. Si potrebbe creare un documento informativo che spiega chiaramente l'uso previsto dei dati sanitari, le misure di sicurezza adottate e i benefici attesi per la ricerca medica. I pazienti sarebbero quindi invitati a firmare questo consenso per autorizzare l'utilizzo dei loro dati.

3.2.2 Benefici per i partecipanti

Per incentivare la partecipazione, si potrebbero offrire vantaggi significativi ai cittadini che acconsentono a condividere i propri dati. Questi benefici potrebbero includere servizi sanitari gratuiti, monitoraggio costante della salute, accesso prioritario a trattamenti innovativi o la possibilità di partecipare a studi clinici particolari.

3.2.3 Sicurezza dei dati

Il centro deve implementare misure di sicurezza avanzate per proteggere i dati sanitari. Queste misure possono comprendere l'uso di crittografia avanzata per proteggere i dati in transito e a riposo, firewall avanzati per proteggere l'infrastruttura IT e l'accesso limitato solo al personale autorizzato.

3.3 Promozione della partecipazione

La promozione alla partecipazione attiva dei cittadini è un elemento cruciale per il successo del Centro Nazionale di Medicina Computazionale e Tecnologica. Affinché i cittadini acconsentano all'utilizzo dei loro dati sanitari, è fondamentale instaurare una comunicazione chiara, trasparente e informativa che metta in luce i benefici della ricerca medica basata sui dati e le misure di protezione dei dati implementate.

3.3.1 Campagne di sensibilizzazione

Il centro dovrebbe condurre campagne di sensibilizzazione su vasta scala per informare i cittadini sull'importanza dell'utilizzo dei dati sanitari per la ricerca medica. Queste campagne dovrebbero essere facilmente accessibili attraverso vari mezzi di comunicazione, tra cui pubblicità online, annunci televisivi, social media, opuscoli informativi e eventi pubblici.

Le campagne di sensibilizzazione dovrebbero mettere in evidenza i successi passati della ricerca medica basata sui dati, sottolineando come tale ricerca abbia contribuito a salvare vite, sviluppare nuovi trattamenti e migliorare la

salute pubblica. Inoltre, dovrebbero sottolineare l'importanza della privacy dei dati e come il centro sia impegnato a proteggere le informazioni personali dei partecipanti.

3.3.2 Approccio "opt-in"

Per rispettare le preferenze dei cittadini e garantire un consenso informato, il centro potrebbe adottare un approccio "opt-in". Invece di presumere il consenso dei cittadini, si dovrebbe chiedere loro attivamente se desiderano partecipare all'utilizzo dei dati per scopi di ricerca medica. Questo approccio mette in risalto il controllo dei cittadini sulla propria privacy e crea una base di partecipazione volontaria.

Un esempio ispiratore di come promuovere la partecipazione dei cittadini potrebbe essere preso dalle iniziative italiane per i donatori di sangue che sono incentivati a contribuire al bene comune attraverso la predisposizione ad esami del sangue gratuiti trimestrali e annuali.

3.3.3 Trasparenza continua

La trasparenza è fondamentale per mantenere la fiducia dei cittadini nel tempo. Il centro dovrebbe impegnarsi a fornire aggiornamenti regolari sulla ricerca e sui progressi compiuti. Questi aggiornamenti dovrebbero essere accessibili al pubblico e dovrebbero dimostrare il valore dell'utilizzo dei dati nella promozione della salute e nella ricerca medica.

Inoltre, il centro dovrebbe impegnarsi a rendere pubbliche le politiche sulla privacy dei dati, in modo che i cittadini possano comprenderne appieno le misure di protezione adottate.

La promozione attiva della partecipazione dei cittadini attraverso queste strategie può aiutare a costruire una solida base di fiducia nella comunità e garantire che la ricerca medica basata sui dati possa avanzare in modo etico e con l'approvazione dei partecipanti.

4 La sfida comunicativa

Al fine di comprendere con un approccio olistico l'uomo con l'ambizione di prevenire e curare tempestivamente, è fondamentale avere un team multidisciplinare costituito non solo da figure professionali come medici, chimici o biologi, ma anche da ingegneri, fisici, matematici e informatici. Come si intende già del nome, la medicina computazionale si trova nell'intersezione di questi domini scientifici e deve, quindi, affrontare il problema della comunicazione tra i due mondi.

La caratteristica interdisciplinare del progetto del Centro Nazionale di Medicina Computazionale e Tecnologica implica inevitabilmente una sfida comunicativa tra i vari ambiti.

4.1 Un team eterogeneo

Nel background di ciascun ricercatore c'è sempre una minima conoscenza informatica, ma al fine di impiegare l'intelligenza artificiale nell'ambito sanitario sono necessarie figure specializzate in ambiti molto diversi tra loro, che ora devono interagire. Da una parte team di bioinformatici e ricercatori computazionali elaborano complessi set di dati eterogenei; dall'altra biologi e medici trasferiscono le loro conoscenze per insegnare alla macchina il loro lavoro. Avere figure professionali esperte in entrambi gli ambiti è prezioso e ambizioso, ma molto spesso lavorare a stretto contatto implica in qualche modo uno scambio e condivisione di competenze.

Nel momento dell'assunzione, è quindi preferibile selezionare personale che ha già avuto esperienze negli ambiti in questione, in modo da favorire la comprensione, oppure scegliere figure professionali ancora da specializzare nel settore più utile.

4.2 Una piattaforma centralizzata

Non solo tra colleghi, la collaborazione è necessaria anche con partner esterni, a volte più specializzati e, di conseguenza, richiesti. La conoscenza va condivisa internamente ma anche esternamente, in modo tale da permettere ai collaboratori uno sviluppo più facile e autonomo delle soluzioni. Si richiede, dunque, l'uso di una piattaforma centralizzata, in questo modo è possibile raccogliere in modo omogeneo i dati e condurre analisi in qualunque tempo e luogo, con le limitazioni analizzate nella precedente sezione. Inoltre, ciò consentirebbe di condividere le informazioni in modo rapido a tutti i collaboratori, permettendo una lettura accessibile e interazioni rapide. In quest'ottica, risulta fondamentale una ragionata gestione dei privilegi, in modo da poter condividere solo una parte di informazioni, grazie a filtri altamente granulari. È consigliabile che la piattaforma implementi la politica del minimo privilegio, così da non diffondere dati non necessari e causare problemi di privacy. L'investimento in questa struttura informatica non è indifferente, ma sarebbe molto significativo e strategico, in caso di grandi moli di dati e di un grande numero di collaboratori.

4.3 Una comunicazione efficace

La visualizzazione dei risultati dell'intelligenza artificiale ha un ruolo significativo in quest'ottica, in modo tale che sia comprensibile al personale sanitario. Comunicare gli esiti di uno studio non è sempre facile, ma diventa cruciale in un contesto multidisciplinare. L'utilizzo di dashboard il più possibile interattive e grafici chiari ben evidenziati consentono una trasparenza comunicativa necessaria. Potrebbe tornare utile una figura specializzata nel graphic design e nella visualizzazione dati, che realizzi report completi e, contemporaneamente, semplici. Tuttavia, l'assunzione di personale specializzato implica un costo. Quindi, in base alle conoscenze dell'interlocutore, bisogna valutare se è più importante la chiarezza dei risultati oppure la disponibilità economica.

5 In conclusione

Si può concludere trovando una soluzione ai problemi riportati inizialmente.

La carenza di dati può essere risolta digitalizzando almeno parte delle banche dati di cui il Sistema Sanitario Nazionale è in possesso. La necessaria potenza di calcolo può essere ottenuta con servizi di cloud computing oppure costruendo un'infrastruttura propria; entrambi gli approcci presentano i propri vantaggi e le proprie criticità.

La gestione della privacy dei dati sanitari costituisce il fondamento etico e legale su cui si basa il successo del Centro di Medicina Computazionale e Tecnologica. Attraverso consensi informati, benefici, e una comunicazione trasparente, è possibile costruire una solida fiducia nella comunità, permettendo alla ricerca medica di progredire in armonia con i valori dei partecipanti.

La comunicazione si può intendere come una sfida imprescindibile dal carattere interdisciplinare proprio del progetto, ma è una challenge verosimilmente affrontabile. La collaborazione e l'esperienza dei ricercatori consentono condivisione di conoscenza. La realizzazione di una piattaforma centralizzata è necessaria in vista di collaborazioni interne ed esterne. Infine, una chiara e semplice esposizione dei risultati permette la comprensione per tutte le figure a prescindere dal background.