Ricerca di soggetti disponibili a supportare attività aziendali con contributi privati da erogarsi mediante contratti di sponsorizzazione o liberalità

Bando di riferimento:

P17 deliberazione n. 2156 del 15 dicembre 2016

Tipologia di riferimento: **P17-03**

☐ **Formazione e ottimizzazione dei processi aziendali**

☒ **Attività di sperimentazione studio, ricerca e sviluppo in campo medico scientifico, clinico, di perfezionamento**

- ☐ **Acquisito parere Comitato etico in data _____**
- ☐ **Parere Comitato etico da richiedere**
- ☒ **Non necessario parere Comitato etico**

Scheda di progetto P17-03-87

Data di emissione	Novembre 2026
Titolo	Validazione di dataset sintetici basati su Intelligenza Artificiale per il supporto alla ricerca clinica nel carcinoma mammario luminale
Periodo riferimento	Novembre 2026 – Maggio 2028
Struttura proponente	SC Oncologia
Responsabile e Referente del progetto	Prof. Alberto Zambelli SC Oncologia e-mail: azambelli@asst-pg23.it
Descrizione progetto	Il carcinoma mammario luminale rappresenta il sottotipo biologico più frequente di tumore della mammella ed è caratterizzato da un'elevata eterogeneità clinica e molecolare, che rende fondamentale la disponibilità di dataset clinici ampi, completi e di elevata qualità per lo sviluppo di modelli predittivi, studi osservazionali e strumenti di medicina di precisione. Negli ultimi anni, la crescente digitalizzazione dei percorsi diagnostico-terapeutici ha determinato un'importante disponibilità di dati sanitari. Tuttavia, il loro utilizzo a fini di ricerca continua a essere limitato da problematiche legate alla tutela della privacy, all'interoperabilità dei

	sistemi informativi, alla frammentazione delle fonti e alle restrizioni normative che regolano la condivisione dei dati clinici. In questo contesto, i dati sintetici costituiscono una delle innovazioni più promettenti. Generati mediante modelli avanzati di Intelligenza Artificiale, essi hanno potenzialmente la capacità di riprodurre fedelmente le proprietà statistiche e le relazioni presenti nei dati reali senza contenere informazioni riconducibili ai singoli pazienti, offrendo quindi una soluzione efficace per conciliare esigenze di ricerca, protezione dei dati personali e condivisione delle informazioni. Nonostante il crescente interesse internazionale verso questa tecnologia, rimangono ancora limitate le evidenze riguardanti la reale capacità dei dataset sintetici di preservare le caratteristiche clinicamente rilevanti delle popolazioni oncologiche. Pertanto, la validazione di tali modelli rappresenta un passaggio fondamentale per favorire l'adozione di approcci innovativi alla ricerca basata sui dati e per contribuire allo sviluppo di metodologie in grado di coniugare robustezza scientifica, tutela della riservatezza e facilità di accesso alle informazioni per finalità di ricerca. In questo contesto, il presente progetto si propone di valutare l'affidabilità e l'equivalenza dei dati sintetici rispetto ai dati clinici reali di pazienti affette da carcinoma mammario luminale, analizzandone la capacità di mantenere distribuzioni statistiche, correlazioni tra variabili cliniche e coerenza informativa. L'obiettivo finale è fornire evidenze metodologiche che supportino l'utilizzo dei dati sintetici come strumento sicuro, etico e scientificamente affidabile per la ricerca oncologica, favorendo lo sviluppo di studi multicentrici, modelli predittivi e applicazioni di Intelligenza Artificiale nel rispetto della normativa sulla protezione dei dati personali. Il progetto verrà articolato in due fasi: • Fase 1 — Preparazione e acquisizione ◦ definizione del dataset di riferimento (selezionato dai database relativi ai percorsi diagnostico-terapeutici delle pazienti affette da carcinoma mammario di tipo luminale); ◦ attivita di data cleaning e standardizzazione; ◦ acquisizione del software necessario per la generazione dei dati sintetici. • Fase 2 — Elaborazione e analisi ◦ generazione del dataset sintetico tramite piattaforma dedicata; ◦ valutazione statistica e metodologica della coerenza dei dati sintetici (inclusa l'analisi comparativa delle principali variabili cliniche e delle correlazioni rilevanti presenti nei dataset reali e nei dataset sintetici); ◦ predisposizione dei report conclusivi con le evidenze emerse. In caso di successo, il progetto potrà favorire l'adozione dei dati sintetici come risorsa a supporto della ricerca clinica, contribuendo allo sviluppo di nuove progettualità nazionali e internazionali e alla creazione di ecosistemi

	di ricerca più aperti, interoperabili e collaborativi. La disponibilità di un framework metodologico replicabile consentirà di estendere l'approccio a differenti popolazioni di pazienti, ad altre neoplasie e/o ad altri contesti clinici, favorendo collaborazioni tra istituzioni, sviluppo di modelli di Intelligenza Artificiale e valorizzazione del patrimonio informativo sanitario in un contesto di innovazione sostenibile. Il presente progetto ha finalità metodologica e propedeutica allo sviluppo di futuri studi di ricerca clinica e osservazionale che potranno avvalersi dei dati sintetici generati e validati nell'ambito di questa progettualità. L'eventuale utilizzo di dati clinici reali nell'ambito dei futuri studi sarà effettuato esclusivamente previa acquisizione delle necessarie approvazioni del Comitato Etico e nel rispetto della normativa vigente; per alcuni di questi studi, il relativo iter di approvazione è già in corso.
Obiettivi del progetto	Sponsorizzazioni destinate al finanziamento e allo sviluppo di progetti di ricerca finalizzati alla validazione di dataset sintetici generati mediante Intelligenza Artificiale, a supporto della ricerca clinica sul carcinoma mammario luminale. Nello specifico i progetti che verranno attivati avranno come obiettivi: • Valutare il grado di equivalenza tra dataset clinici reali e dataset sintetici in ambito oncologico e, nello specifico, nel carcinoma mammario di tipo luminale; • Verificare la capacita dei dati sintetici di preservare distribuzioni statistiche, correlazioni e coerenza informativa rilevanti dal punto di vista clinico; • Esplorare il potenziale utilizzo dei dati sintetici quale strumento per supportare attività di ricerca clinica e osservazionale in maniera etica, sicura e sostenibile e, in prospettiva, supportare futuri studi clinici e osservazionali.
Fasi e tempi di realizzazione stimati	Durata stimata della progettualità: 18 mesi
Collaborazioni con altre strutture aziendali o altri soggetti esterni	Richiesta di supporto a enti finanziatori esterni per l'avvio e la realizzazione della progettualità. Non è previsto il coinvolgimento di altre strutture aziendali. Qualora necessario, verranno coinvolti dei consulenti bioinformatici esterni, tra cui il Dott. Saverio D'Amico di Train Srl.
Risorse professionali	Personale Medico Specialista in Oncologia Data-manager Studente PhD
Strumentazione	Acquisizione di Server e Software ad alta potenza computazionale necessari per la generazione dei dati sintetici, in accordo con ICT aziendale

Contropartita per i finanziatori	Come contropartita per gli enti finanziatori è previsto un ritorno di immagine sul progetto in essere. Gli enti finanziatori verranno infatti citati in tutte le pubblicazioni scientifiche derivanti dal progetto (inclusi gli abstract congressuali). Inoltre, potranno essere predisposti: (i) totem espositivo che reca informazione circa il progetto da collocarsi presso gli ambulatori oncologici PG23, (ii) indicazione della progettualità sul sito aziendale e sui social, laddove possibile (iii) articoli divulgativi su riviste di settore (per esempio: rivista "Insieme" dell'Associazione Oncologica Bergamasca - AOB). Il logo degli enti finanziatori verrà riportato sui materiali di progetto e sui canali utilizzati dall'Oncologia di PG23 per la presentazione dello studio.
Finanziamento richiesto	€ 90.000 Il supporto richiesto è finalizzato a porre le basi metodologiche e tecnologiche per future progettualità di ricerca che utilizzeranno dati clinici, nel rispetto delle necessarie procedure di approvazione etica.
Criteri ed indicatori per la verifica del raggiungimento degli obiettivi	Il raggiungimento degli obiettivi sarà verificato attraverso la dimostrazione dell'equivalenza tra dataset reali e sintetici mediante analisi statistiche delle principali variabili cliniche e delle loro correlazioni. Gli indicatori comprenderanno il grado di sovrapposizione delle distribuzioni dei dati, il mantenimento delle relazioni tra le variabili cliniche rilevanti e la produzione di report scientifici (report intermedio e report finale) che documentino l'affidabilità dei dati sintetici quale supporto alla ricerca clinica nel carcinoma mammario luminale a partire da dataset clinici reali.