IRCCS "Giovanni Paolo II"
PugliaSalute

DATA PROTECTION IMPACT ASSESSMENT

ai sensi dell'art. 35 del Reg. UE 2016/679

VALUTAZIONE DI IMPATTO SULLA PROTEZIONE DEI DATI STUDI RETROSPETTIVI

Codice	Descrizione
DPIA-001	PEERAD - PrEdicting Endopredict score with RADiomics: a novel radiomics model based on artificial intelligence to drive adjuvant treatment in patients with early-stage, intermediate-risk, hormone-receptor positive HER2 negative breast cancer
ELABORAZIONE DPIA PER	☒ **Nuova attività trattamento** ☐ Aggiornamento DPIA ☐ Revisione periodica DPIA

Attività	Struttura/Funzione	Responsabile	data	firma
Redazione	**Principal Investigator**	Annarita Fanizzi		
Verifica	**DPO**	Iris Mannarini		
Approvazione	**Direttore Generale**	Alessandro Delle Donne		

DATA PROTECTION IMPACT ASSESSMENT

ai sensi dell'art. 35 del Reg. UE 2016/679

SOGGETTI COINVOLTI NELLO STUDIO	
TITOLARE promotore	**IRCCS ISTITUTO TUMORI GIOVANNI PAOLO II DI BARI (Principal Investigator: Dott.ssa Annarita Fanizzi)**
Centri partecipanti quali Titolari del trattamento	• Azienda Ospedaliero Universitaria Consorziale Policlinico di Bari • Azienda Ospedaliero Universitaria - Ospedali Riuniti, Foggia • Ospedale San Paolo, ASL Bari • ASL Taranto - Ospedale SS Annunziata di Taranto
RESPONSABILE DEL TRATTAMENTO	**Non presente**
COORDINATORE E SPERIMENTATORI	**Coordinatore:** **Dott.ssa Annarita Fanizzi** (IRCCS Istituto Tumori "Giovanni Paolo II- Bari) **All'interno della steering committee sono da considerarsi:** • Dott. Alessandro Rizzo, U.O.C Oncologia Medica 'Don Tonino Bello', IRCCS Istituto Tumori 'Giovanni Paolo II', Bari – co-PI • Dott. Davide Quaresmini, U.O.C Oncologia Medica, IRCCS Istituto Tumori 'Giovanni Paolo II', Bari – Ricercatore Collaboratore • Dr. Michele Telegrafo, Radiodiagnostica a indirizzo senologico, Azienda Ospedaliero Universitaria Consorziale Policlinico di Bari – Responsabile Unità Operativa Collaborante • Dott. Magnasco Salvatore, F.F. Direttore U.O.C. Anatomia Patologica – ASL Taranto

	• Dott.ssa Massafra Raffaella, Responsabile Laboratorio Biostatistica e Bioinformatica, IRCCS Istituto Tumori 'Giovanni Paolo II', Bari • Dott.ssa Comes Maria Colomba, Laboratorio Biostatistica e Bioinformatica, IRCCS Istituto Tumori 'Giovanni Paolo II', Bari • Dott.ssa Samantha Bove, Laboratorio Biostatistica e Bioinformatica, IRCCS Istituto Tumori 'Giovanni Paolo II', Bari • Dott. Alfredo Zito, U.O.C. Anatomia Patologica, IRCCS Istituto Tumori 'Giovanni Paolo II', Bari • Dott.ssa Simona De Summa, Laboratorio Biostatistica e Bioinformatica, IRCCS Istituto Tumori 'Giovanni Paolo II', Bari

FASI DPIA	☒	**Conduzione DPIA**	
	☐	Parere del DPO	
	☐	Validazione del Titolare	
	☐	Consultazione Preventiva	
	☐	Revisione DPIA	
MODALITA' CONDUZIONE	☒	DPIA OBBLIGATORIA	
	☐	DPIA VOLONTARIA	

IRCCS "Giovanni Paolo II"
PugliaSalute

DATA PROTECTION IMPACT ASSESSMENT

ai sensi dell'art. 35 del Reg. UE 2016/679

Sommario

DATA PROTECTION IMPACT ASSESSMENT

ai sensi dell'art. 35 del Reg. UE 2016/679

Informazioni sulla DPIA

La DPIA, acronimo di *Data Protection Impact Assessment*, è una valutazione preliminare, eseguita dal Titolare del trattamento dei dati personali, relativa agli impatti a cui andrebbe incontro un trattamento laddove dovessero essere violate le misure di protezione dei dati. In linea con l'approccio basato sul rischio adottato dal Regolamento generale sulla protezione dei dati, è necessario realizzare una valutazione d'impatto sulla protezione dei dati soltanto quando la tipologia di trattamento" può presentare un rischio elevato per i diritti e le libertà delle persone fisiche" (articolo 35 del Regolamento 2016/679).

Ai sensi dell'articolo 35, paragrafo 3 del Regolamento 2016/679 la valutazione è effettuata nei casi in cui un trattamento può presentare rischi elevati, ossia quando:
a. una valutazione sistematica e globale di aspetti personali relativi a persone fisiche, basata su un trattamento automatizzato, compresa la profilazione, e sulla quale si fondano decisioni che hanno effetti giuridici o incidono in modo analogo significativamente su dette persone fisiche;
b. il trattamento, su larga scala, di categorie particolari di dati personali di cui all'articolo 9, paragrafo 1, o di dati relativi a condanne penali e a reati di cui all'articolo 10;
c. la sorveglianza sistematica su larga scala di una zona accessibile al pubblico.

In particolare, preso atto della tipologia di Studio (retrospettivo) in argomento, è condotta e resa pubblica una valutazione d'impatto sulla protezione dei dati, in forza dell'art. 35 del Reg. UE 2016/679 riguardo al trattamento dei dati ai sensi e per gli effetti del combinato disposto degli artt. 9, par. 2, lett. j) del GDPR, 110 e 110 bis, comma 4 del Codice Privacy.

La presente valutazione contiene:
a) una descrizione sistematica dei trattamenti previsti e delle finalità del trattamento
b) una valutazione della necessità e proporzionalità dei trattamenti in relazione alle finalità;
c) una valutazione dei rischi per i diritti e le libertà degli interessati;
d) le misure previste per affrontare i rischi, includendo le garanzie, le misure di sicurezza e i meccanismi per garantire la protezione dei dati personali e dimostrare la conformità al presente regolamento, tenuto conto dei diritti e degli interessi legittimi degli interessati e delle altre persone in questione.

ACCETTABILITA' DEL RISCHIO

Tenuto conto della natura, del contesto, delle finalità e dell'ambito di applicazione del trattamento in esame, il **livello di rischio residuo**, considerato accettabile indicato dal Titolare, sentito anche il parere del DPO, è risultato ☒ BASSO ☐ MEDIO ☐ ALTO

Di seguito sono illustrati i dettagli della valutazione d'impatto sulla protezione dei dati.

1 Descrizione sistematica del trattamento

1.1 Contesto

Diversi test genomici sono stati recentemente sviluppati per valutare il beneficio fornito dall'aggiunta della chemioterapia alla terapia ormonale in pazienti con carcinoma mammario (BC) radicalmente resecato, positivo per i recettori ormonali e HER2-negativo. Il decreto del Ministero della Salute del maggio 2021 ha stabilito che la popolazione di pazienti suscettibili di eseguire test genomici include pazienti con carcinoma mammario (BC) endocrino-positivo HE2-negativo in fase iniziale, per i quali il beneficio dell'aggiunta della chemioterapia alla terapia endocrina adiuvante è controverso.
Tuttavia, questi strumenti sono costosi e la loro efficacia in termini di costi deve essere considerata. L'analisi radiomica di immagini radiologiche diagnostici e di vetrini interi istopatologici mediante tecniche innovative di intelligenza artificiale (AI) può rappresentare un'alternativa ai test genomici. L'obiettivo primario del progetto è sviluppare un sofisticato modello di intelligenza artificiale opportunamente addestrato sui dati prodotti nella pratica clinica quotidiana in grado di riprodurre accuratamente il punteggio di un test genomico, riducendo così i tempi di valutazione e i costi dell'assistenza sanitaria senza compromettere sulla cura del paziente. A tale scopo, un'ampia raccolta di dati radiomici da pazienti "reali" affetti da BC sarà ingegnerizzata.
Sul mercato sono disponibili diversi test genomici che possono essere eseguiti utilizzando un campione di tessuto prelevato dalla resezione chirurgica del tumore o dalla biopsia. Tra questi il test EndoPredict (EP) è un classificatore di prognosi molecolare basato sulla valutazione di 12 geni nelle cellule BC e, a differenza di Oncotype-DX, l'unico pannello multigenico incluso nell'ultimo American Joint Committee on Cancer Classification (8th Edition), può essere eseguito in modo affidabile nei laboratori istituzionali e periferici; EP utilizza inoltre parametri patologici nel calcolo del punteggio di rischio ed è certificata CE-IVD. Nonostante le raccomandazioni dell'Istituto Superiore di Sanità e di Eccellenza Clinica siano favorevoli all'utilizzo dei test genomici,

il rimborso di questi test è stato approvato solo in alcune regioni italiane, oltre che non tutti gli Istituti di cura dispongono di laboratori attrezzati per effettuare in loco il test genomico. Pertanto, vi è l'urgente necessità di esplorare nuovi strumenti prognostici affidabili e meno costosi in questo contesto. La radiogenomica rappresenta la nuova frontiera della diagnostica per immagini. Studi recenti hanno mostrano una correlazione tra i dati quantitativi estratti dalle immagini biomediche (radiomica) e le categorie di rischio stimate dai test genomici per alcuni tipi di tumore. Lo studio si ripropone principalmente di sviluppare e validare, mediante tecniche di intelligenza artificiale, un modello predittivo del risultato di un test genomico analizzando separatamente e congiuntamente caratteristiche cliniche e feature quantitative estratte da immagini radiologiche diagnostiche e vetrini istologici digitalizzati. A tal fine il contributo dei centri partecipanti allo studio si riassume di seguito:

- IRCCS Istituto Tumori 'Giovanni Paolo II' di Bari (Principal Investigator: Dott. Annarita Fanizzi): (1) collazionamento retrospettivo di dati e immagini radiologiche e digital pathology di pazienti afferenti all'Istituto di appartenenza che hanno eseguito test genomico EndoPredict; (2) gestione e revisione periodica sistematica (ogni trimestre) del dataset multicentrico di dati, immagini e materiale biologico (vetrini) di pazienti afferenti ai diversi centri partecipanti al progetto; (3) analisi dei dati secondo le finalità dello studio;
- Azienda Ospedaliero Universitaria Consorziale Policlinico di Bari (Responsabile Scientifico: Dr. Michele Telegrafo): (1) supporto alla definire della semiotica per l'estrazione di indicatori radiologici dai referti radiologici, (2) collazionamento retrospettivo di dati, immagini radiologiche e materiale biologico (vetrini) di pazienti afferenti all'Istituto di appartenenza che hanno eseguito test genomico EndoPredict;
- Azienda Ospedaliero Universitaria - Ospedali Riuniti, Foggia (Responsabile Scientifico: Dott.ssa Maria Iole Natalicchio): (1) collazionamento retrospettivo di dati, immagini radiologiche e materiale biologico (vetrini) di pazienti afferenti all'Istituto di appartenenza che hanno eseguito test genomico EndoPredict;
- Ospedale San Paolo, ASL Bari (Responsabile Scientifico: Dott.ssa Stefania Bruno): (1) collazionamento retrospettivo di dati immagini radiologiche di pazienti afferenti all'Istituto di appartenenza che hanno eseguito test genomico EndoPredict;
- ASL Taranto - Ospedale SS Annunziata di Taranto (Referente: Dott. Michele Pirelli, Direttore U.O.C. Anatomia Patologica (Responsabile: Dott. Michele Pirelli): (1) fornitura

del vetrino biologico da digitalizzare presso l'Istituto proponente sul quale il Laboratorio di Diagnostica Molecolare e Farmacogenetica dell'Istituto proponente ha già eseguito il test Endopredict così come da pratica clinica definita dal medico oncologo di riferimento del centro Asl Taranto.

Tutti i dati collezionati verranno quindi centralizzati presso i sistemi di raccolta dati predisposti dal centro promotore come dettagliato nella sezione 2.1.2.

1.2 Panoramica del trattamento

1.2.1 Quale è il trattamento in considerazione?

Lo studio non prevede alcuna modifica dell'iter diagnostico e terapeutico del paziente. Lo scopo principale dell'attività del progetto è la definizione di un modello radiomico che consenta di prevedere il rischio di recidiva stimato dal test genomico Endopredict a partire da immagini radiologiche di primo livello, vetrini digitalizzati e caratteristiche cliniche baseline (alla diagnosi) di pazienti con carcinoma mammario in stadio iniziale con positività dei recettori ormonali e fattore di crescita epidermico umano 2 negativo (HER2-) per i quali è dubbia l'effettiva utilità della chemioterapia adiuvante post-operatoria, in aggiunta alla terapia ormonale.
I dati clinici retrospettivi, riferiti a tali pazienti, saranno raccolti direttamente dalle cartelle cliniche. Altre informazioni clinico-patologiche disponibili in cartella cliniche e ritenute di interesse da parte dei clinici in corso di studio saranno altresì collezionate. Saranno inoltre collezionate le immagini ecografiche e/o mammografiche del tumore primario della mammella, esame diagnostico comunemente effettuato nella pratica clinica, nonché vetrini digitalizzati del tumore primario.
Il team di ricerca metterà a punto un Common Data Model (CDM). Il CDM raccoglierà dati trasversali da pazienti con carcinoma mammario in stadio iniziale positivi al recettore ormonale e negativi al fattore di crescita epidermico umano 2 (HER2). I dati saranno poi sottoposti ad (i) una fase di pulizia e ad (ii) una fase di controllo qualità.
Saranno inoltre collezionate immagini mammografiche e/o ecografiche diagnostiche pre-operatorie e le immagini digitalizzate del pezzo operatorio che saranno oggetto di elaborazione mediante algoritmi di deep-learning per l'estrazione di caratteristiche quantitative (radiomiche) potenzialmente informative rispetto all'end-point dello studio.

I dati clinici e le immagini radiologiche baseline (pre-operatori), nonché le immagini digitalizzate del pezzo operatorio riferite a pazienti per i quali è noto il risultato del test genomico EP saranno utilizzate per la messa a punto di un modello di intelligenza artificiale predittivo del Recurrence Risk score generato da EP.

Per le finalità del progetto, saranno collezionati due campioni. Un primo campione (s1) riferisce a pazienti che hanno effettuato test genomico Endopredict e sulla scorta dei quali sarà sviluppato un modello predittivo del Recurrence Risk score di tale test genomico, sulla scorta dei dati clinici e di immagini collezionati. Un secondo campione (s2), invece, riferisce a pazienti che hanno avuto un primo tumore della mammella prima del 2014 (anno in cui il test genomico Endopredict è entrato nel mercato italiano), ma di cui è noto il follow-up, con caratteristiche tali da essere oggi candidabili a test genomico e sulle quali sarà effettuato in retrospettivo il test genomico Endopredict con il duplice scopo di (i) validare il test genomico in retrospettivo e (ii) definire una ulteriore campione di validazione del modello sviluppato sul campione s.

I *criteri di eleggibilità* dei pazienti del campione *s1* sono:

- carcinoma della mammella HER2 negative ormono-responsive early stage a rischio intermedio a rischio intermedio di recidiva, candidabili a test genomico come indicato nel decreto del Ministero della salute del 18 Maggio 2021,
- aver effettuato test genomico Endopredict presso il nostro Istituto,
- disponibilità delle immagini diagnostiche di primo livello (ecografia e/o mammografia) e/o vetrino digitalizzabile del pezzo operatorio.

I *criteri di eleggibilità* dei pazienti del campione *s2* sono:

- carcinoma della mammella HER2 negative ormono-responsive early stage a rischio intermedio a rischio intermedio di recidiva candidabili a test genomico come indicato nel decreto del Ministero della salute del 18 Maggio 2021,
- follow-up di almeno 10 anni,
- disponibilità del preparato istologico per l'esecuzione del test genomico Endopredict in retrospettivo presso i nostri laboratori,
- disponibilità delle immagini diagnostiche di primo livello (ecografia e/o mammografia) e/o vetrino digitalizzabile del pezzo operatorio.

Secondo recenti studi epidemiologici circa il 70% dei tumori al seno è HER2 negativo e recettori ormonali positivi (D.M. 18 Maggio 2021). Data tale incidenza, la numerosità minima campionaria del campione s1 necessaria per le analisi oggetto di studio con un errore di primo tipo (alfa) tollerato pari a 0.05 (intervallo di confidenza del 95%), una precisione della stima di 0.07, ed una potenza statistica del 90%, deve essere almeno di 164 pazienti. Considerando i risultati di una prima ricognizione sui dati dei pazienti seguiti dalla Breast Unit del nostro Istituto, al fine di rendere più robuste le analisi è possibile aumentare la numerosità campionaria a 200 pazienti.
Con riferimento al campione di studio s2 con almeno 10 anni di follow-up per i quali si intende effettuare in retrospettivo il test genomico Endopredict su preparato istologico, vista la disponibilità economica del budget finanziato, si ipotizza una numerosità di 80 pazienti.
Su ciascuno dei due campioni saranno collezionate variabili cliniche e immagini radiologiche baseline, e immagini digitalizzate del pezzo operatorio.
I dati clinici retrospettivi, riferiti a tali pazienti, saranno raccolti direttamente dalle cartelle cliniche e sono riassunti nella scheda CRF allegata. Altre informazioni clinico-patologiche disponibili in cartella cliniche e ritenute di interesse da parte dei clinici in corso di studio saranno altresì collezionate. Saranno inoltre collezionate le immagini ecografiche e/o mammografiche del tumore primario della mammella, esame diagnostico comunemente effettuato nella pratica clinica, nonché vetrini digitalizzati del tumore primario.
Partendo da immagini radiologiche, immagini di vetrini del pezzo tumorale e note caratteristiche cliniche incluse nei referti medici, il team di ricerca metterà a punto un Common Data Model (CDM). Il CDM sarà progettato secondo il modello relazionale. I dati saranno poi sottoposti ad (i) una fase di pulizia e ad (ii) una fase di controllo qualità. Il CDM raccoglierà dati trasversali da pazienti con carcinoma mammario in stadio iniziale positivi al recettore ormonale e negativi al fattore di crescita epidermico umano 2 (HER2). Per popolare il CDM, le variabili clinicamente rilevanti saranno estratte oltre che dalle cartelle cliniche, anche da report di testo libero utilizzando algoritmi basati sia sull'elaborazione del linguaggio naturale (PNL) che sull'apprendimento automatico. Per fare ciò, gli investigatori definiranno in precedenza l'insieme delle caratteristiche semantiche rilevanti che la PNL rileverà accuratamente dai referti ecografici e mammografici. Infatti, l'ecografia e la mammografia sono due diverse indagini diagnostiche che

possono evidenziare diverse caratteristiche del tessuto mammario e reperti patologici. Per l'esame mammografico i segni più comuni di neoplasia sono rappresentati da opacità, distorsioni architettoniche, asimmetrie focali e microcalcificazioni come riportato dalla Quinta Edizione di ACR BIRAD Atlas. Le opacità sono classificate in base a densità, forma, margini, presenza di microcalcificazioni. I risultati degli ultrasuoni sono divisi in mass-like e non mass -like. I reperti ecografici con aspetto di mass-lick sono classificati in base all'ecogeneità, alla forma, ai margini e alla presenza di spot microcalcificati iperecogeni. Pertanto, per ciascuna delle due tecniche saranno definiti set di caratteristiche semantiche rilevanti.
I dati clinici e le immagini radiologiche baseline (pre-operatori), nonché le immagini digitalizzate del pezzo operatorio riferite a pazienti per i quali è noto il risultato del test genomico EP saranno utilizzate per la messa a punto di un modello di intelligenza artificiale predittivo del Recurrence Risk score generato da EP.
Com'è noto, l'analisi radiomica delle immagini radiologiche che sarà implementata nel corso del progetto consente di caratterizzare le lesioni estraendo informazioni quantitative non rilevabili attraverso la semplice osservazione visiva delle immagini radiologiche da parte dell'operatore. Le caratteristiche radiomiche saranno estratte da ciascuna Regione di Interesse (ROI) identificata da radiologi esperti acquisite al tempo della diagnosi. Quindi, le ROI selezionate verranno analizzate mediante tecniche di deep-learning. Saranno inoltre sviluppate tecniche di segmentazione automatica, al fine di rendere l'algoritmo meno operatore dipendente.
Inoltre, recentemente, l'introduzione della patologia digitale, ovvero la digitalizzazione di vetrini di campioni biologici, sta riscuotendo un forte interesse nella comunità clinica e scientifica. Con la pratica di Whole-Slide Imaging (WSI) i vetrini vengono digitalizzati e possono essere visualizzati, gestiti, condivisi e analizzati sul monitor di un computer. Il campo della patologia digitale è in crescita e trova applicazioni in medicina con l'obiettivo di ottenere automaticamente informazioni diagnostiche e prognostiche e predire il corso della malattia grazie all'Intelligenza Artificiale. Il vetrino in ematossilina/eosina del blocchetto in FFPE (Formalin-Fixed Paraffin embedded) più rappresentativo del tumore asportato chirurgicamente verrà selezionato dall'anatomo patologo di riferimento. Tale vetrino verrà poi sottoposto a scansione digitale tramite apposito scanner già presente nel nostro Istituto. Allo stato dell'arte, le immagini H&E sono le immagini patologiche digitali più utilizzate per prevedere determinati risultati. Tuttavia,

altri punti istologici potrebbero essere valutati dai nostri esperti patologi. Sui WSI, il rilevamento e la classificazione delle cellule, ad esempio nelle cellule tumorali, nelle cellule immunitarie e nello stroma, verranno eseguiti per estrarre caratteristiche morfologiche, quali eccentricità, circolarità, che verranno poi valutate insieme alle caratteristiche quantitative estratte dalle stesse immagini mediante algoritmi di deep learning. In particolare, il ground truth relativo alla classificazione dei tipi cellulari, necessario per l'addestramento, sarà opera di annotazione da parte dei patologi coinvolti.
In questo progetto, verranno utilizzate CNN pre-addestrate per l'estrazione di caratteristiche radiomiche da US e/o mammografia e WSI. Per set di dati limitati nell'imaging medico, viene spesso utilizzata la tecnica del transfer learning. Tale tecnica, consiste nel pre-addestramento della rete neurale convoluzionale (CNN) per aggirare il requisito della mole massiccia di dati necessari per il processo di addestramento da zero di una rete convoluzionale. Le CNN sono state precedentemente addestrate (CNN pre-addestrate) su un numero enorme (milioni) di immagini naturali non mediche per apprendere come estrarre automaticamente caratteristiche di diverso tipo che sono, di basso livello caratteristiche, ad esempio bordi e punti, che rappresentano strutture locali dell'immagine e caratteristiche di alto livello, ad esempio forme e oggetti, che catturano caratteristiche globali dalle immagini. Le conoscenze che le reti hanno acquisito durante la formazione possono essere trasferite e applicate su immagini mai inedite attraverso diversi campi di ricerca (transfer learning), Quindi, algoritmi di machine learning saranno addestrati sulle caratteristiche radiomiche estratte ed elaborati congiuntamente con tutte le caratteristiche clinico-patologiche considerate rilevanti dai clinici per predire l'esito del test EP.L'associazione statistica tra le variabili cliniche e radiomiche e il risultato EP sarà valutata attraverso l'implementazione di test statistici appropriati (es. T-Student's test, Chi squared test). L'integrazione di informazioni da più fonti (ad es. dati clinici e di imaging) per prevedere l'esito del test EP sarà effettuata attraverso modelli statistici multivariati, algoritmi AI noti allo stato dell'arte (ad es. Random Forest, Support Vector Machine) o Reti neurali artificiali o algoritmi innovativi (es. XGboost, Gradient Boosting). Verranno implementate adeguate tecniche di importanza e selezione delle caratteristiche nidificate all'interno dei modelli di classificazione per identificare le caratteristiche più significative per il problema di previsione. Le metriche standard, come l'area sotto la curva (AUC) della caratteristica operativa di ricezione (ROC), nonché

l'accuratezza, la precisione, la sensibilità e la specificità verranno calcolate per valutare le prestazioni del modello utilizzando schemi di convalida incrociata e test indipendenti.
Verrà, inoltre, sviluppato un modello multimodale integrando i dati WSI in ematossilina-eosina, il risultato del test Endopredict e le caratteristiche anamnestiche delle pazienti con lo scopo di migliorare le performance di predizione del rischio di recidiva, con particolare focus sul setting Luminal B considerata la sua aggressività.
I modelli sviluppati saranno ottimizzati e validati mediante tecniche di validazione incrociata e set di dati indipendenti raccolti durante questo progetto di ricerca.
I risultati di questo studio, se di interesse, potranno costituire la base per uno studio di validazione di più grandi dimensioni, coinvolgendo altri centri interessati allo studio.
Si sottolinea che, lo scopo dello studio è attualmente lo "sviluppo di un algoritmo predittivo" e non prevede alcuna implicazione nella pratica clinica.
Ciò nonostante, in linea con i principi di conoscibilità, non esclusività e non discriminazione algoritmica richiamati da ultimo anche dal Garante Privacy, in materia di utilizzo di algoritmi di intelligenza artificiale, si sottolinea quanto segue: (1) tutti i risultati della ricerca e i dettagli tecnici dei modelli implementati saranno resi pubblici mediante pubblicazioni scientifica e divulgazione in conferenze nazionali ed internazionali, (2) saranno collezionate anche informazioni riferite a soggetti deceduti e non contattabili in assenza delle quali il campione selezionato sarebbe incompleto creando di conseguenza, come rappresentato, possibili bias nello sviluppo dell'algoritmo, (3) fermi restando i vantaggi offerti dai sistemi automatizzati di computazione dei dati che si intendono implementare, tali operazioni non sono unicamente delegate ad elaborazioni automatizzate, ma al contrario, necessitano di una integrazione con l'intervento umano che, sulla base delle competenze ed expertise tecnico-specifiche, monitorano, correggono o mitigano le operazioni effettuate mediante algoritmi automatizzati. Pertanto, le tecniche di intelligenza artificiale e di apprendimento automatizzato utilizzate nell'ambito delle attività di elaborazione dello Studio non portano a un processo decisionale automatizzato.

Tipologia di Studio

Trattasi di studio osservazionale retrospettivo multicentrico.

1.2.2 Quali sono le responsabilità connesse al trattamento?

Gli sperimentatori coinvolti nello Studio sono appositamente autorizzati al trattamento dei dati, ai sensi dell'art. 29 del Reg. UE 2016/679 e dell'art. 2 quaterdecies del Dlgs 196/2003, così come novellato dal D.lgs 101/2018. Nell'ambito dello Studio non risultano designati soggetti terzi in qualità di Responsabili del trattamento dati ai sensi dell'art. 28 del GDPR.

1.2.3 Ci sono standard applicabili al trattamento?

- La linea guida di Buona Pratica Clinica [Good Clinical Practice (GCP)] è uno standard internazionale di etica e qualità scientifica per progettare, condurre, registrare e relazionare gli studi clinici che coinvolgono soggetti umani. La GCP ha l'obiettivo di fornire, in conformità con i principi per la tutela dei diritti dell'uomo stabiliti dalla Dichiarazione di Helsinki, uno standard comune ad Unione Europea, Giappone e Stati Uniti per facilitare la mutua accettazione dei dati clinici da parte delle autorità regolatorie di queste aree geografiche;

- La linea guida recepita dall'Italia (G.U.R.I. n.191 del 18 agosto 1997) è stata messa a punto sulla base delle GCP attualmente adottate da Unione Europea, Giappone e Stati Uniti, oltre che da Australia, Canada, Paesi Nordici e dall'Organizzazione Mondiale della Sanità (OMS);

- Il trattamento di dati personali per scopi di ricerca scientifica è effettuato nel rispetto del Regolamento UE 2016/679, del Codice, delle Prescrizioni relative al trattamento dei dati genetici e delle Prescrizioni relative al trattamento dei dati personali effettuato per scopi di ricerca scientifica, nonché delle Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica, che costituiscono condizione essenziale di liceità e correttezza dei trattamenti.

1.3 Dati, processi e risorse di supporto

1.3.1 Quali sono i dati trattati e gli asset a supporto?

DATA PROTECTION IMPACT ASSESSMENT

ai sensi dell'art. 35 del Reg. UE 2016/679

Tipologia di dati personali	Categoria interessati
☒ **Dati identificativi comuni** (es. nome, cognome,indirizzo) ☒ **Dati di contatto** (recapiti email, telefono, cellulare, etc.) ☒ **Dati sanitari già presenti negli archivi** ☒ **Dati raccolti da archivi cartacei** ☒ **Dati raccolti da archivi informatici** ☐ Credenziali di autenticazioni, chiavi di accesso ☐ Dati raccolti da strumenti audiovisivi, videosorveglianza ☐ Dati raccolti da tecnologie traccianti e/o di monitoraggio ☐ Dati raccolti da tecnologie IoT ☐ Dati su abitudini di vita, consumi e comportamento ☐ Dati su familiari/stato familiari ☐ Dati bancari ☐ Dati sulla localizzazione ☐ Dati sulla solvibilità economica	• Pazienti deceduti o non reperibili • Pazienti in vita (in follow-up)
☐ Appartenenza sindacale ☐ Convinzioni politiche, religiose o filosofiche ☐ Origine razziale o etnica ☒ **Dati sulla salute** ☐ Orientamento e vita sessuale ☐ Dati genetici ☐ Dati biometrici ☐ Dati "giudiziari" (diritto penale)	
☐ Dati soggetti a maggior tutela	

Altro: età alla diagnosi, immagini mammografiche/ecografiche diagnostiche, vetrini istologici digitalizzati, campioni biologici	

COMPONENTI ORGANIZZATIVE	
Soggetti interni	Lo staff dello studio è composto dal Principal Investigator, ricercatori e data manager opportunamente individuati in fase di sottomissione dello studio e nel corso dello stesso. Al Principal Investigator viene conferita la delega per la gestione delle attività di trattamento dei dati personali per i compiti relativi alla protezione dei dati personali necessari per la conduzione dello studio. Gli altri componenti dello staff sono autorizzati al trattamento di dati personali da parte del P.I. tramite apposito atto di nomina individuale.
Soggetti esterni	Ciascun centro partecipante allo studio fornisce i dati pseudonimizzati mediante attribuzione di un codice. Il Responsabile del Trattamento riceve ed elabora, pertanto solo i da dati pseudonimizzati. Tra IRCCS e centri partecipanti i rapporti sono regolati dal protocollo, dal Data Transfer Agreement (DTA) e dal Material Transfer Agreement (MTA).
COMPONENTI TECNOLOGICHE	
Applicazioni	Per l'elaborazione dei dati sono utilizzati sistemi di office automation quali RedCap, Microsoft Word, Excel, Python (versione >3.9.2), Matlab (versione >2023a), QuPath, utilizzati esclusivamente presso il Centro Promotore (IRCCS Istituto Tumori di Bari)
Infrastrutture ICT	Per la conservazione dei dati in formato elettronico sono utilizzati i sistemi di storage aziendali opportunamente protetti sia per quanto riguarda l'accesso fisico che l'accesso ai database che sono opportunamente criptati secondo le regole tecniche usuali e politiche di backup specifiche.
Reti informatiche	I computer utilizzati per il trattamento dei dati si trovano su rete dedicata e messa in sicurezza su apposita VLAN.
COMPONENTI FISICHE	
Asset	Per l'elaborazione dei dati sono utilizzati appositi strumenti software aziendali. I PC su cui sono installati tali software sono muniti di idonei sistemi di autenticazione, autorizzazione e tracciabilità delle operazioni
Sedi	Il trattamento dei dati avviene attraverso postazioni di lavoro presso la sede aziendale della ricerca scientifica con accesso riservato

Archivi	I dati personali sono conservati in sicurezza presso l'archivio corrente aziendale o su IaaS (Infrastructure as a Service) in cloud opportunamente protetta.

1.4 Finalità del trattamento

Il trattamento dei dati personali identificativi risulta necessario per la ricerca scientifica e, nel dettaglio, per le seguenti finalità dello Studio:

1. progettazione e definizione di un CDM per la raccolta sistematica dei dati relativi alla popolazione oggetto di studio;
2. definizione e validazione di un sistema automatizzato di previsione del Recurrence Score Risk generato dal test genomico EP;
3. studio di validazione del risultato del test genomico EP valutato su dati retrospettivi (pazienti con più di 10 anni di follow-up per le quali non era possibile effettuare il test genomico al momento del primo tumore della mammella);
4. valutazione dell'impatto clinico ed economico dell'utilizzo del modello PEERAD sul governo clinico e sulla qualità della vita (QoL) dei pazienti;
5. messa a punto di un modello di predizione multi-modale basato su imaging istologico, risultato di Endopredict e caratteristiche cliniche per la predizione del rischio di recidiva nella coorte di pazienti retrospettiva.

2 Principi Fondamentali

2.1 Valutazione della necessità e proporzionalità del trattamento del trattamento

Il trattamento è effettuato nel rispetto delle prescrizioni previste dall'art. 5 del GDPR e pertanto saranno trattati secondo i principi di:

1. liceità, correttezza e trasparenza
2. limitazione della finalità
3. minimizzazione dei dati
4. esattezza
5. limitazione della conservazione
6. integrità e riservatezza

Lo Studio in argomento comporta il trattamento di dati personali riconducibili allo stato di salute

degli assistiti in cura presso l'IRCCS, secondo i criteri di inclusione dello Studio.

2.1.1 Gli scopi del trattamento sono specifici, espliciti e legittimi?

Il trattamento correlato allo Studio è effettuato nel rispetto del principio di liceità e trasparenza. A tal proposito è stata predisposta e pubblicata sul sito internet istituzionale, l'informativa Privacy sugli studi retrospettivi. Lo scopo dello Studio è esplicito ed è descritto dettagliatamente nella documentazione di presentazione del medesimo Studio approvato dal Comitato Etico competente.

2.1.2 Quale è il flusso dei dati durante il ciclo di vita del trattamento?

Il ciclo di vita del dato ha origine dall'acquisizione dei dati relativi alla salute dalla documentazione sanitaria e archivi presenti presso le Unità Operative dell'IRCCS ai sensi dell'art. 110Bis, 4 comma, Cod. Privacy, e dei singoli centri partecipanti.

Su tali dati verranno effettuate le attività delle elaborazioni statistiche peculiari dello Studio. La trasmissione dei dati cifrati, dai centri partecipanti al Promotore, avverrà solo tramite la piattaforma RedCAP. Ad ogni centro partecipante saranno assegnate credenziali specifiche.

I dati clinici sono introdotti su sistema informativo di raccolta delle eCRF "RedCAP" ospitato sulla IAAS di InnovaPuglia (Cloud Sanità) che espone la propria interfaccia web su Internet con protocollo https e separa la Web Application dal server database.

La Web Application è sita nel layer DMZ, non raccoglie dati, ma consente l'accesso agli utenti autorizzati ed è protetta dalla rete internet esterna tramite firewall. Il database che colleziona i dati è accessibile esclusivamente dalla DMZ, la connessione è protetta da un apposito firewall.

I dati relativi all'identità del paziente sono sottoposti a pseudonimizzazione eseguita secondo una delle due procedure descritte nel paragrafo relativo alle misure tecniche.

I dati provenienti da centri collaboranti avranno accesso alla eCRF che espone i propri servizi secondo quanto descritto nel paragrafo delle misure tecniche.

Per quanto attiene alle immagini esse vengono trasferite mediante collegamento in VPN (con protocollo IPSec) su server FTPS ospitato su rete interna all'Istituto con acceso controllato tramite apposite credenziali.

L'infrastruttura prevede uno spazio di archiviazione dedicato (NAS), cu cui viene effettuato un backup completo dei dati con frequenza giornaliera, dal provider del servizio cloud (InnovaPuglia s.p.a.).
L'accesso al sistema cloud è basato su connessione VPN con protocollo IPsec, che garantisce l'autenticazione dell'utente, preservando l'integrità dei dati e la confidenzialità.

Al sistema cloud possono accedere solo i ricercatori individuati per lo studio in oggetto, per i quali verranno messe a disposizione delle apposite credenziali di accesso.

Per quanto attiene ai vetrini di H&E, i vetrini oggetto di studio saranno trasferiti dai centri collaboranti al centro coordinatore utilizzando servizi qualificati (corriere) per il trasferimento di materiale biologico. Ciascun vetrino di H&E sarà opportunamente identificato a cura del centro collaborante con indicazione del codice pseudo-anonimizzato definito in fase di collazionamento del paziente. Il centro coordinatore si occuperà della digitalizzazione del vetrino mediante gli scanner già disponibili presso il centro coordinatore e della successiva restituzione del vetrino con lo stesso servizio qualificato per il trasferimento di materiale biologico. L'immagine generata relativa al singolo vetrino digitalizzato sarà denominata con l'identificato individuato per lo stesso paziente in fase di reclutamento.

2.1.3 Quali sono le basi legali che rendono lecito il trattamento?

Basi giuridiche del trattamento di dati.

Paziente in vita e rintracciabile

Art. 6 par. 1, lett. A) e Art. 9 par. 2 lett. a) del GDPR (**acquisizione del consenso**).

Pazienti deceduti o non rintracciabili

Art. 9 par. 2 lett. j) del GDPR e artt. 110-110 bis c. 4 del d.lgs 196/03 e Aut. Gen. 9/2016 e ss aggiornamenti:
Il trattamento è necessario a fini di ricerca scientifica in campo medico, biomedico o epidemiologico, effettuata in base a disposizioni di legge o di regolamento o al diritto dell'Unione europea, ed è condotta e resa pubblica una valutazione d'impatto sulla protezione dei dati.

L'interessato è qualificato come non rintracciabile dopo almeno 3 tentativi (tracciati) di contatto non riusciti.
L'interessato deceduto viene rilevato dalla CC (in caso di decesso durante il periodo di degenza) o dal sistema TS (tessera sanitaria).

Ulteriori garanzie:
art. 8, comma 5-bis del d.lgs. n. 288 del 2003

2.1.4 I dati sono esatti e aggiornati?

I dati personali ed i campioni biologici sono acquisiti dagli archivi aziendali con ulteriori controlli interni in caso di omonimie o omocodie.

2.1.5 Qual è il periodo di conservazione dei dati?

Tipologia di dati personali	Tempi di conservazione
Dati pseudonimizzati	I dati pseudonimizzati saranno conservati per la durata dello Studio, decorsi i quali saranno anonimizzati.
Tabella di corrispondenza Codice ID/Paziente	La tabella di corrispondenza sarà cancellata in modalità permanente al termine della durata dello Studio in parola.

2.2 Misure a tutela dei diritti degli interessati

2.2.1 Come sono informati del trattamento gli interessati?

Con riferimento ai pazienti viventi, saranno rese le informazioni sul trattamento dei dati **ai sensi dell'art. 13** del Reg. UE 2016/679, nella fase di arruolamento.
A beneficio dei pazienti deceduti (o per quelli irreperibili) sono pubblicate nell'apposita sezione del sito internet istituzionale, le informazioni sul trattamento dei dati, **ai sensi dell'art. 14, par. 5, lett. b)** del Reg. UE 2016/679. È altresì pubblicato l'informativa al trattamento dei dati personali con relativa valutazione d'impatto.

2.2.2 Ove applicabile: come si ottiene il consenso degli interessati?

Paziente in vita e rintracciabile

Art. 6, par. 1, lett. a) e Art. 9 par. 2 lett. a) del GDPR

Il consenso, ove possibile, è raccolto dal paziente in fase di arruolamento, tramite l'acquisizione di firma autografa su modello "Consenso Informato Privacy specifico della Ricerca rev 3.0 del 30 ottobre 2024".

Paziente deceduti/non rintracciabili

Per tale categoria di pazienti, il consenso non può essere raccolto pertanto per il nostro IRCCS ci si avvale dell'art. **110 e 110 bis, comma 4 del Codice Privacy e art. 9, par 2, lett j), GDPR.**

2.2.3 Come fanno gli interessati a esercitare i loro diritti?

I diritti dei pazienti in vita e/o ricontattabili di cui agli artt. 15-22 del GDPR sono sempre garantiti nelle modalità indicate nell'informativa ex artt. 13-14 del GDPR rese al momento dell'arruolamento. Altresì sono resi disponibili sul sito internet istituzionale (https://www.sanita.puglia.it/web/irccs/privacy1) i modelli da poter utilizzare per l'esercizio di tali diritti.

I diritti di cui agli articoli da 15 a 22 del Reg. UE 2016/679 riferiti ai dati personali concernenti persone decedute possono essere esercitati da chi ha un interesse proprio, o agisce quale avente diritto o per ragioni familiari meritevoli di protezione. Le informazioni sul trattamento dei dati circa gli Studi condotti in assenza del consenso sono rese pubbliche sul sito internet istituzionale, nell'apposita sezione dedicata alla ricerca scientifica, unitamente alla valutazione di impatto sulla protezione dei dati personali.

2.2.4 In caso di trasferimento di dati al di fuori dell'Unione europea, i dati godono di una protezione equivalente?

I dati personali non saranno trasferiti verso Paesi Terzi extra UE.

2.3 Misure esistenti o pianificate per la protezione del dato

• **garanzie** (adozione di tecniche di pseudonimizzazione, minimizzazione, implementazione della

privacy by design e by default, previsione di procedure volte a testare, verificare e valutare l'efficacia delle garanzie e misure adottate).

• **misure di sicurezza organizzative** (es: norme e procedure che disciplinano l'aspetto organizzativo della sicurezza)

• **misure di sicurezze fisiche** (es: misure di protezione di aree, apparecchiature, dati)

• **misure di sicurezza logiche** (backup, piano di continuità operativa, piano di disaster recovery) sia in relazione al corretto utilizzo degli strumenti elettronici, sia in relazione alla loro gestione e manutenzione

Di seguito le principali misure tecniche applicate, ai sensi dell'art. 32 del Reg. UE 2016/679:

- Endpoint protection: Antivirus e *firewall* sulle singole postazioni di lavoro costantemente aggiornati mediante server ed associazioni a dominio. L'IRCCS ha acquisito un sistema di sicurezza integrato che comprende la gestione del *firewall* e del SOC.
  Per il monitoraggio e il controllo della rete viene utilizzato lo strumento Manage Engine Central che individua, gestisce e tiene traccia delle risorse poste sotto il dominio. Il SOC funge da primo soccorso in caso di incidente di sicurezza. Si possono eseguire operazioni come: isolare gli *endpoint,* terminare i processi dannosi, impedire l'esecuzione di processi dannosi ed eliminare i *files*.
- Implementazione di un Piano Operativo del servizio di sicurezza;
- Adozione del *cloud* di Regione Puglia, gestito dalla società *in-house* Innovapuglia, come *cloud* aziendale. Per i dati migrati sui menzionati *cloud* sono garantiti ridondanza dei dati e *backup*.
- Accesso alla postazione di lavoro mediante password a dominio aggiornata secondo i criteri di sicurezza adeguati al trattamento dei dati sensibili.
- Collocazione del database su postazione di lavoro isolata dalla rete. L'accesso ai locali come i centri di controllo è consentito solo al personale di manutenzione che detiene le chiavi di accesso. In tema di sicurezza fisica, viene altresì garantita la continuità elettrica da parte dei sistemi UPS, la refrigerazione attraverso impianti centralizzati e sistemi SPLIT local.
- Database criptato e protetto da password adeguato al trattamento dei dati sensibili
- Tecniche di pseudonimizzazione dell'identità dei pazienti realizzate alla volta con:
  1. tabella fisica di associazione pseudonimo/identità custodita in armadio a chiave dal PI e solo da questi accessibile
  2. esecuzione di algoritmi di hashing non reversibili a chiave
- Registrazione dei log di accesso al server applicativo e *database*. È altresì prevista la possibilità di verificare i log dall'Event Viewer di ciascuna postazione degli utenti abilitati come Amministratori di Sistema.

- Aggiornamento costante dei sistemi operativi e dei software di sistema e di ambiente. Predisposizione di un *asset inventory* tecnologico attraverso lo strumento di Manage Engine Central che individua, gestisce e tiene traccia delle risorse poste sotto dominio.
- Backup quotidiano della base dei dati su supporto ottico custodito separatamente in armadio ad accesso fisico ad uso esclusivo del PI. È presente un data center virtuale con servizi Backup As A Service presso Innovapuglia e PSN.
- Utilizzo di utenze nominative
- Meccanismi di identificazione ed autenticazione degli utenti
- Classificazione strutturata delle informazioni che tenga conto delle informazioni riservate/contenenti particolari categorie di dati *ex* art. 9 GDPR, attraverso sw dotati di certificazione di sicurezza.
- Password Policy adeguate al trattamento dei dati sensibili. L'ente ha sviluppato una policy sull'assegnazione delle password e che prescriva come tutte le macchine sotto dominio dell'ente richiedano periodicamente l'aggiornamento delle password.
- Erogazione di contenuti formativi per i dipendenti dell'ente che operano nel campo della ricerca.

Misure di sicurezza specifiche per campioni biologici:

Per la custodia e la sicurezza dei campioni biologici sono adottate le seguenti cautele:

a) L'accesso ai locali avviene previa identificazione delle persone, preventivamente autorizzate, che accedono a qualunque titolo dopo l'orario di chiusura;
b) la conservazione, l'utilizzo e il trasporto dei campioni biologici avvengono con modalità volte anche a garantire la qualità, l'integrità, la disponibilità e la tracciabilità;
c) la consultazione dei dati biologici trattati con strumenti elettronici è consentita tramite sistemi di autentificazione multi-fattore;
d) i campioni biologici contenuti in banche dati, sono trattati con tecniche di cifratura /pseudonimizzazione.

Con specifico riferimento alle operazioni di elaborazione dei dati dello Studio memorizzati in database centralizzato presso l'IRCCS Bari, sono implementate le seguenti misure di garanzia:

a) sistemi di autenticazione e di autorizzazione per il personale preposto al trattamento in funzione dei ruoli ricoperti e delle esigenze di accesso al trattamento;

b) procedure per la verifica periodica della qualità e coerenza delle credenziali di autenticazione e dei profili di autorizzazione assegnati ai soggetti designati al trattamento (sperimentatori);
c) sistemi di audit log per il controllo degli accessi al database e per il rilevamento di eventuali anomali.

3 Rischi

3.1 Panoramica dei rischi per diritti e libertà

Il processo di ***valutazione del rischio*** parte dalla determinazione dell'impatto sull'interessato (cioè sulla persona fisica a cui il dato si riferisce) in caso di distruzione, perdita, modifica, divulgazione non autorizzata o altri avvenimenti negativi che possono compromettere la sicurezza del trattamento.

L'impatto derivante dalla perdita di una o più delle caratteristiche della sicurezza delle informazioni, ossia riservatezza, integrità e disponibilità, rappresenta la gravità del danno diretto o indiretto causato agli interessati.

Nel valutare i rischi per le libertà e diritti degli interessati, però, come suggerisce la norma ISO/IEC 29134 si dovrebbero considerare anche altri aspetti, oltre alla sicurezza dei dati; e che pertanto devono essere considerati gli effetti complessivi del trattamento.

I rischi pertanto sono identificati in base ai seguenti quattro parametri:

1) conformità ai principi applicabili al trattamento dei dati (art. 5 del Reg. UE 2016/679)
2) riservatezza
3) integrità
4) disponibilità.

A tal fine, nella determinazione del livello di impatto sono incluse valutazioni sulle possibili conseguenze derivanti da mancanza di trasparenza, mancato rispetto dei tempi di conservazione dei dati, o dalla violazione degli altri principi fondamentali applicabili alla protezione dei dati personali.

- **Quali sono le principali minacce che potrebbero concretizzare il rischio?**

Una minaccia potrebbe concretizzarsi solo al momento dell'acquisizione dei dati durante la consultazione della documentazione sanitaria, che però è effettuata da personale esercente la professione sanitaria, tenuta al segreto professionale, ed istruita in materia di protezione dei dati

personali.

- **Quali sono le fonti di rischio?**

Una fonte di rischio potrebbe essere rappresentata dalla tabella di transcodifica che è gestita separatamente e che se sottratta insieme al database centralizzato dello Studio, consentirebbe di risalire allo stato di salute ed alle patologie dei soggetti inclusi nello Studio.

Non si ravvisano rischi per l'assistito in merito alla perdita di disponibilità del dato in quanto, in caso di evento avverso, non saranno compromessi i dati acquisiti e conservati per finalità di diagnosi, assistenza e cura. Anche in caso di perdita di integrità, non saranno compromessi dati acquisiti e conservati per finalità di diagnosi, assistenza e cura, ma solo per la finalità dello Studio.

- **Quali misure fra quelle individuate contribuiscono a mitigare il rischio?**

Oltre alle istruzioni operative fornite agli sperimentatori, è implementato un sistema crittografico sull'archivio centralizzato che prevede crittografia AES 256 bit con 14 round o cicli di elaborazione crittografica.

- **Come stimereste la gravità del rischio, specialmente alla luce degli impatti potenziali e delle misure pianificate?**

Il rischio residuo calcolato, dopo l'adozione delle misure di sicurezza pianificate e di quelle tecniche generali dell'Ente, visto anche il rispetto del principio di non esclusività della decisione algoritmica in quanto la validazione resta degli specialisti che intervengono nelle fasi di studio (radiologi e ricercatori clinici) anche al fine di correggere eventuali output non conformi, considerata l'applicazione delle misure di sicurezza dirette sul dato come la pseudonimizzazione e cifratura, nonché dell'assenza del trasferimento dei dati personali verso Paesi extra UE, è BASSO.

Le fonti di rischio possono essere categorizzate in:

- **Violazioni dei principi applicabili ai trattamenti di dati personali**
- **Minacce alla sicurezza dei trattamenti**
- **Eventi con danni fisici/materiali**
- **Eventi naturali**
- **Perdita o indisponibilità di servizi essenziali**
- **Compromissione di dati e informazioni**

- **Problemi tecnici**
- **Azioni non autorizzate**
- **Compromissione di funzioni / servizi per errori o azioni malevole**

Il livello di rischio è direttamente proporzionale alla probabilità che si verifichino le diverse minacce e alla gravità dell'impatto per gli interessati. Può essere mitigato con l'applicazione delle necessarie misure di mitigazione.

Se l'applicazione delle misure di mitigazione riduce il livello di rischio, fornendo un primo livello di rischio residuo, il governo dei processi e il presidio di controlli efficaci può fornire un ulteriore livello di ponderazione. Ecco perché oltre alle specifiche contromisure, la metodologia utilizzata inserisce, mediante un self assesment, degli obiettivi di controllo specifici per diverse categorie e ambiti, e dei controlli sullo svolgimento del processo di Valutazione di impatto.

Per la data protection si fa riferimento ai controlli della ISO/IEC 29151, estensione di quelli della ISO/IEC 27001 Annex A, a quelli della ISO/IEC 27701:2019 e della ISDP10003:2018.

3.2 Accesso illegittimo ai dati

3.2.1 Quali potrebbero essere i principali impatti sugli interessati se il rischio si dovesse concretizzare?

Danno immateriale, perdita dignità, perdita di controllo sui propri dati personali, irritazione, perdita della fiducia nella sanità pubblica, perdita finanziaria

3.2.2 Quali sono le principali minacce che potrebbero concretizzare il rischio?

Accessi esterni non autorizzati, Uso improprio del software, Corruzione dei dati, Comunicazione illegale dei dati e dei documenti, Uso non autorizzato dei dati, attacco hacker.

3.2.3 Quali sono le fonti di rischio?

Fonti di rischio umane interne, fonti di rischio umane esterne

3.2.4 Quali misure fra quelle individuate contribuiscono a mitigare il rischio?

Crittografia AES 256 bit sugli archivi elettronici dello Studio e dei relativi backup, Controllo degli accessi logici, Tracciabilità, Lotta contro il malware, Gestione postazioni, Politica di tutela della privacy, Firewalling, EDR, registrazione dei log di accesso al server mediante applicativo e *database*.

3.2.5 Come stimereste la gravità del rischio, specialmente alla luce degli impatti potenziali e delle misure pianificate?

Significativa. La gravità del rischio potenziale di accesso illecito ai dati è stimata come ALTA, in considerazione della tipologia di dati raccolti.

3.2.6 Come stimereste la probabilità del rischio, specialmente con riguardo alle minacce, alle fonti di rischio e alle misure pianificate?

La probabilità di accadimento è stimata come BASSA in considerazione delle misure di garanzia implementate con particolare riferimento alle tecniche di pseudonimizzazione e crittografia applicate, oltre che a tutte le misure di natura tecnica e organizzativa implementate dall'ente.

3.3 Modifiche indesiderate dei dati

3.3.1 Quali sarebbero i principali impatti sugli interessati se il rischio si dovesse concretizzare?

Modifiche ai dati raccolti per finalità di ricerca non comportano un impatto diretto all'interessato.

3.3.2 Quali sono le principali minacce che potrebbero consentire la concretizzazione del rischio?

Accessi esterni non autorizzati, Azione di virus informativi o di codici malefici, Uso non autorizzato dei dati, Sabotaggio, Alterazione dolosa o colposa dati.

3.3.3 Quali sono le fonti di rischio?

Fonti di rischio umane interne e fonti di rischio umane esterne.

3.3.4 Quali misure, fra quelle individuate, contribuiscono a mitigare il rischio?

Crittografia, Controllo degli accessi logici, Tracciabilità, Lotta contro il malware, Gestione postazioni, Backup, Manutenzione, Politica di tutela della privacy, Firewalling, EDR, Registrazione dei log di accesso al server applicativo e *database*.

3.3.5 Come stimereste la gravità del rischio, in particolare alla luce degli impatti potenziali e delle misure pianificate?

Limitata. La gravità del rischio potenziale di modifica illecita dei dati è stimata come BASSA, in considerazione della presenza di dati originali già raccolti per finalità di diagnosi e cura.

3.3.6 Come stimereste la probabilità del rischio, specialmente con riguardo a minacce, fonti di rischio e misure pianificate?

La probabilità di accadimento è stimata come BASSA in considerazione delle misure di garanzia implementate.

3.4 Perdita di dati

3.4.1 Quali potrebbero essere gli impatti principali sugli interessati se il rischio dovesse concretizzarsi?

Perdita di fiducia, irritazione, perdita reputazione, perdita di controllo sui propri dati personali

3.4.2 Quali sono le principali minacce che potrebbero consentire la materializzazione del rischio?

Azione di virus informativi o di codici malefici, Sabotaggio, attacco hacker, Uso non autorizzato dei dati, Uso improprio del software, Accessi esterni non autorizzati

3.4.3 Quali sono le fonti di rischio?

Fonti di origine naturale, fonti di rischio umane esterne, fonti di rischio umane interne

3.4.4 Quali misure, fra quelle individuate, contribuiscono a mitigare il rischio?

Backup, Disaster Recovery plan, Manutenzione, Politica di tutela della privacy, Controllo degli accessi logici, Crittografia, Tracciabilità, Lotta contro il malware, Firewalling, EDR, Registrazione dei log di accesso al server applicativo e *database*.

3.4.5 Come stimereste la gravità del rischio, specialmente alla luce degli impatti potenziali e delle misure pianificate?

La probabilità di accadimento è stimata come BASSA in considerazione delle misure di garanzia implementate.

3.5 METRICHE PER ANALISI RISCHIO

Valori dei livelli di rischio

Livello	Descrizione

BASSO	Il rischio per gli interessati è accettabile dall'organizzazione mediante misure organizzative e tecniche idonee, ma deve continuare ad essere monitorato per controllare che cambiamenti non incrementino il livello di rischio
MEDIO	Il rischio medio per gli interessati potrebbe essere accettabile ma l'adozione delle misure tecnico-organizzative deve essere monitorata su base regolare, e il trattamento può essere sottoposto a ulteriori considerazioni
ALTO	Il rischio per le persone interessate al trattamento è ad un livello non accettabile enecessita un rafforzamento delle misure di mitigazione
ELEVATO	Il rischio per gli interessati si presenta elevato o molto critico, mantenendo un livello non accettabile per l'organizzazione e necessitando l'aggiunta di ulteriori controlli a prevenzione/mitigazione dello stesso

Valori dei livelli di probabilità

Livello	Descrizione
BASSO	Evento/Minaccia poco probabile/frequente, o raro; è improbabile che la minaccia si concretizzi in condizioni normali o può verificarsi con frequenza inferiore rispetto alle tendenze riportate da studi, ricerche, statistiche di settore
MEDIO	Evento/Minaccia possibile; è un evento che si è già verificato o che può verificarsi con frequenza in media con le tendenze riportate da studi, ricerche, statistiche di settore
ALTO	Evento/Minaccia probabile; è un evento che si è già verificato o che può verificarsi con frequenza superiore rispetto alla media con riferimento alle tendenze riportate da studi, ricerche, statistiche di settore

Valori dei livelli di impatto

Livello	Descrizione
IRRILEVANTE	Gli interessati possono incontrare alcuni piccoli inconvenienti, che supereranno senza troppi problemi
LIMITATO	Gli interessati possono incontrare disagi significativi, che riusciranno comunque a superare a dispetto di alcuni problemi
SIGNIFICATIVO	Gli interessati possono incontrare conseguenze significative, che dovrebbero essere ingrado di superare anche se con gravi difficoltà
CRITICO	Gli interessati possono avere conseguenze gravi, o addirittura irreversibili, che potrebbero non superare

4 Panoramica dei rischi

Rischio Privacy	Descrizione delle conseguenze per gli interessati derivanti dalla vulnerabilità del trattamento	Livello di impatto
Perdita dei dati personali	**La perdita dei dati potrebbe comportare un danno agli interessati in termini di perdita di controllo sui propri dati**	**MEDIO**
Distruzione non autorizzata o indisponibilità	**La distruzione dei dati o l'indisponibilità degli archivi dello Studio non comporta un impatto diretto sugli interessati**	**BASSO**
Modifica non autorizzata	**La modifica dei dati per finalità di ricerca non comporta un impatto diretto sugli interessati**	**BASSO**

Divulgazione non autorizzata	**La divulgazione di dati personali dei partecipanti allo Studio potrebbe comportare un danno rilevante agli interessati**	**ALTO**
Accesso ai dati non autorizzato	**L'accesso illecito ai dati personali dei partecipanti allo Studio potrebbe comportare un danno rilevante agli interessati**	**ALTO**
Eccessiva raccolta di dati personali	**Utilizzare più dati personali del dovuto implicherebbe un'esposizione di dati personali all'utilizzo per scopi non pertinenti e non compatibili**	**BASSO**
Collegamenti o raffronti inappropriati o non autorizzati a dati personali	**Collegamenti o raffronti con altre banche dati potrebbe comportare danni immateriali agli interessati**	**BASSO**

Rischio Privacy	**Descrizione delle conseguenze per gli interessati derivanti dalla vulnerabilità del trattamento**	**Livello di impatto**
Perdita di controllo dei dati da parte degli interessati	**La mancanza di trasparenza e sicurezza dei trattamenti potrebbe comportare un impatto per gli interessati**	**BASSO**

Riuso per finalità diverse dei dati personali senza la consapevolezza e/o il consenso degli interessati	**I dati personali potrebbero essere utilizzati per altre finalità sconosciute all'interessato con danno immateriale agli interessati (mancanza di trasparenza e consenso)**	**BASSO**
Disequità o difettosità dell'elaborazione o del processo	**In caso di errata elaborazione delle informazioni, errori di registrazione etc. gli interessati potrebbero subire nocumento**	**BASSO**
Conservazione immotivatamente prolungata dei datipersonali	**La conservazione dei dati oltre il periodo prestabilito e motivato potrebbe comportare un danno immateriale agli interessati**	**BASSO**
Inesattezza o perdita di qualità dei dati personali	**Eventuali inesattezze o perdita della qualità dei dati raccolti non presenta un impatto diretto sui pazienti**	**BASSO**
Re-identificazione dei soggetti interessati	**Il processo di anonimizzazione potrebbe non eliminare la probabilità di re-identificazione dei partecipanti allo Studio, con particolare riferimento a malattie rare, con conseguente nocumento agli interessati**	**BASSO**

DATA PROTECTION IMPACT ASSESSMENT

ai sensi dell'art. 35 del Reg. UE 2016/679

CATEGORIE DI MINACCE CONSIDERATE	Livello MAX Prob.
Minacce alla conformità del trattamento	BASSO
Eventi con danni fisici	BASSO
Eventi naturali	BASSO
Indisponibilità dei servizi essenziali	BASSO
Violazioni di dati per azioni deliberate	MEDIO
Problemi tecnici	BASSO
Violazioni di dati per azioni involontarie	BASSO

CATEGORIE DI MINACCE	EFFICACIA MISURA ESISTENTE
Minacce alla conformità del trattamento	MISURE ESISTENTI ADEGUATE
Eventi con danni fisici/materiali/immateriali	MISURE ESISTENTI ADEGUATE
Eventi Naturali	MISURE ESISTENTI ADEGUATE
Indisponibilità di Servizi essenziali	MISURE ESISTENTI ADEGUATE
Compromissione di dati e informazioni per azioni deliberate	MISURE ESISTENTI ADEGUATE
Problemi tecnici	MISURE ESISTENTI ADEGUATE

DATA PROTECTION IMPACT ASSESSMENT

ai sensi dell'art. 35 del Reg. UE 2016/679

Compromissione di dati o servizi per azioni involontarie	**MISURE ESISTENTI ADEGUATE**

A seguito della ponderazione del livello di rischio calcolata mediante l'applicazione della mitigazione delle misure tecniche ed organizzative, il **rischio residuo** risulta **BASSO**, pertanto

ACCETTABILE ☒	**NON ACCETTABILE** ☐

DATA PROTECTION IMPACT ASSESSMENT

ai sensi dell'art. 35 del Reg. UE 2016/679

Il Titolare del trattamento, in persona del direttore Generale *pro tempore*, preso atto delle valutazioni sopra riportate in ordine all'analisi del potenziale impatto per i diritti e le libertà degli interessati, con l'adozione della VIP, dispone che il documento:

a) sia reso pubblico sul sito internet istituzionale nell'apposita sezione della Ricerca Scientifica (pubblicazione obbligatoria se lo Studio rientra nell'ambito del programma di ricerca nazionale, ai sensi dell'articolo 12-bis del decreto legislativo 30 dicembre 1992, n. 502)
b) sia resa disponibile agli interessati, su istanza dei medesimi.

Data……..

Firma del Direttore Generale……