

SINOSSI DEL PROGETTO “PsAicare Multimodal-AI”

Progetto presentato ammesso a cofinanziamento dall'ATS PuntoUnimol, costituita dall'impresa mandataria Punto Exe S.r.l. e dall'Università degli Studi del Molise sul Programma Regionale Molise FESR FSE+ 2021-2027 – Priorità 1 – “Un Molise più intelligente” - Obiettivo specifico RS01.1 - Sviluppare e rafforzare le capacità di ricerca e di innovazione e l'introduzione di tecnologie avanzate - Azione 1.1.1 – “Sostegno a progetti di ricerca, sviluppo sperimentale, trasferimento tecnologico e innovazione – ivi inclusi quelli incentrati sull'economia circolare – nelle imprese” – Avviso Pubblico Aiuti alle imprese per progetti di Ricerca e Sviluppo – I Edizione – CUP J39J25000350009

Contesto e motivazioni

La pandemia da Covid-19 ha accelerato l'adozione della psicoterapia online, mettendo in evidenza i limiti delle piattaforme generaliste di videocomunicazione rispetto ai requisiti di sicurezza, privacy e supporto al lavoro clinico degli psicologi. Per rispondere a questa esigenza è stata sviluppata una piattaforma dedicata alla psicologia online, che offre videochiamate sicure, gestione del setting terapeutico e strumenti digitali specifici per la pratica clinica.

L'integrazione di algoritmi di Intelligenza Artificiale per l'analisi delle emozioni rappresenta il passo successivo: non per sostituire il terapeuta, ma per fornirgli indicatori oggettivi e continui sullo stato emotivo del paziente, difficilmente osservabili in modo sistematico nel corso di molte sedute. Il progetto PsAicare Multimodal-AI nasce per superare i limiti delle soluzioni IA esistenti, in particolare l'uso di dataset video generici e non specifici per il contesto psicologico, e l'assenza di una reale integrazione multimodale tra informazioni facciali e vocali.

Descrizione della soluzione tecnologica

PsAicare Multimodal-AI si basa su una piattaforma cloud per la psicologia online in cui il paziente accede a una stanza virtuale dedicata e il terapeuta dispone di strumenti di gestione del percorso (agenda, note, questionari, materiale clinico). All'interno di questo ambiente vengono integrati due moduli IA distinti ma coordinati: il **motore di analisi facciale delle emozioni** e il **motore di analisi vocale emotiva**.

Il modulo video analizza il flusso della webcam durante la seduta, rilevando espressioni facciali e micro-espressioni associate a emozioni primarie e ad alcune emozioni secondarie clinicamente rilevanti; produce nel tempo indicatori quantitativi e grafici di andamento emotivo. Il modulo audio elabora invece il segnale vocale (prosodia, intonazione, ritmo, pause, pattern di discorso), estraendo feature che descrivono lo stato emotivo e la presenza di pattern non convenzionali potenzialmente associati a condizioni psicologiche specifiche.

Le informazioni provenienti dai due canali vengono combinate tramite tecniche di **data fusion** per generare una rappresentazione multimodale robusta, visualizzata in una dashboard per il terapeuta con timeline emotiva, indicatori di stress, livelli di congruenza/incongruenza tra volto e voce e altri indici configurabili. In ogni momento il clinico può confrontare tali indicatori con le proprie note, i temi emersi e gli strumenti psicometrici utilizzati.

Struttura del programma di ricerca

Il programma è articolato in sei Work Package (WP), distinti fra Ricerca industriale (RI) e Sviluppo Sperimentale (SS) che coprono l'intero ciclo di vita della soluzione, dalla ricerca di base fino alla preparazione

al mercato. I **WP1–3** riguardano prevalentemente il canale **audio**, mentre i **WP4–6** si concentrano sulla componente **video** e sull'integrazione finale multimodale.

- **WP1 – Analisi vocale: ricerca e selezione tecnologica (RI) - Università del Molise**
Svolge una ricognizione approfondita sullo stato dell'arte delle tecnologie per il riconoscimento delle emozioni dal parlato, selezionando framework, librerie e modelli di IA (ad esempio reti neurali ricorrenti, Transformer) più adatti all'ambito psicologico. Definisce le caratteristiche vocali chiave (prosodia, intonazione, ritmo, variazioni di frequenza) e produce linee guida per l'implementazione del motore audio.
- **WP2 – Ricerca innovativa su pattern di discorso e segnali vocali (RI) - Università del Molise**
Approfondisce l'analisi semantica e dei pattern di discorso (pause, interruzioni, cambi di velocità, flussi narrativi) per identificare algoritmi capaci di cogliere sfumature emotive non convenzionali. In collaborazione con psicologi, basato su dataset vocali generici (commerciali) o simulati, clinicamente orientati.
- **WP3 – Scrittura e ottimizzazione del modulo di analisi vocale di base (RI) – Punto Exe S.r.l.**
Traduce le indicazioni scientifiche dei WP1–2 in una prima architettura software del modulo vocale, curando la pulizia del segnale, l'ottimizzazione del codice e la scalabilità. Il risultato atteso è un prototipo di motore di analisi audio.
- **WP4 – Dataset proprietario e motore di analisi facciale (SS) - Università del Molise**
Costruisce un dataset di immagini e video provenienti da sedute reali o simulate in contesto universitario, etichettato da psicologi secondo protocolli rigorosi e nel pieno rispetto delle normative privacy. Il Dipartimento di Informatica addestra un nuovo algoritmo di analisi emozionale facciale, più accurato e specifico rispetto alle soluzioni basate su dataset pubblici generici.
- **WP5 – Ingegnerizzazione del sistema multimodale (SS) – Punto Exe S.r.l.**
Progetta l'architettura a microservizi e la containerizzazione per ospitare i moduli di analisi vocale, facciale e la loro fusione; realizza la dashboard clinica con timeline emotiva e indicatori di congruenza volto–voce. In questa fase avviene l'integrazione dei moduli IA nella piattaforma e la preparazione ai test su scala limitata con professionisti.
- **WP6 – Verifica, validazione e preparazione al mercato (SS) - Punto Exe S.r.l.**
Conduce test approfonditi su performance, sicurezza, scalabilità e usabilità del sistema integrato, coinvolgendo psicologi in scenari reali di seduta e servizi di supporto psicosociale. Raccoglie feedback clinici per ottimizzare gli algoritmi e l'interfaccia, garantendo la piena conformità alle normative sui dati sanitari e biometrici e ponendo le basi per un futuro utilizzo su larga scala.

Obiettivi generali e specifici

L'obiettivo generale del progetto è sviluppare un sistema multimodale audio–video basato su IA, integrato nella piattaforma di psicologia online, in grado di riconoscere e rappresentare in modo affidabile le emozioni espresse dai pazienti durante sedute psicologiche o psicoterapeutiche a distanza. Il sistema deve fornire al clinico indicatori sintetici, comprensibili e clinicamente rilevanti, mantenendo il pieno controllo decisionale nelle mani del professionista.

Tra gli obiettivi specifici rientrano:

- la costruzione di **dataset proprietari e annotati** da professionisti della salute mentale per il canale facciale;

- lo sviluppo di **nuovi algoritmi di analisi vocale emotiva** (prosodia, ritmo, pause) e di **riconoscimento facciale delle emozioni** basati su metodi di IA avanzati;
- la progettazione di un'architettura software scalabile e sicura che integri i due canali in un unico motore di analisi multimodale;
- la validazione tecnica, d'uso e preliminarmente clinica del sistema in contesti reali, in collaborazione con psicologi, servizi di supporto psicosociale e mondo accademico.

Endpoint e disegno dello studio osservazionale (WP4)

Nell'ambito del progetto complessivo, la presente sperimentazione è configurata come studio osservazionale, non interventistico, finalizzato alla costruzione e alla caratterizzazione di un dataset proprietario di frame facciali provenienti da sedute psicologiche/psicoterapeutiche reali e annotati da psicologi/psicoterapeuti secondo criteri standardizzati definiti nel protocollo.

Lo studio non prevede modifiche al trattamento psicologico o psicoterapeutico dei partecipanti, né interventi sperimentali sulla terapia. Le sedute si svolgono secondo la normale pratica clinica e l'attività di ricerca riguarda esclusivamente la raccolta, selezione, anonimizzazione e annotazione dei frame ottenuti dalle registrazioni video.

Endpoint primario

Considerata la finalità principale dello studio, volta alla realizzazione e alla caratterizzazione di un dataset di frame facciali etichettati, l'endpoint primario sarà il numero complessivo di frame validi e inclusi nel dataset finale, fissato a minimo 90.000.

Endpoint secondari

- Disponibilità di un dataset strutturato e caratterizzato quantitativamente, con informazioni relative alla numerosità dei frame, alla loro utilizzabilità e alla distribuzione delle categorie emotive;
- Definizione della percentuale di frame utilizzabili rispetto ai frame complessivamente acquisiti, nonché della frequenza dei frame esclusi o non classificabili;
- Verifica della fattibilità del flusso operativo complessivo, comprendente acquisizione, selezione, controllo, etichettatura e anonimizzazione dei frame in un contesto di pratica clinica reale;
- Disponibilità di una base dati strutturata per le successive attività di addestramento, validazione e valutazione di modelli di intelligenza artificiale e per supportare la lettura dei segnali emotivi non verbali durante le sedute.

Risultati attesi

Il dataset realizzato nell'ambito dello studio costituirà una base metodologica e tecnica per le successive attività di sviluppo e validazione di strumenti di analisi automatizzata delle espressioni facciali. I risultati potranno contribuire, in fasi successive, all'eventuale integrazione di un modulo di analisi delle espressioni facciali nella piattaforma PsyCare, concepito come strumento di supporto al professionista e non come sostituto del giudizio clinico.

Le potenziali ricadute dello studio si inseriscono nell'ambito della strategia di innovazione digitale per la salute mentale promossa dal progetto PsAicare Multimodal-AI e cofinanziato dalla Regione Molise, con l'obiettivo di valutare la possibilità di fornire ai professionisti uno strumento di supporto nell'osservazione e nell'interpretazione di segnali non verbali durante le sedute psicoterapeutiche.

Distinzione tra componente audio e video nella sperimentazione clinica

All'interno del progetto complessivo, la componente **audio** ha una forte connotazione di ricerca metodologica e di sviluppo preclinico: gran parte delle attività riguardano la selezione di tecnologie, la definizione delle feature, la costruzione di dataset vocali e la creazione di un motore di analisi da validare progressivamente. La componente **video**, grazie alla presenza di una piattaforma già operativa con un primo modulo di riconoscimento facciale, è invece maggiormente orientata alla creazione di un dataset proprietario e alla sperimentazione applicata in contesti clinici controllati.

Nelle attività che richiedono il coinvolgimento diretto di pazienti o utenti in setting psicologici regolati dal Comitato Etico, viene testato **esclusivamente** il modulo video di analisi facciale e la relativa dashboard; le funzioni di analisi vocale e di fusione multimodale restano confinate alle fasi di ricerca non clinica. In questo modo si garantisce una chiara separazione tra lo sviluppo tecnologico dei moduli IA e l'utilizzo clinico sugli assistiti.

Aspetti etici, privacy e sicurezza

Lo studio è stato sottoposto alla valutazione del Comitato Etico Territoriale (CET) ASReM ed è stato approvato il 21/10/2026 a seguito di verbale riunione CET del 28/09/2026. L'avvio delle attività di ricerca è lunedì 26/10/2026. Lo studio è stato redatto e sarà condotto in conformità al seguente codice etico: Helsinki, Art.3; (2) Helsinki, Art 1; (3) Norimberga, Art.2; (4) Norimberga, Art.2; (5) Helsinki, Art. 4, 5, 7; (6) Helsinki, Art.6; (7) Helsinki, Art. 9, 10, 11. Lo studio sarà condotto in accordo con le norme di Buona Pratica Clinica (ICH/GCP), come da D.M. 15 Luglio 1997 e in accordo con le altre normative vigenti in Italia in materia di sperimentazioni cliniche. In nessun caso è previsto che le procedure del protocollo in oggetto possano arrecare pregiudizio alla normale attività d'assistenza prestata. Le condizioni di salute, psicologiche e sociali dei pazienti su cui si svolge la ricerca, dovranno essere tali da consentire l'espressione di un libero consenso richiesto e ottenuto secondo le procedure stabilite.

Lo studio non prevede interventi sperimentali sulla terapia, né modifiche dei protocolli clinici; inoltre, il progetto adotta standard elevati di protezione dei dati e di etica nella ricerca. Tutte le attività che prevedono l'uso di dati audio o video di soggetti umani sono accompagnate da procedure di consenso informato, pseudonimizzazione dei dati, minimizzazione delle informazioni raccolte e conservazione su infrastrutture sicure, nel rispetto del Regolamento (UE) 2016/679 (GDPR) e della normativa applicabile in materia di protezione dei dati personali. Le decisioni cliniche non sono mai automatizzate: gli output del sistema hanno esclusivamente funzione di supporto al giudizio del professionista.

Sono previste procedure di audit, log degli accessi e controlli periodici di sicurezza informatica, nonché momenti di formazione per i professionisti coinvolti sull'uso appropriato degli strumenti IA e sui loro limiti. Il progetto promuove un approccio di "IA responsabile", in cui trasparenza, controllabilità e tracciabilità degli algoritmi sono condizioni essenziali per l'adozione in ambito di salute mentale.

Impatti attesi

PsAicare Multimodal-AI mira a migliorare la qualità della presa in carico psicologica online attraverso strumenti di Intelligenza Artificiale (IA) progettati per supportare e potenziare il lavoro del terapeuta, senza sostituirne il ruolo, il giudizio clinico o la relazione terapeutica. Questa soluzione può offrire al terapeuta nuovi strumenti di monitoraggio del decorso emotivo, potenzialmente utili per individuare precocemente cambiamenti significativi, segnali di rischio o momenti cruciali del percorso, lasciando sempre al professionista l'interpretazione delle informazioni e le decisioni cliniche. La piattaforma consente inoltre di sistematizzare informazioni che oggi possono rimanere implicite o affidate alla memoria del clinico, offrendo un ulteriore livello di supporto all'attività professionale e favorendo una presa in carico più consapevole e personalizzata.

Nell'ambito della presente sperimentazione, il principale risultato atteso è la realizzazione di un dataset proprietario di frame facciali, anonimizzati e annotati secondo procedure standardizzate, che costituisca la base per le successive attività di addestramento, sviluppo e validazione dei modelli di Intelligenza Artificiale (IA).

Dal punto di vista industriale, il progetto punta a generare una piattaforma innovativa e competitiva sul mercato nazionale e internazionale della salute mentale digitale, valorizzando competenze e dataset proprietari difficilmente replicabili. Per il sistema sanitario e per i servizi di supporto psicosociale, la soluzione può contribuire a una maggiore efficienza nella gestione dei percorsi e alla diffusione di buone pratiche di telepsicologia supportata da IA, in un modello human-in-the-loop in cui la tecnologia affianca il professionista senza sostituire la componente umana, relazionale e clinica della cura.