RELAZIONE TECNICA · COMPUTER VISION APPLICATA ALLA RIABILITAZIONE

Analisi del movimento in tempo reale da webcam

Una piattaforma web che trasforma una comune webcam in uno strumento di misurazione biomeccanica, senza marker, senza sensori indossabili e senza che un solo fotogramma video lasci il computer del paziente.

Piattaforma Physio Portal

Motore CV Google MediaPipe Tasks Vision

Elaborazione locale sul dispositivo

INDICE

01 Sintesi

02 Che cosa fa il sistema

03 Acquisizione da webcam

04 Il motore MediaPipe

05 La pipeline di elaborazione

06 I punti di repere

07 Metriche biomeccaniche

08 Esercizi su misura

09 Doppio schermo

10 Dati e archiviazione

11 Privacy by design

12 Requisiti operativi

13 Quadro di sintesi

01 Sintesi

La piattaforma affronta un problema concreto della riabilitazione: **la valutazione del movimento è oggi quasi sempre soggettiva**. Il fisioterapista osserva, stima un angolo articolare a occhio, annota una sensazione. Il progresso tra una seduta e l'altra è difficile da quantificare e quasi impossibile da mostrare al paziente.

La risposta tecnologica adottata è la **stima della posa umana da singola telecamera** (*markerless motion capture*): il flusso video della webcam viene analizzato fotogramma per fotogramma da una rete neurale che individua trentatré punti anatomici del corpo, e da quelle coordinate il sistema calcola in tempo reale angoli articolari, velocità, stabilità e simmetria del movimento.

Il risultato è una misura numerica, ripetibile e archiviabile, ottenuta con hardware che il paziente possiede già. Nessun sensore indossabile, nessuna tuta con marker, nessuna telecamera di profondità.

33
punti corporei tracciati per fotogramma

21
punti aggiuntivi per ciascuna mano

25–30
fotogrammi analizzati al secondo

0
fotogrammi video trasmessi in rete

---

02 Che cosa fa il sistema, in concreto

Il fisioterapista accede a una propria area riservata, registra i pazienti e costruisce per ciascuno un programma di esercizi. Ogni esercizio non è una semplice descrizione testuale: è una **configurazione geometrica**. Il terapista seleziona su un diagramma anatomico i punti del corpo che quel movimento coinvolge — per esempio spalla, gomito e polso per una flessione del gomito — e stabilisce che cosa il sistema deve misurare: un angolo, una distanza, un'ampiezza di movimento.

Il paziente si mette davanti alla webcam. Sullo schermo vede la propria immagine con sovrapposto lo scheletro digitale: i punti tracciati e i segmenti che li uniscono si muovono con lui in tempo reale. La postazione del terapista dispone della stessa vista, affiancata dal pannello di controllo e dalle metriche numeriche aggiornate in continuo.

Il modulo di registrazione archivia la sessione: non un video, ma la **traiettoria numerica completa del movimento**, insieme alle metriche calcolate. È il materiale su cui costruire il confronto tra la seduta di oggi e quella di sei settimane fa.

03 Acquisizione da webcam

Il livello di acquisizione è costruito sulle API `MediaDevices` del browser e gestisce l'intero ciclo di vita del flusso video: enumerazione delle telecamere disponibili, richiesta dei permessi, negoziazione della risoluzione, avvio, cambio dispositivo a caldo e rilascio pulito delle risorse all'uscita.

PARAMETRO	VALORE	MOTIVAZIONE
`width / height`	`1280 × 720` `max 1920 × 1080`	Risoluzione sufficiente al riconoscimento della posa mantenendo basso il carico di inferenza.
`frameRate`	`30 fps` `max 60`	Cattura movimenti riabilitativi senza scatti percepibili.
`facingMode`	`user`	Camera frontale: il paziente si vede come allo specchio.
`audio`	`false`	Il microfono non viene mai richiesto né attivato.
`zoom digitale`	`0,5× – 3,0×`	Adatta l'inquadratura a stanze piccole o a distretti corporei specifici.

Particolare cura è stata dedicata ai **casi di errore**, che in ambulatorio sono la norma e non l'eccezione. Permesso negato, nessuna telecamera collegata e telecamera già occupata da un altro programma vengono distinti e restituiti all'operatore con un messaggio che dice cosa fare, non un codice di errore. Ogni volta che il componente viene chiuso, tutte le tracce del flusso vengono esplicitamente arrestate: la spia della webcam si spegne.

Il sistema supporta inoltre configurazioni **multi-camera**: se sono presenti più dispositivi, l'operatore li seleziona da un menu e può passare dall'uno all'altro senza ricaricare la pagina — utile per affiancare a una camera frontale una vista laterale.

04 Il motore MediaPipe

Il riconoscimento della posa è affidato a **MediaPipe Tasks Vision**, la libreria di computer vision di Google, integrata nella versione `0.10.22`. È la stessa famiglia di modelli impiegata nei prodotti Google per il tracciamento del corpo umano, distribuita per l'esecuzione in ambiente browser.

L'aspetto architetturalmente rilevante è che **l'inferenza avviene interamente sul dispositivo del paziente**. I modelli vengono compilati in WebAssembly ed eseguiti nel browser: il video non viene inviato a un server per essere analizzato. Questa scelta ha tre conseguenze dirette — latenza nulla di rete, funzionamento indipendente dalla qualità della connessione, e assenza strutturale di dati video in transito.

I due modelli impiegati

MODELLO	COMPITO	CONFIGURAZIONE
`PoseLandmarker` `lite / float16`	Rilevamento dei 33 punti corporei sull'intera figura.	`numPoses: 1` `runningMode: VIDEO` `soglie: 0,3`
`HandLandmarker` `float16`	Rilevamento dei 21 punti articolari della mano, attivato solo per gli esercizi che riguardano il distretto mano-dita.	`numHands: 2` `runningMode: VIDEO` `soglie: 0,3`

La variante *lite* del modello di posa è una scelta deliberata: privilegia la fluidità su hardware modesto rispetto alla precisione marginale delle varianti maggiori. In riabilitazione, un tracciamento continuo a 25 fps è clinicamente più utile di uno più preciso ma a scatti.

Il modello delle mani viene caricato **solo quando serve**: il sistema esamina la configurazione dell'esercizio in corso e, se questo coinvolge punti della mano, inizializza il secondo modello e fonde i due insiemi di coordinate. Un esercizio di mobilizzazione della spalla non paga il costo computazionale del tracciamento delle dita.

Modalità video e coerenza temporale

Entrambi i modelli operano in `runningMode: VIDEO` anziché su immagini isolate. È una distinzione sostanziale: in questa modalità MediaPipe mantiene lo stato tra fotogrammi successivi e applica tracciamento anziché rilevamento indipendente, ottenendo traiettorie temporalmente coerenti — la condizione necessaria perché le derivate del movimento, cioè velocità e fluidità, siano misure sensate anziché rumore.

05 La pipeline di elaborazione

Ogni fotogramma percorre cinque stadi. La sequenza si ripete circa venticinque volte al secondo, sincronizzata al ciclo di rendering del browser.

01	02	03	04	05
Cattura	**Inferenza**	**Filtraggio**	**Metriche**	**Resa e memoria**
Il fotogramma corrente viene prelevato dal flusso della webcam e disegnato su una superficie di lavoro.	La rete neurale individua i punti anatomici e restituisce per ciascuno coordinate e grado di attendibilità.	I punti sotto la soglia di visibilità vengono scartati: articolazioni occluse non producono misure inventate.	Su una finestra scorrevole degli ultimi trenta fotogrammi si calcolano angoli, velocità, stabilità e simmetria.	Lo scheletro viene sovrapposto al video e le coordinate vengono registrate nella cartella della sessione.
~30 fps	tipico < 30 ms	soglia 0,4–0,6	finestra 1 s	disegno alternato

Governo del carico di calcolo

Un sistema che deve girare tanto su una workstation d'ambulatorio quanto su un portatile domestico non può avere un solo punto di funzionamento. Sono quindi implementati tre **profili prestazionali** selezionabili durante la sessione, che agiscono contemporaneamente su frequenza di analisi, soglia di attendibilità e risoluzione di lavoro:

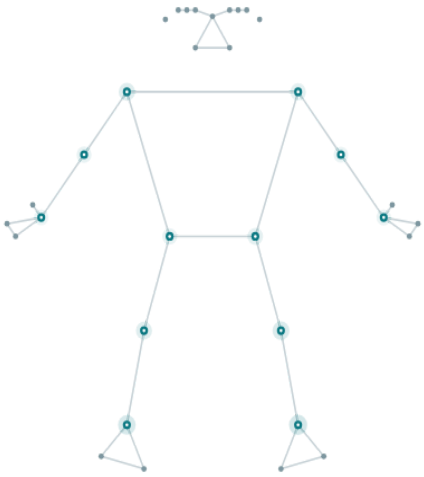
PROFILO	ANALISI	SOGLIA	SCALA DI LAVORO	DESTINAZIONE
High Quality	30 fps	0,6	90 %	Postazione ambulatoriale performante
Balanced	25 fps	0,5	80 %	Predefinito
Fast	20 fps	0,4	70 %	Portatili e hardware datato

A questi si aggiungono due accorgimenti: il **disegno dello scheletro è disaccoppiato dall'analisi** e avviene a fotogrammi alterni, perché l'occhio umano non distingue la differenza mentre il processore sì; e il ciclo è agganciato a `requestAnimationFrame`, così l'elaborazione si sospende automaticamente quando la finestra passa in secondo piano.

Durante la sessione un pannello diagnostico riporta in continuo **fotogrammi al secondo, tempo di inferenza, attendibilità media del rilevamento e fotogrammi scartati**: l'operatore ha una lettura oggettiva della qualità della sessione in corso e può correggere illuminazione o inquadratura prima che i dati siano compromessi.

06 I punti di repere

Il modello di posa restituisce trentatré punti numerati, dal naso alle punte dei piedi. Per ciascuno il sistema riceve tre coordinate — orizzontale, verticale e profondità stimata — più un valore di visibilità che esprime quanto la rete sia sicura di quel rilevamento.



Modello dei 33 punti. Il diagramma riproduce la disposizione dei reperi così come è definita nel selettore anatomico della piattaforma. I punti evidenziati sono le articolazioni maggiori su cui poggia il calcolo degli angoli.

0–10	volto: naso, occhi, orecchie, bocca
11–22	arti superiori: spalle, gomiti, polsi, dita
23–32	arti inferiori: anche, ginocchia, caviglie, piedi

Accanto alle coordinate normalizzate sull'immagine, MediaPipe fornisce anche i **world landmarks**: le stesse posizioni espresse in unità metriche con origine al centro dei fianchi. Il sistema archivia entrambe le rappresentazioni. È una scelta che guarda avanti: le coordinate metriche sono la base su cui poggiano le misurazioni angolari indipendenti dalla distanza del paziente dalla telecamera.

Per il distretto della mano, il secondo modello aggiunge ventuno reperi per lato — polso, e per ogni dito metacarpo e falangi — permettendo di descrivere esercizi di motricità fine che il solo modello corporeo non potrebbe rappresentare.

07 Metriche biomeccaniche

Le coordinate grezze non sono ancora informazione clinica. Il motore di analisi le trasforma in quattro famiglie di misure, calcolate su una finestra scorrevole degli ultimi trenta fotogrammi — circa un secondo di movimento.

ROM	VELOCITÀ
8 articolazioni Escursione articolare in gradi, calcolata come angolo tra tre reperi consecutivi. Spalle, gomiti, anche e ginocchia, destra e sinistra.	**media / picco** Spostamento dei punti visibili nell'unità di tempo. Dalla dispersione dei valori si deriva la fluidità del gesto.
STABILITÀ	**SIMMETRIA**
0 – 100 Oscillazione del baricentro pelvico e indice di tremore, come deviazione standard della posizione del centro corpo.	**%** Confronto fra coppie di reperi speculari, distinto tra cingolo superiore e inferiore. Indicatore di compenso.

Il **Range of Motion** è la misura clinicamente più immediata: sostituisce la stima a occhio o la goniometria manuale con un valore numerico continuo. L'angolo del gomito, per esempio, è ricavato dalla terna spalla–gomito–polso; quello del ginocchio dalla terna anca–ginocchio–caviglia.

Le tre famiglie successive descrivono la **qualità** del movimento più che la sua ampiezza, ed è qui che l'approccio strumentale si distacca dall'osservazione clinica tradizionale. Un paziente può raggiungere l'escursione richiesta compensando con il tronco, oppure con un gesto a scatti: sono informazioni che l'ampiezza da sola non registra, e che il rapporto fra dispersione delle velocità e oscillazione del baricentro rende invece esplicite.

> NOTA METODOLOGICA
>
> Le misure prodotte sono **indicatori di andamento**, progettati per il confronto longitudinale dello stesso paziente nelle stesse condizioni di ripresa. Non sostituiscono la strumentazione diagnostica certificata: il valore clinico sta nella curva nel tempo, non nel singolo numero assoluto.

08 Esercizi su misura, senza scrivere codice

È la componente che distingue la piattaforma da un catalogo chiuso di esercizi preconfezionati. Il fisioterapista dispone di un **selettore anatomico interattivo**: una figura umana vettoriale su cui i trentatré reperi sono rappresentati come punti cliccabili, ciascuno con la propria denominazione anatomica.

La costruzione di un esercizio segue tre passaggi:

1. **Selezione dei punti** — il terapista clicca sui reperi coinvolti; il sistema traccia i collegamenti nell'ordine di selezione.
2. **Scelta del tipo di misura** — angolo su tre punti, distanza fra due punti, o ampiezza di movimento su un insieme libero.
3. **Descrizione clinica** — nome, istruzioni di esecuzione, distretto corporeo, difficoltà e durata consigliata.

L'esercizio così definito viene archiviato come **configurazione strutturata** accanto alla scheda descrittiva: l'elenco dei reperi coinvolti, il tipo di calcolo da applicare e i parametri di riferimento. Accanto alla descrizione testuale, ogni esercizio porta con sé la propria definizione geometrica in forma leggibile dalla macchina.

Lo stesso strumento opera in modalità **mano** sul modello a ventun reperi, con la nomenclatura clinica delle articolazioni digitali — carpo-metacarpale, metacarpo-falangea, interfalangea — per la costruzione di protocolli di motricità fine.

A supporto è presente una funzione sperimentale di **interpretazione del linguaggio naturale**: da una descrizione testuale dell'esercizio il sistema propone una preselezione dei reperi pertinenti, che il terapista conferma o corregge.

09 L'impianto a doppio schermo

La sessione è progettata attorno a una constatazione pratica: paziente e terapista non hanno bisogno delle stesse informazioni sullo schermo, e mostrarle insieme danneggia entrambi.

Il sistema apre quindi **due viste distinte**, destinate a due monitor diversi:

- **Vista Paziente** — a tutto schermo, sul monitor grande davanti a chi si allena. Solo la propria immagine speculare con lo scheletro sovrapposto e un indicatore di sessione attiva. Nessun numero, nessun grafico: l'attenzione resta sul movimento.
- **Vista Terapista** — sulla postazione dell'operatore. Video ridimensionato più il pannello di controllo: avvio e arresto sessione, statistiche live, escursione articolare, velocità, stabilità e simmetria aggiornate in continuo, e lo stato tecnico del sistema.

L'apertura delle due finestre è gestita con un comando unico, con una temporizzazione studiata per non incorrere nel blocco popup dei browser. L'inquadratura si adatta al formato del monitor su cui la finestra viene trascinata.

10 Dati e archiviazione

L'archivio è un database **PostgreSQL** gestito su infrastruttura serverless, organizzato attorno al percorso clinico: profili e ruoli, anagrafica dei fisioterapisti, cartelle dei pazienti con diagnosi e piano terapeutico, catalogo degli esercizi per categoria, assegnazioni personalizzate, sedute, obiettivi terapeutici con stato di avanzamento.

Le due tabelle rilevanti per la parte di computer vision sono quelle che ricevono l'esito dell'elaborazione:

ARCHIVIO	CONTENUTO	CADENZA
`dati_movimento`	Coordinate complete dei reperi per fotogramma, in forma normalizzata e metrica, con numero progressivo e attendibilità del rilevamento.	`ogni fotogramma`
`metriche_sessione`	Valori calcolati di escursione articolare, velocità, stabilità e simmetria, con unità di misura e articolazione di riferimento.	`ogni ~1 secondo`

Le coordinate sono archiviate in formato **JSON strutturato** e indicizzate. La scelta è motivata dall'evoluzione attesa del sistema: il formato dei reperi cambierà al variare dei modelli di riconoscimento, e una struttura documentale assorbe queste evoluzioni senza migrazioni distruttive dello schema.

Il dato conservato è la **traiettoria completa del movimento**. Significa che una sessione archiviata può essere rianalizzata a posteriori con metriche che oggi non sono ancora implementate, senza chiedere al paziente di ripetere l'esercizio. È il patrimonio informativo che si accumula nel tempo.

L'accesso è regolato da **tre ruoli** — amministratore di sistema, fisioterapista, paziente — con autenticazione a token firmato su cookie protetto e password cifrate con algoritmo bcrypt a livello di database. I pazienti accedono con codice fiscale e credenziali generate dal terapista, che può consegnarle stampate.

11 Privacy by design

È la caratteristica architetturale che merita più attenzione in fase di valutazione, perché non è una policy dichiarata ma una **proprietà strutturale del sistema**.

L'analisi è integralmente locale. Il flusso della webcam è collegato a un elemento video nel browser e da lì passa direttamente al motore di riconoscimento, eseguito nella stessa pagina. **Non esiste, in tutto il progetto, alcun componente di registrazione video**: nessuna funzione di cattura su file, nessuna esportazione di immagini, nessun caricamento verso un server. Il fotogramma esiste in memoria per i millisecondi necessari all'inferenza e viene sovrascritto dal successivo.

Ciò che raggiunge il server sono **esclusivamente coordinate numeriche**: liste di terne di numeri, prive di qualunque informazione di aspetto, sfondo, abbigliamento o identità visiva. Da esse è impossibile ricostruire l'immagine della persona o dell'ambiente.

> IMPLICAZIONE OPERATIVA
>
> Un sistema che non trasmette né conserva immagini elimina alla radice la categoria di rischio più delicata nel trattamento di dati sanitari da videoripresa. Il paziente può esercitarsi nella propria abitazione senza che nulla del suo ambiente domestico lasci il dispositivo. Nessun archivio video da proteggere significa anche nessun archivio video che possa essere violato.

Il microfono non viene mai richiesto. La spia della telecamera si spegne alla chiusura della sessione perché il rilascio delle risorse è esplicito e verificabile. I dati biometrici derivati restano collegati alla cartella del paziente sotto il controllo del fisioterapista responsabile.

12 Requisiti operativi

Il sistema è un'applicazione web: non richiede installazione né diritti di amministratore sulle macchine.

- **Postazione** — computer con browser aggiornato (Chrome, Edge, Firefox o Safari) con supporto WebAssembly.
- **Telecamera** — qualunque webcam a 720p o superiore, integrata o USB.
- **Rete** — connessione necessaria all'accesso e all'archiviazione, non all'analisi del movimento, che resta locale.
- **Spazio** — inquadratura che comprenda il distretto corporeo interessato; illuminazione frontale uniforme, evitando controluce.
- **Doppio schermo** — consigliato in ambulatorio per l'impianto a due viste; non obbligatorio per l'uso domiciliare.

L'inferenza è configurata per l'esecuzione su **processore**: una scelta di massima compatibilità, che garantisce il funzionamento anche su macchine prive di scheda grafica dedicata. Il sistema non pone quindi requisiti hardware particolari oltre a quelli di un normale computer da ufficio.

13 Quadro di sintesi

Le capacità tecnologiche descritte nelle pagine precedenti si riassumono nei seguenti termini.

AMBITO	SOLUZIONE ADOTTATA
Acquisizione	Webcam standard a 720p o superiore via API native del browser, con gestione multi-camera, zoom digitale e rilascio esplicito delle risorse.
Riconoscimento	Google MediaPipe Tasks Vision: modello di posa a 33 reperi, modello mani a 21 reperi per lato, in modalità video con coerenza temporale.
Esecuzione	Inferenza WebAssembly interamente sul dispositivo, con tre profili prestazionali e diagnostica di qualità in tempo reale.
Misurazione	Escursione articolare su otto articolazioni, velocità e fluidità, stabilità posturale e tremore, simmetria fra emilati.
Personalizzazione	Selettore anatomico interattivo per la definizione di esercizi su misura, per il corpo intero e per la motricità fine della mano.
Archiviazione	PostgreSQL su infrastruttura serverless: traiettorie complete del movimento in formato documentale indicizzato, con metriche associate.
Accesso	Tre ruoli distinti, autenticazione a token firmato su cookie protetto, password cifrate con bcrypt a livello di database.
Riservatezza	Nessuna registrazione, trasmissione o conservazione di immagini: al server giungono esclusivamente coordinate numeriche.

Considerazione conclusiva

Il nodo tecnologicamente più impegnativo di un sistema di questo tipo è il riconoscimento affidabile del corpo umano a partire da una singola telecamera non calibrata, in ambiente non controllato, a una frequenza sufficiente a misurare il movimento mentre avviene.

È il punto su cui la piattaforma è costruita, ed è verificabile dal vivo: basta una webcam e un browser.

Relazione tecnica

Physio Portal — piattaforma di riabilitazione assistita

Analisi condotta sul codice sorgente della piattaforma