

POLITECNICO DI TORINO



Dipartimento di Ingegneria Meccanica e Aerospaziale

Corso di Laurea Magistrale in Ingegneria Biomedica

TESI DI LAUREA MAGISTRALE

*Sviluppo e validazione di un sistema markerless basato su IA
per la valutazione posturale a supporto della riabilitazione
protesica degli arti inferiori*

Relatore

Prof. Taian Martins

Co-relatore

PhD Marco Gagliardi

Candidato

Salvatore Rapisarda

Matricola: 319720

Anno Accademico 2025–2026



Il presente lavoro di tesi, svolto in collaborazione con ROGA SpA (Enna, Italia), propone lo sviluppo e la validazione di un sistema markerless basato su IA per valutare la postura di pazienti con protesi meccanica di arto inferiore. Il sistema, pensato come strumento di telemedicina, consente al fisioterapista di verificare la correttezza posturale in stazione ortostatica tramite semplici fotografie da smartphone, eliminando la necessità di infrastrutture cliniche specializzate. A seguito di una revisione sistematica dei principali modelli di pose estimation markerless (quali DeepPoseKit, MediaPipe, DensePose), è stato selezionato MediaPipe (MP), per efficienza computazionale, regressione diretta di 33 keypoints e compatibilità con dispositivi a bassa potenza di calcolo, ideale per usi remoti. I keypoints, ossia punti anatomici di riferimento identificati automaticamente dal modello MP, costituiscono la base per l'estrazione delle metriche posturali. L'obiettivo è validare il sistema confrontandolo con VICON, gold standard optoelettronico marker-based, che ricostruisce con precisione millimetrica la cinematica 3D tramite marker riflettenti e telecamere IR. La validazione si focalizza su metriche lineari e angolari definite dal fisioterapista collaboratore: allineamento rotula-secondo metatarso, angolo tra femore e tibia, linea alba, linea degli apici delle scapole, linea virtuale sagittale, angoli di estensione di anca (valore ideale 0°-10°), ginocchio (0°) e caviglia (85°-90°). Tali parametri riproducono il protocollo clinico ortostatico adottato nella pratica fisioterapica per identificare asimmetrie o compensi posturali. La sperimentazione è stata condotta presso il laboratorio di analisi del movimento del Politecnico di Torino su tre pazienti con protesi di arto inferiore (amputazioni transfemorali o transtibiali). Per ciascun paziente, il protocollo VICON ha previsto: posizionamento in stazione ortostatica, raccolta dei dati antropometrici (quali altezza, peso e distanza medio-laterale tra i malleoli), applicazione dei marker sui repere anatomici secondo il modello Plugin Gait Full Body adattato al contesto protesico, e tre acquisizioni statiche della durata di 1 secondo. Il protocollo MediaPipe ha invece previsto tre acquisizioni per ciascuna vista (frontale o sagittale) tramite la fotocamera di uno smartphone posizionato a distanza fissa dal soggetto. Le metriche derivate dal modello markerless sono state quindi confrontate con quelle ottenute dal sistema VICON. I risultati preliminari mostrano una buona coerenza nei segmenti non protesici, mentre persistono criticità nella localizzazione dei keypoints in corrispondenza della protesi, assente nel dataset di addestramento del modello. Gli errori, nel contesto dei range clinici di riferimento adottati da ROGA, indicano un potenziale applicativo del sistema, con margini di miglioramento attraverso l'ampliamento del dataset e tecniche di adattamento del modello. In conclusione, lo studio dimostra la fattibilità di soluzioni markerless in ambito protesico. Il compromesso tra accessibilità e accuratezza, inevitabile rispetto a sistemi come VICON, appare giustificato dall'abbattimento delle barriere cliniche e logistiche che limitano la continuità riabilitativa oltre i confini del laboratorio.



Sommario

1. Stato dell'arte.....	7
1.1. Introduzione.....	7
1.2. Tecnologie principali alla base della stima markerless	8
1.3. Applicazioni.....	11
1.4. Dataset di riferimento per la stima della posa umana	11
1.5. Sfide specifiche per i portatori di protesi.....	12
1.6. Confronto tra modelli markerless per la stima della posa umana	13
1.6.1. AlphaPose	14
1.6.2. DeepPoseKit	14
1.6.3. DensePose	14
1.6.4. Detectron2	15
1.6.5. High-Resolution Network.....	15
1.6.6. MediaPipe (MP).....	16
1.6.7. OpenPifPaf	16
1.6.8. OpenPose	16
1.6.9. PoseNet.....	17
2. Materiali e metodi	17
2.1. Analisi comparativa	18
2.1.1. Accuratezza nella stima dei keypoints	18
2.1.2. Velocità di inferenza	18
2.1.3. Adattabilità.....	18
2.1.4. Implementazione	18
2.2. Approfondimento su DeepPoseKit.....	19
2.2.1. Architettura.....	19
2.2.2. Addestramento e ottimizzazione	20
2.2.3. Applicazioni	21
2.2.4. Limiti e considerazioni	21
2.2.5. Il tentativo con DeepPoseKit.....	21
2.3. Approfondimento su MediaPipe	22
2.3.1. Architettura.....	23
2.3.2. Specifiche tecniche	25
2.3.3. Personalizzazione tramite ModelMaker	26
2.3.4. Integrazione con strumenti di debug.....	26
2.3.5. Applicazioni	27
2.3.6. Integrazione clinica.....	27



2.3.7. Normativa.....	27
2.3.8. Limiti e considerazioni	27
2.4. Soggetti	28
2.4.1. Criteri di inclusione ed esclusione.....	28
2.4.2. Popolazione dello studio.....	29
2.4.3. Considerazioni etiche	29
2.5. VICON.....	29
2.5.1. Descrizione del sistema	29
2.5.2. Plug-in Gait Full Body	30
2.5.3. Posizionamento dei marker	31
2.5.4. Misure antropometriche.....	32
2.5.5. Calibrazione delle telecamere.....	34
2.5.6. Protocollo di acquisizione.....	35
2.6. MediaPipe	35
2.6.1. Strumentazione e configurazione	35
2.6.2. Protocollo di acquisizione.....	36
2.6.3. Rilevamento dei marker.....	36
2.6.4. Fattore di scala	37
2.7. Metriche posturali.....	39
2.7.1. Vista frontale	40
2.7.2. Vista sagittale.....	41
2.7.3. Metriche non incluse.....	42
2.8. Estrazione metriche da VICON	43
2.9. Metodologia di confronto.....	44
2.9.1. Struttura del dataset di confronto	44
2.9.2. Metriche di errore	44
2.9.3. Strumenti software.....	44
2.9.4. Motivazione dell'approccio descrittivo	44
3. Elaborazione Python: gli script nel dettaglio	45
3.1. Implementazione.....	45
3.2. Architettura	45
3.3. Calcolo dei fattori di scala	46
3.4. Elaborazione MediaPipe	46
3.4.1. Configurazione.....	46
3.4.2. Organizzazione cartelle e rilevamento automatico della vista	47
3.4.3. Calcolo delle metriche posturali	48



3.4.4. Output.....	48
3.5. Elaborazione VICON	50
3.5.1. Configurazione.....	50
3.5.2. Estrazione angoli articolari e traiettorie	50
3.5.3. Output.....	50
3.6. Notebook validazione.....	51
4. Prototipo applicativo per la telemedicina	51
4.1. Motivazione e obiettivi	51
4.2. Architettura del sistema.....	51
4.3. Calcolo automatico del fattore di scala	52
4.4. Endpoint API.....	53
4.5. Report clinico PDF	53
4.6. Android Studio	54
4.7. Considerazioni	54
5. Risultati	55
5.1. Struttura del dataset di confronto	55
5.2. Convenzione segni.....	55
5.3. Metriche della vista frontale	57
5.3.1. Angolo asse femore-tibia (valgismo/varismo)	57
5.3.2. Metriche lineari frontali.....	58
5.4. Metriche della vista sagittale	60
5.4.1. Deviazioni dalla plumb line virtuale sagittale	60
5.4.2. Angoli articolari sagittali	63
6. Discussione.....	68
6.1. Considerazioni generali sulla popolazione.....	68
6.2. Angolo asse femore-tibia nel piano frontale (valgismo/varismo).....	69
6.3. Metriche lineari frontali.....	69
6.3.1. Allineamento rotula–II metatarso.....	69
6.3.2. Deviazione della linea alba	70
6.3.3. Tilt delle scapole	70
6.4. Plumb line virtuale sagittale	70
da spalla	71
gran trocantere	71
ginocchio	71
caviglia	71
6.5. Angoli articolari sagittali	72



6.5.1. Angolo di estensione dell'anca.....	72
6.5.2. Angolo di estensione del ginocchio.....	72
6.5.3. Angolo di dorsi-flessione della caviglia.....	72
6.6. Analisi delle criticità sistematiche e prospettive di miglioramento.....	73
7. Conclusioni.....	74
7.1. Considerazioni.....	74
7.2. Il contributo applicativo.....	74
7.3. Interpretazione clinica e limiti.....	75
Appendice A Protocollo di acquisizione.....	77
Appendice B Report clinico generato dal prototipo applicativo.....	80
Bibliografia.....	82



1. Stato dell'arte

1.1. Introduzione

Nel campo della medicina riabilitativa e protesica, la precisione nella rilevazione delle pose umane costituisce un fattore determinante per il successo della personalizzazione delle soluzioni riabilitative e protesiche. Studi clinici hanno evidenziato come un'accurata analisi cinematica del movimento consenta di definire protocolli riabilitativi personalizzati, migliorando l'integrazione biomeccanica delle protesi e riducendo il rischio di complicanze secondarie [1].

Le tecniche per la stima della posa umana hanno percorso un lungo cammino nel corso degli ultimi decenni. I metodi tradizionali si basano sull'uso di marker fisici posizionati su punti di articolazione strategici del corpo, come quelli utilizzati nei sistemi di *motion capture*, ad esempio il sistema VICON. Queste soluzioni, sebbene estremamente precise, presentano numerosi limiti in termini di intrusività, costi elevati e necessità di contesti controllati [2]. La necessità di un approccio meno invasivo ha portato allo sviluppo di tecniche *markerless*, che inizialmente facevano affidamento su algoritmi di visione computazionale classici e modelli statistici per l'estrazione di *features*. L'introduzione di tecniche di deep learning ha modificato radicalmente gli approcci alla stima della postura umana, consentendo l'utilizzo di modelli basati su reti neurali addestrate su grandi dataset annotati. L'utilizzo di reti neurali convoluzionali (*Convolutional Neural Networks, CNN*) ha permesso di sfruttare enormi quantità di dati per addestrare modelli in grado di riconoscere in maniera robusta ed efficiente i *keypoints* del corpo umano. Quest'ultimi, sono punti di riferimento che rappresentano le posizioni significative, come la sede della testa, delle mani, del tronco, delle estremità e di altre aree rilevanti, e costituiscono la base per descrivere la configurazione posturale e il movimento del soggetto. Tra i pionieri di questa innovazione vi sono modelli come OpenPose, che ha introdotto il concetto delle *Part Affinity Fields* per riconoscere le connessioni tra *keypoints* su immagini bidimensionali. In parallelo, altre architetture, come HRNet, hanno puntato a mantenere le informazioni ad alta risoluzione lungo tutta la pipeline computazionale, garantendo una maggiore accuratezza nella stima. Questi progressi hanno fornito la base teorica e pratica per l'adozione di soluzioni *markerless* in applicazioni cliniche, aprendo la strada alla personalizzazione dei dispositivi protesici [3].

Gli sviluppi in tale settore hanno permesso di realizzare sistemi *markerless* capaci di:



- acquisire dati in tempo reale: oggi è possibile elaborare immagini con frequenze elevate (30-60 fps) anche su CPU standard, rendendo la tecnologia idonea per applicazioni dinamiche e per il monitoraggio continuo nei contesti clinici.
- adattarsi a dataset limitati: l'implementazione di tecniche di *data augmentation* e *transfer learning* consente di perfezionare i modelli anche quando si dispone di una quantità limitata di dati annotati, caratteristica preziosa in ambito clinico, ove i dataset specifici possono essere contenuti.
- mantenere un'elevata accuratezza: l'innovazione negli algoritmi di regressione diretta ha ridotto il carico computazionale e migliorato la precisione nella stima dei *keypoints* del corpo.

1.2. Tecnologie principali alla base della stima markerless

L'evoluzione delle tecniche di deep learning ha portato allo sviluppo di differenti architetture utilizzate nella stima markerless della postura.

- CNN: sono architetture neurali progettate per l'analisi di dati con struttura spaziale, come immagini o mappe di pixel. La loro architettura gerarchica è composta da strati convoluzionali, di *pooling* e completamente connessi che consentono l'estrazione automatica di caratteristiche visive, dalle più semplici (bordi, contorni) alle più complesse (forme, pattern corporei). Inoltre, durante l'addestramento, i parametri delle CNN vengono aggiornati iterativamente tramite l'algoritmo di *backpropagation*, che consente di minimizzare la funzione di perdita attraverso metodi di ottimizzazione basati sul gradiente. Questo approccio ottimizza i parametri della rete confrontando le previsioni del modello con i dati annotati (*ground truth*), consentendo l'apprendimento di rappresentazioni robuste dei *keypoints* corporei. Per ridurre i tempi di addestramento, l'allenamento delle CNN viene generalmente eseguito su GPU, particolarmente efficienti nell'esecuzione di operazioni matriciali parallele.

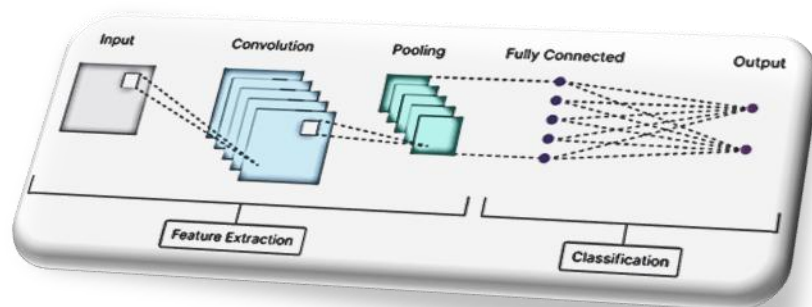


Fig.1.1 Architettura CNN

- Reti ricorrenti (*Recurrent Neural Networks, RNN*): sono progettate per l'elaborazione di dati sequenziali, come le serie temporali o le sequenze di immagini, grazie alla loro capacità di mantenere memoria degli stati precedenti attraverso connessioni ricorsive. Questa caratteristica le rende ideali per modellare dipendenze temporali nei dati, un aspetto fondamentale nella stima della posa umana, dove la coerenza tra fotogrammi consecutivi è essenziale per ricostruire traiettorie di movimento fluide e accurate. L'architettura di una RNN si basa su unità ricorrenti che elaborano ogni elemento di una sequenza, aggiornando uno stato nascosto che incorpora informazioni dai passi temporali precedenti. Nei sistemi di *pose estimation* sequenziale, lo stato nascosto della rete può essere utilizzato per stimare le coordinate dei *keypoints* mantenendo coerenza temporale tra fotogrammi consecutivi. L'addestramento delle RNN avviene tramite *backpropagation through time (BPTT)*, un'estensione della *backpropagation* che propaga i gradienti attraverso la sequenza temporale, ottimizzando i pesi della rete. Tuttavia, i gradienti possono ridursi esponenzialmente man mano che si propagano all'indietro attraverso molti passi temporali, rendendo difficile l'apprendimento di dipendenze a lungo termine (*vanishing gradient*). Ciò compromette la capacità delle RNN di modellare sequenze lunghe, come quelle tipiche di video clinici che registrano movimenti complessi su intervalli estesi. Inoltre, la complessità computazionale della BPTT può rendere l'addestramento lento, specialmente per sequenze lunghe.

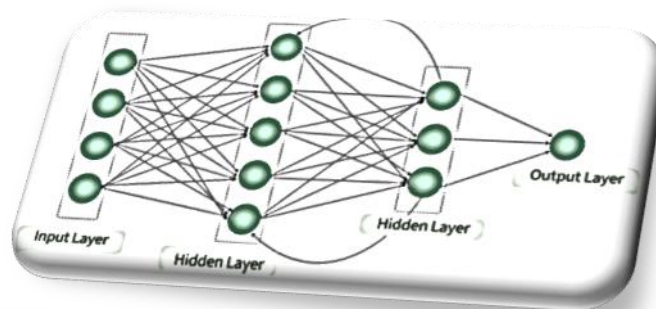


Fig.1.2 Architettura RNN

- *Long Short-Term Memory (LSTM)*: costituiscono un'evoluzione delle reti neurali ricorrenti tradizionali, progettata per mitigare il problema del *vanishing gradient* nelle sequenze temporali lunghe. Le LSTM si distinguono per l'introduzione di una cella di memoria a tre porte: *input gate*, che stabilisce quali nuove informazioni vengono memorizzate; *forget gate*, la quale decide quali informazioni memorizzate scartare; *output gate*, che controlla quali informazioni contenute nella cella vengono trasmesse all'output. Tali porte sono implementate come funzioni sigmoidali e operazioni puntuali, permettendo di preservare le informazioni

rilevanti per lunghi intervalli temporali, riducendo il *vanishing gradient* e garantendo un addestramento più stabile.

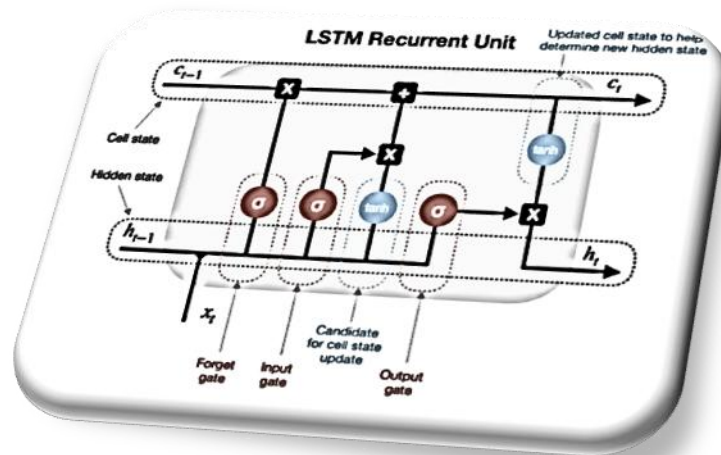


Fig.1.3 Architettura LSTM

- *Transformer*: realizzati originariamente per l'elaborazione del linguaggio naturale, hanno rivoluzionato il campo della visione computazionale. Tali strumenti elaborano intere sequenze in parallelo, superando i limiti delle RNN, sequenziali, e riducendo drasticamente i tempi di calcolo. Il transformer si basa su due componenti principali: *self-attention* e *feed-forward* con normalizzazione: vengono calcolati i pesi di importanza per ogni elemento della sequenza rispetto a tutti gli altri, consentendo al modello di concentrarsi su relazioni rilevanti, come la connessione tra *keypoints* corporei. L'addestramento è tipico e avviene tramite la minimizzazione della funzione di perdita. Infine, l'architettura parallela consente di sfruttare pienamente la GPU, rendendo i transformer efficienti anche su dataset di grandi dimensioni come quelli tipici per *pose estimation*.

- *Data augmentation*: è una strategia di apprendimento automatico che consiste nell'ampliare artificialmente un dataset attraverso l'applicazione di trasformazioni ai dati esistenti, al fine di migliorare la robustezza e la generalizzazione dei modelli. Nella stima della posa umana, dove l'accuratezza nella rilevazione dei *keypoints* dipende dalla capacità del modello di adattarsi a variazioni ambientali, posturali e antropometriche, la *data augmentation* svolge un ruolo cruciale, specialmente in contesti clinici caratterizzati da dataset limitati e altamente specifici, come quelli relativi a pazienti con protesi. Aumentando la variabilità dei dati di addestramento, questa tecnica riduce il rischio di *overfitting*, consentendo ai modelli di generalizzare meglio su nuovi scenari, come immagini acquisite in condizioni di illuminazione non ideale o con occlusioni parziali. Ai dati si possono applicare: trasformazioni geometriche (rotazioni, traslazioni, ridimensionamenti, riflessioni o deformazioni prospettiche) e trasformazioni

fotometriche (modifiche in termini di luminosità, contrasto, saturazione o rumore). In particolare, nella *human pose estimation (HPE)*, l'*augmentation* include occlusioni sintetiche (vengono aggiunte delle maschere al fine di simulare occlusioni parziali, come quelle causate dagli indumenti), variazioni temporali (alterazione della velocità dei movimenti o inserimento di fotogrammi interpolati) e la generazione di dati sintetici (dataset con scenari realistici ma non presenti nei dati originali).

- *Transfer learning*: tecnica di apprendimento automatico che consiste nell'utilizzare un modello pre-addestrato su un dataset ampio e generico, come fonte di conoscenza iniziale, per adattarlo a un compito specifico mediante un processo di *fine-tuning* su un dataset più piccolo e mirato. Questa tecnica è particolarmente efficace nella stima della posa, dove la disponibilità di dataset annotati di alta qualità, specialmente in contesti clinici, è spesso limitata da vincoli logistici, etici e finanziari. In campo biomedico, il *transfer learning* consente di superare la scarsità di dati, come immagini di pazienti con protesi, sfruttando caratteristiche apprese da dataset più ampi, quali COCO [4] o MPII [5], che contengono migliaia di immagini di pose umane in contesti vari.

1.3. Applicazioni

L'impiego della tecnologia *markerless* nella stima della posa umana ha portato numerosi benefici in ambito clinico.

- o Personalizzazione delle protesi: la possibilità di acquisire dati precisi sul movimento degli arti permette la progettazione di protesi su misura, migliorando l'adattamento del dispositivo e l'efficienza funzionale [6].
- o Riabilitazione motoria: in ambito riabilitativo, il monitoraggio continuo dei movimenti consente di valutare la progressione terapeutica e di adattare gli esercizi in base ai cambiamenti funzionali del paziente [7].
- o Analisi del cammino e diagnosi delle disfunzioni: la stima *markerless* consente di rilevare anomalie nel cammino e di diagnosticare precocemente potenziali disfunzioni muscoloscheletriche, supportando i medici nel definire protocolli terapeutici personalizzati [8].

1.4. Dataset di riferimento per la stima della posa umana

Lo sviluppo e la validazione dei modelli di stima della posa umana sono stati resi possibili dalla disponibilità di dataset pubblici di grandi dimensioni, che hanno fornito le annotazioni necessarie per l'addestramento supervisionato e la definizione di benchmark di riferimento per il confronto tra metodi. I principali dataset utilizzati in letteratura sono riportati in Tabella 1.1.



Dataset	Anno	Immagini/Sequenze	Keypoints	Modalità	Utilizzo
MPII Human Pose	2014	~25.000	16	2D, RGB	Training e benchmark 2D
MS-COCO	2014	~200.000	17	2D, RGB	Training e benchmark 2D
Human3.6M	2014	3,6M	32	3D, RGB + MoCap	Benchmark 3D
HumanEva	2010	Sequenze video	15	3D, RGB + MoCap	Benchmark 3D con ground truth
SUMediPose	2025	~51.000 pose, 71 soggetti	53	2D+3D, RGB + MoCap	Benchmark markerless-MoCap

Tabella 1.1 – Principali dataset pubblici per la stima della posa umana.

Il dataset MPII [5] e MS-COCO [4] rappresentano i riferimenti storici per la stima 2D della posa, e sono stati impiegati per il pre-addestramento della maggior parte dei modelli analizzati, incluso il modello di *pose estimation* MediaPipe. Human3.6M costituisce lo standard per la valutazione 3D. HumanEva [21], invece, ha introdotto la metodologia di acquisizione sincronizzata video-MoCap, anticipando il paradigma di validazione adottato nel presente lavoro. Mentre il recente dataset SUMediPose [22] rappresenta un riferimento specifico per studi di validazione markerless-MoCap, comprendendo 71 soggetti che eseguono esercizi di forza con acquisizione simultanea RGB e sistema VICON.

È importante sottolineare che nessuno di questi dataset include soggetti con protesi di arto inferiore, il che costituisce una lacuna significativa nella letteratura e una delle motivazioni principali del presente studio.

1.5. Sfide specifiche per i portatori di protesi

Una sfida specifica e finora poco affrontata in letteratura riguarda l'applicazione dei sistemi HPE a soggetti portatori di protesi di arto inferiore. Modelli come MediaPipe sono addestrati esclusivamente su soggetti normodotati, con morfologia degli arti inferiori omogenea e prevedibile. La presenza di una protesi altera in modo significativo l'aspetto visivo dell'arto, modificandone il profilo, la texture e la colorazione, introducendo compensazioni posturali atipiche che possono confondere il modello nella localizzazione dei *keypoints* distali. Zhou et al. [6] hanno quantificato questa criticità, dimostrando che OpenPose applicato su immagini di amputati transtibiali e transfemorali produce errori di localizzazione dei *keypoints* significativamente superiori rispetto ai soggetti normodotati, con riduzioni dell'errore del 37%



e 76% rispettivamente solo dopo trasformazione delle immagini tramite modelli di diffusione generativi [6]. Questo evidenzia la necessità di studi di validazione dedicati a questa popolazione.

1.6. Confronto tra modelli markerless per la stima della posa umana

Esistono numerosi modelli che impiegano approcci differenti per estrarre i *keypoints*: la scelta del modello ideale è determinata dalla capacità di bilanciare accuratezza, velocità di inferenza, adattabilità a dataset limitati e requisiti hardware contenuti.

I modelli per la stima della posa possono essere classificati in due tipologie,

- o *Top-Down*: il sistema parte dal rilevamento dell'intera persona e, successivamente, stima i *keypoints* all'interno delle regioni individuate, permettendo una maggior precisione nella stima. Tuttavia, la complessità computazionale cresce con l'aumentare del numero di persone presenti nell'immagine. Un esempio è AlphaPose;
- o *Bottom-Up*: il sistema rileva tutti i *keypoints* presenti nell'immagine indipendentemente dalle persone, influenzando sulla precisione complessiva e, successivamente, li raggruppa in base a criteri spaziali e relazionali per ricostruire le pose. Un esempio è OpenPose.

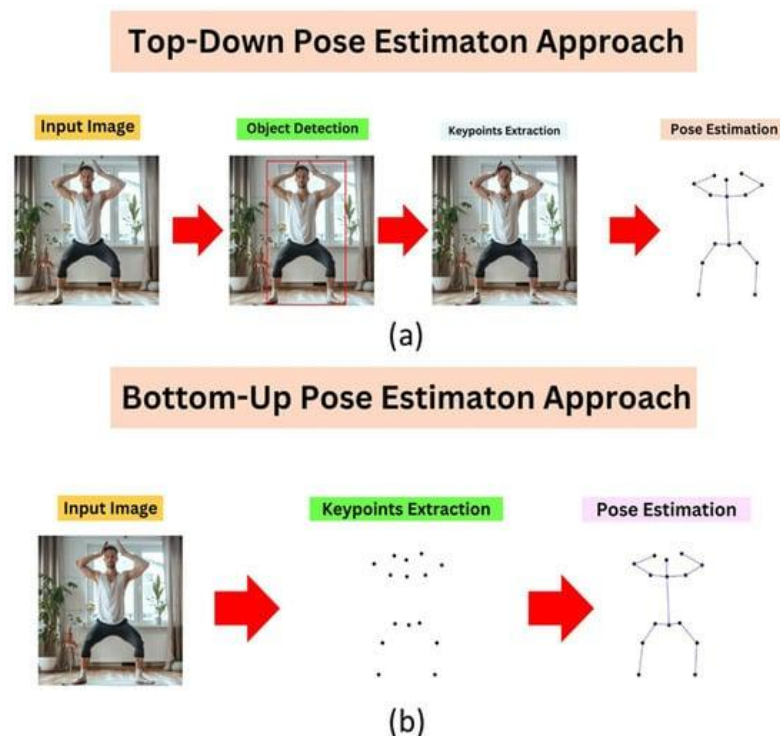


Figura 1.4 - Confronto tra approccio top-down (a) e bottom-up (b) [27]

Sono stati analizzati diversi modelli evidenziando le loro caratteristiche tecniche e i relativi limiti.

1.6.1. AlphaPose

Modello *markerless multi-body* che si distingue per la sua elevata accuratezza nella stima dei *keypoints* e la capacità di lavorare in tempo reale in scenari complessi [9].

- Caratteristiche

Basata sull'architettura CNN, utilizza la tecnica della *Symmetric Integral Keypoint Regression (SIKR)* per stimare le posizioni dei *keypoints*, combinando le mappe di confidenza con i *Part Affinity Fields (PAFs)*, i quali collegano i *keypoints* tra loro.

- Pro

- o Accuratezza elevata, in particolare per piccoli *keypoints* come mani e viso;
- o Adatto per il rilevamento multi-persona grazie all'uso di tecniche come la *Non-Maximum Suppression (NMS)* per eliminare posizioni ridondanti.

- Contro

- o Rischia un alto consumo computazionale a causa delle architetture complesse utilizzate.

1.6.2. DeepPoseKit

Framework *markerless* basato su TensorFlow/Keras, che si caratterizza per la sua elevata flessibilità e capacità di personalizzazione [10].

- Caratteristiche

- o Si basa su architetture modulari, come *Stacked DenseNet*, che consentono di configurare il modello in base alle specifiche esigenze cliniche;
- o Supporta tecniche di *data augmentation* e *transfer learning* per addestrare il modello anche su dataset limitati.

- Pro

- o Flessibile e personalizzabile.

- Contro

- o In fase di addestramento può richiedere hardware più potenti per ottenere performance ottimali.

1.6.3. DensePose

A differenza degli altri modelli esaminati, non rientra nelle categorie *Top-Down* o *Bottom-Up*, bensì utilizza un approccio *direct-to-3D* ove mappa i pixel di un'immagine bidimensionale su una rappresentazione tridimensionale della superficie corporea, fornendo una segmentazione densa e dettagliata [11].



- Caratteristiche
 - o Utilizza l'architettura *Mask-RCNN* combinata con una *Feature Pyramid Network (FPN)* per estrarre e mappare le informazioni in modo molto fine;
 - o Per stabilire le corrispondenze tra lo spazio 2D e quello 3D, sfrutta modelli tridimensionali come lo *Skinned Multi-Person Linear Model (SMPL)*.
- Pro
 - o Estrema precisione nella segmentazione a livello del pixel.
- Contro
 - o Elevate richieste computazionali.

1.6.4. Detectron2

Framework flessibile per la rilevazione di oggetti e la stima dei *keypoints* [12].

- Caratteristiche
 - o Basato su algoritmi come *Mask-RCNN* e *Faster-RCNN*, offre la possibilità di personalizzare la pipeline in base ai dati disponibili;
 - o Si distingue dagli altri modelli per la sua capacità di gestire diversi formati di input, come immagini, annotazioni o maschere.
- Pro
 - o Elevata modularità.
- Contro
 - o Richiede GPU potenti e un *tuning* avanzato per ottenere prestazioni in tempo reale, limitando la sua applicazione in contesti clinici di piccola scala.

1.6.5. High-Resolution Network

Progettata per risolvere problemi di visione artificiale sensibili alla posizione, opera ad alta risoluzione lungo tutta la pipeline di stima della posa [13].

- Caratteristiche
 - o Gestisce simultaneamente flussi ad alta e bassa risoluzione, permettendo uno scambio iterativo di informazioni che preserva i dettagli spaziali.
- Pro
 - o Elevata accuratezza e livello di dettaglio.
- Contro
 - o Elevati requisiti hardware e tempo di elaborazione.



1.6.6. MediaPipe (MP)

Sviluppato da Google, con il modulo *PoseLandmarker Heavy*, è uno dei modelli *markerless* più popolari per la sua efficienza [2].

- Caratteristiche
 - o Utilizza un approccio di regressione continua per stimare direttamente le coordinate (x, y, z) dei *keypoints*, riducendo la necessità di *heatmap* intermedie nell'ultimo stadio di stima;
 - o Consente di eseguire inferenza ad alta frequenza (30-60 fps) su hardware standard e supporta l'uso, seppur limitato, di tecniche come il *transfer learning*.
- Pro
 - o Efficienza operativa;
 - o Semplicità di implementazione;
 - o Facilità di integrazione in sistemi clinici.
- Contro
 - o In situazioni che richiedono un perfezionamento estremo dei *keypoints*, potrebbe essere necessario eseguire un *tuning* aggiuntivo.

1.6.7. OpenPifPaf

Utilizza il *Temporal Composite Association Field (TCAF)* per costruire pose spazio-temporali rappresentandole come grafi connessi [14].

- Caratteristiche
 - o Permette di associare i *keypoints* tra fotogrammi consecutivi, migliorando la coerenza temporale in analisi video.
- Pro
 - o Gestisce efficacemente sequenze video e possiede una velocità di inferenza competitiva.
- Contro
 - o Richiede una configurazione più complessa e può risultare sovradimensionato per applicazioni cliniche basilari.

1.6.8. OpenPose

Tra i modelli *markerless* più completi, è in grado di rilevare fino a 135 *keypoints* [15].

- Caratteristiche



- o Utilizza un approccio *Bottom-Up* per generare mappe di confidenza e il *Part Affinity Fields (PAFs)* per collegare tra loro i *keypoints* rilevati;
- o Sfrutta tecniche come la *Spatial Transformer Network (SSTN)*, la *Single Person Pose Estimator (SPPE)* e la *Spatial De-Transformer Network (SDTN)* per affinare la stima delle pose.
- Pro
 - o Elevata completezza ed elevato dettaglio, ideale per scenari complessi e multi-persona.
- Contro
 - o Alto costo computazionale.

1.6.9. PoseNet

- Caratteristiche
 - o Basato sull'architettura *VGG-16*, stima i 6 gradi di libertà della fotocamera per determinare posizione e orientamento nello spazio rispetto all'oggetto fotografato a partire da una singola immagine RGB [16].
- Pro
 - o Velocità elevata e facile implementazione.
- Contro
 - o Precisione inferiore rispetto a modelli specializzati per la stima dei *keypoints* del corpo umano.

2. Materiali e metodi

Il disegno sperimentale dello studio comprende: selezione del modello, i soggetti arruolati, le strumentazioni impiegate, i protocolli di acquisizione adottati per i sistemi VICON e MediaPipe e la pipeline di calcolo delle metriche posturali.

L'obiettivo è valutare il grado di accordo tra le metriche posturali estratte dal sistema *markerless* basato su MP e quelle calcolate dal sistema opto-elettronico VICON, assunto come *gold standard*. Entrambi i sistemi hanno acquisito dati nella medesima sessione sperimentale, in modo da eliminare la variabilità inter-sessione dovuta a possibili variazioni posturali del paziente.

Lo studio è stato condotto in collaborazione con ROGA, azienda specializzata nella fornitura di ausili per la mobilità e nella riabilitazione protesica, che ha reclutato i pazienti e ha definito le metriche posturali clinicamente rilevanti per la valutazione dell'adattamento protesico. Le



acquisizioni sono state effettuate presso il Laboratorio di Analisi del Movimento del Politecnico di Torino, dove è installato il sistema VICON.

2.1. Analisi comparativa

I criteri principali per selezionare il modello opportuno includono: accuratezza nella stima dei *keypoints*, velocità di inferenza, adattabilità e implementazione. Tali parametri sono stati valutati in base all'obiettivo clinico, ove risulta fondamentale la stima dei *keypoints* del paziente che indossa la protesi [1][8].

2.1.1. Accuratezza nella stima dei keypoints

Bisogna che il modello sia in grado di rilevare con precisione la posizione dei *keypoints*, anche in presenza di oclusioni o variazioni di abbigliamento, per garantire una corretta calibrazione protesica. A tale scopo, modelli come AlphaPose, HRNet e DensePose offrono elevata accuratezza, ma possono risultare troppo dispendiosi dal punto di vista computazionale. Per contro, MP e DeepPoseKit garantiscono un buon compromesso tra accuratezza e flessibilità [17].

2.1.2. Velocità di inferenza

L'utilizzo di modelli che operano rapidamente non è il requisito principale, ma è stato considerato per garantire risposte tempestive. Questo è possibile perché, nel nostro contesto, dobbiamo stimare i *keypoints* a partire da un'immagine 2D statica. Modelli come MP e PoseNet sono progettati per inferenze veloci su dispositivi economici. Soluzioni complesse come OpenPose e HRNet, invece, richiedono GPU ad alte prestazioni.

2.1.3. Adattabilità

Si è optato per modelli che permettano il *fine-tuning* utilizzando tecniche di *data augmentation* e *transfer learning*, in maniera tale da addestrare il modello con un numero limitato di immagini. DeepPoseKit, grazie alla sua flessibilità architettonica, eccelle in tale contesto. MP, inoltre, supporta l'adattamento tramite strumenti come *Model Maker*.

2.1.4. Implementazione

Disporre di una documentazione chiara e aggiornata è un requisito non banale al fine di implementare correttamente il modello in relazione all'obiettivo. MP, per esempio, è fortemente documentato e in continuo aggiornamento, rendendolo facilmente integrabile.

L'analisi comparativa dei modelli *markerless* (sez. 1.6) evidenzia come, nonostante numerose soluzioni offrano elevate prestazioni in termini di accuratezza, ciascuna presenta compromessi specifici. Ad esempio, modelli come OpenPose e HRNet raggiungono livelli elevatissimi di



dettaglio, ma richiedono hardware avanzato e risultano complessi da integrare. Al contrario, soluzioni più leggere come PoseNet sono ottimizzate per la velocità e l'efficienza computazionale, pur sacrificando parte dell'accuratezza necessaria per analisi cliniche dettagliate.

Nell'ambito del progetto, in cui si lavora a partire da immagini bidimensionali statiche e adattarsi a dataset ridotti è cruciale, emerge un candidato particolarmente idoneo: MediaPipe.

A supporto di questa analisi qualitativa, la letteratura offre evidenze quantitative: Aguilar-Ortega et al. [17] hanno confrontato sette metodi HPE, tra cui MediaPipe, OpenPose e AlphaPose, su un dataset di esercizi fisioterapici acquisiti con cinque telecamere RGB e validati con il sistema di tracking ottico OptiTrack. I risultati mostrano che i metodi 2D sono sufficienti per la stima degli angoli articolari nelle viste frontale e sagittale, e che MP raggiunge prestazioni competitive rispetto a soluzioni più complesse, con il vantaggio di non richiedere GPU dedicata. O ancora, Mercadal-Baudart et al. [24] hanno validato un modello 3D di stima della posa da una singola telecamera rispetto al sistema VICON su esercizi di forza, ottenendo errori RMS inferiori a 10° per angoli di ginocchio, anca e tronco. Infine, Lafayette et al. [25] hanno confrontato MP con un sistema RGB-D (Kinect V2) e con il *gold standard* Qualisys, dimostrando che MP produce errori paragonabili a quelli dei goniometri clinici standard per la misurazione degli angoli articolari degli arti superiori e inferiori, con il vantaggio di non richiedere sensori di profondità. Nel complesso, questi dati supportano la scelta di MP come strumento idoneo per applicazioni cliniche a basso costo computazionale, pur nella consapevolezza dei limiti discussi nelle sezioni successive.

2.2. Approfondimento su DeepPoseKit

Framework open-source per la stima *markerless* della posa umana, sfrutta le potenzialità di TensorFlow e Keras. Il suo design è orientato a progetti che richiedono una forte personalizzazione e la capacità di addestrarsi con dataset di dimensioni ridotte.

2.2.1. Architettura

DeepPoseKit si fonda su un'architettura modulare estremamente flessibile, che consente di adattare il modello in base alle esigenze specifiche del progetto clinico. La pipeline si articola in,

- *pre-processing*: le immagini acquisite vengono normalizzate e preparate per l'estrazione di *features*. Questo step è essenziale per assicurare uniformità e integrità nei dati in ingresso, riducendo il rumore e le variazioni indesiderate;



- reti convoluzionali personalizzate: attraverso modelli come *Stacked DenseNet* o *DeepLabCut* [20] è possibile modulare il numero di *layer*, la dimensione dei filtri e il numero di neuroni per ciascuno strato;
- estrazione dei *keypoints*: mediante tecniche di regressione vengono estratte le coordinate dei punti. In questa fase, l'applicazione di *fine-tuning* e *transfer learning* è fondamentale per affinare la precisione del modello;
- *post-processing*: attraverso meccanismi di *smoothing* e filtri temporali si mitigano fluttuazioni indesiderate ed errori dovuti alle variazioni ambientali.

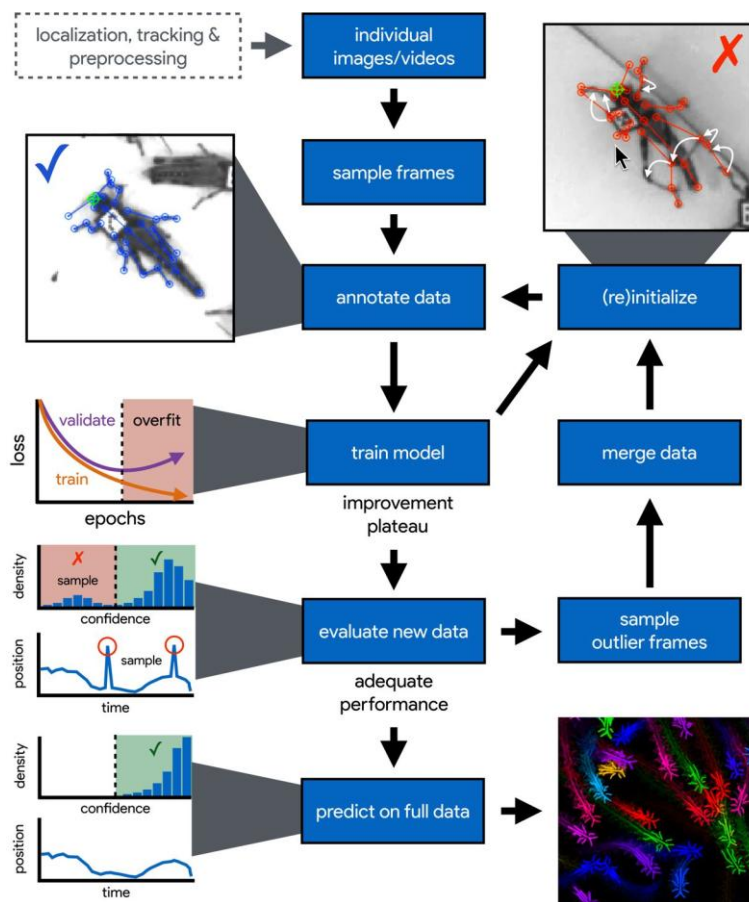


Fig. 2.1 – Flusso di lavoro generico DeepPoseKit

2.2.2. Addestramento e ottimizzazione

Il modello è in grado di addestrarsi efficacemente anche con dataset limitati grazie a,

- *data augmentation*: l'integrazione di librerie come *imgaug* consente di generare variazioni artificiali nelle immagini di training. Le trasformazioni, che includono rotazioni, scalature, traslazioni e variazioni di luminosità, simulano diverse condizioni ambientali e posturali, amplificando virtualmente il dataset e migliorando la robustezza del modello;

- *transfer learning*: l'adozione di modelli pre-addestrati riduce in maniera significativa i tempi di addestramento e permette di ottenere una stima accurata dei *keypoints* anche in presenza di pochi dati annotati.

2.2.3. Applicazioni

Con opportune ottimizzazioni, DeepPoseKit può essere eseguito su CPU, garantendo tempi di risposta adeguati ad applicazioni in tempo reale. Questo è particolarmente utile laddove le risorse hardware possono essere limitate. Inoltre, il modello supporta attività di monitoraggio continuo del movimento, fornendo dati fondamentali per realizzare analisi cinematiche personalizzate. Queste informazioni permettono di adattare e calibrare i dispositivi protesici in modo preciso, migliorando il comfort e l'efficacia terapeutica.

2.2.4. Limiti e considerazioni

Nonostante i numerosi vantaggi, DeepPoseKit presenta alcune criticità:

- complessità di implementazione: la forte modularità, pur offrendo grandi possibilità di personalizzazione, rende il *tuning* del modello complesso;
- risorse computazionali: la configurazione può richiedere un uso intensivo delle risorse, specialmente durante le fasi di addestramento. Questo può rappresentare un ostacolo in contesti clinici in cui l'hardware disponibile è limitato;
- sensibilità al *pre-processing*: l'efficacia del modello dipende fortemente dalla qualità dei dati di input. Condizioni ambientali non ideali o difformità nei dataset, come variazioni di illuminazione o rumore, possono influenzare negativamente la precisione della stima dei *keypoints*.

In sintesi, DeepPoseKit presenta elevata flessibilità e personalizzazione, efficacia in termini di inferenza ed applicazioni cliniche mirate.

2.2.5. Il tentativo con DeepPoseKit

Prima di adottare MP come strumento principale, è stato condotto un tentativo di utilizzo di DeepPoseKit, motivato dalla sua capacità teorica di addestrare un modello di stima dei *keypoints* su immagini personalizzate di soggetti con protesi, superando il limite dei modelli addestrati esclusivamente su soggetti normodotati.

La pipeline implementata ha previsto le seguenti fasi: annotazione manuale di un sottoinsieme di immagini raffiguranti soggetti con amputazioni trans-tibiali o trans-femorali (fornite dall'azienda collaboratrice), tramite l'interfaccia di DeepPoseKit, con salvataggio delle annotazioni in formato .h5; costruzione di un *DataGenerator* e di un *TrainingGenerator* a



partire dal file annotato; addestramento del modello LEAP (*Lightweight Efficient Architecture for Pose estimation*) incluso nel *framework*; predizione dei *keypoints* su immagini di test e visualizzazione dei risultati.

L'implementazione ha richiesto la risoluzione di numerose incompatibilità software. In particolare, sono stati riscontrati i seguenti problemi:

- Incompatibilità NumPy/h5py: le versioni recenti di NumPy hanno rimosso l'attributo *np.sctypes* e la proprietà *.value* dei dataset h5py, utilizzati internamente da DeepPoseKit, rendendo necessaria l'implementazione di patch manuali prima dell'importazione delle librerie.
- Incompatibilità TensorFlow/Keras: le versioni recenti di Keras hanno rimosso i metodi *predict_generator* e *fit_generator* dagli oggetti *Functional*, sostituiti da *predict* e *fit* con interfacce diverse. DeepPoseKit chiama internamente questi metodi deprecati, rendendo necessaria la sovrascrittura dei metodi `__init_predict_model__` e *fit* della classe base *BaseModel* tramite *monkey-patching*.
- Compilazione manuale del modello: il framework non gestisce correttamente la compilazione automatica del modello di addestramento nelle versioni recenti di Keras, richiedendo un'ulteriore patch per forzare la compilazione con l'ottimizzatore Adam e la funzione di perdita MSE.

Nonostante la risoluzione di queste criticità abbia permesso di completare un ciclo di addestramento, i risultati ottenuti sulla predizione dei *keypoints* si sono rivelati insoddisfacenti: i punti predetti risultano posizionati in modo casuale e non convergono verso le posizioni anatomiche attese, anche con immagini di test simili a quelle annotate. Questo comportamento è attribuibile alla dimensione estremamente ridotta del dataset di addestramento disponibile (inferiore alla soglia minima raccomandata dagli autori, di almeno 50 campioni), all'instabilità del processo di addestramento e alla mancanza di un dataset specifico per soggetti con protesi su cui effettuare pre-addestramento.

Alla luce di queste criticità, si è optato per l'adozione di MP, che non richiede fase di addestramento, è immediatamente operativo su immagini RGB standard e ha dimostrato in letteratura risultati validati su popolazioni cliniche [1][8].

2.3. Approfondimento su MediaPipe

MP rappresenta una piattaforma open-source sviluppata da Google, che si è affermata come strumento di riferimento per la stima *markerless* della posa umana. All'interno di questa



piattaforma, il modulo *PoseLandmarker Heavy* è composto da due modelli di Machine Learning: un *Detector* ed un *Estimator*. Il primo, isola la regione corrispondente alla persona nell'immagine di input, mentre il secondo, elabora l'immagine ridimensionata a 256x256 px della persona rilevata per estrarne i *keypoints* [2].

Inoltre, il modello si distingue per la sua capacità di rilevare fino a 33 *keypoints* in tempo reale, individuando i punti di riferimento essenziali, come la sede della testa, le estremità, il tronco e le mani, che sono fondamentali per descrivere la configurazione posturale e monitorare il movimento. Tale accuratezza e velocità lo rendono particolarmente indicato per applicazioni cliniche, in particolare per la personalizzazione delle protesi per arti inferiori, in cui la precisione nella calibrazione del movimento costituisce un requisito imprescindibile [18].

2.3.1. Architettura

L'efficacia operativa di *PoseLandmarker Heavy* si fonda su una pipeline altamente modulare e flessibile (Fig. 2.2). Inoltre, la suddivisione in fasi garantisce che ogni operazione venga eseguita in modo efficiente e personalizzabile.

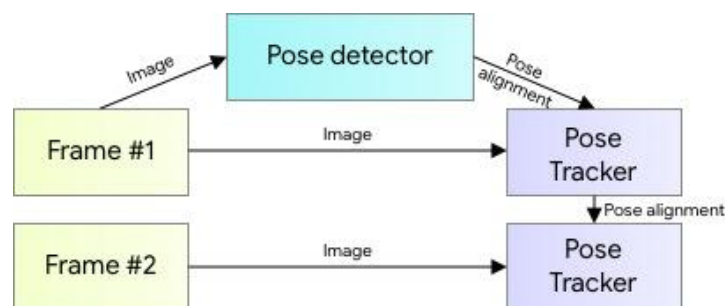


Fig. 2.2 - HPE pipeline [2]

Il processo inizia con il *pre-processing* delle immagini, ove i dati di input vengono normalizzati e sottoposti a operazioni di ridimensionamento, conversione dei colori e normalizzazione dell'intensità luminosa. Queste operazioni garantiscono l'uniformità dei dati, riducendo il rischio di errori dovuti a variazioni ambientali.

Successivamente, la fase di estrazione di *features* si avvale di CNN ottimizzate per individuare automaticamente bordi, forme e texture. *PoseLandmarker*, in particolare, integra tecniche basate sulla regressione continua e sulla classificazione, che permettono di stimare direttamente le coordinate dei *keypoints* dai dati estratti dalla CNN. A differenza dei sistemi tradizionali che utilizzano le *heatmap* per indicare la probabilità di presenza dei *keypoints*, *PoseLandmarker* impiega un approccio ibrido di regressione diretta: il modello apprende una funzione che

associa *features* estratte alle coordinate spaziali dei *keypoints* (x, y e, opzionalmente, z), riducendo il carico computazionale e aumentando il numero di frame processabili al secondo.

Il processo di inferenza termina con una fase di *post-processing*: vengono applicate tecniche di *smoothing* e filtri temporali. Tali operazioni permettono di mitigare eventuali oscillazioni e inconsistenze nei risultati, fenomeni che possono derivare da errori momentanei, oclusioni parziali o condizioni di illuminazione non ideali.

Nello specifico, il *Detector* è basato su un'architettura *Single-Shot Detector (SSD)*. Quindi, la rete è in grado di rilevare il soggetto nell'immagine con una sola passata, senza la necessità di avere regioni candidate ripetute, garantendo velocità ed efficienza [18].

Data un'immagine in input di dimensioni (1, 224, 224, 3), dove 1 indica il *batch-size*, 224 il numero di pixel di larghezza e altezza della *bounding box* e 3 il numero di canali (RGB), il modello restituisce una *bounding box* di dimensioni (1, 2254, 12). In pratica, per ogni immagine vengono generate 2254 proposte, ognuna con una posizione differente dell'immagine stessa e descritta da 12 parametri. Questi ultimi sono organizzati come (x, y, w, h, kp^1_x , kp^1_y , ..., kp^n_x , kp^n_y); dove (x, y) rappresentano la posizione, (w, h) la larghezza e l'altezza della *bounding box* e $kp^n_{x,y}$ le coordinate dei *keypoints*. Inoltre, per ogni proposta, il modello fornisce un valore che indica la probabilità che quella proposta contenga effettivamente il soggetto di interesse.

Esistono due modalità d'uso del *Detector*:

- *Box*: a partire dalla posizione (x, y) e dalla dimensione (w, h) viene determinata la *bounding box*. Tale modalità è utile quando la rotazione o l'orientamento del soggetto non è rilevante;
- *Alignment*: oltre alla posizione e alle dimensioni, si sfruttano due *keypoints* per determinare il fattore di scala e l'angolo di rotazione, i quali migliorano l'interpretazione della geometria e dell'orientamento del soggetto rilevato. Tale modalità è conveniente quando la persona è inclinata.

L'*Estimator*, invece, durante la fase di addestramento sfrutta le *heatmap* per apprendere la posizione dei *keypoints* e durante l'inferenza ne calcola direttamente le coordinate. Il modello restituisce un tensore di dimensione (1, 195) contenente le informazioni relative ai 33 *keypoints* (Fig. 2.3). Per ciascun punto vengono forniti i seguenti valori:

- Le coordinate spaziali x e y;



- La profondità z , che rappresenta la distanza tra i fianchi della persona e la telecamera. Un valore positivo indica che il punto si trova dietro i fianchi, un valore negativo che si trova davanti;
- La visibilità, ossia la probabilità che sia visibile e non occluso;
- La presenza, ovvero la probabilità che sia effettivamente presente.

Oltre a tali valori, detti landmarks, l'Estimator restituisce un secondo output rappresentato dal *flag* (1, 1), utilizzato per logiche decisionali interne del modello.

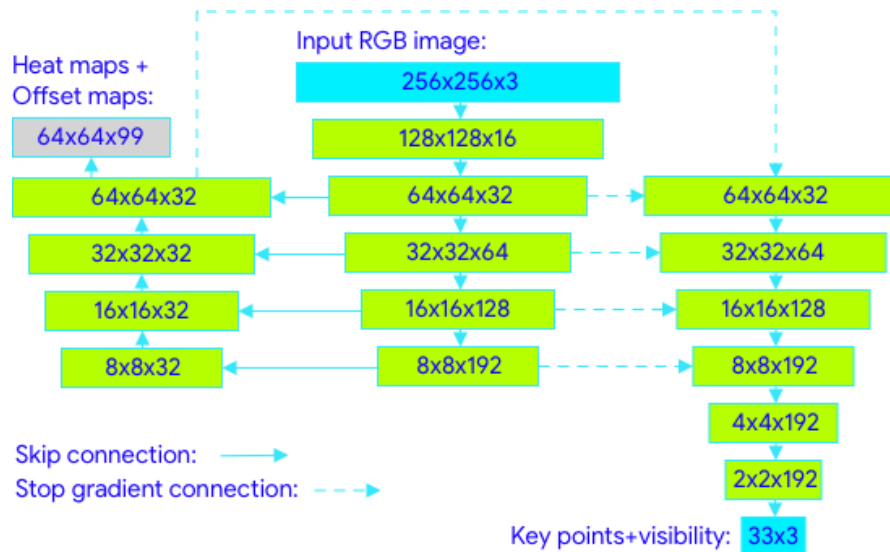


Fig. 2.3 - Architettura della rete di tracking: regressione con supervisione tramite heatmap [2].

2.3.2. Specifiche tecniche

Il modulo *PoseLandmarker* utilizza una tecnica di convoluzione che suddivide il processo in due fasi,

1. *Depthwise convolution*: applica un kernel separato 3×3 a ciascun canale dell'immagine in ingresso per rilevare pattern locali, riducendo il numero di operazioni rispetto a una convoluzione tradizionale. Quindi, invece di utilizzare un singolo filtro che processa contemporaneamente tutti i canali, vengono applicati filtri separati per ogni canale;

2. *Pointwise convolution*: successivamente, viene applicato un kernel 1×1 per combinare le informazioni tra i diversi canali, ottimizzando il numero di parametri senza comprometterne il valore informativo. In pratica, per ogni posizione spaziale, il kernel combina le informazioni provenienti da tutti i canali, fondendoli in nuovi canali che rappresentano una sintesi più ricca dal punto di vista informativo delle *features* estratte.

Il primo passo effettivo eseguito dal modulo è la riduzione progressiva della dimensione spaziale dell'immagine, attraverso *strided convolutions* o *pooling*. Così facendo, si riduce il numero di dati elaborati in ogni fase, permettendo al modello di osservare informazioni su scale diverse: mentre una porzione di dati fornisce dettagli fini, altre porzioni catturano informazioni più generali. Una volta ricavate queste informazioni multi-scala, il modello utilizza gli strati finali, completamente connessi, per trasformare le *features* ottenute in coordinate che indicano la posizione dei *keypoints*. In questo modo il modello non solo riconosce le caratteristiche salienti dell'immagine ma le mappa direttamente alle posizioni spaziali dei punti di interesse.

2.3.3. Personalizzazione tramite ModelMaker

Un elemento distintivo di MP è la possibilità di utilizzare *ModelMaker*, un tool integrato che consente di personalizzare il modello per applicazioni specifiche. Tale framework permette di integrare facilmente dati, anche quando il numero di immagini annotate è limitato, attraverso tecniche di *transfer learning*. Così facendo, il modello può essere ri-addestrato su dataset specifici.

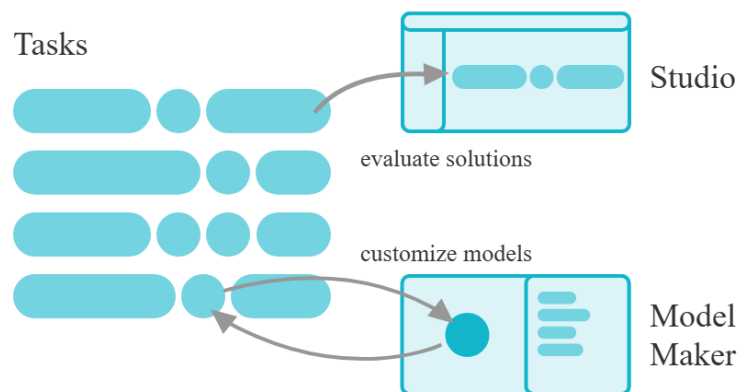


Fig. 2.4 - Model Maker, MP

Nella pratica, *ModelMaker* non supporta la *keypoint detection*, ma solo classificazione di immagini, *object detection* e segmentazione semantica. Quindi, al momento non esiste un flusso specifico per insegnare al modello a restituire i landmarks 2D/3D sulla protesi [2].

2.3.4. Integrazione con strumenti di debug

Per una valutazione accurata è fondamentale poter monitorare il comportamento interno del framework. In questo senso, strumenti come *TensorBoard* possono essere integrati per visualizzare in tempo reale le metriche di *training*, come ad esempio l'andamento della funzione di perdita e le distribuzioni dei pesi nei vari strati della rete. Queste visualizzazioni

non solo facilitano l'ottimizzazione dei parametri ma permettono anche di identificare eventuali anomalie legate all'*overfitting* o ad altre problematiche durante il *training*.

2.3.5. Applicazioni

MP è efficace nel monitoraggio del movimento: la sua capacità di rilevare in tempo reale la postura e la configurazione dei *keypoints* lo rende uno strumento ideale per,

- la calibrazione e personalizzazione delle protesi, dove la precisione nella rilevazione dei movimenti degli arti inferiori è fondamentale [6];
- il monitoraggio progressivo dei pazienti in fase di riabilitazione, consentendo interventi tempestivi in caso di anomalie [7][8].

2.3.6. Integrazione clinica

Un altro aspetto riguarda la capacità di MP di integrarsi con altri sistemi e piattaforme. Le API RESTful e le soluzioni cloud fornite da Google, per esempio, consentono di sviluppare interfacce utente e dashboard personalizzate che possono essere connesse ai sistemi informativi ospedalieri. Questa interoperabilità permette di trasmettere dati in tempo reale ai clinici e di integrarli con *Electronic Health Records (EHR)* o software di gestione clinica, favorendo un approccio sistemico alla riabilitazione. L'integrazione con strumenti di realtà aumentata e virtuale apre inoltre la strada a interfacce immersive per la visualizzazione dei movimenti, migliorando il feedback visivo durante le sessioni di riabilitazione.

2.3.7. Normativa

Nel contesto clinico, la sicurezza dei dati e la conformità alle normative rappresentano aspetti cruciali. L'utilizzo di MP, essendo una piattaforma *open-source*, richiede l'adozione di accortezze per il trattamento dei dati sensibili. È pertanto opportuno integrare nei flussi di lavoro sistemi di cifratura e autenticazione, oltre a garantire la conformità alle normative locali, quali il GDPR in Europa o l'HIPAA negli Stati Uniti.

2.3.8. Limiti e considerazioni

Nonostante i numerosi vantaggi descritti, MP presenta criticità rilevanti che devono essere considerate nel contesto specifico della valutazione posturale di pazienti con protesi di arto inferiore.

Il primo limite riguarda il dataset di addestramento. MP è stato addestrato su immagini di soggetti normodotati; la morfologia visiva di un arto protesico, in termini di profilo, colorazione, texture e assenza di repere anatomici naturali, non è rappresentata nel *training set*. Questo può causare localizzazione imprecisa o errata dei *keypoints* distali dell'arto protesico,



in particolare della caviglia e del piede. Zhou et al. [6] hanno quantificato questo fenomeno su amputati transtibiali e transfemorali, evidenziando errori sistematici nei *keypoints* dell'arto protesico che richiedono approcci correttivi specifici.

Il secondo limite riguarda la stima della profondità. La coordinata z restituita da MP dalla singola telecamera monoculare è una stima relativa, non metrica: rappresenta la distanza approssimativa rispetto ai fianchi del soggetto, non una profondità assoluta calibrata. Questo rende le metriche tridimensionali non direttamente comparabili con le misure del sistema VICON senza una calibrazione aggiuntiva. Nel presente studio, l'analisi è stata pertanto limitata alle metriche calcolabili nel piano frontale e sagittale dalle viste 2D calibrate.

Il terzo limite riguarda la ricostruzione di landmark non rilevati direttamente. Infatti, alcuni landmark anatomici clinicamente rilevanti per la valutazione posturale dei pazienti protesici, come le SIAS, le SIPS, il solco gluteo, l'occipite e C7, non sono inclusi nei 33 *keypoints* di MP e devono essere ricostruiti a partire da quelli esistenti. Ruescas-Nicolau et al. [19] hanno dimostrato che questa ricostruzione introduce errori sistematici nella stima degli angoli articolari, con valore dipendente dal tipo di movimento e dal landmark considerato. Pertanto, nel nostro caso vengono esclusi.

Il quarto limite riguarda la sensibilità alle condizioni ambientali: variazioni di illuminazione, sfondo non uniforme e abbigliamento aderente o atipico possono ridurre l'accuratezza della localizzazione dei *keypoints*. Questo aspetto è particolarmente rilevante in contesti ambulatoriali reali, dove le condizioni di acquisizione sono meno controllate rispetto a un laboratorio dedicato.

Per superare questi limiti, le prospettive future includono: l'integrazione di approcci di *domain adaptation* per adattare il modello a soggetti con protesi [6]; l'utilizzo di telecamere RGB-D per la stima metrica della profondità [16]; e lo sviluppo di dataset annotati specifici per pazienti con amputazione di arto inferiore, attualmente carenti in letteratura.

2.4. Soggetti

2.4.1. Criteri di inclusione ed esclusione

I criteri di inclusione adottati per la selezione dei partecipanti sono: amputazione monolaterale di arto inferiore (transtibiale o transfemorale); utilizzo della protesi da almeno sei mesi; capacità di mantenere la stazione eretta autonomamente per almeno 60 secondi; assenza di patologie ortopediche o neurologiche concomitanti all'arto controlaterale tali da alterare significativamente la postura. I criteri di esclusione, invece: incapacità di comprendere e



seguire le istruzioni del protocollo; presenza di ferite aperte o lesioni cutanee in sede di posizionamento dei marker VICON; rifiuto di fornire il consenso informato.

2.4.2. Popolazione dello studio

Lo studio ha coinvolto tre soggetti adulti (S1, S2, S3), tutti di sesso maschile, reclutati tramite l'azienda collaboratrice ROGA. Le caratteristiche cliniche sono riportate in Tabella 2.1.

Soggetto	Sesso	Altezza (cm)	Peso (Kg)	Amputazione	Lato
S1	M	174	76	Transtibiale	Destro
S2	M	177	75	Transtibiale	Sinistro
S3	M	183	89	Transfemorale	Sinistro

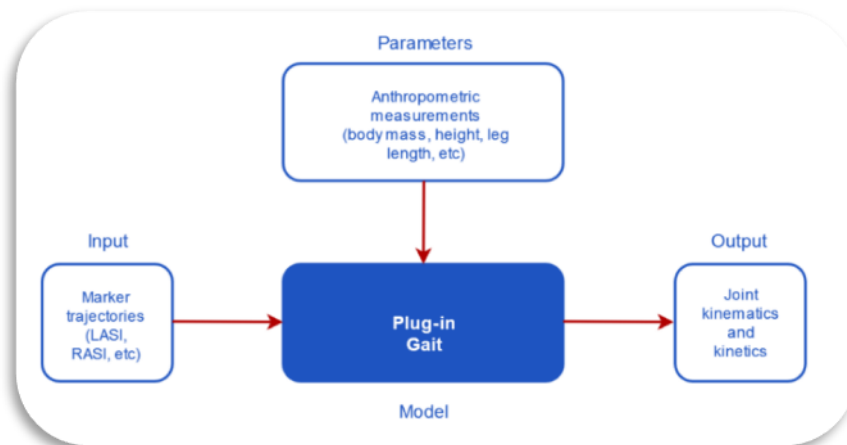
Tabella 2.1 - Parametri clinici soggetti S1, S2, S3.

I soggetti S1 e S2 presentano amputazione transtibiale, rispettivamente al lato destro e sinistro. Il Soggetto S3 presenta amputazione transfemorale al lato sinistro. Tutti i soggetti, inoltre, indossavano la propria protesi abituale durante le acquisizioni.

2.4.3. Considerazioni etiche

Lo studio è stato condotto nel rispetto dei principi della Dichiarazione di Helsinki e della normativa europea in materia di protezione dei dati personali (GDPR 2016/679). I dati sono stati anonimizzati e archiviati in un database sicuro accessibile ai soli ricercatori coinvolti nel progetto.

2.5. VICON



2.5.1. Descrizione del sistema

VICON è un sistema opto-elettronico di cattura del movimento *marker-based*, considerato il *gold standard* per la misura quantitativa della cinematica del corpo umano in biomeccanica clinica e sperimentale [23]. Il sistema è composto da dodici telecamere a infrarossi ad alta risoluzione che rilevano la posizione tridimensionale di marker semiriflettenti sferici (diametro 14 mm) posizionati su repere anatomici specifici del soggetto. La frequenza di acquisizione

adottata nel presente studio è di 100 Hz. La ricostruzione tridimensionale delle traiettorie dei marker avviene tramite il software *Vicon Nexus 2.17*, che integra algoritmi di triangolazione stereovisiva, filtraggio e *gap-filling* delle traiettorie.

Il laboratorio è dotato di un volume di acquisizione calibrato di circa $3 \times 3 \times 2$ metri. Prima di ogni sessione di acquisizione, il sistema viene acceso con almeno 30 minuti di anticipo per consentire la stabilizzazione termica delle telecamere, condizione necessaria per garantire la qualità della calibrazione. La qualità della calibrazione viene valutata tramite l'errore residuo di ricostruzione 3D (*World Error*) e l'errore di immagine (*Image Error*), espressi rispettivamente in millimetri e pixel.

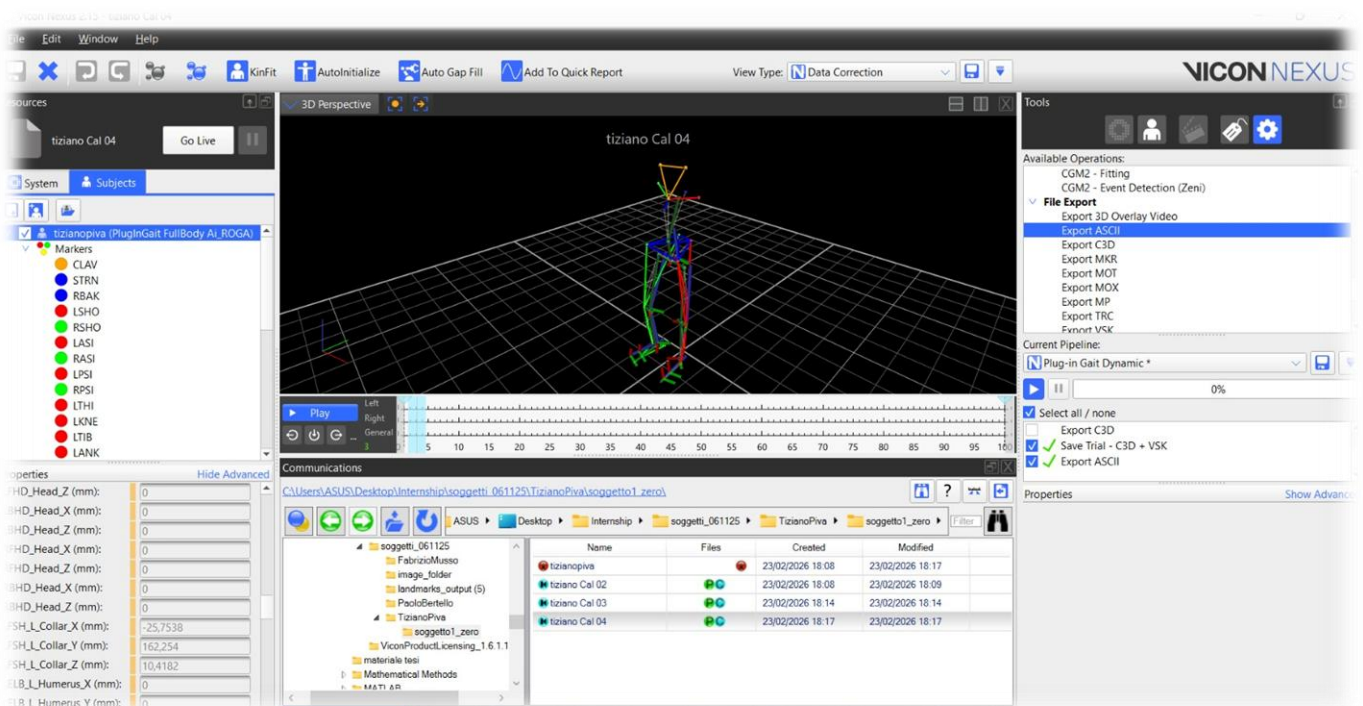


Figura 2.5 - Interfaccia Vicon Nexus 2.17.

2.5.2. Plug-in Gait Full Body

La cinematica articolare è stata calcolata mediante il modello biomeccanico *Plug-in Gait Full Body*, che costituisce il riferimento standard per l'analisi quantitativa del movimento in ambito clinico [23]. Il modello prevede l'applicazione di un set di marker riflettenti su punti anatomici specifici di tutto il corpo, la raccolta di misure antropometriche soggetto-specifiche e l'esecuzione di un trial statico di calibrazione in posizione standardizzata, a partire dal quale vengono calcolati gli offset articolari soggetto-dipendenti.

Nel presente studio è stato adottato il modello *full body*, che include sia il modulo *lower body* (v. fig. 2.6, 2.7, 2.8) che quello *upper body* (v. fig. 2.9, 2.10, 2.11). Il primo calcola angoli e

momenti articolari di anca, ginocchio e caviglia, nonché gli angoli pelvici. Il secondo, invece, estende la modellazione al tronco, alle spalle, ai gomiti e ai polsi. I parametri di output utilizzati per il confronto con MediaPipe includono anche gli angoli articolari valutati nel piano frontale e sagittale.



Da sinistra verso destra, dall'alto verso il basso: Fig.2.6, 2.7, 2.8 set marker lower body; Fig. 2.9, 2.10, 2.11 set marker upper body. Viste frontali e sagittali. Guida ufficiale VICON,

2.5.3. Posizionamento dei marker

Il set di marker adottato comprende 23 marker funzionali, in numero ridotto rispetto al set completo previsto dal modello *Plug-in Gait Full Body*, con l'aggiunta di due marker auricolari (LEAR, REAR) per l'implementazione della *plumb line* virtuale sagittale richiesta dalla fisioterapista collaboratrice. I marker e le loro posizioni anatomiche sono descritti nelle Tabelle 2.2 e 2.3.

Marker	Definizione	Posizionamento anatomico
LASI / RASI	Spina iliaca antero-superiore	Cresta iliaca superiore anteriore, sinistra e destra
LPSI / RPSI	Spina iliaca postero-superiore	Al di sotto delle articolazioni sacro-iliache, bilateralmente
LTHI / RTHI	Coscia	Superficie laterale, terzo distale della coscia (asimmetrici: L inferiore, R superiore)
LKNE/RKNE	Ginocchio	Sull'asse di flessione-estensione del ginocchio, lateralmente
LTIB / RTIB	Tibia	Superficie laterale, terzo distale della gamba (asimmetrici: L inferiore, R superiore)
LANK/RANK	Caviglia	Malleolo laterale, lungo l'asse trasversale
LHEEL/RHEEL	Tallone	Sul calcagno, alla stessa altezza del marker LTOE/RTOE
LTOE / RTOE	Punta del piede	Testa del secondo metatarso (punto di transizione avampiede-mesopiede)

Tabella 2.2 – Set marker del modello lower body.

Marker	Definizione	Posizionamento anatomico
CLAV	Clavicola	Fossa giugulare (giunzione clavicola-sterno)
STRN	Sterno	Processo xifoideo dello sterno
LSHO / RSHO	Spalla	Articolazione acromio-claveare, bilateralmente
RBAK	Schiena destra	Punto arbitrario sulla scapola destra (solo per autolabeling)
LEAR / REAR	Orecchio	In corrispondenza del trago, approssimazione dell'articolazione temporo-mandibolare

Tabella 2.3 – Set marker del modello upper body con LEAR e REAR aggiuntivi.

I marker della coscia (THI) e della tibia (TIB) vengono posizionati asimmetricamente tra lato destro e sinistro per facilitare il riconoscimento automatico da parte del software di etichettatura.

2.5.4. Misure antropometriche

Prima dell'acquisizione, per ogni soggetto sono state rilevate le misure antropometriche richieste dal modello *Plug-in Gait*, utilizzando metro e calibro. Le misure sono state inserite nel software Nexus nel nodo soggetto (*Subject Node*) in millimetri. Le misure rilevate sono riportate in Tabella 2.4, 2.5 e 2.6.

Misura	Descrizione	Lato	Strumento	Valore
<i>Body mass (Kg)</i>	Massa corporea	—	Bilancia	76
<i>Height (mm)</i>	Altezza	—	Metro	1740
<i>Inter-ASIS distance (mm)</i>	Distanza tra le due ASIS	—	Metro	320
<i>Leg Length (mm)</i>	Distanza da ASIS al malleolo mediale	L / R	Metro	850
<i>ASIS-Trochanter Distance (mm)</i>	Proiezione sagittale della distanza verticale tra ASIS e grande trocantere	L / R	Metro	151
<i>Knee Width (mm)</i>	Larghezza medio-laterale del ginocchio lungo l'asse di flessione	L / R	Calibro	85
<i>Ankle Width (mm)</i>	Distanza medio-laterale tra i malleoli	L / R	Calibro	70
<i>Shoulder Offset (mm)</i>	Distanza cranio-caudale tra centro articolare gleno-omeroale e base marker acromion-claveare	L / R	Metro	65

Tabella 2.4 – Misure antropometriche rilevate per il modello Plug-in Gait Full Body per S1.

Misura	Descrizione	Lato	Strumento	Valore
<i>Body mass (Kg)</i>	Massa corporea	—	Bilancia	75
<i>Height (mm)</i>	Altezza	—	Metro	1770
<i>Inter-ASIS distance (mm)</i>	Distanza tra le due ASIS	—	Metro	340
<i>Leg Length (mm)</i>	Distanza da ASIS al malleolo mediale	L / R	Metro	940
<i>ASIS-Trochanter Distance (mm)</i>	Proiezione sagittale della distanza verticale tra ASIS e grande trocantere	L / R	Metro	140
<i>Knee Width (mm)</i>	Larghezza medio-laterale del ginocchio lungo l'asse di flessione	L / R	Calibro	104
<i>Ankle Width (mm)</i>	Distanza medio-laterale tra i malleoli	L / R	Calibro	71
<i>Shoulder Offset (mm)</i>	Distanza cranio-caudale tra centro articolare gleno-omeroale e base marker acromion-claveare	L / R	Metro	77

Tabella 2.5 – Misure antropometriche rilevate per il modello Plug-in Gait Full Body per S2.



Misura	Descrizione	Lato	Strumento	Valore
<i>Body mass (Kg)</i>	Massa corporea	—	Bilancia	89
<i>Height (mm)</i>	Altezza	—	Metro	1830
<i>Inter-ASIS distance (mm)</i>	Distanza tra le due ASIS	—	Metro	360
<i>Leg Length (mm)</i>	Distanza da ASIS al malleolo mediale	L / R	Metro	960
<i>ASIS-Trochanter Distance (mm)</i>	Proiezione sagittale della distanza verticale tra ASIS e grande trocantere	L / R	Metro	135
<i>Knee Width (mm)</i>	Larghezza medio-laterale del ginocchio lungo l'asse di flessione	L / R	Calibro	130
<i>Ankle Width (mm)</i>	Distanza medio-laterale tra i malleoli	L / R	Calibro	67
<i>Shoulder Offset (mm)</i>	Distanza cranio-caudale tra centro articolare gleno-omeroale e base marker acromion-claveare	L / R	Metro	105

Tabella 2.6 – Misure antropometriche rilevate per il modello Plug-in Gait Full Body per S3.

Per la localizzazione del marker del ginocchio (KNEE), si flette ed estende passivamente il ginocchio identificando il punto della superficie laterale che rimane più stabile durante il movimento: in un adulto in piedi tale punto si trova circa 1.5 cm sopra la rima articolare.

Nei soggetti con protesi di arto inferiore, i marker dell'arto protesico vengono posizionati direttamente sul rivestimento della protesi in corrispondenza dei repere anatomici equivalenti, adattando la procedura alla morfologia specifica del dispositivo.

Per i soggetti con amputazione, la misura di *Leg Length* dell'arto protesico è stata rilevata sulla protesi, dalla proiezione dell'ASIS al malleolo protesico equivalente. La *Knee Width* e l'*Ankle Width*, invece, sono state misurate solo sul lato sano e riportate anche nel lato protesico.

2.5.5. Calibrazione delle telecamere

La calibrazione del sistema VICON viene effettuata una sola volta, prima delle sessioni di acquisizione e si articola in due fasi. Nella prima fase (*Mask Cameras*), vengono mascherati i riflessi parassiti delle telecamere, al fine di migliorare la robustezza della calibrazione e ridurre i falsi positivi nel riconoscimento dei marker. Nella seconda fase (*Calibrate Cameras*), l'operatore esegue movimenti oscillatori dell'asta a T calibrata (*Active Wand v2*) all'interno del volume di acquisizione, orientandola verso ciascuna telecamera fino all'illuminazione del LED



verde di conferma. Il sistema fornisce in output il *World Error* (errore di calibrazione espresso in mm) e l'*Image Error* (errore RMS in pixel tra coordinate 2D dei marker e la loro ricostruzione 3D).

La definizione del sistema di riferimento globale viene effettuata tramite la funzione *Set Volume Origin*, posizionando l'asta a T in corrispondenza del centro dello spazio sperimentale. L'offset tra la base dell'asta e il suolo è stato impostato a $Z = -9$ mm (con base di spessore 2 mm e marker di diametro 14 mm).

2.5.6. Protocollo di acquisizione

Il soggetto viene preparato con l'applicazione dei marker secondo il posizionamento descritto nella sezione 2.5.3 e vengono rilevate le misure antropometriche. Successivamente, il soggetto viene invitato a posizionarsi al centro del volume di acquisizione in posizione ortostatica standardizzata (stazione eretta, arti superiori rilassati lungo i fianchi, sguardo orizzontale).

Per ciascun soggetto vengono effettuate tre ripetizioni di acquisizione statica (Cal_02, Cal_03, Cal_04), ciascuna della durata di 1 secondo. La fase di calibrazione, richiesta dal sistema Vicon, viene eseguita in “posa da motociclista”, con piedi leggermente divaricati e arti superiori abdotti di circa 45° con gomiti flessi. Le successive acquisizioni sperimentali, invece, vengono effettuate in posizione ortostatica con gli arti superiori distesi lungo i fianchi. A causa di un problema tecnico del sistema, per il soggetto S3 è stato possibile acquisire soltanto una ripetizione (Cal_02), mentre la terza e la quarta (Cal_03 e Cal_04) sono state scartate.

Al termine di ogni acquisizione, viene eseguita la pipeline di post-processing nel software Nexus nella seguente sequenza: *Reconstruct and Label*, per la ricostruzione 3D delle traiettorie marker e l'etichettatura automatica; *Auto Initialize Labeling*, che include *Autolabel Static Frame*, *Scale Subject* e *Marker-only Subject Calibration*; *Plug-in Gait Static*, per il calcolo degli offset soggetto-specifici; *Plug-in Gait Dynamic*, per il calcolo della cinematica articolare; *Export ASCII*, per l'esportazione delle traiettorie marker e degli output del modello (angoli articolari) in formato .csv. Inoltre, i file vengono salvati nei formati C3D e VSK per garantire la tracciabilità dei dati.

Si consiglia la visione dell'APPENDICE A per il protocollo completo.

2.6. MediaPipe

2.6.1. Strumentazione e configurazione

Le immagini posturali sono state acquisite con un iPhone 16 (fotocamera principale con lunghezza focale equivalente di 26 mm, 2556×1179 pixel a 460 ppi) montato su un treppiede



regolabile. La fotocamera è stata posizionata con la lente alla stessa altezza dell'ombelico del soggetto rispetto al pavimento, orientata parallelamente al piano di acquisizione (frontale o sagittale), ad una distanza di circa 2.5 metri dal soggetto. Quest'ultima è stata calcolata moltiplicando l'altezza del soggetto per un fattore di scala empiricamente determinato in funzione della lunghezza focale della fotocamera, in modo da includere l'intero corpo nell'immagine (cfr. Appendice A).

Per la conversione delle coordinate da pixel a millimetri è stato utilizzato un riferimento metrico fisico: una borraccia di altezza nota $h = 24.8$ cm, posizionata sul pavimento in prossimità del soggetto durante ogni acquisizione. Specificatamente, per le acquisizioni in vista frontale la borraccia è stata posizionata sul pavimento a 50 cm dal piede in direzione medio-laterale, controlateralmente all'arto protesico. Per le acquisizioni in vista sagittale, la borraccia è stata posizionata a 50 cm dai piedi in direzione antero-posteriore.

2.6.2. Protocollo di acquisizione

Il protocollo di acquisizione MediaPipe è stato eseguito nella medesima sessione del sistema VICON, a garanzia della comparabilità delle misure. Prima dell'acquisizione, il soggetto ha mantenuto la propria protesi e l'abbigliamento adottato per l'acquisizione VICON (abiti aderenti che risaltano la silhouette corporea, con scarpe).

Per ogni soggetto sono state acquisite tre immagini per ciascuna delle tre viste: vista frontale (fotocamera parallela al piano coronale del soggetto); vista sagittale sinistra (fotocamera parallela al piano sagittale, lato sinistro); vista sagittale destra (fotocamera parallela al piano sagittale, lato destro). In totale, per ogni soggetto sono state acquisite 3 immagini. Ciascuna acquisizione corrisponde alla stessa ripetizione del trial VICON.

Durante l'acquisizione, il soggetto manteneva la posizione ortostatica standardizzata con arti superiori rilassati lungo i fianchi. Per ogni soggetto, le acquisizioni VICON hanno preceduto quelle MediaPipe, puntando a minimizzare il tempo intercorso tra le due misurazioni.

Si consiglia la visione dell'APPENDICE A per il protocollo completo.

2.6.3. Rilevamento dei marker

Il modulo *PoseLandmarker Heavy* di MediaPipe è stato eseguito in modalità immagine statica (*static_image_mode=True*) con soglia di confidenza minima per il rilevamento pari a 0.5 (*min_detection_confidence=0.5*). Il modello restituisce, per ciascuna delle 33 posizioni landmark, le seguenti informazioni: coordinate normalizzate (x, y) rispetto alle dimensioni dell'immagine nell'intervallo [0, 1]; coordinata di profondità z relativa (normalizzata rispetto



alla larghezza dei fianchi); punteggio di visibilità (probabilità che il landmark sia visibile e non occluso); punteggio di presenza (probabilità che il landmark sia effettivamente presente nell'immagine). Le coordinate normalizzate sono state convertite in coordinate pixel moltiplicando per le dimensioni dell'immagine acquisita (larghezza W e altezza H in pixel). I 33 landmark e le loro corrispondenze anatomiche sono descritti in Tabella 2.7.

ID	Nome landmark	Anatomia approssimata
0	NOSE	Naso
7	LEFT_EAR	Orecchio sinistro (trago)
8	RIGHT_EAR	Orecchio destro (trago)
11	LEFT_SHOULDER	Spalla sinistra (articolazione gleno-omeroale)
12	RIGHT_SHOULDER	Spalla destra
23	LEFT_HIP	Anca sinistra (appros. grande trocantere in vista lat.)
24	RIGHT_HIP	Anca destra
25	LEFT_KNEE	Ginocchio sinistro (centro articolare)
26	RIGHT_KNEE	Ginocchio destro
27	LEFT_ANKLE	Caviglia sinistra (centro articolare)
28	RIGHT_ANKLE	Caviglia destra
29	LEFT_HEEL	Tallone sinistro
30	RIGHT_HEEL	Tallone destro
31	LEFT_FOOT_INDEX	Indice piede sinistro (II metatarso distale)
32	RIGHT_FOOT_INDEX	Indice piede destro

Tabella 2.7 – Landmark di MP utilizzati per il calcolo delle metriche posturali.

2.6.4. Fattore di scala

Le coordinate dei landmark estratte mediante MediaPipe sono inizialmente espresse nel sistema di riferimento dell'immagine, quindi in unità di pixel. Per consentire il confronto diretto con le misure acquisite tramite il sistema di *motion capture* VICON, espresse in millimetri, è stata quindi necessaria una conversione delle coordinate in unità metriche. A tal fine è stato introdotto un fattore di scala soggetto-specifico, determinato a partire da un riferimento metrico fisico presente nell'immagine di acquisizione.

Il riferimento adottato è costituito da una borraccia di altezza nota pari a $h_{ref} = 248$ mm. Il fattore di scala SF , espresso in millimetri per pixel, è definito come il rapporto tra l'altezza reale del riferimento e la sua altezza misurata nell'immagine:



$$SF = \frac{h_{ref}}{\Delta y_{px}} \left[\frac{mm}{pixel} \right]$$

dove Δy_{px} rappresenta la differenza di coordinata y , in pixel, tra il punto superiore e quello inferiore del riferimento fisico nell'immagine. Tale misura è stata ottenuta mediante selezione manuale assistita dei bordi della borraccia su immagine ingrandita. Poiché la distanza tra fotocamera e soggetto può variare leggermente tra le diverse acquisizioni, il fattore di scala viene calcolato indipendentemente per ciascuna immagine, garantendo una calibrazione metrica specifica per ogni acquisizione.

Una volta determinato il fattore di scala, le coordinate metriche dei landmark vengono ottenute moltiplicando le coordinate in pixel per SF . Il sistema di riferimento delle immagini elaborate con MP presenta l'asse x orientato da sinistra verso destra, in accordo con la convenzione adottata dal sistema VICON, mentre l'asse y risulta orientato dall'alto verso il basso rispetto al soggetto esaminato, mantenendo quindi la stessa direzione ma verso opposto rispetto al sistema VICON.

Poiché il fattore di scala rappresenta un termine moltiplicativo comune a tutte le misure lineari espresse in millimetri, è necessario valutarne le principali sorgenti di incertezza e la loro propagazione nelle metriche finali. Considerando l'espressione $SF = h_{ref}/\Delta y_{px}$, l'incertezza relativa del fattore di scala può essere stimata mediante propagazione dell'incertezza al primo ordine:

$$\frac{u_{SF}}{SF} = \sqrt{\left(\frac{u_h}{h_{ref}}\right)^2 + \left(\frac{u_{px}}{\Delta y_{px}}\right)^2}$$

Le principali sorgenti di incertezza casuale sono due. La prima è associata alla tolleranza dimensionale del riferimento fisico utilizzato, stimata pari a $u_h = \pm 5$ mm. La seconda deriva dall'incertezza nella determinazione della distanza in pixel tra i bordi superiore e inferiore della borraccia, dovuta alla selezione manuale assistita, stimata pari a $u_{px} = \pm 3$ pixel. Assumendo un valore rappresentativo $\Delta y_{px} \approx 200$ pixel nelle immagini acquisite, i contributi relativi risultano pari a circa 2.00% e 1.50%. La combinazione delle due componenti porta a un'incertezza casuale relativa complessiva di circa 2.50%.

La fonte di errore più rilevante è di natura sistematica ed è legata alla parallasse fotografica. Il riferimento metrico non si trova infatti esattamente sul piano del soggetto, ma è posizionato



lateralmente o anteriormente/posteriormente ad esso ad una distanza di circa $d = 50$ cm. Considerando una distanza nominale fotocamera-soggetto pari a $D \approx 2500$ mm, la distanza effettiva tra fotocamera e riferimento risulta:

$$D_{ref} = \sqrt{D^2 + d^2} \approx \sqrt{2500^2 + 500^2} \approx 2550 \text{ mm}$$

Poiché la dimensione apparente di un oggetto nell'immagine dipende inversamente dalla distanza dalla fotocamera, il fattore di scala calcolato sul riferimento risulta leggermente sovrastimato rispetto a quello valido per il piano del soggetto. L'entità relativa di tale errore può essere stimata come:

$$\varepsilon_{parallasse} = \frac{D_{ref} - D}{D} \approx \frac{2550 - 2500}{2500} \approx 1.98\%$$

Questo errore comporta quindi una sovrastima sistematica delle distanze metriche ricavate dalle immagini di circa il 2%. In termini assoluti, ciò si traduce in scostamenti inferiori a 1 mm per distanze inferiori a 50 mm e inferiori a circa 3 mm per distanze fino a 150 mm, valori che rimangono contenuti rispetto ad altre sorgenti di errore presenti nel processo di misura, come il posizionamento dei marker sui soggetti esaminati.

È importante osservare che tale problematica non influenza invece le metriche angolari. Gli angoli articolari vengono infatti calcolati a partire dal prodotto scalare tra vettori normalizzati. La normalizzazione, quindi, elimina qualsiasi fattore moltiplicativo comune nelle componenti dei vettori; di conseguenza il fattore di scala si annulla matematicamente nel calcolo degli angoli. Parametri clinicamente rilevanti quali l'angolo femore-tibia, l'estensione dell'anca, l'estensione del ginocchio e la dorsi-flessione della caviglia risultano pertanto indipendenti dalla calibrazione metrica dell'immagine.

Alla luce di queste considerazioni, la procedura di calibrazione basata su fattore di scala adottata nel presente studio può essere considerata adeguata rispetto agli obiettivi della validazione.

2.7. Metriche posturali

Le metriche posturali calcolate a partire dai landmark MP sono state definite in accordo con le indicazioni cliniche fornite dalla fisioterapista collaboratrice, descritte nelle sezioni successive, e con la letteratura biomeccanica di riferimento [8][19][24]. Le metriche sono suddivise in due gruppi in base alla vista di acquisizione: vista frontale e vista sagittale. Per ogni metrica vengono descritti il significato clinico, la formula di calcolo e i *keypoints* MP utilizzati. Le



metriche secondarie (linea spinosa vertebrale, linea occipite-C7-solco gluteo, linea mammaria), richieste dalla fisioterapista, non sono state incluse poiché i *keypoints* restituiti da MP non consentono una ricostruzione affidabile di tali repere anatomici, come discusso nella sezione 2.5.3.

2.7.1. Vista frontale

Angolo tra asse femore e asse della tibia (valgismo/varismo ginocchio)

Valuta l'allineamento del ginocchio nel piano frontale. Un angolo nullo corrisponde ad allineamento perfetto degli assi femorali e tibiali; valori positivi indicano varismo, valori negativi valgismo. Clinicamente, deviazioni superiori a 5-7° sono considerate significative.

Definiti il vettore femorale $f = (\text{HIP} - \text{KNEE})$ e il vettore tibiale $t = (\text{KNEE} - \text{ANKLE})$, l'angolo è:

$$\vartheta = \cos^{-1}\left(\frac{f \cdot t}{|f| \cdot |t|}\right)$$

con segno determinato dalla direzione laterale della deviazione.

Keypoints: LEFT_HIP, LEFT_KNEE, LEFT_ANKLE (lato sinistro); RIGHT_HIP, RIGHT_KNEE, RIGHT_ANKLE (lato destro).

Il centro della rotula deve cadere in corrispondenza del II metatarso

Valuta il disallineamento orizzontale tra la rotula (approssimata con KNEE) e il secondo metatarso distale (approssimato con FOOT_INDEX). Un disallineamento eccessivo indica sovraccarico del compartimento mediale o laterale del ginocchio durante la deambulazione.

$$d = X_{\text{FOOT_INDEX}} - X_{\text{KNEE}}$$

espresso in pixel e, dopo l'applicazione del fattore di scala, in millimetri.

Keypoints: LEFT_KNEE, LEFT_FOOT_INDEX, RIGHT_KNEE, RIGHT_FOOT_INDEX.

Linea alba dell'addome (plumb line virtuale frontale)

La linea alba è la linea verticale immaginaria che passa per il centro dell'addome e che dovrebbe essere allineata con il centro di gravità. Una deviazione laterale indica asimmetria posturale del tronco, frequente nei pazienti con protesi d'arto inferiore per compensare il peso.

$$x_{\text{centro_bacino}} = \frac{X_{\text{LEFT_HIP}} + X_{\text{RIGHT_HIP}}}{2}$$



$$x_{centro_spalle} = \frac{X_{LEFT_SHOULDER} + X_{RIGHT_SHOULDER}}{2}$$

$$l = x_{centro_spalle} - x_{centro_bacino}$$

Keypoints: LEFT_SHOULDER, RIGHT_SHOULDER, LEFT_HIP, RIGHT_HIP.

Linea apici scapole

Valuta l'inclinazione della linea che congiunge gli apici scapolari approssimati con i landmark delle spalle. Idealmente la linea dovrebbe essere orizzontale. Un valore positivo indica che la spalla destra è più alta; negativo, che lo è la sinistra.

$$\text{tilt} = Y_{RIGHT_SHOULDER} - Y_{LEFT_SHOULDER}$$

Keypoints: LEFT_SHOULDER, RIGHT_SHOULDER.

2.7.2. Vista sagittale

Linea virtuale passante per orecchio, spalla, gran trocantere del femore, posteriormente alla rotula, davanti al malleolo laterale della caviglia (plumb line virtuale laterale)

In vista sagittale, la *plumb line virtuale laterale* è definita dall'asse verticale passante per il landmark dell'orecchio (EAR), che costituisce il riferimento anatomico standard per l'analisi posturale sagittale. Le deviazioni antero-posteriori dei principali landmarks rispetto a questa linea vengono calcolate come:

$$\Delta x_{spalla} = X_{LEFT/RIGHT_SHOULDER} - X_{LEFT/RIGHT_EAR}$$

$$\Delta x_{anca} = X_{LEFT/RIGHT_HIP} - X_{LEFT/RIGHT_EAR}$$

$$\Delta x_{ginocchio} = X_{LEFT/RIGHT_KNEE} - X_{LEFT/RIGHT_EAR}$$

$$\Delta x_{caviglia} = X_{LEFT/RIGHT_ANKLE} - X_{LEFT/RIGHT_EAR}$$

Valori positivi indicano spostamento anteriore rispetto alla *plumb line*, valori negativi uno spostamento posteriore. In una postura ideale, tutti i valori dovrebbero essere prossimi a zero.

Keypoints: LEFT/RIGHT_EAR, LEFT/RIGHT_SHOULDER, LEFT/RIGHT_HIP,
LEFT/RIGHT_KNEE, LEFT/RIGHT_ANKLE.

Angolo di estensione dell'anca (valore ideale 0°-10°)

Misura l'angolo tra l'asse del femore e l'asse del tronco in vista sagittale. In ortostatica ideale il valore è prossimo a 0°, con un intervallo fisiologico di 0°-10°. Definiti il vettore tronco $w = (\text{SHOULDER} - \text{HIP})$ e il vettore femorale $f = (\text{HIP} - \text{KNEE})$:



$$\alpha = \cos^{-1}\left(\frac{w \cdot f}{|w| \cdot |f|}\right)$$

Keypoints: LEFT_SHOULDER, LEFT_HIP, LEFT_KNEE (lato sinistro); RIGHT_SHOULDER, RIGHT_HIP, RIGHT_KNEE (lato destro).

Angolo di estensione del ginocchio (valore ideale 0°)

Misura l'angolo tra l'asse del femore e l'asse della tibia in vista sagittale. In ortostatica il ginocchio dovrebbe essere completamente esteso (0°). Valori positivi indicano flessione, valori negativi iperestensione.

$$\beta = \cos^{-1}\left(\frac{f \cdot t}{|f| \cdot |t|}\right)$$

Keypoints: LEFT_HIP, LEFT_KNEE, LEFT_ANKLE; RIGHT_HIP, RIGHT_KNEE, RIGHT_ANKLE.

Angolo di flessione della caviglia (valore ideale 85°-90°)

Misura l'angolo tra l'asse della tibia e l'asse del piede. In ortostatica il valore fisiologico è compreso tra 85° e 90° (piede quasi perpendicolare alla tibia). Deviazioni indicano dorsiflessione eccessiva (>90°) o plantarflexione (<85°).

$$\gamma = \cos^{-1}\left(\frac{t \cdot p}{|t| \cdot |p|}\right)$$

dove il vettore piede $p = (\text{FOOT_INDEX} - \text{ANKLE})$.

Keypoints: LEFT_KNEE, LEFT_ANKLE, LEFT_FOOT_INDEX; RIGHT_KNEE, RIGHT_ANKLE, RIGHT_FOOT_INDEX.

2.7.3. Metriche non incluse

Alcune metriche posturali secondarie non sono state incluse nella pipeline di valutazione, in quanto i landmark restituiti da MP non consentono una loro ricostruzione affidabile.

La linea spinosa vertebrale (allineamento dei processi spinosi nel piano frontale, indicativa di scoliosi e squilibri posturali del rachide), ad esempio, non è calcolabile in quanto MP non fornisce landmark vertebrali. Le curve fisiologiche della colonna non sarebbero rappresentabili e ogni ricostruzione comporterebbe variabilità e scarsa robustezza clinica.

O ancora, l'allineamento SIAS-SIPS (differenza di quota tra spina iliaca antero-superiore e postero-superiore, indicativa di antiversione o retroversione pelvica) non è direttamente calcolabile poiché non vengono forniti tali landmark. È stato verificato se esistesse una relazione antropometrica stabile tra SIAS/SIPS e i *keypoints* di HIP e SHOULDER disponibili;



tuttavia, tale relazione risulta soggetto-dipendente e non sufficientemente robusta per essere generalizzata senza un dataset di calibrazione specifico [26].

La intra/extra rotazione dell'anca richiederebbe una valutazione nel piano trasversale, mentre le acquisizioni disponibili sono limitate al piano frontale e sagittale. Pertanto, la stima indiretta dall'angolo tra asse pelvico e catena cinematica distale è stata esclusa per l'elevata variabilità introdotta dalla proiezione 2D di un movimento 3D.

Infine, la linea occipite-C7-solco gluteo in vista sagittale e la linea mammaria in vista frontale richiedono repere anatomici (occipite, C7, solco gluteo) non rilevati direttamente da MP. La loro ricostruzione artificiale da combinazioni di *keypoints* disponibili introdurrebbe errori sistematici non quantificabili senza una calibrazione soggetto-specifica dedicata [26].

2.8. Estrazione metriche da VICON

Per il confronto con le metriche MediaPipe, dal sistema VICON sono stati estratti gli angoli e le distanze corrispondenti, calcolati a partire dalle traiettorie tridimensionali dei marker e dagli output del modello *Plug-in Gait*. Le metriche VICON sono espresse in gradi per gli angoli articolari e in millimetri per le distanze.

L'angolo asse femore-tibia nel piano frontale, ad esempio, corrisponde all'angolo di abduzione/adduzione del ginocchio calcolato dal modello *Plug-in Gait: KneeAngles*, componente Y nel sistema di riferimento JCS, *Joint Coordinate System*. Quest'ultimo è il sistema di riferimento locale dell'articolazione utilizzato per descrivere in modo biomeccanicamente corretto i movimenti relativi tra due segmenti corporei. L'allineamento rotula-secondo metatarso, metrica lineare, è stato calcolato come differenza di coordinata X tra il marker KNEE e il marker TOE nel sistema di riferimento del laboratorio, proiettata sul piano coronale. L'allineamento degli apici delle scapole, invece, è stato derivato dalla differenza di quota (coordinata Z nel sistema laboratorio) tra i marker RSHO e LSHO. Infine, la deviazione della linea alba è stata calcolata come differenza di coordinata X tra il punto medio delle spalle (media tra LSHO e RSHO) e il punto medio del bacino (media tra LASI e RASI).

Per la vista sagittale, le deviazioni dalla *plumb line virtuale laterale* sono state calcolate come differenze di coordinata antero-posteriore (asse Y del laboratorio) tra il landmark dell'orecchio (LEAR o REAR) e i landmark di spalla, anca, ginocchio e caviglia. In conclusione, gli angoli di estensione dell'anca, ginocchio e caviglia corrispondono agli angoli sagittali calcolati dal modello *Plug-in Gait (HipAngles, KneeAngles, AnkleAngles)*, componente X in relazione alla convenzione JCS adottata).



Inoltre, gli angoli sono stati calcolati anche a partire dalle coordinate tridimensionali dei marker.

2.9. Metodologia di confronto

2.9.1. Struttura del dataset di confronto

Per ogni soggetto e per ogni ripetizione disponibile, le metriche estratte dai due sistemi sono state organizzate in un dataset strutturato con le seguenti variabili: soggetto (S1, S2, S3), numero di calibrazione (02, 03, 04), vista (frontale, sagittale), nome della metrica MediaPipe, nome della metrica VICON corrispondente, lato (destro, sinistro, centrale), valore MediaPipe, valore VICON ed errore assoluto. In totale, il dataset comprende 140 coppie di misure.

2.9.2. Metriche di errore

Per ogni coppia di misure è stato calcolato l'errore assoluto (EA) come:

$$EA = |Valore_MediaPipe - Valore_VICON|$$

Per ogni metrica e per ogni soggetto sono stati calcolati la media e la deviazione standard dell'errore assoluto, aggregando le misure sulle ripetizioni disponibili. L'analisi è stata condotta separatamente per vista frontale e sagittale, e separatamente per lato destro e sinistro.

2.9.3. Strumenti software

Tutta l'elaborazione dei dati, inclusa l'esecuzione di MediaPipe, il calcolo del fattore di scala, il calcolo delle metriche posturali, l'estrazione delle metriche VICON dai file CSV esportati da Nexus e il calcolo degli errori, è stata implementata in Python tramite le librerie *MediaPipe*, *OpenCV*, *NumPy*, *Pandas* e *Matplotlib*. I notebook utilizzati per l'elaborazione automatizzata sono descritti nella sezione successiva.

2.9.4. Motivazione dell'approccio descrittivo

L'obiettivo del presente studio non è quello di dimostrare l'equivalenza statistica tra il sistema MediaPipe e il sistema VICON su una popolazione, ma piuttosto caratterizzare il comportamento del sistema *markerless* in condizioni specifiche per la valutazione posturale protesica, individuare i pattern sistematici di errore e valutare la plausibilità clinica delle metriche estratte. A tale scopo, la caratterizzazione descrittiva dell'errore assoluto metrica per metrica, distinguendo tra soggetti e lato, appare lo strumento più adatto e informativo rispetto alla statistica formale. Quest'ultima è rimandata alla successiva fase di validazione prospettica, che dovrà prevedere un campione adeguato, stimabile per almeno 30 soggetti con amputazione di arto inferiore, bilanciati per livello di amputazione e sesso, su cui sarà possibile applicare strumenti statistici appropriati.



3. Elaborazione Python: gli script nel dettaglio

3.1. Implementazione

L'intera pipeline è stata implementata in Python 3, su ambiente Google Colaboratory, ed è organizzata in quattro notebook Jupyter distinti e sequenziali: calcolo manuale assistito dei fattori di scala delle immagini in esame; elaborazione batch delle immagini con MediaPipe PoseLandmarker e calcolo automatico delle metriche posturali; elaborazione batch dei file CSV esportati da VICON Nexus e calcolo delle metriche corrispondenti; confronto automatico tra le misure dei due sistemi. Il codice sorgente completo è disponibile, su richiesta, nel repository del progetto su GitHub (<https://github.com/salvatorerapisardaai-max/Validation-AI-Approach>”).

3.2. Architettura

La pipeline è stata progettata con un approccio modulare: ogni notebook opera in modo indipendente e riceve come input i file prodotti dal notebook precedente, consentendo il riutilizzo dei singoli moduli e la sostituzione di componenti senza alterare il flusso complessivo. Tutti i dati intermedi e i risultati finali vengono salvati su Google Drive in formato CSV, garantendo la tracciabilità e la riproducibilità dell'elaborazione. È stato sviluppato un ulteriore script (*automatic.py*) che esegue in sequenza automatica tutti e tre i notebook di elaborazione (VICON, MediaPipe, Validazione), consentendo la rielaborazione completa dell'intero dataset con un singolo avvio.

La Tabella 3.1 schematizza il flusso complessivo dell'elaborazione, dal dato grezzo di ingresso (immagini RGB e file CSV VICON) all'output finale (confronto MediaPipe-VICON).

Fase	Input	Notebook	Output
1. Fattori di scala	Immagini RGB (9 immagini, 3 sogg. × 3 viste)	<i>fattori_di_scala_mediapipe.ipynb</i>	scale_factors.csv
2. Elaborazione e MediaPipe	Immagini RGB + scale_factors.csv	<i>MediaPipe_processing.ipynb</i>	landmarks_S*.csv, metriche_S*.csv, immagini annotate, maschere di segmentazione
3. Elaborazione e VICON	File CSV ASCII da Nexus (Trajectories + Model Outputs)	<i>VICON_processing.ipynb</i>	soggetto{n}_metriche_{cal}.csv, v, soggetto{n}_risultati_{cal}.txt
4. Validazione	metriche MediaPipe + metriche VICON	<i>Validazione_processing.ipynb</i>	confronto_trial_per_trial.csv, statistiche_per_metrica.csv, grafici

Tabella 3.1 – Architettura a quattro moduli della pipeline di elaborazione. S* = soggetto (1, 2 o 3).



3.3. Calcolo dei fattori di scala

Come discusso in 2.6.4., il modello MP restituisce le coordinate dei landmark in forma normalizzata rispetto alle dimensioni dell'immagine, prive di qualsiasi informazione metrica assoluta. Per poter confrontare le misure con quelle del sistema VICON, espresse in millimetri, è necessario determinare un fattore di conversione pixel → millimetri specifico per ogni immagine, dipendente dalla distanza fotocamera-soggetto, dalla lunghezza focale e dalla risoluzione. A tale scopo è stato posizionato nell'inquadratura un riferimento fisico di dimensioni note: una borraccia di altezza $h = 24.8$ cm. Il fattore di scala SF (mm/px) è calcolato come:

$$SF = \frac{h_{ref}}{\Delta y_{px}} \left[\frac{mm}{pixel} \right]$$

Il processo è semi-automatico: l'operatore identifica visivamente i bordi superiore e inferiore della borraccia sull'immagine ingrandita (circa 1 minuto per tutte le immagini), e il notebook calcola e verifica il fattore in tempo reale. I fattori vengono esportati in `scale_factors.csv` (colonne: "Image_Name", "Scale_mm_per_px") e condivisi con i moduli successivi.

3.4. Elaborazione MediaPipe

3.4.1. Configurazione

Il rilevamento dei landmark corporei è stato realizzato tramite il modulo `PoseLandmarker` dell'API MediaPipe Tasks, variante heavy ad alta accuratezza. Il modello viene scaricato automaticamente dal repository Google al primo avvio. La configurazione del rilevatore è definita tramite l'oggetto `PoseLandmarkerOptions` e sono stati selezionati i seguenti parametri:

Parametro	Valore	Significato
<code>running_mode</code>	<code>RunningMode.IMAGE</code>	Modalità immagine statica
<code>num_poses</code>	1	Numero massimo di pose rilevabili
<code>min_pose_detection_confidence</code>	0.5	Soglia minima di confidenza per il rilevamento
<code>min_pose_presence_confidence</code>	0.5	Soglia minima di confidenza per la presenza della posa
<code>min_tracking_confidence</code>	-	Soglia minima di confidenza per il tracking (non attiva in modalità IMAGE)
<code>output_segmentation_masks</code>	<code>True</code>	Abilita l'output della maschera di segmentazione del corpo

Tabella 3.4 – Parametri di configurazione del modello `PoseLandmarker Heavy`.

La soglia di confidenza pari a 0.5 per il rilevamento e la presenza è stata scelta come compromesso tra la sensibilità del rilevamento (soglie più basse aumentano i falsi positivi) e la robustezza nelle condizioni di acquisizione adottate. In particolare, per i soggetti con protesi di



arto inferiore, le estremità distali dell'arto protesico possono presentare punteggi di visibilità inferiori ai soggetti normodotati a causa della morfologia atipica, come si evince dalla figura 2.12:

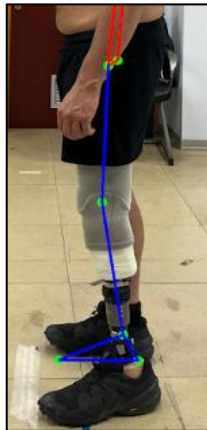


Figura 2.12 - Protesi S2

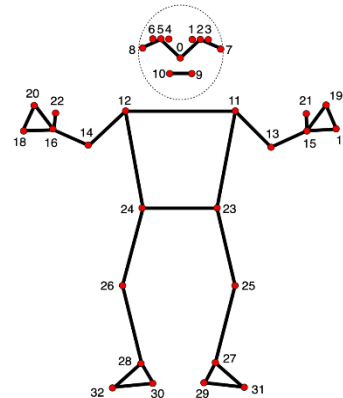


Figura 2.13 - 33 Landmarks restituiti da MP

Dei 33 landmark restituiti dal modello (figura 2.13), la pipeline utilizza un sottoinsieme di 14 landmark:

ID	Nome	Utilizzo
7, 8	LEFT_EAR, RIGHT_EAR	Riferimento plumb line virtuale sagittale
11, 12	LEFT_SHOULDER, RIGHT_SHOULDER	Linea alba, tilt scapole, angolo est. anca
23, 24	LEFT_HIP, RIGHT_HIP	Vettore femorale, linea alba, angolo est. anca
25, 26	LEFT_KNEE, RIGHT_KNEE	Varismo/valgismo, angolo est. ginocchio, allineamento rotula - II metatarso
27, 28	LEFT_ANKLE, RIGHT_ANKLE	Angolo fless. caviglia
29, 30	LEFT_HEEL, RIGHT_HEEL	Connessione piede
31, 32	LEFT_FOOT_INDEX, RIGHT_FOOT_INDEX	Approssimazione II metatarso, angolo fless. caviglia

Tabella 3.5 – Landmark MediaPipe utilizzati nella pipeline di calcolo delle metriche posturali.

3.4.2. Organizzazione cartelle e rilevamento automatico della vista

Il notebook scansiona ricorsivamente tutte le sottocartelle di Google Drive, identifica i file immagine supportati e processa ciascuna immagine in sequenza, recuperando il fattore di scala corrispondente da *scale_factors.csv*. La vista di acquisizione (frontale, laterale sinistra, laterale destra) viene determinata automaticamente dalla funzione *detect_view()* in base ai punteggi di visibilità delle ginocchia: entrambe visibili → frontale; solo la sinistra → laterale sx; solo la destra → laterale dx. L'approccio è robusto per deviazioni dall'asse ideale fino a 15-20°.

3.4.3 Calcolo delle metriche posturali

Le coordinate normalizzate vengono convertite in pixel (moltiplicando per la larghezza dell'immagine W_{img} e l'altezza H_{img} rispettivamente), quindi in millimetri tramite il fattore di scala SF. Il calcolo degli angoli, adimensionale, avviene in pixel; quello delle distanze lineari in millimetri. I landmark con visibilità < 0.5 producono un valore NaN per la metrica dipendente. Le metriche della vista frontale sono calcolate da *calculate_frontal_metrics()*, quelle sagittali da *calculate_sagittal_metrics()*. Per la definizione clinica e le formule analitiche si rimanda alla sezione 2.7. La Tabella 3.2 riassume la corrispondenza tra metriche MP e metriche VICON utilizzata nel confronto, con indicazione dell'unità di misura e del lato di riferimento.

Metrica MP	Titolo clinico	Corrispondenza VICON	Lato	Unità
<i>ang_fem_tib_dx / sx</i>	Ang. femore-tibia frontale	<i>KneeAngles Y</i> (valgismo/varismo)	Destro / Sinistro	°
<i>delta_rotula_metatarso_dx / sx mm</i>	Rotula → II metatarso	Marker RKNE/LKNE – RTOE/LTOE	Destro / Sinistro	mm
<i>dev_linea_alba_mm</i>	Deviazione linea alba	(LSHO+RSHO)/2 – (LASI+RASI)/2, asse X	Frontale	mm
<i>tilt_scapole mm</i>	Tilt apici scapole	Z RSHO – Z LSHO	Frontale	mm
<i>delta_plumb_sh / hip / knee / ankle_dx / sx_mm</i>	Plumb line sagittale (spalla, anca, ginocchio, caviglia)	Δ coordinata Y: marker – EAR1/EAR2	Destro / Sinistro	mm
<i>ang_est_anca_dx / sx</i>	Angolo estensione anca	<i>HipAngles X</i> (flesso-estensione)	Destro / Sinistro	°
<i>ang_est_gin_dx / sx</i>	Angolo estensione ginocchio	<i>KneeAngles X</i> (flesso-estensione)	Destro / Sinistro	°
<i>ang_est_cav_dx / sx</i>	Angolo flessione caviglia	<i>AnkleAngles X</i> (dorsi/plantarflessione)	Destro / Sinistro	°

Tabella 3.2 – Mappatura completa tra metriche MediaPipe e metriche VICON.

3.4.4. Output

Per ogni immagine elaborata, il notebook produce un'immagine annotata con i landmark rilevati e le connessioni scheletriche disegnate, ed un'immagine segmentata ottenuta applicando la maschera di segmentazione del corpo. Entrambe le immagini vengono salvate nella cartella di output.



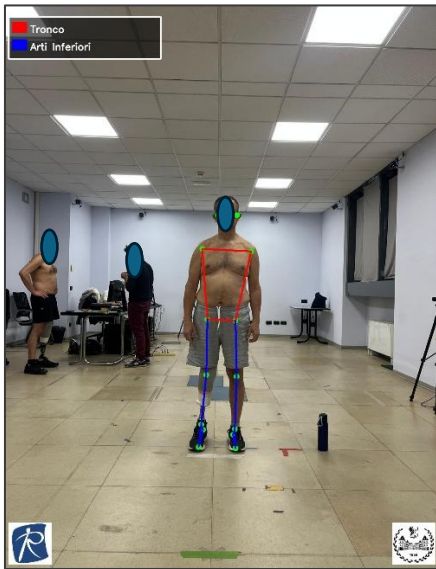


Figura 2.13 – Esempio di landmark restituiti dal modello - Vista Frontale

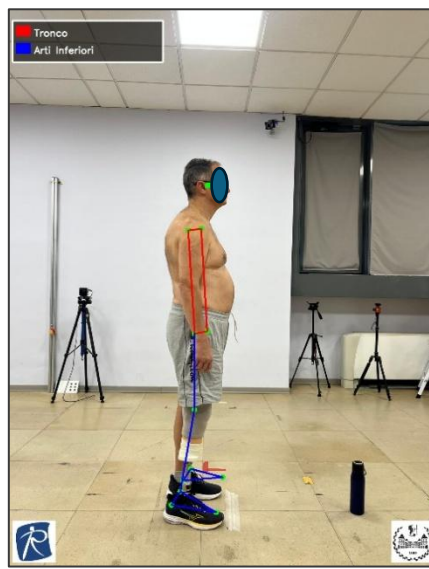


Figura 2.14 – Landmark. Vista sagittale, lato destro

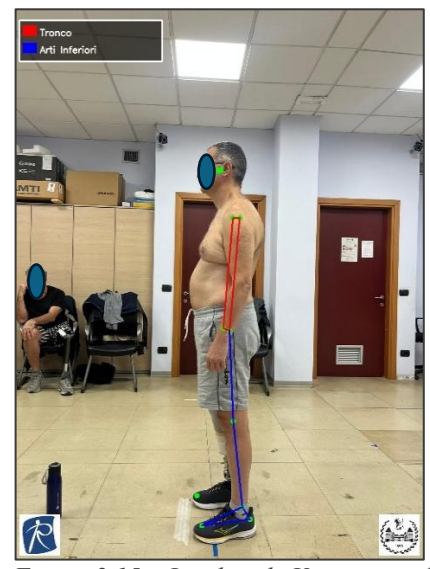


Figura 2.15 – Landmark. Vista sagittale, lato sinistro



Figura 2.16 - Esempio maschera di segmentazione restituita dal modello - Vista Frontale



Figura 2.17 – Maschera - Vista sagittale, lato destro



Figura 2.18 - Maschera - Vista sagittale, lato sinistro

Il notebook, inoltre, produce due file CSV per ogni soggetto: uno contenente le coordinate grezze di tutti i landmark rilevati, l'altro le metriche calcolate.

File di output	Contenuto	Colonne principali
landmarks_Sx.csv	Coordinate grezze di tutti i landmark	image_name, view, landmark_id, landmark_name, x, y, z, visibility, x_px, y_px, x_mm, y_mm
soggetto {n}_metriche.csv	Metriche posturali	cfr. 2.7
annotated_*.jpg	Immagine con landmark sovrainpressi	-
segmented_*.jpg	Immagine con maschera di segmentazione	-

Tabella 3.6 – File di output tramite elaborazione MP.

3.5. Elaborazione VICON

3.5.1. Configurazione

Il software VICON Nexus esporta i dati di acquisizione in formato ASCII (CSV) tramite la pipeline *Export ASCII*. Il file prodotto presenta due sezioni: la prima, (introdotta dall'intestazione *Model Outputs*) contiene i dati cinematici calcolati dal modello *Plug-in Gait Full Body* (angoli articolari, momenti, potenze), organizzati per frame temporale; la seconda sezione (intestazione *Trajectories*), invece, contiene le coordinate tridimensionali XYZ dei marker in mm per ogni frame.

Il notebook implementa un parser dedicato che ricerca le intestazioni della sezione di interesse, estrae il numero di frame disponibili, i nomi delle variabili e i dati numerici, e ricostruisce strutture dati NumPy.

3.5.2. Estrazione angoli articolari e traiettorie

Le metriche angolari sono estratte da due sorgenti distinte, entrambe incluse nell'output per consentire un confronto più robusto con MediaPipe:

- *Model Outputs (MO)*: angoli articolari pre-calcolati dal modello *Plug-in Gait Full Body* secondo la convenzione JCS (*Joint Coordinate System*). Forniscono stime biomeccanicamente corrette ma basate sulla ricostruzione tridimensionale completa con marker.
- *Traiettorie 3D (Traj.)*: angoli calcolati direttamente dalle coordinate 3D dei marker, proiettando i vettori anatomici sul piano di interesse (piano frontale XZ per valgismo/varismo, piano sagittale YZ per angoli all'anca, ginocchio e caviglia). Questo approccio riproduce concettualmente il calcolo bidimensionale di MP e consente un confronto metodologicamente più diretto.

Le metriche lineari (allineamenti e deviazioni in mm) vengono calcolate esclusivamente dalle traiettorie 3D nel sistema di riferimento del laboratorio VICON (X: mediolaterale, Y: antero-posteriore, Z: verticale).

3.5.3. Output

Per ogni coppia soggetto-calibrazione, *vicon_processing.ipynb* produce un file CSV delle metriche (*soggetto{n}_metriche_{cal}.csv*) con 18 righe (una per metrica) e le colonne "Metric", "Right (or frontal)", "Left".



3.6. Notebook validazione

Il notebook *validazione_processing.ipynb* confronta automaticamente le metriche MP con le metriche VICON mediante una mappatura esplicita (dizionario “METRIC_MAPPING”). Per ogni coppia di misure viene calcolato l'errore assoluto $EA = |\text{Valore_MP} - \text{Valore_VICON}|$. I risultati vengono aggregati per metrica e per soggetto, calcolando media e deviazione standard dell'errore assoluto sulle ripetizioni disponibili. Il dataset di confronto finale comprende 140 coppie di misure e viene esportato in *confronto_trial_per_trial.csv*.

4. Prototipo applicativo per la telemedicina

4.1. Motivazione e obiettivi

Parallelamente allo sviluppo e alla validazione scientifica della pipeline di analisi, è stato realizzato un prototipo applicativo volto a trasferire il sistema di valutazione posturale in un contesto di utilizzo clinico remoto.

L'obiettivo principale è dimostrare la fattibilità tecnica di un flusso operativo in cui il fisioterapista, mediante un comune smartphone, acquisisce e invia tre immagini del paziente e riceve, nell'arco di pochi secondi, un report clinico strutturato contenente le metriche posturali calcolate automaticamente dal sistema. Tale processo avviene senza la necessità di strumentazione dedicata o infrastrutture hardware specialistiche, configurandosi come una soluzione potenzialmente integrabile nei percorsi di telemedicina.

L'approccio proposto intende quindi coniugare rigore metodologico e accessibilità tecnologica, con particolare riferimento ai pazienti portatori di protesi di arto inferiore, per i quali il monitoraggio posturale periodico rappresenta un elemento clinico di rilievo.

4.2. Architettura del sistema

Il prototipo è stato progettato secondo un'architettura di tipo *client-server*, un modello molto diffuso nelle applicazioni web. In questo schema, il *client* (ad esempio un'applicazione su smartphone) invia una richiesta a un *server*, che elabora i dati e restituisce una risposta con i risultati. Nel caso specifico, il sistema è composto da:

- Backend (*server*): sviluppato in Python mediante il framework FastAPI, consente di creare servizi web in modo strutturato ed efficiente, definendo degli “indirizzi” (*endpoint*) ai quali il *client* può inviare richieste (ad esempio per analizzare le immagini) e dai quali riceve i risultati elaborati.



- Accesso sicuro (HTTPS): la comunicazione tra app e *server* avviene tramite protocollo HTTPS, ovvero una modalità di trasmissione dati cifrata. Questo significa che le informazioni scambiate, comprese le immagini e i dati del paziente, vengono protette durante il trasferimento, riducendo il rischio di intercettazioni.
- Tunnel *ngrok*: durante la fase di sviluppo, il *server* era eseguito in ambiente locale (Google Colaboratory) e quindi non direttamente accessibile da Internet. Per consentire all'app Android di comunicare con esso, è stato utilizzato *ngrok*, che crea temporaneamente un collegamento sicuro tra il *server* locale e la rete pubblica.
- *Client* (app Android): è stata sviluppata un'applicazione prototipale in Android Studio, testata su emulatore. L'app consente al fisioterapista di inserire i dati del paziente, caricare le immagini e inviarle al *server* per l'analisi.

Per quanto riguarda l'invio dei dati, viene utilizzato un formato di richiesta chiamato *multipart/form-data*. Tale meccanismo permette di trasmettere, all'interno della stessa richiesta, sia file, come le fotografie nel nostro caso, sia informazioni testuali, quali altezza del paziente e dati antropometrici. Il *server* riceve tutti questi elementi insieme, li elabora e restituisce le metriche calcolate, le immagini annotate e il report clinico in formato PDF.

4.3. Calcolo automatico del fattore di scala

Per trasformare le distanze misurate in pixel nelle immagini in misure reali (millimetri), è necessario calcolare un fattore di scala. Nella fase sperimentale di ricerca, come nel presente progetto di tesi, questo passaggio viene effettuato utilizzando un riferimento fisico noto (nel nostro caso, una bottiglietta di altezza nota; cfr. 2.6.4.). Tuttavia, in un contesto clinico remoto tale soluzione risulterebbe poco pratica.

Per questo motivo, nel prototipo è stato introdotto un metodo alternativo che non richiede oggetti di riferimento esterni. Il sistema calcola automaticamente la scala utilizzando:

- l'altezza reale del paziente, inserita dal fisioterapista una sola volta;
- la distanza in pixel tra due punti anatomici rilevati automaticamente dall'algoritmo: l'orecchio (EAR) e il tallone (HEEL).

Il fattore di scala (SF), espresso in millimetri per pixel, è calcolato come:



$$SF = \frac{altezza_{cm} \times 10}{d_{pixel}(EAR - HEEL)}$$

La scelta del segmento orecchio-tallone è motivata dal fatto che:

- è quasi sempre visibile nelle tre viste richieste (frontale e sagittali);
- è stabile e ben identificabile dall'algoritmo;
- appartiene al corpo del paziente e non dipende da oggetti esterni.

Il calcolo viene effettuato separatamente per ciascuna delle tre immagini, così da compensare eventuali differenze di inquadratura.

4.4. Endpoint API

Nel sistema sono definiti tre principali “indirizzi” del *server* (endpoint), ciascuno con una funzione specifica:

Endpoint	Metodo	Funzione
<i>/health</i>	GET	Verifica che il server sia attivo e funzionante
<i>/analyze/static</i>	POST	Riceve le tre immagini e i dati del paziente, calcola le metriche posturali e genera il report
<i>/report/pdf</i>	POST	Genera il report clinico in formato PDF

L'endpoint principale è */analyze/static*. Quando il fisioterapista preme il pulsante “Analizza” nell'app:

1. le tre fotografie e i dati del paziente vengono inviati al *server*;
2. il *server* elabora le immagini;
3. calcola automaticamente le metriche posturali;
4. genera le immagini annotate e il report clinico;
5. restituisce tutti i risultati all'app.

4.5. Report clinico PDF

Al termine dell'elaborazione, il sistema genera automaticamente un report clinico in formato PDF (A4), utilizzando la libreria software dedicata *ReportLab* (cfr. APPENDICE B). L'obiettivo è fornire al fisioterapista uno strumento immediato per la consultazione e



l'archiviazione digitale, facilitando anche il confronto nel tempo tra valutazioni successive dello stesso paziente.

4.6. Android Studio

L'applicazione *client* è stata sviluppata come prototipo su Android Studio e testata su emulatore Nexus 5X (API 30). L'interfaccia è stata progettata per essere essenziale e intuitiva, così da non interferire con il normale flusso di lavoro clinico.

Il fisioterapista può:

- inserire i dati anagrafici e antropometrici del paziente;
- acquisire o caricare tre fotografie standardizzate (frontale, sagittale sinistra, sagittale destra);
- avviare l'analisi tramite il pulsante "Analizza";
- visualizzare i risultati;
- scaricare il report PDF generato automaticamente.

Le fotografie devono essere scattate replicando il protocollo del seguente studio: la fotocamera deve riprendere l'intero soggetto e deve essere posizionata il più possibile parallela al piano del soggetto fotografato, al fine di minimizzare distorsioni prospettiche ed errori.

4.7. Considerazioni

È importante evidenziare che il sistema rappresenta un *proof-of-concept*, finalizzato a dimostrare la fattibilità del prototipo in telemedicina. Le prestazioni cliniche dipendono in modo significativo da:

- qualità delle immagini (illuminazione, distanza, corretto allineamento della camera);
- accuratezza dell'altezza inserita;
- corretta esecuzione delle viste richieste.

Analisi Posturale
Seleziona le 3 foto e compila i dati paziente

FOTOGRAFIE

Frontale Lat. Destra Lat. Sinistra

DATI PAZIENTE

Nome Cognome

Fisioterapista

Altezza paziente (cm) *

Obbligatorio — usato per calcolare automaticamente la scala mm/px

Note cliniche (opzionale)

PARAMETRI ANATOMICI

Defaulti clinici preimpostati. Modificabili solo su misura sul paziente.

Semilargh. malle. 35.0 Dist. inter-SIAS L. 170.0

Default: 35.0 Default: 170.0

Offset patella (m.) 21.25 Offset ginocchio... 52.0

Default: 21.25 Default: 52.0

ANALIZZA POSTURA

Figura 4 - Interfaccia App su smartphone

Pertanto, prima di un eventuale impiego clinico strutturato, sarà necessario condurre uno studio prospettico di validazione per:

- valutare accuratezza e riproducibilità del metodo di scala automatico;
- confrontare le misurazioni con standard strumentali di riferimento;
- analizzare la robustezza del sistema in condizioni reali non controllate.

5. Risultati

5.1. Struttura del dataset di confronto

Il dataset finale comprende 140 coppie di misure, distribuite su 3 soggetti (S1, S2, S3), 3 ripetizioni di acquisizione per S1 e S2 (Cal_02, Cal_03, Cal_04) e 1 sola ripetizione per S3 (Cal_02), 3 viste e 20 metriche posturali. Le metriche risultano espresse in gradi (°) per le misure angolari e in millimetri (mm) per le misure lineari. Per ogni coppia di misure è calcolato l'errore assoluto $EA = |\text{Valore_MP} - \text{Valore_VICON}|$. Ove disponibili, vengono riportati separatamente i valori angolari da *Model Outputs* di VICON. Le tabelle che seguono riportano, per ciascuna metrica, la media dei valori MP, VICON e dell'errore assoluto aggregando le ripetizioni disponibili per ogni soggetto.

5.2. Convenzione segni

La corretta interpretazione clinica dei risultati richiede una comprensione precisa della convenzione dei segni adottata per ciascuna metrica. Le metriche lineari sono distanze con segno (mm) che indicano la direzione del disallineamento rispetto a una retta di riferimento; le metriche angolari sono angoli con segno (°) che indicano il tipo e il verso della deviazione dal valore fisiologico ideale. La Tabella 5.1. riporta, per ogni metrica, la definizione del segno positivo, del segno negativo e il valore di riferimento clinico adottato dal protocollo ROGA per gli angoli sagittali di anca, ginocchio e caviglia; mentre, sono stati scelti i range più diffusi in letteratura per le restanti metriche. Sia MediaPipe che VICON adottano la medesima convenzione geometrica per le metriche derivate dalla proiezione 2D; per gli angoli articolari, il riferimento è il *Model Outputs* del *Plug-in Gait*, che esprime gli angoli nel sistema di riferimento articolare del *Joint Coordinate System (JCS)*.



Metrica	Vista	Unità	Valore POSITIVO (+)	Valore NEGATIVO (-)	Valore ideale
VISTA FRONTALE — METRICHE ANGOLARI					
Ang. femore-tibia DX	Frontale	°	Varismo (ginocchio varo): tibia deviata lateralmente rispetto al femore	Valgismo (ginocchio valgo): tibia deviata medialmente rispetto al femore	0° (allineamento neutro) Range normale: ±5°
Ang. femore-tibia SX	Frontale	°	Varismo lato sinistro	Valgismo lato sinistro	0° Range normale: ±5°
VISTA FRONTALE — METRICHE LINEARI					
Rotula-II metatarso DX	Frontale	mm	Rotula laterale rispetto al II metatarso (piede varo)	Rotula mediale rispetto al II metatarso (piede valgo)	0 mm (allineamento verticale)
Rotula-II metatarso SX	Frontale	mm	Rotula laterale rispetto al II metatarso (lato SX)	Rotula mediale rispetto al II metatarso (lato SX)	0 mm
Dev. linea alba	Frontale	mm	Deviazione verso destra del punto medio ombelicale rispetto alla verticale testa-pelvi	Deviazione verso sinistra	0 mm (baricentro centrato) Tolleranza clinica: ±10 mm
Tilt scapole	Frontale	mm	Spalla destra più alta della sinistra (dislivello positivo DX-SX)	Spalla sinistra più alta della destra	0 mm (scapole simmetriche) Tolleranza: ±10 mm
VISTA SAGITTALE — DEVIAZIONI DALLA PLUMB LINE					
Spalla / Anca / Ginocchio / Caviglia	Sagittale	mm	Punto anatomico anteriore alla plumb line (proiezione verticale del meato acustico esterno): postura in ante-flessione o	Punto anatomico posteriore alla plumb line: postura in retro-flessione o spostamento del carico verso dietro	0 mm (allineamento verticale ideale)



Metrica	Vista	Unità	Valore POSITIVO (+)	Valore NEGATIVO (-)	Valore ideale
			spostamento del carico verso avanti		
VISTA SAGITTALE — ANGOLI ARTICOLARI (°)					
Ang. estensione anca	Sagittale	°	Estensione d'anca: coscia arretrata rispetto al tronco (fisiologico in stazione ortostatica)	Flessione d'anca: coscia avanzata rispetto al tronco (postura in flessione)	0°-10° (stazione ortostatica ideale per amputati)
Ang. estensione ginocchio	Sagittale	°	Flessione residua del ginocchio (ginocchio flesso in stazione ortostatica: compenso frequente)	Iperestensione del ginocchio (compenso frequente nel lato sano degli amputati)	0° (estensione completa ideale) Tolleranza clinica: ±5°
Ang. flessione caviglia	Sagittale	°	Dorsi-flessione (avampiede in alto: configurazione fisiologica in stazione ortostatica)	Plantar-flessione (avampiede in basso: protesi con angolo fisso in plantarflessione)	85°-90° (lieve dorsi-flessione fisiologica)

Tabella 5.1 - Convenzione dei segni adottata per ogni metrica posturale. Verde = segno fisiologico o atteso in stazione ortostatica normale; giallo = deviazione che richiede valutazione clinica; rosso = compenso posturale patologico frequente in amputati. La convenzione è coerente tra il sistema MediaPipe e il sistema VICON per tutte le metriche, con l'eccezione degli angoli articolari sagittali per i quali VICON MO incorpora la correzione JCS del Plug-in Gait.

5.3. Metriche della vista frontale

5.3.1. Angolo asse femore-tibia (valgismo/varismo)

La Tabella 5.2 riporta i valori medi degli angoli femore-tibia frontali con la deviazione standard (DS) calcolata sulla variabilità intra-sessione di VICON (n=3 trial per S1 e S2; n=1 per S3). La DS di EA coincide con la DS di VICON in quanto MP è costante (unico valore). Il confronto include entrambi i riferimenti VICON (*Traj.* e *MO*): per questa metrica, VICON *Traj.* è il riferimento metodologicamente corretto (confronto geometrico 2D omogeneo), mentre VICON *MO* incorpora correzioni JCS 3D che ne rendono i valori incompatibili con la stima 2D di MP (es. S3 SX: EA vs MO = 13.52° vs EA vs *Traj.* = 2.35°).



Sogg.	Lato	MP (°)	VICON Traj (°) media ± DS	VICON MO (°)	EA vs Traj (°) media ± DS	EA vs MO (°)	Tipo deviazione
<i>Lato destro Range neutro: 0° ± 5°</i>							
<i>+= Varismo -= Valgismo</i>							
S1	DX*	-7.99	-1.27 ± 0.19	-0.30	6.73 ± 0.19	7.70	Valgismo (-)
S2	DX	-4.20	-7.87 ± 0.64	-5.72	3.67 ± 0.64	1.52	Valgismo (-)
S3	DX	+2.90	+4.63 (n=1)	-5.14	1.73 (n=1)	8.04	Varismo (+)
<i>Lato sinistro Range neutro: 0° ± 5°</i>							
S1	SX	+0.18	+0.82 ± 0.10	+2.42	0.64 ± 0.10	2.24	Varismo (+)
S2	SX*	+3.84	+5.85 ± 1.05	+10.75	2.00 ± 1.05	6.91	Varismo (+)
S3	SX*	+2.51	+4.86 (n=1)	-11.01	2.35 (n=1)	13.52	Varismo (+)
Media glob.	DX	-	-	-	4.70 ± 2.05°	5.10 ± 3.37°	
Media glob.	SX	-	-	-	1.47 ± 0.99°	5.85 ± 4.18°	

EA ≤ 5° Ottimo	EA 5-10° Accettabile	EA > 10° Critico	Protesi o anomalia	In range clinico
-----------------------	-----------------------------	----------------------------	---------------------------	-------------------------

Tabella 5.2 – Angolo asse femore-tibia nel piano frontale. EA = errore assoluto rispetto a VICON Traiettorie (VICON Traj.) o VICON Model Outputs (VICON MO). DS = deviazione standard inter-trial di VICON e di EA. Valore fisiologico di riferimento: 0° ± 5°. * = lato corrispondente all'arto protesico.

5.3.2. Metriche lineari frontali

La Tabella 5.3 riporta le tre metriche lineari della vista frontale. La DS di EA è disponibile per S1 (n=3) e per S2 dove applicabile. Da notare che, per la deviazione della linea alba e il tilt delle scapole, la DS di VICON tra trial è elevata per S2 (rispettivamente 8.95 mm e 7.97 mm), indicando oscillazioni posturali reali del soggetto tra le acquisizioni e non errori sistematici del



sistema di misura. Le anomalie di S3 (valori VICON negativi di grande entità) sono ritenute artefatti strumentali e non sono incluse nella sintesi S1+S2.

Sogg.	Metrica	Lato	MP (mm)	VICON Traj (mm) media \pm DS	EA (mm) media \pm DS	Note cliniche
Rotula – II Metatarso Valore ideale: 0 mm + = rotula laterale - = rotula mediale						
S1	Rotula-II meta	DX*	+79.41	+27.98 \pm 1.33	51.43 \pm 1.33	MP sistematicamente lateralizzato
S2	Rotula-II meta	DX	+82.85	+41.12 \pm 4.95	41.73 \pm 4.95	Bias sistematico MP
S3	Rotula-II meta	DX	+53.15	+59.16 (n=1)	6.01 (n=1)	Eccellente; morfologia protesi compatibile
S1	Rotula-II meta	SX	-24.86	+36.50 \pm 1.12	61.36 \pm 1.12	Discordanza di segno: MP mediale, VICON laterale
S2	Rotula-II meta	SX*	-56.22	-26.89 \pm 5.49	29.33 \pm 5.49	Concordanza di segno; sovrastima MP lateralità
S3	Rotula-II meta	SX*	-8.48	-176.30 (n=1)	167.83 (n=1)	Anomalia VICON; EA non interpretabile
Deviazione linea alba Valore ideale: 0 mm + = deviazione verso DX - = verso SX F = Frontale						
S1	Dev. linea alba	F	+33.34	+31.84 \pm 2.02	2.05 \pm 1.07	Eccellente accordo
S2	Dev. linea alba	F	+12.93	+9.27 \pm 8.95	5.64 \pm 7.23	Alta DS VICON: instabilità tronco tra trial



Sogg.	Metrica	Lato	MP (mm)	VICON Traj (mm) media \pm DS	EA (mm) media \pm DS	Note cliniche
S3	Dev. linea alba	F	+31.16	-109.25 (n=1)	140.41 (n=1)	Anomalia VICON; EA non interpretabile
Tilt scapole Valore ideale: 0 mm + = spalla DX più alta - = spalla SX più alta F = Frontale						
S1	Tilt scapole	F	+16.33	+13.87 \pm 2.14	2.47 \pm 2.13	Ottimo accordo
S2	Tilt scapole	F	+9.45	+13.45 \pm 7.97	7.32 \pm 2.69	Alta DS VICON; accordo accettabile
S3	Tilt scapole	F	+1.92	-8.94 (n=1)	10.86 (n=1)	Discordanza di segno
Sintesi S1+S2	(escluse anomalie S3)				Tilt: 4.89 \pm 3.6 mm Alba: 3.84 \pm 4.5 mm Rotula DX: 46.6 mm Rotula SX: 45.3 mm	

EA \leq 10 mm Ottimo	EA 10-30 mm Accettabile	EA > 30 mm Critico	Accordo eccellente	Anomalia/discordanza segno
---------------------------	----------------------------	-----------------------	--------------------	----------------------------

Tabella 5.3 - Metriche lineari frontali (F) per soggetto. I valori negativi indicano la direzione del disallineamento secondo la convenzione di segno adottata. * = lato corrispondente all'arto protesico.

5.4. Metriche della vista sagittale

5.4.1. Deviazioni dalla plumb line virtuale sagittale

La Tabella 5.4 riporta le deviazioni antero-posteriori dalla plumb line per quattro segmenti corporei. La DS di EA per S1 (n=3 trial) è generalmente contenuta (0.80-2.75 mm), confermando l'elevata ripetibilità di VICON intra-sessione. Per S2 lato DX il riferimento



VICON è disponibile per il solo trial T04 (n=1): la DS non è calcolabile e il dato va interpretato con cautela. I segmenti DX di S3 sono privi di riferimento VICON per impossibilità di ricostruire i marker sul segmento protesico (indicati con '-'). Le discordanze di segno per l'anca (VICON negativo, MP positivo) sono sistematiche e riflettono un bias geometrico di MediaPipe nella localizzazione del gran trocantere rispetto al punto ricostruito da VICON.

Sogg.	Segmento	Lato	MP (mm)	VICON Traj (mm) media ± DS	EA (mm) media ± DS	Note cliniche
Spalla Valore ideale: ~0 mm (spalla sulla verticale dell'orecchio) + = anteriore - = posteriore						
S1	Spalla	DX*	+86.39	+85.63 ± 0.80	0.76 ± 0.80	<i>Eccellente; spalla quasi sulla plumb</i>
S1	Spalla	SX	+73.80	+39.61 ± 0.89	34.19 ± 0.89	<i>MP sovrastima proiezione ant. spalla SX</i>
S2	Spalla	DX	+101.20	+87.28 (n=1)	13.92 (n=1)	<i>Solo T04; spalla ant. alla plumb</i>
S2	Spalla	SX*	+62.40	+54.47 ± 12.57	9.75 ± 10.47	<i>Alta DS VICON accordo medio</i>
S3	Spalla	DX	+119.17	-	-	<i>Nessun riferimento VICON disponibile</i>
S3	Spalla	SX*	+97.83	+155.23 (n=1)	57.40 (n=1)	<i>MP sottostima; shift ant. reale di S3</i>
Anca Valore ideale: 0 mm Bias sistematico MP: anca sempre positiva vs VICON spesso negativa						
S1	Trocantere	*DX	+57.17	+26.48 ± 1.69	30.69 ± 1.69	<i>MP sistematicamente ant. al punto corrispondente VICON</i>
S1	Trocantere	SX	+27.59	-1.10 ± 1.60	28.69 ± 1.60	<i>Discordanza segno: bias sistematico MP sull'anca</i>



Sogg.	Segmento	Lato	MP (mm)	VICON Traj (mm) media ± DS	EA (mm) media ± DS	Note cliniche
S2	Trocantere	DX	+67.71	+23.16 (n=1)	44.55 (n=1)	<i>Solo T04; errore molto elevato</i>
S2	Trocantere	*SX	+15.11	-22.91 ± 13.79	38.01 ± 13.79	<i>Discordanza segno; alta DS VICON tra trial</i>
S3	Trocantere	DX	+86.42	-	-	<i>Nessun riferimento VICON disponibile</i>
S3	Trocantere	SX*	+48.49	+141.77 (n=1)	93.28 (n=1)	<i>MP sottostima drasticamente</i>
Ginocchio Valore ideale: 0 mm Frequente discordanza di segno per ginocchio post. plumb (VICON -, MP +)						
S1	Ginocchio	DX*	+58.33	+7.80 ± 2.30	50.53 ± 2.30	<i>MP molto ant. alla plumb;</i>
S1	Ginocchio	SX	+47.10	+51.66 ± 2.24	4.56 ± 2.24	<i>Ottimo accordo lato sano</i>
S2	Ginocchio	DX	+73.63	+49.27 (n=1)	24.36 (n=1)	<i>Solo T04; errore moderato</i>
S2	Ginocchio	SX*	+0.04	-34.00 ± 14.57	34.05 ± 14.57	<i>Discordanza segno; VICON: ginocchio post. plumb</i>
S3	Ginocchio	DX	+96.34	-	-	<i>Nessun riferimento VICON disponibile</i>
S3	Ginocchio	SX*	+27.70	+106.55 (n=1)	78.85 (n=1)	<i>MP sottostima; ginocchio molto post. in S3</i>
Caviglia Valore ideale: ~0-20 mm ant. Meno influenzato da compensi protesici						
S1	Caviglia	DX*	+120.24	+94.90 ± 2.43	25.34 ± 2.43	<i>MP ant. alla plumb; errore moderato</i>



Sogg.	Segmento	Lato	MP (mm)	VICON Traj (mm) media \pm DS	EA (mm) media \pm DS	Note cliniche
S1	Caviglia	SX	+84.01	+65.66 \pm 2.75	18.35 \pm 2.75	<i>Errore sistematico contenuto</i>
S2	Caviglia	DX	+127.03	+146.17 (n=1)	19.14 (n=1)	<i>Solo T04; VICON > MP; errore contenuto</i>
S2	Caviglia	SX*	+50.56	+59.05 \pm 10.94	10.13 \pm 8.61	<i>Accordo accettabile; DS VICON moderata</i>
S3	Caviglia	DX	+149.01	–	–	<i>Nessun riferimento VICON disponibile</i>
S3	Caviglia	SX*	+41.08	+94.91 (n=1)	53.83 (n=1)	<i>MP sottostima; caviglia molto post. in S3</i>

EA \leq 10 mm Ottimo	EA 10-30 mm Accettabile	EA > 30 mm Critico	Accordo eccellente	Dato mancante
---------------------------	----------------------------	-----------------------	--------------------	---------------

Tabella 5.4 - Deviazioni dalla plumb line virtuale sagittale per segmento corporeo e lato. Valori positivi = spostamento anteriore rispetto alla proiezione verticale dell'orecchio; valori negativi = spostamento posteriore.

* = lato corrispondente all'arto protesico. Per S3 lato DX, il riferimento VICON non era disponibile per la maggior parte dei segmenti a causa dell'assenza del marker sul segmento protesico corrispondente.

5.4.2. Angoli articolari sagittali

La Tabella 5.5 riporta i valori medi degli angoli articolari sagittali (anca, ginocchio, caviglia) per ciascun soggetto e lato, confrontati con entrambe le stime VICON: la proiezione geometrica sul piano sagittale (VICON Traj) e il Model Output del Plug-in Gait (VICON MO). Per varismo/valgismo, le due stime VICON differiscono e il confronto con entrambe consente di identificare quale riferimento sia più appropriato per ciascuna metrica. La codifica cromatica degli errori assoluti (verde $\leq 5^\circ$, giallo 5–10°, rosso $> 10^\circ$) è applicata indipendentemente alle due colonne EA, evidenziando i casi in cui il sistema beneficia dalla scelta del riferimento VICON MO rispetto al Traj.



Soggetto	Articolazione	Lato	MP (°)	VICON Traj (°)	VICON MO (°)	EA vs Traj (°)	EA vs MO (°)	Note cliniche
Anca Range clinico ROGA: 0°–10° (estensione)								
S1	Anca	DX*	12.09	19.69	19.69	7.60	7.60	<i>Sottostima; in range 0–10°</i>
S1	Anca	SX	8.90	11.18	11.18	2.28	2.28	<i>In range 0–10°</i>
S2	Anca	DX	4.51	6.83	6.83	2.31	2.31	<i>In range 0–10°</i>
S2	Anca	SX*	6.84	12.81	12.81	5.97	5.97	<i>MO fuori range; MP in range</i>
S3	Anca	DX	8.39	2.82	2.82	5.56	5.56	<i>Lato protesico; in range</i>
S3	Anca	SX*	35.92	-2.42	-2.42	38.34	38.34	<i>Anomalia protesi transfem.</i>
Ginocchio Range clinico ROGA: 0° (estensione completa)								
S1	Ginocchio	DX*	9.29	19.28	21.29	9.98	12.00	<i>Flessione residua; sottostima MP</i>
S1	Ginocchio	SX	2.85	9.29	5.87	6.44	3.02	<i>MP vicino a MO</i>
S2	Ginocchio	DX	7.38	15.59	6.19	8.21	1.19	<i>Ottimo accordo con MO</i>



Soggetto	Articolazione	Lato	MP (°)	VICON Traj (°)	VICON MO (°)	EA vs Traj (°)	EA vs MO (°)	Note cliniche
S2	Ginocchio	SX*	10.59	17.10	13.79	6.51	3.20	Accordo migliore con MO
S3	Ginocchio	DX	6.08	4.13	0.81	1.95	5.27	Protesi in estensione completa
S3	Ginocchio	SX*	5.48	-5.79	-1.97	11.27	7.45	VICON rileva l'estensione, MP non la rileva
Caviglia Range clinico ROGA: 85°-90° (lieve dorsiflessione)								
S1	Caviglia	DX*	90.09	83.62	82.62	6.47	7.47	Range 85-90°; MP in range
S1	Caviglia	SX	88.04	87.55	89.66	0.49	1.62	Eccellente; in range
S2	Caviglia	DX	91.06	77.04	81.96	14.02	9.10	Protesi: MP sovrastima
S2	Caviglia	SX*	81.47	83.10	83.60	1.62	2.13	Lieve deficit; in range
S3	Caviglia	DX	87.50	82.29	87.94	5.22	0.44	Ottimo vs MO; protesi allineata
S3	Caviglia	SX*	83.88	85.73	84.78	1.85	0.90	In range; ottimo



EA \leq 5° Ottimo	EA 5–10° Accettabile	EA > 10° Critico	In range clinico	Anomalia protesi
---------------------------------------	---------------------------------	--------------------------------	-------------------------	-------------------------

Tabella 5.5 – Angoli articolari sagittali: confronto MediaPipe vs VICON Traiettorie e VICON Model Outputs. I valori MP e VICON riportati sono medie aggregate sui trial disponibili (T02-T04 per S1 e S2; solo T02 per S3).

Il range clinico di riferimento è definito dal protocollo ROGA. * = lato corrispondente all'arto protesico.

Si riportano di seguito i landmark restituiti da MP per S1, S2 ed S3 al fine di facilitare l'interpretazione dei risultati. Si consiglia inoltre, la visione parallela degli script Python nel repository di GitHub, disponibile su richiesta (<https://github.com/salvatorerapisardaai-max/Validation-AI-Approach>”).

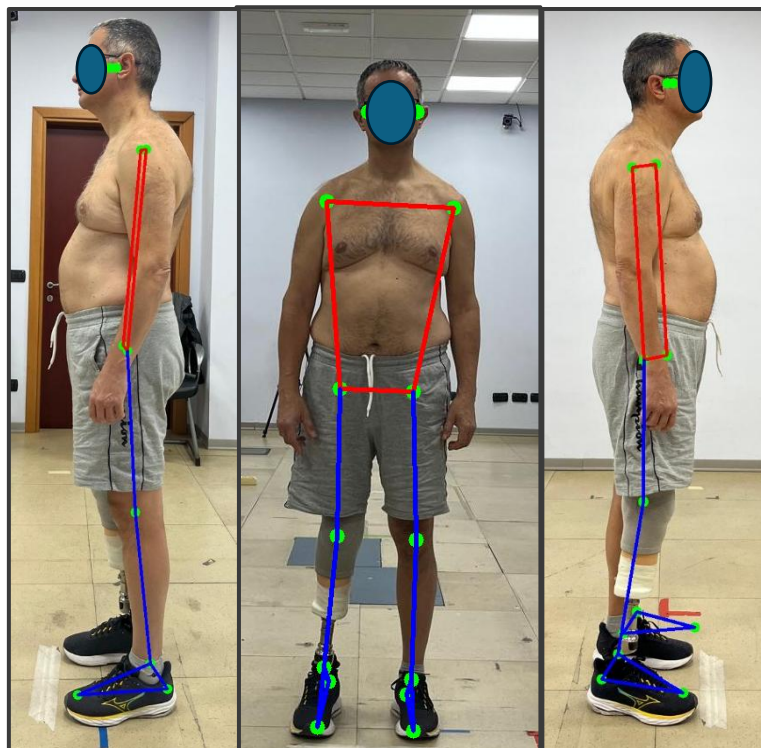


Figura 6.1 - Landmark restituiti da MP | S1

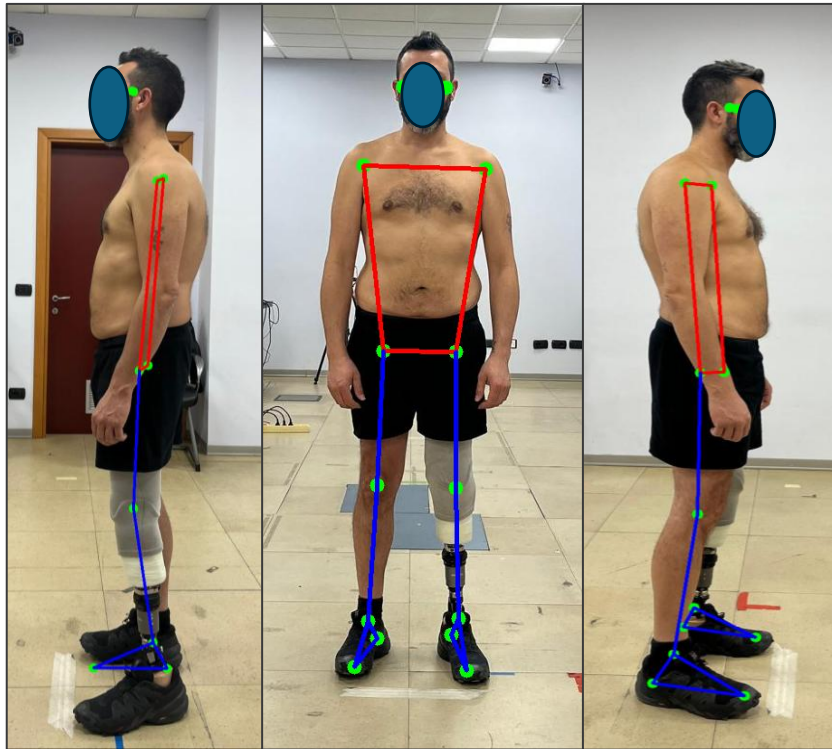


Figura 6.2 - Landmark restituiti da MP | S2



Figura 6.3 - Landmark restituiti da MP | S3

6. Discussione

6.1. Considerazioni generali sulla popolazione

La sperimentazione è stata condotta su un campione di tre soggetti con amputazione di arto inferiore, due transtibiali e un transfemorale, selezionati come popolazione clinicamente rappresentativa degli obiettivi dello studio.

Nell'ambito della ricerca protesica, la disponibilità di soggetti amputati per protocolli sperimentali che richiedano simultaneamente l'acquisizione con sistemi optoelettronici e la stima *markerless* è strutturalmente limitata. Le ragioni sono molteplici: la necessità di un laboratorio di analisi del movimento certificato e operativo, la complessità procedurale dell'applicazione dei marker anatomici su morfologie non standard, e il tempo di sessione richiesto per ciascun partecipante. Campioni di tre soggetti sono pertanto ricorrenti nella letteratura di validazione preliminare in questo dominio specifico e non costituiscono un'anomalia metodologica: ne è testimonianza il lavoro discusso di Zhou *et al.*, che valida un sistema di stima della posa basato su modelli di diffusione generativi su portatori di protesi di arto inferiore impiegando tre soggetti.

L'obiettivo primario del presente lavoro non è la generalizzazione statistica dei risultati all'intera popolazione di amputati, bensì la valutazione della robustezza metodologica del sistema proposto e l'identificazione di pattern sistematici di errore nel confronto tra le metriche estratte dal sistema *markerless* e quelle ottenute mediante *motion capture* optoelettronico.

Un aspetto di rilievo riguarda l'eterogeneità intenzionale dei soggetti inclusi. I tre partecipanti presentano caratteristiche antropometriche differenti, quali statura, massa corporea, proporzioni segmentali, e configurazioni protesiche non sovrapponibili per livello di amputazione, tipologia di componenti e morfologia esterna del dispositivo. Questa variabilità non è un limite del disegno sperimentale: è una scelta voluta. Permette di testare il comportamento dell'algoritmo in presenza di morfologie corporee e geometrie protesiche diverse, mantenendo invariate le formulazioni geometriche per il calcolo delle metriche posturali. Il metodo è così sottoposto a condizioni che riflettono la variabilità tipica della pratica clinica reale, anziché essere valutato in un contesto artificialmente omogeneo.

Per il soggetto S3, la disponibilità di una sola acquisizione valida (Cal_02), dovuta a un malfunzionamento tecnico del sistema nella sessione primaria, richiede una lettura prudente dei dati corrispondenti. In particolare, i marker della coscia e dell'anca non vengono ricostruiti correttamente. In assenza di repliche intrasessione, alcune anomalie non possono essere



attribuite con certezza alla variabilità posturale del soggetto o a ad imprecisioni nel posizionamento dei marker anatomici. Nel sistema VICON, differenze anche minime nella collocazione dei marker, dell'ordine di pochi millimetri, possono introdurre bias sistematici nelle misure cinematiche, per via della propagazione dell'errore nelle trasformazioni rigide che definiscono i sistemi di riferimento segmentali nel modello *Plug-in Gait*. Nonostante questa limitazione, i dati di S3 conservano un valore informativo rilevante: il soggetto presenta un livello di amputazione più prossimale e una configurazione protesica morfologicamente più distante dall'anatomia standard rispetto agli altri due partecipanti, rappresentando una condizione limite, ma clinicamente tutt'altro che rara, utile per esplorare i margini di applicabilità del metodo proposto.

6.2. Angolo asse femore-tibia nel piano frontale (valgismo/varismo)

L'angolo femore-tibia nel piano frontale rappresenta la metrica angolare frontale con le prestazioni più solide del sistema. Il **lato sinistro** mostra un EA medio (globale) di $1.67^\circ \pm 0.90^\circ$, un risultato clinicamente eccellente che si posiziona ampiamente al di sotto della soglia di significatività clinica di 5° . Il caso migliore è S1 lato sinistro (EA = 0.64°): MP stima 0.18° contro un valore VICON Traj. di 0.82° , mostrando un allineamento quasi perfetto.

Il **lato destro** presenta un EA medio di $4.04^\circ \pm 2.52^\circ$, ancora compatibile con la soglia clinica per S2 e S3. L'errore massimo si registra per S1 lato destro (EA = 6.73°): MediaPipe stima -7.99° mentre VICON Traj restituisce -1.27° . Questo lato corrisponde all'arto protesico di S1, e la discrepanza è coerente con la difficoltà del modello nel localizzare correttamente i keypoints ANKLE e KNEE su una superficie protesica di morfologia diversa da quella biologica.

I valori VICON MO per questa metrica mostrano differenze sistematiche rispetto alle traiettorie 3D (es. S2 SX: MO = 10.75° vs Traj = 5.85°), confermando che la convenzione JCS *del Plug-in Gait* incorpora correzioni anatomiche, inclusi l'angolo di torsione tibiale e l'inclinazione dell'asse meccanico del femore, non presenti nel calcolo geometrico diretto su proiezione frontale. Per la validazione clinica, solo per tale metrica, il confronto con **VICON Traj** rimane il riferimento scelto.

6.3. Metriche lineari frontali

6.3.1. Allineamento rotula-II metatarso

La metrica rotula-II metatarso misura il disallineamento lineare frontale tra la proiezione della rotula e quella della testa del II metatarso. Il valore ideale è 0 mm, ove si ha il perfetto



allineamento verticale. Per il **lato destro** l'EA medio è **33.06 mm ± 23.92 mm**: S1 e S2 presentano errori sistematici di circa 41-51 mm, attribuibili alla tendenza di MediaPipe a spostare il *keypoint* KNEE medialmente rispetto alla posizione anatomica della rotula. S3 lato destro mostra invece un EA eccellente di 6.01 mm (MP = 53.15 mm, VICON = 59.16 mm).

Il **lato sinistro** presenta un **EA medio di 62.85 mm ± 49.09 mm** con variabilità molto elevata. S2 mostra un errore più contenuto (29.33 mm), mentre S3 registra il valore anomalo di 167.83 mm: VICON Traj riporta -176.30 mm (ossia la proiezione del II metatarso è quasi 176 mm a sinistra della rotula), un valore inusualmente elevato attribuibile alla morfologia protesica di S3.

6.3.2. Deviazione della linea alba

La deviazione della linea alba è la metrica lineare frontale con le prestazioni più omogenee per S1 e S2. S1 mostra un EA medio di 2.05 mm e S2 di 5.64 mm, entrambi clinicamente trascurabili e compatibili con l'incertezza di posizionamento dei marker stessi, tipicamente di circa 5-10 mm. MediaPipe individua la linea alba come segmento tra il punto medio delle spalle e il punto medio delle anche, il che corrisponde a una buona approssimazione del vettore assiale del tronco nella proiezione frontale.

S3 presenta nuovamente un valore anomalo: VICON Traj restituisce -109.25 mm mentre MediaPipe stima 31.16 mm, con EA = 140.41 mm. Il valore VICON negativo di tale entità (oltre 10 cm di deviazione laterale della linea alba) non è coerente con la stazione ortostatica di un soggetto amputato e suggerisce fortemente un errore di ricostruzione 3D dei marker nell'unica Cal_02 di S3. Escludendo il terzo soggetto, la metrica mostra una performance eccellente (**EA medio = 3.84 mm**).

6.3.3. Tilt delle scapole

Il tilt delle scapole, ossia l'inclinazione della linea che congiunge gli apici delle scapole rispetto all'orizzontale, presenta un **EA medio globale di 5.75 mm ± 3.86 mm**. S1 ottiene il valore migliore (EA = 2.47 mm), S2 un errore accettabile (7.32 mm), e S3 un errore di 10.86 mm. Per S3, il valore VICON è negativo (-8.94 mm) mentre MediaPipe stima +1.92 mm. Complessivamente, l'errore rientra nell'incertezza strumentale della misura VICON stessa.

6.4. Plumb line virtuale sagittale

Le deviazioni dalla plumb line sagittale misurano lo spostamento antero-posteriore di ciascun punto anatomico (spalla, gran trocantere, ginocchio, caviglia) rispetto alla proiezione verticale del trago. Valori positivi indicano spostamento anteriore, negativi posteriore. La qualità del



confronto è limitata per S2 e S3 lato destro dall'assenza del riferimento VICON su alcuni segmenti, dimostrando, un'errata ricostruzione tridimensionale dovuta ad un errore tecnico.

da spalla: la spalla destra mostra un EA eccellente per S1 (0.76 mm), con MediaPipe che stima 86.39 mm vs VICON 85.63 mm. S2 ha un errore di 13.92 mm (disponibile solo per T04). La spalla sinistra presenta errori più elevati: S1 ha EA = 34.19 mm (MP = 73.80 mm vs VICON = 39.61 mm), con MediaPipe che sovrastima sistematicamente la proiezione anteriore della spalla. S3 lato sinistro presenta EA = 57.40 mm, con VICON che registra 155.23 mm. La stima MP di 97.83 mm è sottostimata rispetto al riferimento.

gran trocantere: le deviazioni mostrano errori medi di 34.15 mm (DX) e 41.91 mm (SX), con grande variabilità tra soggetti. Un dato di rilievo è che per S1 e S2, il valore VICON Traj dell'anca sinistra è negativo (rispettivamente -1.10 mm e -22.91 mm), indicando che in questi soggetti il trocantere è leggermente posteriore alla plumb line. MediaPipe stima invece valori positivi (27.59 mm e 15.11 mm): una differenza sistematica che suggerisce come il *keypoint* HIP di MP, in vista laterale, localizzi un punto anatomicamente diverso (più anteriore) rispetto al gran trocantere ricostruito da VICON (LTHI). Questo bias sistematico, coerente tra S1 e S2, rappresenta un'informazione preziosa per future calibrazioni geometriche, suggerendo la necessità di un set completo di offset soggetto-specifici.

ginocchio: il ginocchio sinistro di S1 ottiene EA = 4.56 mm, ma S2 e S3 mostrano errori elevati (34–79 mm). Per S2, il valore VICON negativo del ginocchio sinistro (-22 a -50 mm nei tre trial) indica un ginocchio posteriore alla plumb line, mentre MP stima 0.04 mm: la discrepanza è attribuibile a una difficoltà del modello nell'individuare il ginocchio quando il soggetto mantiene una leggera flessione compensatoria sul lato sano. Il ginocchio destro mostra errori sistematicamente elevati per S1 (50.53 mm), con MP che sovrastima la proiezione anteriore rispetto a VICON (58.33 mm vs 7.80 mm).

caviglia: la caviglia mostra errori medi di 23.79 mm (DX) e 19.89 mm (SX), i più contenuti tra le metriche della plumb line. Infatti, le posizioni dei marker nei due sistemi, VICON e MP, coincidono quasi perfettamente. Per S2 caviglia sinistra l'EA è 10.13 mm, per S1 destra 25.34 mm. Per S3 caviglia sinistra l'EA è 53.83 mm, con VICON che registra 94.91 mm: la grande distanza posteriore dell'articolazione tibio-tarsica dal plumb è coerente con una postura compensatoria presente in S3, ma MP (41.08 mm) non cattura completamente questo spostamento.



6.5. Angoli articolari sagittali

6.5.1. Angolo di estensione dell'anca

L'anca presenta un **EA medio di $5.16^\circ \pm 2.67^\circ$ (DX) e $15.01^\circ \pm 19.84^\circ$ (SX)**. Il range clinico di riferimento ROGA è 0° - 10° in stazione ortostatica. L'enorme variabilità del lato sinistro è quasi interamente determinata da S3 (EA = 38.34°): MediaPipe stima $+35.92^\circ$ per l'anca sinistra di S3, mentre VICON MO restituisce -2.42° (un valore di lieve iperestensione). Questo caso rappresenta la limitazione più evidente del sistema per gli arti protesici prossimali: la protesi transfemorale di S3 posiziona il *keypoint* HIP in una sede anatomica non corrispondente a quella in VICON, inducendo un errore angolare sistematico. Per S1 e S2, escludendo S3, l'EA medio dell'anca sinistra è 4.12° , comparabile con il lato destro.

Tutti i valori MediaPipe per l'anca, eccetto per il lato protesico di S3, rientrano nel range fisiologico 0° - 10° , confermando che il sistema è in grado di discriminare correttamente la postura eretta dalla postura in flessione d'anca. S1 DX mostra il bias sistematico più elevato nel lato sano: MP stima 12.09° vs VICON MO 19.69° (EA = 7.60°), indicando una sottostima dell'estensione d'anca. Questo è attribuibile alla proiezione 2D del vettore femorale in immagini con prospettiva laterale: una leggera rotazione del soggetto rispetto al piano sagittale puro introduce un errore di proiezione.

6.5.2. Angolo di estensione del ginocchio

Il ginocchio mostra un **EA medio di $6.15^\circ \pm 5.46^\circ$ (DX) e $4.56^\circ \pm 2.51^\circ$ (SX)**. Il valore clinico di riferimento è 0° (estensione completa): tutti i soggetti mostrano una leggera flessione residua in stazione ortostatica, maggiore per S1 e S2 che per S3. S1 lato destro presenta l'EA massimo: MP stima 12.0° di flessione vs VICON Traj. 19.28° . La sottostima della flessione del ginocchio da parte di MediaPipe è un fenomeno ricorrente nella letteratura *markerless* e riflette la difficoltà del modello nel discriminare la posizione del marker *knee* nella vista laterale. S2 lato destro (sano) ottiene invece l'EA migliore (1.19°), con MP a 7.38° vs VICON MO a 6.19° .

6.5.3. Angolo di dorsi-flessione della caviglia

Per la caviglia, infine, si ottengono ottime prestazioni. Il **lato sinistro** ottiene un **EA medio di soli $1.54^\circ \pm 0.62^\circ$** : un risultato di precisione clinica, con tutti e tre i soggetti che mostrano errori inferiori a 2° . I valori MediaPipe (81.47° - 88.04°) e VICON MO (83.60° - 89.66°) si collocano entrambi nel range fisiologico 85° - 90° , confermando che il sistema coglie correttamente l'angolo in esame sul lato sano per S1 e su quello protesico per S2 ed S3. Il **lato destro** presenta un **EA medio di $5.67^\circ \pm 4.60^\circ$** . S2 lato destro mostra l'errore massimo (9.10°): MP stima 91.06°



mentre VICON MO è a 81.96° . Nel complesso, l'angolo di dorsi-flessione della caviglia è catturato correttamente dal modello, sovrastimandone leggermente il valore.

6.6. Analisi delle criticità sistematiche e prospettive di miglioramento

L'analisi complessiva dei risultati consente di identificare tre categorie di criticità sistematiche:

Errori da morfologia protesica. MediaPipe localizza i *keypoints* sulle superfici visibili della protesi, le cui geometrie differiscono sostanzialmente dall'anatomia biologica. Questo introduce bias sistematici in tutte le metriche che coinvolgono l'arto protesico, con entità variabile in funzione del tipo di protesi e del suo allineamento in stazione ortostatica. Il fenomeno è atteso e documentato in letteratura [6], ed è superabile con dataset di addestramento specifici per portatori di protesi, o introducendo offset soggetto-specifici da applicare alle metriche MP al fine di approssimare meglio i landmark di MP con i marker di VICON.

Errori proiezione 2D. Metriche che dipendono dalla proiezione sagittale di vettori articolari (anca, ginocchio) sono soggette a errori sistematici quando il soggetto non è perfettamente orientato nel piano sagittale. Piccole rotazioni assiali (5° - 10°) del soggetto rispetto alla telecamera introducono errori di proiezione geometrica non controllabili in assenza di informazioni 3D. Questo limite, intrinseco all'approccio monoculare, è parzialmente mitigabile con un'attenta standardizzazione del protocollo di acquisizione.

Anomalie nei dati VICON di riferimento per S3. Diversi valori VICON Traj per S3 (linea alba = -109.25 mm, rotula SX = -176.30 mm) sono incoerenti con la stazione ortostatica e suggeriscono errori di ricostruzione 3D nella singola acquisizione disponibile. In assenza di repliche, non è possibile verificare se questi valori rappresentino la postura reale del soggetto o artefatti strumentali. Questa fonte di incertezza rende l'interpretazione degli EA per S3 sulle metriche lineari frontali necessariamente cauta.

In prospettiva, le strategie di miglioramento più promettenti includono: il *fine-tuning* di opportuni modelli IA (cfr. 1.6) su dataset di soggetti con protesi, per addestrare il modello a localizzare i *keypoints* anatomici corretti indipendentemente dalla superficie protesica; l'introduzione di un step di calibrazione paziente-specifica per stimare i bias sistematici delle metriche lineari e angolari e correggerli in post-processing; lo sviluppo di un indice di fiducia per i *keypoints* che consenta di segnalare automaticamente al clinico i segmenti anatomici con bassa affidabilità di stima.



7. Conclusioni

7.1. Considerazioni

La sperimentazione è stata condotta su un campione di tre soggetti con amputazione di arto inferiore, con configurazioni protesiche eterogenee, presso il Laboratorio di Analisi del Movimento del Politecnico di Torino. Il dataset comprende 140 coppie di misure distribuite su 20 metriche posturali, analizzate separatamente per vista (frontale e sagittale), per lato (destro e sinistro) e per sistema di riferimento VICON (traiettorie 3D e *Model Outputs* del *Plug-in Gait*).

Nel complesso, i risultati indicano che il sistema MediaPipe è in grado di fornire stime posturali clinicamente significative per le metriche angolari della vista frontale (femore-tibia: EA medio 1.5°-4.7°) e per l'angolo di dorsi-flessione della caviglia (EA medio 1.17° sul lato sinistro), raggiungendo livelli di accuratezza comparabili con quelli riportati in letteratura per sistemi *markerless* di validazione clinica su soggetti normodotati [25][26]. Le metriche lineari frontali di tilt delle scapole (EA medio 5.75 mm) e deviazione della linea alba per S1 e S2 (EA medio ~3.8 mm) mostrano prestazioni compatibili con l'incertezza clinica accettabile stabilita.

Le maggiori limitazioni persistono nelle metriche della plumb line sagittale (EA medio 20-44 mm), nell'angolo di estensione dell'anca (EA medio ~5°-9°, con un *outlier* a 38° per S3), e nell'angolo di ginocchio (EA medio ~7-8°). Tali valori, pur non raggiungendo la precisione di VICON, possono comunque fornire indicazioni qualitative utili al fisioterapista nell'ambito di un sistema di telemedicina, dove l'obiettivo primario è il rilevamento di asimmetrie grossolane e la verifica dell'andamento della riabilitazione nel tempo piuttosto che la misurazione cinematica di precisione.

Il compromesso tra accessibilità e accuratezza, inevitabile rispetto a sistemi come VICON, appare giustificato dall'abbattimento delle barriere cliniche e logistiche che limitano la continuità riabilitativa oltre i confini del laboratorio. Il sistema proposto dimostra la fattibilità di un approccio *markerless* basato su immagini di smartphone per la valutazione posturale in ambito protesico, con un potenziale applicativo concreto che potrà essere ulteriormente ampliato attraverso le strategie di miglioramento identificate.

7.2. Il contributo applicativo

Il prototipo applicativo sviluppato nel Capitolo 4 trascende il perimetro della validazione accademica per collocarsi in uno spazio applicativo concreto. L'architettura *client-server*, dal backend FastAPI alla generazione automatica del report PDF, costituisce una dimostrazione



end-to-end di un flusso clinico realmente percorribile: il fisioterapista acquisisce tre fotografie dallo smartphone, le carica sull'applicazione, e riceve in pochi secondi un report strutturato con le metriche posturali calcolate automaticamente. L'introduzione del fattore di scala automatico basato sull'altezza del paziente elimina la necessità di oggetti di riferimento esterni, rendendo il sistema compatibile con l'uso domiciliare. Sebbene il prototipo richieda un'ulteriore fase di validazione prospettica in condizioni non controllate, rappresenta un precedente tecnico chiaro e replicabile.

7.3. Interpretazione clinica e limiti

L'analisi complessiva dello studio condotto consente di distinguere chiaramente due tipologie di limitazioni.

Limiti strutturali: questi sono intrinseci all'approccio monoculare 2D. MediaPipe localizza i *keypoints* sulla superficie visibile del corpo senza informazioni di profondità affidabili. Inoltre, piccole rotazioni del soggetto rispetto al piano di acquisizione generano errori di proiezione geometrici. O ancora, la morfologia protesica, assente nei dataset di addestramento, introduce bias sistematici nella localizzazione dei *keypoints* distali. Infine, l'assenza di misure 3D impedisce la stima di metriche come la rotazione interna/esterna dell'anca. Tali limitazioni sono comuni a qualsiasi sistema monoculare 2D e non eliminabili senza opportune modifiche architettureali.

Criticità: la principale criticità è la mancanza di un dataset specifico per soggetti con protesi. Il fine-tuning di MP, o di altri modelli *markerless* (cfr. 1.6.) su immagini di amputati, accessibile tramite strumenti come *Model Maker* (cfr. 2.3.3.), rappresenta la strategia più promettente per superare questa limitazione. Tale approccio, insieme all'introduzione di calibrazioni paziente-specifiche sarebbe in grado di compensare i bias sistematici incontrati.

Inoltre, tutti e tre i soggetti arruolati sono di sesso maschile. Questa uniformità, determinata dalla disponibilità dei pazienti reclutati tramite ROGA, non consente di valutare se le prestazioni del sistema varino in funzione del sesso, che può influenzare sia la morfologia corporea e protesica sia i pattern posturali compensatori. Studi futuri dovranno includere soggetti femminili per verificare il metodo proposto.

In conclusione, il presente lavoro dimostra che la valutazione posturale *markerless* basata su immagini ottenute da smartphone non è un'aspirazione futuristica, bensì una realtà tecnica conseguibile con gli strumenti disponibili oggi. Eliminare la necessità di laboratori specializzati e sistemi optoelettronici da centinaia di migliaia di euro significa restituire





continuità riabilitativa a quei pazienti, con mobilità limitata, che nella frequenza del controllo posturale trovano il presupposto del successo della propria protesi. È proprio in questa discontinuità logistica e clinica che si colloca la lacuna che il presente lavoro mira a colmare, offrendo un contributo concreto e innovativo alla riabilitazione protesica al fine di migliorare la continuità dei controlli clinici e, di conseguenza, la qualità della vita dei pazienti.



Appendice A | Protocollo di acquisizione

Si riporta un esempio del protocollo di acquisizione revisionato dal Prof. Taian Martins, relatore del progetto, e comprovato dall'azienda collaboratrice ROGA SpA. Il documento è riportato integralmente nella forma originale (Figure A.1-A.3).



Protocollo sperimentale di acquisizione | MediaPipe e VICON

Lo scopo del seguente protocollo è descrivere gli strumenti e le metodologie per l'acquisizione di immagini bidimensionali di soggetti con protesi di arto inferiore, al fine di rilevare correttamente i *keypoints* restituiti dal modello *markerless* e *marker-based*. Successivamente, verranno calcolate le metriche MediaPipe da confrontare con quelle calcolate tramite VICON.

Considerazioni etiche

- Ottenere il consenso informato dai partecipanti prima dell'acquisizione dei dati.
- Garantire l'anonimato dei dati raccolti, memorizzando le acquisizioni in un database sicuro.
- Rispettare le normative GDPR per la gestione dei dati personali.

Materiali

- **Fotocamera:** fotocamera digitale ad alta risoluzione (consigliata superiore a 720p).
- **Mediapipe:** libreria IA per il rilevamento dei *keypoints*.
- **Riferimento fisico:** oggetto di dimensioni note per la conversione da *pixel* a *cm* (es. bottiglietta alta 20 cm).
- **Sistema opto-elettronico** Vicon gestito dal software *Vicon Nexus 2.17*.
- **Marker semiriflettenti sferici** (diametro 14 mm).
- **Metro e Calibro.**
- **Illuminazione:** si richiede un ambiente illuminato uniformemente per una visualizzazione chiara del soggetto.
- **Spazio sperimentale:** area di 3x5 metri, priva di ostacoli e soggetti non di interesse, con sfondo e pavimento che risaltino l'individuo fotografato.
- **Abbigliamento:** il paziente deve indossare vestiti che risaltano la silhouette del corpo. Non deve indossare scarpe.

Metodi

ACQUISIZIONE VICON

- **Accensione del sistema VICON (warm-up):** attendere che i sensori arrivino a temperatura.
- Il soggetto si prepara ad indossare vestiti che risaltano la silhouette del corpo e si toglie le scarpe.
- **Raccolta dati antropometrici**
 - Altezza (mm) e peso (Kg)
 - Distanza tra i due ASIS (mm)
 - Distanza, lungo la direzione cranio-caudale, tra l'ASIS e il grande trocantere, proiettata sul piano sagittale del soggetto in posizione supina (mm)

Figura A.1 - Protocollo operativo (pagina 1)





- Spessore del ginocchio lungo l'asse medio-laterale (mm)
 - Distanza medio-laterale tra i malleoli (mm)
 - Componente cranio-caudale della distanza tra il centro articolare gleno-omerale e la base del marker acromion-clavicolare proiettata sul piano sagittale (mm)
 - Lunghezza della gamba, definita come la distanza da ASIS al malleolo mediale, passando per l'articolazione del ginocchio (mm)
 - Distanza tra ombelico e la sua proiezione sul pavimento (mm)
 - **Disporre i 23 marker** utili all'analisi seguendo il *Plug-in Gait Full Body* (ASIS, PSIS, LTHI, RTHI, LKNEE, RKNEE, LTIB, RITB, LANK, RANK, LHEEL, RHEEL, LTOE, RTOE, RBAK, CLAV, STRN, LSHO, RSHO). Aggiungere due marker sull'orecchio sinistro e destro, LEAR e REAR, circa in corrispondenza del centro dell'articolazione tempomandibolare e in prossimità del trago.
 - La fase di warm up è terminata.
 - **Calibrazione videocamere VICON:** eseguire movimenti oscillatori dell'asta a T (*wand*) nel volume di prelievo interessato. Bisogna eseguire tali movimenti orientando l'asta verso le telecamere e fino a quando i LED di ogni camera non diventano di colore verde.
 - **Mascherare le camere** ove è presente rumore.
 - **Posizionare l'asta** in corrispondenza del centro dello spazio sperimentale (ove è presente la pedana di forza) per definire il **sistema di riferimento globale**.
 - **Creare** un nuovo soggetto in Nexus (*Create new Patient Classification -> Create new Patient -> Create new Session*).
 - Inserire le misure antropometriche precedentemente raccolte in *Properties*.
 - Effettuare la **calibrazione statica (*Process Static Plug In Gait Model*)** con il soggetto in posa da motociclista.
 - Eseguire il **Process Dynamic Plug-In Gait Model**. Si acquisisce per circa 1 secondo il soggetto in posizione ortostatica (in piedi con braccia rilassate lungo i fianchi durante tutto l'arco dell'acquisizione). Per ogni soggetto esaminato, si eseguono tre acquisizioni.
 - Per ciascun trial si effettua il salvataggio nei formati **C3D** e **VSK** (*Save Trial – C3D + VSK*), così da conservare sia le traiettorie marker che il modello scheletrico del soggetto. Inoltre, nella pipeline di post-processing si include l'opzione **Export ASCII**, selezionando esclusivamente:
 - **3D Marker Trajectories**, contenenti le coordinate tridimensionali dei marker;
 - **Model Outputs**, comprendenti le variabili cinematiche/articolari di interesse.
- Ogni prova viene infine rinominata in modo univoco, così da garantirne la tracciabilità durante le fasi successive di analisi.

Figura A.2 - Protocollo operativo (pagina 2)





ACQUISIZIONE MEDIAPIPE

- **Posizione fotocamera:** su un treppiede, con la lente principale posizionata alla stessa altezza dell'ombelico del soggetto rispetto al pavimento. La lente è orientata e parallela al piano in esame (frontale o sagittale), ad una distanza di circa 2.5-3 metri dal soggetto considerando l'utilizzo di uno smartphone dotato di fotocamera con lunghezza focale 26 mm (distanza camera-soggetto=altezza soggetto * fattore di scala).
- **Collocazione oggetto di riferimento:** posizionare l'oggetto scelto sul pavimento a 50 centimetri dal piede sinistro o destro in direzione medio-laterale, controlateralmente all'arto protesico per l'acquisizione frontale. Posizionare l'oggetto sul pavimento a 50 centimetri dai piedi in direzione antero-posteriore per le acquisizioni sagittali.
- **Posizione paziente:** il paziente deve mantenere una posizione ortostatica (in piedi con braccia rilassate lungo i fianchi) durante tutto l'arco dell'acquisizione.
- **Acquisizione:** scattare tre foto; una foto con la camera parallela al piano frontale del soggetto, due foto con la camera parallela al piano sagittale del soggetto (laterale sinistro e destro). Si ripete il processo di acquisizione tre volte per ogni soggetto.

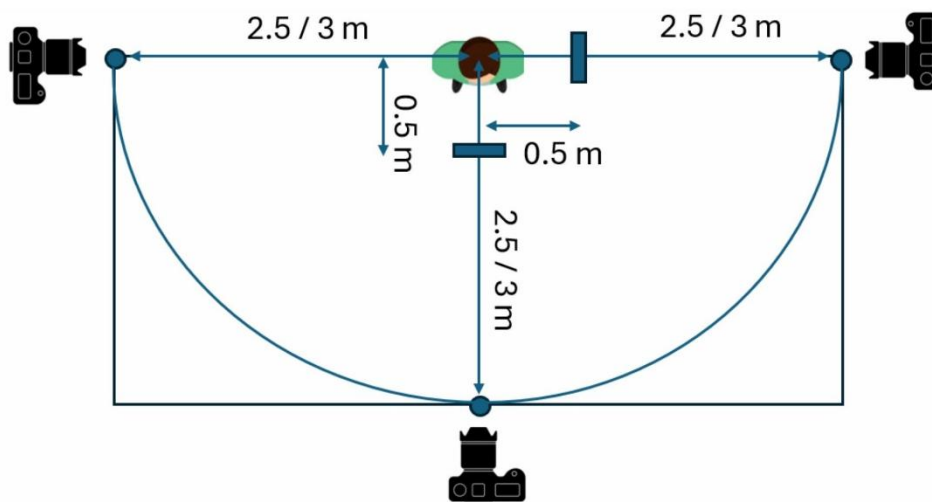


Figura 1 - Setup acquisizione Mediapipe

Figura A.3 - Protocollo operativo (pagina 3)



Appendice B | Report clinico generato dal prototipo applicativo

In questa appendice è riportato un esempio del report clinico generato automaticamente dal sistema descritto nel Capitolo 4. Il documento è presentato a scopo illustrativo per mostrare la struttura dell'output, l'organizzazione delle metriche posturali e l'integrazione delle immagini annotate con i landmark rilevati.

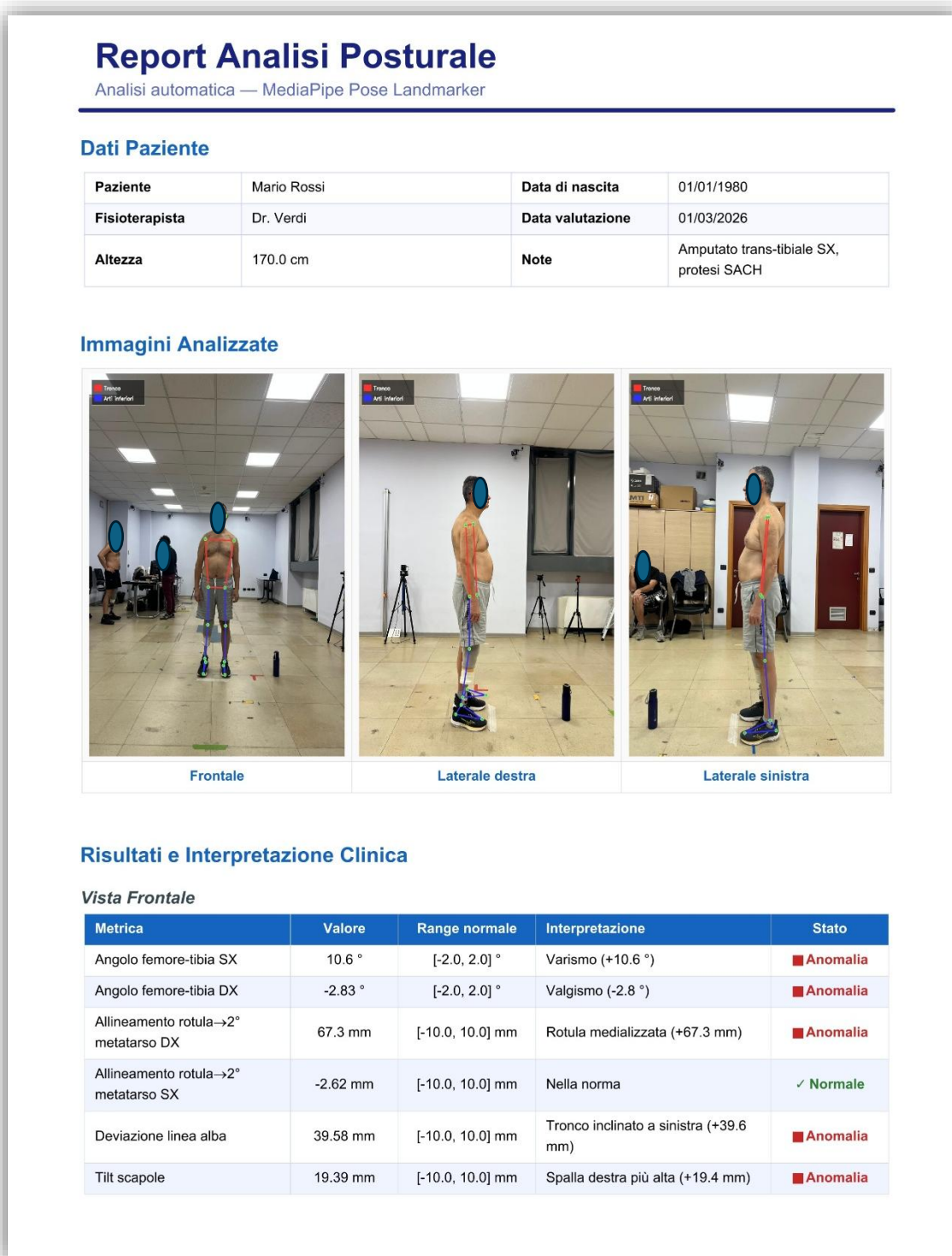


Figura B.1 – Fac-simile del report clinico prodotto dal sistema prototipale di analisi posturale (pagina 1). Il report è stato generato su dati di test a fini dimostrativi e non costituisce valutazione clinica reale. Nello specifico, i range fisiologici sono arbitrari.

Vista Laterale destra

Metrica	Valore	Range normale	Interpretazione	Stato
Plumb line spalla DX	-39.63 mm	[-20.0, 20.0] mm	Spalla post. alla caviglia (-39.6 mm)	■ Anomalia
Plumb line anca DX	-73.83 mm	[-20.0, 20.0] mm	Anca post. alla caviglia (-73.8 mm)	■ Anomalia
Plumb line ginocchio DX	-72.47 mm	[-20.0, 20.0] mm	Ginocchio post. alla caviglia (-72.5 mm)	■ Anomalia
Flessione/estensione anca DX	-0.19 °	[-5.0, 5.0] °	Nella norma	✓ Normale
Flessione/estensione ginocchio DX	2.23 °	[-5.0, 5.0] °	Nella norma	✓ Normale
Angolo caviglia DX	95.62 °	[85.0, 95.0] °	Flessione plantare (+95.6 °)	■ Anomalia

Vista Laterale sinistra

Metrica	Valore	Range normale	Interpretazione	Stato
Plumb line spalla SX	-11.38 mm	[-20.0, 20.0] mm	Nella norma	✓ Normale
Plumb line anca SX	-62.91 mm	[-20.0, 20.0] mm	Anca post. alla caviglia (-62.9 mm)	■ Anomalia
Plumb line ginocchio SX	-41.16 mm	[-20.0, 20.0] mm	Ginocchio post. alla caviglia (-41.2 mm)	■ Anomalia
Flessione/estensione anca SX	-2.82 °	[-5.0, 5.0] °	Nella norma	✓ Normale
Flessione/estensione ginocchio SX	-8.47 °	[-5.0, 5.0] °	Estensione ginocchio (recurvatum) (-8.5 °)	■ Anomalia
Angolo caviglia SX	93.9 °	[85.0, 95.0] °	Nella norma	✓ Normale

Report generato il 28/02/2026 alle 23:59 | Sistema di analisi posturale automatica — MediaPipe Pose Landmarker (heavy)

■ Questo report è uno strumento di supporto clinico. La diagnosi finale è di esclusiva competenza del professionista sanitario abilitato.

Figura B.2 - Fac-simile report clinico (pagina 2).

Bibliografia

- [1] J. Stenum, K. M. Cherry-Allen, C. O. Pyles, R. D. Reetzke, M. F. Vignos, R. T. Roemmich, "Applications of Pose Estimation in Human Health and Performance across the Lifespan", *Sensors*, vol. 21, no. 21, art. 7315, 2021.
- [2] C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines", arXiv:1906.08172, Google Research, 2019.
- [3] R. B. Neupane, K. Li, T. F. Boka, "A survey on deep 3D human pose estimation", *Artificial Intelligence Review*, vol. 58, art. 24, 2025.
- [4] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context", in *Proc. ECCV*, pp. 740–755, 2014.
- [5] M. Andriluka, L. Pishchulin, P. Gehler, B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis", in *Proc. IEEE CVPR*, pp. 3686–3693, 2014.
- [6] T. Zhou, M. N. S. Iskandar, K.-H. Chiam, "Diffusion models enable zero-shot pose estimation for lower-limb prosthetic users", *PLOS Digital Health*, vol. 4, no. 3, art. e0000745, 2025.
- [7] J. Stenum, M. M. Hsu, A. Y. Pantelyat, R. T. Roemmich, "Clinical gait analysis using video-based pose estimation: Multiple perspectives, clinical populations, and measuring change", *PLOS Digital Health*, vol. 3, no. 3, art. e0000467, 2024.
- [8] C. S. T. Hii, K. B. Gan et al., "Automated Gait Analysis Based on a Marker-Free Pose Estimation Model", *Sensors*, vol. 23, no. 13, art. 6489, 2023.
- [9] H.-S. Fang, S. Xie, Y.-W. Tai, C. Lu, "RMPE: Regional Multi-person Pose Estimation", in *Proc. IEEE ICCV*, pp. 2334–2343, 2017.
- [10] J. M. Graving et al., "DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning", *eLife*, vol. 8, art. e47994, 2019.
- [11] R. A. Güler, N. Neverova, I. Kokkinos, "DensePose: Dense Human Pose Estimation In The Wild", in *Proc. IEEE CVPR*, pp. 7297–7306, 2018.
- [12] Wu Y. et al., "Detectron2", GitHub repository, Facebook AI Research, 2019. [Online]. Available: <https://github.com/facebookresearch/detectron2>.
- [13] J. Wang et al., "Deep High-Resolution Representation Learning for Visual Recognition", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, 2020.
- [14] S. Kreiss, L. Bertoni, A. Alahi, "OpenPifPaf: Composite Fields for Semantic Keypoint Detection and Spatio-Temporal Association", *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 13498–13511, 2022.
- [15] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, 2021.
- [16] A. Kendall, M. Grimes, R. Cipolla, "PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization", in *Proc. IEEE ICCV*, pp. 2938–2946, 2015.



- [17] R. Aguilar-Ortega et al., "UCO Physical Rehabilitation: New Dataset and Study of Human Pose Estimation Methods on Physical Rehabilitation Exercises", *Sensors*, vol. 23, no. 21, art. 8862, 2023.
- [18] Bazarevsky V. et al., "BlazePose: On-device Real-time Body Pose tracking", arXiv:2006.10204, 2020. Notare che PoseLandmarker Heavy è un'estensione successiva di BlazePose, documentata nella guida Google MediaPipe (non un paper autonomo).
- [19] A. V. Ruescas-Nicolau et al., "A Deep Learning Model for Markerless Pose Estimation Based on Keypoint Augmentation: What Factors Influence Errors in Biomechanical Applications", *Sensors*, vol. 24, no. 6, art. 1923, 2024.
- [20] A. Mathis et al., "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning," *Nature Neuroscience*, vol. 21, pp. 1281–1289, 2018.
- [21] L. Sigal, A. O. Balan, M. J. Black, "HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion," *International Journal of Computer Vision*, vol. 87, pp. 4–27, 2010.
- [22] C.-M. Schreuder, O. Bergh, L. Steyn, R. P. Theart, "SUMediPose: A 2D-3D pose estimation dataset," *Data in Brief*, vol. 60, art. 111579, 2025.
- [23] Vicon Motion Systems Ltd., Plug-in Gait Reference Guide, Oxford, UK, 2023.
- [24] C. Mercadal-Baudart et al., "Exercise quantification from single camera view markerless 3D pose estimation," *Heliyon*, vol. 10, art. e27596, 2024.
- [25] T. B. G. Lafayette et al., " Validation of Angle Estimation Based on Body Tracking Data from RGB-D and RGB Cameras for Biomechanical Assessment," *Sensors*, vol. 23, no. 1, art. 3, 2023.
- [26] P. Bruno e J. Barden, "Comparison of two alternative technical marker sets for measuring 3D pelvic motion during gait," *Journal of Biomechanics*, vol. 48, no. 13, pp. 3876–3882, 2015.
- [27] Haider, Syed Muhammad Sajjad, et al. "AI for Academic Integrity: GPU-Free Pose Estimation Framework for Automated Invigilation." *Automation* 6.4 (2025): 82.

