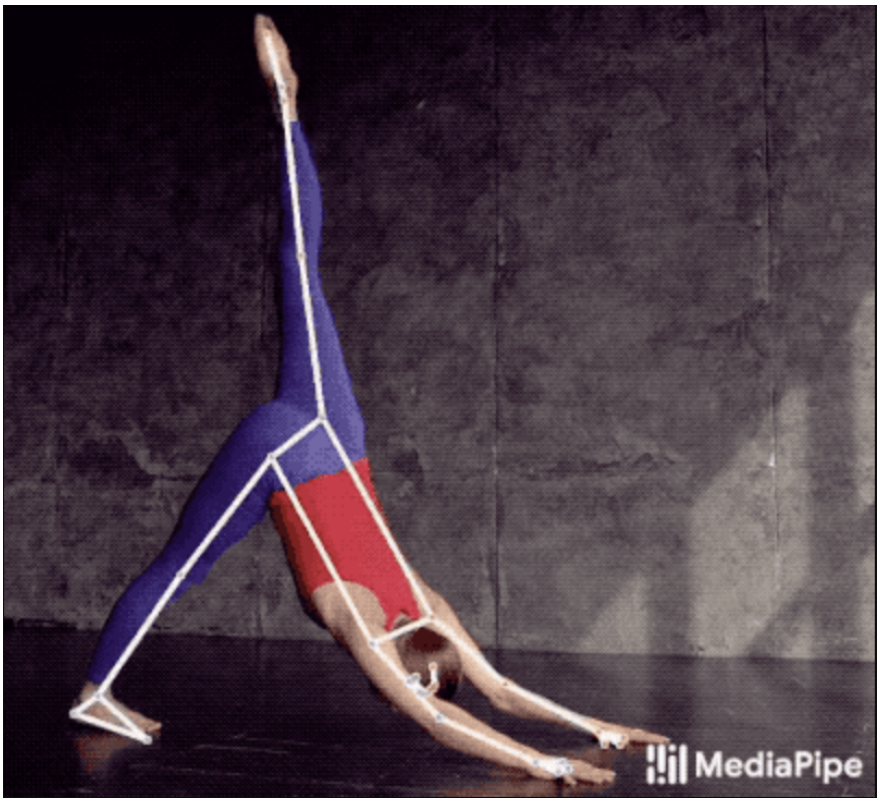
CHALMERS
UNIVERSITY OF TECHNOLOGY



Camera-Based Home Rehabilitation Exercise Monitoring

A Technical Evaluation Against Optical Motion Capture

Master's thesis in Information and Communication Technology, MPICT

TINGTING ZHAO

DEPARTMENT OF ELECTRICAL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Camera-Based Home Rehabilitation Exercise Monitoring

A Technical Evaluation Against Optical Motion Capture

TINGTING ZHAO

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Division of Signal Processing and Biomedical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2026

Camera-Based Home Rehabilitation Exercise Monitoring
A Technical Evaluation Against Optical Motion Capture
TINGTING ZHAO

Supervisor: Xuezhi Zeng, Associate Professor at Division of Signal Processing and Biomedical Engineering, Electrical Engineering, Chalmers
Examiner: Xuezhi Zeng, Associate Professor at Division of Signal Processing and Biomedical Engineering, Electrical Engineering, Chalmers

Master's Thesis 2026
Department of Electrical Engineering
Division of Signal Processing and Biomedical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Example of MediaPipe Pose tracking from RGB video, showing a detected human pose skeleton and segmentation overlay. Adapted from the MediaPipe Pose documentation [1].

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Abstract

Home-based rehabilitation often relies on patients performing prescribed exercises independently, without continuous supervision from a physiotherapist. A digital solution that can monitor the exercise performance and provide feedback to the patients would be valuable to improve the follow-ups and support patient empowerment. Camera-based markerless pose estimation may provide a practical and low-cost way to monitor exercise quality in such settings. This thesis investigates the feasibility of using a single RGB camera and MediaPipe Pose for selected rehabilitation-oriented exercises.

Three exercises were evaluated: single-leg stance, sit-to-stand, and mini-squat. RGB videos were recorded using an iPhone 13, while reference motion data were collected with a Qualisys optical motion capture system. MediaPipe Pose was used to extract body landmarks from the videos, and exercise-specific metrics were computed, including trunk orientation, normalized pelvis displacement, squat depth, movement timing, and knee flexion. To enable comparison with MediaPipe, the three-dimensional motion-capture data were projected onto corresponding frontal or sagittal analysis planes before matched metric definitions were applied.

The results show that the MediaPipe-based pipeline could estimate several selected metrics with small to moderate errors under controlled condition 1 setup which is camera-to-participant at 3 m with normal room lighting. Trunk-orientation and normalized displacement metrics were generally more stable than two-dimensional knee-flexion estimation. The results also showed that performance depended on the exercise, metric definition, participant, and recording setup. Camera distance and lighting affected the metrics differently, and repeatability was generally stronger within the same session than across different days. These findings indicate that MediaPipe-based rehabilitation monitoring should be interpreted at the metric level rather than as a uniformly accurate motion-analysis solution.

A rule-based feedback prototype was also developed to illustrate how pose-derived metrics could be translated into patient-facing feedback and therapist-facing session review outputs. Overall, the findings suggest that MediaPipe Pose can be a useful low-cost component for selected home rehabilitation monitoring tasks, provided that the exercise, camera view, and metric definitions are carefully chosen. Further validation with larger and more realistic datasets is required before clinical or real home deployment.

Keywords: MediaPipe Pose, rehabilitation monitoring, markerless pose estimation, motion capture, exercise assessment.

Acknowledgements

I would like to express my sincere gratitude to my supervisor and examiner, Dr. Xuezhi Zeng, for the guidance, feedback, and support throughout this thesis project. The weekly meetings were very helpful for shaping the direction of the work, improving the evaluation design, and refining the thesis. I am also grateful for the ideas and perspectives shared during the discussions, which helped me understand the project from different angles.

I would also like to thank Yi Zhou for the continuous help with arranging the laboratory sessions and supporting the data collection process. Her assistance made the experimental work much smoother.

I am grateful to Wei Zhang, Weirui Zhao, and Jingcheng Wang for their time and cooperation during the data collection. Without their participation and support, it would not have been possible to complete the evaluation in this thesis.

A very special thanks goes to my boyfriend, Wei Zhang. It was through him that I first learned about Chalmers here in Gothenburg. I never imagined I would one day find myself living and studying in Sweden, and now, these two years of my master's studies have flown by in the blink of an eye. I am deeply grateful for his endless patience and for putting up with all my stress and bad moods over the past two years. His unwavering support was the anchor that helped me push through the tough times and successfully complete this incredible journey abroad.

Finally, I want to express my deepest thanks to my parents for their unconditional support of all my choices. Their belief in me allowed me to pursue this path of self-improvement entirely free of worry. I am also incredibly thankful to my friends, despite the time difference, they were always there to listen, comfort me, and lift my spirits during the times I felt overwhelmed and thought I couldn't keep going.

TINGTING ZHAO, Gothenburg, May 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis, listed in alphabetical order:

2D	Two-Dimensional
3D	Three-Dimensional
AI	Artificial Intelligence
CI	Confidence Interval
CSV	Comma-Separated Values
GHUM	Generative Human Model
JSON	JavaScript Object Notation
LoA	Limits of Agreement
MAE	Mean Absolute Error
MINI	Mini-Squat
MoCap	Motion Capture
MP	MediaPipe
RGB	Red-Green-Blue
RMSE	Root Mean Square Error
SD	Standard Deviation
SLS	Single-Leg Stance
STS	Sit-to-Stand

Nomenclature

Below is the nomenclature of indices, sets, parameters, and variables that have been used throughout this thesis.

Indices

t	Time or timestamp of a video frame or MoCap sample
i	Matched sample or trial-level value index
j, k	Pairwise repeated-recording indices used in the repeatability calculation
q	Landmark, marker, or reference point label
r	Repeated-recording index in the repeatability analysis
s	Body side index, where $s \in \{\mathrm{L}, \mathrm{R}\}$
n	Number of matched samples or trial-level values
m	Number of repeated recordings in a matched repeat group

Sets

Π	Selected projected analysis plane, where Π = front for frontal-plane metrics and Π = sag for sagittal-plane metrics

Parameters

W	Image width in pixels
H	Image height in pixels
f_{MP}	Sampling frequency or frame rate of the MediaPipe video stream
f_{MoCap}	Sampling frequency of the MoCap system

Δ	Difference or change between two quantities

Variables

$(\cdot)^{\mathrm{ref}}$	Quantity derived from the MoCap reference data
$\mathbf{p}_q(t)$	MediaPipe 2D image-plane position of landmark q at time t, before pixel scaling
$\mathbf{p}_q^{\mathrm{img}}(t)$	Pixel-scaled 2D MediaPipe landmark position of landmark q at time t
$\mathbf{p}_q^{\mathrm{ref},3D}(t)$	Original 3D MoCap position of marker or reference point q at time t
$\mathbf{p}_q^{\mathrm{ref},\Pi}(t)$	MoCap reference point q projected onto the selected analysis plane Π
$\mathbf{p}_{\mathrm{pelvis}}(t)$	Pelvis-center position at time t
$\mathbf{p}_{\mathrm{shoulder}}(t)$	Shoulder-center position at time t
$\mathbf{p}_{\mathrm{hip}}(t)$	Hip landmark, marker, or projected hip-side point at time t
$\mathbf{p}_{\mathrm{knee}}(t)$	Knee landmark, marker, or projected knee point at time t
$\mathbf{p}_{\mathrm{ankle}}(t)$	Ankle landmark, marker, or projected ankle point at time t
$x_q(t)$, $y_q(t)$	Normalized MediaPipe image coordinates of landmark q
$x^{\mathrm{img}}(t)$, $y^{\mathrm{img}}(t)$	Pixel-scaled image coordinates
$X_q(t)$, $Y_q(t)$, $Z_q(t)$	3D MoCap coordinates of marker or reference point q
$u_q^{\Pi}(t)$	Selected horizontal coordinate of MoCap point q in the projected analysis plane Π
$u(t)$, $v(t)$	Generic 2D coordinates in a projected comparison plane
$L_{\mathrm{trunk}}(t)$	Trunk length computed from shoulder and pelvis centers at time t
$L_{\mathrm{trunk,ref}}$	Trial-specific reference trunk length obtained from the non-active reference segment and used for MediaPipe displacement normalization
$L_{\mathrm{trunk}}^{\mathrm{ref}}(t)$	Projected MoCap trunk length at time t
$L_{\mathrm{trunk,ref}}^{\mathrm{ref}}$	Trial-specific projected reference trunk length obtained from the non-active reference segment and used for MoCap displacement normalization
$d(t)$	MediaPipe-derived displacement signal

d_{ref}	Reference displacement value obtained from the non-active reference segment of a trial
$d_{\text{norm}}(t)$	Reference-segment-relative displacement signal normalized by reference trunk length
$d^{\text{ref}}(t)$	Projected MoCap displacement signal corresponding to a MediaPipe-derived displacement metric
$d^{\text{ref}}_{\text{ref}}$	Reference displacement value of the projected MoCap signal obtained from the non-active reference segment
$d^{\text{ref}}_{\text{norm}}(t)$	Reference-segment-relative normalized displacement signal derived from the projected MoCap reference
$\theta_{\text{trunk}}(t)$	Trunk-orientation angle computed from MediaPipe image-plane landmarks
$\theta_{\text{knee},s}(t)$	2D MediaPipe knee angle for side s
$\theta^{\text{ref}}_{\text{trunk}}(t)$	Trunk-orientation angle computed from the projected MoCap reference points
$\theta^{\text{ref}}_{\text{knee},s}(t)$	Projected MoCap knee-angle reference for side s
$\mathbf{v}_{\text{thigh},s}(t)$	MediaPipe thigh vector for side s, defined from the knee landmark to the hip landmark
$\mathbf{v}_{\text{shank},s}(t)$	MediaPipe shank vector for side s, defined from the knee landmark to the ankle landmark
$\mathbf{v}^{\text{ref}}_{\text{thigh},s}(t)$	Projected MoCap thigh vector for side s
$\mathbf{v}^{\text{ref}}_{\text{shank},s}(t)$	Projected MoCap shank vector for side s
C_1	Condition 1, the standard recording setup with 3 m camera-to-participant distance and normal room lighting
C_i	A tested recording condition, where $i \in \{2, 3, 4, 5\}$ denotes a distance or lighting variation
x_{MP}	Metric value derived from MediaPipe Pose
x_{MoCap}	Corresponding metric value derived from the MoCap reference
$x_{\text{est},i}$	Estimated value for the i-th matched sample or trial-level value
$x_{\text{ref},i}$	Reference value for the i-th matched sample or trial-level value
e_i	Signed error for the i-th matched value
e_r	Trial-level signed error for repeated recording r

Evaluation Measures

MAE	Mean absolute error between MediaPipe-derived and MoCap-derived measurements
RMSE	Root-mean-square error between MediaPipe-derived and MoCap-derived measurements
Bias	Mean signed difference between MediaPipe and MoCap measurements, defined as MediaPipe minus MoCap where applicable
$SD(e)$	Standard deviation of the signed errors
LoA	Bland–Altman limits of agreement
ΔMAE_{C_i}	Change in MAE between a tested recording condition C_i and the matched Condition 1 trials
Δe_{repeat}	Mean pairwise absolute change in signed MediaPipe–MoCap error within a matched repeat group

Contents

List of Figures

List of Tables

1
Introduction

1.1 Background and Motivation

Rehabilitation often extends over a relatively long period and depends on repeated practice rather than a single clinical visit. In routine clinical practice, exercise quality is commonly assessed through direct observation by physiotherapists. This remains important, but the same level of supervision is difficult to maintain when rehabilitation is transferred to the home environment. Home rehabilitation therefore creates a need for monitoring tools that can provide objective, repeatable, and interpretable information about exercise performance.

Camera-based markerless motion analysis is attractive for home rehabilitation because it can be implemented with low-cost red-green-blue (RGB) devices and does not require reflective markers or laboratory calibration. Recent review studies indicate that single-camera markerless systems have potential in healthcare applications, especially when portability, affordability, and ease of deployment are important for home monitoring and telerehabilitation [2, 3]. However, technical feasibility alone is not sufficient. If such systems are to support rehabilitation assessment, the derived measurements must be accurate and reliable enough to inform follow-up and decision-making [2, 3].

MediaPipe Pose is one accessible markerless solution for real-time body landmark detection from RGB input. It is based on the BlazePose pipeline and has been extended through BlazePose GHUM toward richer three-dimensional (3D) landmark estimation from monocular RGB input [4, 5]. These properties make MediaPipe-based approaches attractive for rehabilitation scenarios, where a single consumer camera is easier to deploy than a laboratory motion-analysis system. At the same time, monocular pose estimation remains sensitive to depth ambiguity, self-occlusion, lighting variation, clothing, and non-ideal camera placement, which may affect rehabilitation-relevant kinematic measures.

1.2 Related Work and Research Gap

Previous research has shown growing interest in markerless motion capture for healthcare and rehabilitation applications. Review studies suggest that single-camera and video-based markerless systems may provide accessible movement measurements for clinical and home-based contexts, but their use as clinical measurement tools is still developing and requires further validation, careful interpretation, and clinician-friendly integration [2, 3, 6]. This is relevant for home rehabilitation be-

cause low-cost camera systems are easier to deploy than laboratory motion-capture systems, but the derived measurements still need to be accurate, repeatable, and meaningful for exercise assessment.
Existing validation studies have usually focused on specific movement tasks or specific types of accuracy. MediaPipe-based gait studies have mainly evaluated temporal or spatiotemporal gait parameters under structured recording conditions [7], while smartphone-based gait analysis has examined how camera placement, distance, and device position influence measurement error [8]. Other markerless studies have investigated postural or exercise-related measurements, including two-dimensional (2D) and frontal-plane measures, and have reported sensitivity to viewpoint, anatomical region, and depth-related distortion [9, 10]. MediaPipe has also been explored for upper-limb movement tracking and rehabilitation-related kinematic analysis, but these studies focus on different tasks and body regions, such as fine upper-limb trajectories or post-stroke reaching movements [11, 12]. Together, these studies show that markerless pose-estimation accuracy is task- and metric-dependent rather than universally valid across all rehabilitation exercises.
Beyond academic validation studies, examples such as QuickPose and OrthoCAP indicate that vision-based movement monitoring is also being explored in practical digital-health and rehabilitation contexts [13, 14]. These examples suggest growing interest in camera-based movement assessment, but they do not remove the need for technical validation of exercise-specific rehabilitation metrics against reference measurements.
A practical gap remains at the level of rehabilitation-specific assessment. Previous studies have often focused on gait parameters, general posture measures, upper-limb tracking, or individual movement tasks.
The contribution of this thesis is therefore to evaluate MediaPipe Pose at the metric level for single-leg stance(SLS), sit-to-stand(STS), and mini-squat(MINI), using matched MediaPipe and projected MoCap definitions for trunk orientation, normalized displacement, movement timing, and knee flexion. In addition, the thesis links the quantitative evaluation to an illustrative rule-based feedback prototype, showing how pose-derived metrics can be translated into patient-facing feedback and physiotherapist-facing review outputs.

1.3 Aim and Research Questions

The aim of this thesis is to investigate the feasibility and technical accuracy of a camera-based home rehabilitation monitoring system based on MediaPipe Pose. More specifically, the thesis examines whether a single RGB camera can provide useful technical measurements for selected rehabilitation exercises when evaluated against a Qualisys MoCap reference.
Based on this aim, the study addresses the following research questions:

- How accurately can selected posture and joint metrics be estimated from a single RGB camera compared with MoCap?
- How sensitive are these measurements to the available practical recording variations: camera distance and lighting?

- How can predefined movement-quality criteria be used to generate rule-based feedback for selected rehabilitation exercises?
- How repeatable are the derived movement measurements across repeated trials under the same recording setup?
- How can quantitative movement metrics be summarized in a way that is understandable for patients and useful for physiotherapist follow-up?

1.4 Scope and Limitations

This thesis is limited to the technical evaluation of a camera-based rehabilitation monitoring pipeline using MediaPipe Pose. It does not develop a new pose-estimation model, but evaluates an application pipeline built on an existing markerless framework. The work focuses on three rehabilitation-oriented exercises: single-leg stance, sit-to-stand, and mini-squat.
The evaluation was carried out using one RGB camera and a laboratory-based Qualisys MoCap system. The selected metrics were defined according to the camera view and the dominant movement plane of each exercise. SLS was evaluated mainly from a frontal view, while STS and MINI were evaluated mainly from a side view. Camera distance and lighting were included as practical recording variations, but camera viewpoint was not treated as a separate robustness condition.
The study was conducted under controlled laboratory conditions with healthy participants. It should therefore be understood as a technical feasibility study rather than a clinical validation study. Real home factors such as background clutter, furniture, uncontrolled camera placement, clothing variation, and partial occlusion were outside the main experimental scope.

1.5 Thesis Outline

The remainder of this thesis is organized as follows. Chapter 2 reviews the theoretical background relevant to markerless pose estimation, rehabilitation assessment, and motion-analysis metrics. Chapter 3 describes the methodology, including the experimental setup, data-processing pipeline, and evaluation strategy against MoCap. Chapter 4 presents the results of the technical evaluation and the prototype feedback design. Chapter 5 discuss findings, limitations, challenges and future work. Finally, Chapter 6 concludes the thesis.

2
Theory

2.1 Markerless Human Pose Estimation

Human pose estimation refers to the task of identifying anatomical body landmarks from visual input, such as images or videos. In markerless pose estimation, these landmarks are estimated without attaching physical markers or sensors to the subject. This makes the approach attractive in settings where low cost, ease of use, and minimal setup are important [15]. Compared with conventional marker-based MoCap systems, single-camera markerless approaches are easier to deploy in homes and clinics, and may therefore support remote rehabilitation monitoring and follow-up [2, 3].

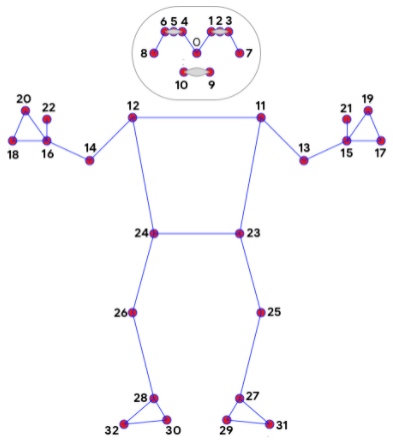
Most current markerless pose-estimation systems use deep-learning methods to detect body landmarks from RGB images or videos [15]. The body is usually represented as a set of keypoints corresponding to major joints or anatomical landmarks. These keypoints can then be used to compute higher-level movement variables, such as joint angles, segment orientations, displacement measures, or timing-related metrics. Depending on the model and data representation, the output may be expressed as 2D image coordinates, pseudo-3D coordinates relative to the camera, or estimated 3D pose representations.

The main advantage of markerless pose estimation is its accessibility, but the measurements are still inferred from image data. Their quality can therefore change with image quality, camera viewpoint, lighting conditions, clothing, body visibility, and self-occlusion. These issues are especially relevant in monocular RGB settings, where depth is not directly measured and out-of-plane motion is difficult to recover reliably [2, 3]. Existing reviews suggest that markerless systems are promising for practical movement monitoring, but that their accuracy may still be limited for detailed 3D biomechanical analysis or highly precise clinical decision-making [2, 3]. Therefore, markerless pose estimation is better understood as a practical and scalable measurement tool whose suitability depends on the task, movement plane, and required measurement precision, rather than as a direct replacement for laboratory MoCap in all situations.

2.2 MediaPipe Pose and BlazePose

MediaPipe Pose is a markerless pose-estimation solution for real-time human-body landmark detection from RGB images or video. Within the broader MediaPipe framework, the pose task is designed for single-person applications and can pro-

cess still images, recorded video, or live-stream input. The model outputs body landmarks in normalized image coordinates and also provides estimated 3D world landmarks, although the latter are inferred from monocular RGB input rather than directly measured in physical space [1, 4, 5].
The pose-estimation pipeline used in MediaPipe Pose is closely related to BlazePose. In this framework, pose estimation is performed in two stages. First, a detector identifies the person and estimates the approximate body region in the image. Second, a landmark model predicts a denser set of body keypoints within the detected region of interest [1, 4]. This two-stage design reduces computational cost while maintaining stable tracking of a single subject, which is useful for real-time applications on consumer devices.
A key characteristic of BlazePose is that it predicts 33 body landmarks for one person, covering the head, upper limbs, trunk, pelvis, and lower limbs [1, 4]. Figure 2.1 illustrates the 33-landmark body representation used in MediaPipe Pose. Compared with simpler skeleton representations, this denser landmark set provides more anatomical information for movement assessment. In this thesis, these landmarks were used to derive rehabilitation-related metrics such as trunk inclination, joint angles, pelvic displacement, and movement timing.



0. nose
1. left_eye_inner
2. left_eye
3. left_eye_outer
4. right_eye_inner
5. right_eye
6. right_eye_outer
7. left_ear
8. right_ear
9. mouth_left
10. mouth_right
11. left_shoulder
12. right_shoulder
13. left_elbow
14. right_elbow
15. left_wrist
16. right_wrist
17. left_pinky
18. right_pinky
19. left_index
20. right_index
21. left_thumb
22. right_thumb
23. left_hip
24. right_hip
25. left_knee
26. right_knee
27. left_ankle
28. right_ankle
29. left_heel
30. right_heel
31. left_foot_index
32. right_foot_index

Figure 2.1: The 33-body-landmark representation provided by MediaPipe Pose. From the MediaPipe Pose documentation [1].

BlazePose GHUM further extends this approach by providing richer 3D human landmark estimation from monocular RGB input [5]. The GHUM-based formulation improves the structural plausibility of the estimated pose and provides output representations that are more informative than standard 2D keypoints alone. However, because the 3D landmarks are still inferred from a single RGB view, they should not be treated as equivalent to directly measured 3D marker trajectories.
For this reason, the quantitative comparison in this thesis was based mainly on metric definitions derived from MediaPipe image-plane landmarks and corresponding projected MoCap trajectories. This allowed the MediaPipe-derived and MoCap-derived measurements to be compared using matched 2D metric definitions. The following evaluation therefore focuses on how well these selected MediaPipe-based metrics agree with MoCap reference metrics under the tested exercise and recording conditions.

2.3 MoCap as Reference Measurement

MoCap is widely used in biomechanics and human movement analysis to obtain structured 3D kinematic data under controlled laboratory conditions. In a typical marker-based setup, reflective markers are attached to predefined anatomical landmarks and tracked simultaneously by multiple infrared cameras. From the captured marker trajectories, the 3D positions of body segments can be reconstructed and further used to estimate joint kinematics and other movement-related variables [16, 17]. Because of its established use in laboratory-based movement analysis, marker-based MoCap is commonly treated as a reference standard when new markerless approaches are evaluated.

The main strength of MoCap lies in its ability to provide structured and quantitatively precise measurements of human movement in 3D space. For rehabilitation-related analysis, this is particularly valuable because posture, joint angles, and segment motions can be derived from a physically defined spatial reference frame rather than inferred only from image appearance. As a result, MoCap systems are often used when detailed biomechanical measurements are required, especially in studies that aim to assess the validity of alternative tracking methods [16, 17].

At the same time, marker-based MoCap also has well-known limitations. The measurement process usually requires careful marker placement, calibration, dedicated equipment, and controlled laboratory conditions. This makes the setup time-consuming and dependent on trained personnel. In addition, since markers are attached to the skin rather than directly to the underlying bones, the measurements may be affected by soft-tissue artefacts and marker-placement variability [16]. These limitations do not remove the value of MoCap as a reference method, but they help explain why more accessible markerless alternatives are being actively studied for rehabilitation and home-monitoring applications [3, 17].

2.4 Rehabilitation Exercise Assessment

In rehabilitation, the objective of exercise assessment is not only to determine whether a patient has completed a prescribed task, but also to evaluate how the movement was performed. A repetition may appear successful at the task level while still being executed with insufficient range of motion, poor postural control, or compensatory movement patterns. For this reason, rehabilitation assessment is closely related to movement quality rather than simple task completion alone. In clinical practice, physiotherapists often judge such quality by observing alignment, control, symmetry, stability, and the presence of compensatory strategies during exercise execution.

This need becomes even more important in home-based rehabilitation. When exercises are performed outside the clinic, direct supervision is reduced and the opportunity for immediate correction is limited. Telerehabilitation research has shown that remote physiotherapy can be a feasible and clinically relevant alternative in many settings, with evidence suggesting outcomes that are often comparable to in-person rehabilitation and satisfaction levels that are generally similar for patients

and professionals [18, 19, 20]. However, these advantages do not remove the need for meaningful exercise assessment. If rehabilitation is to be monitored remotely, the system must provide information that reflects not only whether an exercise was attempted, but whether it was performed in an acceptable way.
In rehabilitation, compensatory strategies may allow a person to complete a task while redistributing effort to other body segments or substituting one movement pattern for another. Such compensation may be clinically important because it can hide limited control in the target movement or alter the therapeutic value of the exercise. Studies on movement-quality assessment have shown that kinematic evaluation should not focus only on the end result of a task, but should also consider the coordination of body segments and the presence of compensatory strategies [21, 22]. This is particularly relevant for exercises such as STS, SLS, or squat-like movements, where trunk motion, pelvic control, and symmetry may be as important as the primary joint movement itself.

2.4.1 Compensatory Movements in the Selected Exercises

Compensatory movement is relevant in rehabilitation-oriented assessment because a task may be completed successfully while the movement strategy still deviates from the intended therapeutic pattern. In this thesis, the selected exercises were therefore considered not only in terms of task completion, but also in terms of common observable compensations. Table 2.1 summarizes the main movement aspects and compensatory patterns used to guide metric selection and feedback design.

Table 2.1: Selected exercises and relevant compensatory movement patterns considered in this thesis.

Exercise	Main movement aspect	Relevant compensatory patterns
SLS	Frontal-plane balance and postural control	Pelvic drop, lateral trunk lean, and arm elevation or upper-limb assistance used as balance strategies. These patterns are related to frontal-plane pelvic control and hip-abductor function [23, 24].
STS	Functional transition from sitting to standing	Excessive trunk momentum, use of arms or hands, and incomplete terminal extension. These may indicate altered movement strategy or reduced lower-limb contribution during the sit-to-stand transition [25, 26].
MINI	Lower-limb control, squat depth, and trunk control	Excessive forward trunk lean, altered knee position, excessive anterior knee translation, and insufficient or excessive squat depth. These patterns are relevant for assessing lower-limb alignment and squat mechanics [27, 28].

Based on these considerations, the feedback prototype in this thesis uses predefined movement criteria to translate pose-derived metrics into simple feedback messages. The criteria are not intended as clinical diagnostic rules, but as a practical demonstration of how camera-based measurements may support patient-facing feedback

and physiotherapist review.

2.5 Kinematic and Postural Metrics

In movement assessment, quantitative metrics are used to translate raw body-motion trajectories into interpretable variables that describe alignment, control, stability, range of motion, and temporal performance. Common metric categories include joint angles, segment inclinations, displacement-based stability measures, range-of-motion measures, and timing-related descriptors [9, 15, 29]. In this thesis, these concepts were used to define exercise-specific measures such as trunk inclination, normalized pelvis displacement, hold duration, repetition duration, squat depth, and knee flexion.

Table 2.2 summarizes the main metric categories that are relevant to the selected rehabilitation exercises.

Table 2.2: Examples of kinematic and postural metric categories relevant to rehabilitation exercise assessment.

Metric category	Example	What it reflects
Joint-angle metrics	Knee flexion angle	Movement strategy and range of motion
Segment-orientation metrics	Trunk inclination	Postural alignment and compensatory movement
Stability metrics	Pelvic or trunk sway	Postural control and balance stability
Range-of-motion metrics	Squat depth or angular excursion	Exercise amplitude
Temporal metrics	Hold or repetition duration	Timing and performance over time

Because these metrics are computed from camera images, their interpretation also depends on the anatomical plane represented by the camera view.

2.5.1 Anatomical Planes and 2D Projection

In human movement analysis, kinematic variables are often interpreted with respect to anatomical planes. The frontal plane represents movement components in the left-right and vertical directions, and is therefore relevant for measures such as lateral trunk lean and mediolateral pelvic displacement. The sagittal plane represents movement components in the anterior-posterior and vertical directions, and is commonly used for movements such as STS, MINI, trunk forward lean, and knee flexion. The transverse plane represents horizontal rotation around the vertical axis. Figure 2.2 illustrates the main anatomical planes used in movement analysis.

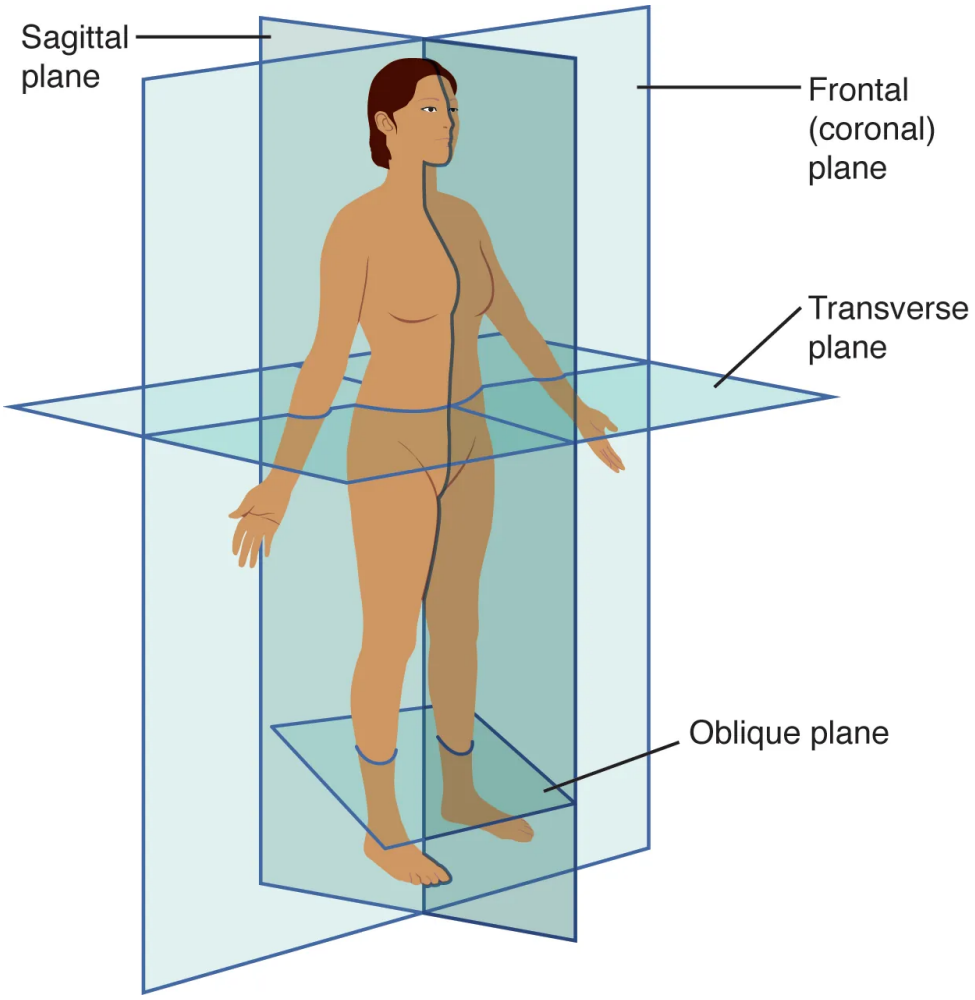


Figure 2.2: Anatomical planes of the human body, including the frontal, sagittal, and transverse planes. Adapted from OpenStax Anatomy and Physiology 2e [30].

For camera-based pose estimation, the anatomical plane is also related to the camera view. A front-view recording mainly represents frontal-plane motion in the image, while a side-view recording mainly represents sagittal-plane motion. This relationship is only approximate, because the camera image is a two-dimensional projection of a three-dimensional body. If the subject rotates or moves out of the intended plane, part of the movement may be lost or distorted in the image-plane representation.

In this context, 2D projection means representing a 3D point or trajectory on a selected 2D plane. For example, a 3D reference point can be projected onto a frontal plane by keeping the horizontal and vertical components, while the depth component is ignored. Similarly, projection onto a sagittal plane keeps the anterior-posterior and vertical components. This allows 3D motion data to be expressed in a form that is conceptually closer to the 2D image-plane landmarks obtained from a single RGB camera.

This projection step is useful for comparing camera-based and reference measurements, because both systems can then be evaluated using matched 2D metric definitions. However, it also introduces an important limitation: movement outside the

selected plane is not fully represented. Therefore, 2D projected metrics should be interpreted as plane-specific approximations rather than complete 3D biomechanical measurements.

2.6 Evaluation Metrics

The quantitative comparison between a markerless pose-estimation system and a reference measurement system can be described using several error-based metrics. In this thesis, the basic error was defined as the difference between an estimated value and the corresponding reference value:

$$e_i = x_{\text{est},i} - x_{\text{ref},i} \tag{2.1}$$

where $x_{\text{est},i}$ is the estimated value, $x_{\text{ref},i}$ is the reference value, and i denotes the i-th matched sample or trial-level value.
The mean absolute error (MAE) describes the average magnitude of the error regardless of direction:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{2.2}$$

The root mean square error (RMSE) is another measure of error magnitude:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}e_i^2} \tag{2.3}$$

Because the error is squared before averaging, RMSE gives more weight to larger errors than MAE.
Agreement between two measurement methods can also be examined using Bland–Altman analysis [31]. The Bland–Altman mean difference corresponds to the bias, and the 95% limits of agreement are commonly computed as

$$\text{LoA} = \text{Bias} \pm 1.96 \cdot SD(e) \tag{2.4}$$

where $SD(e)$ is the standard deviation of the signed errors. The limits of agreement describe the spread of the differences between the two measurement methods.

3
Methods

3.1 Study Design and Overall Workflow

This study was organized as a technical evaluation of the proposed MediaPipe Pose-based monitoring pipeline. The workflow was designed to keep the video-based measurements and the MoCap reference measurements as comparable as possible, from data collection to metric computation and final evaluation.

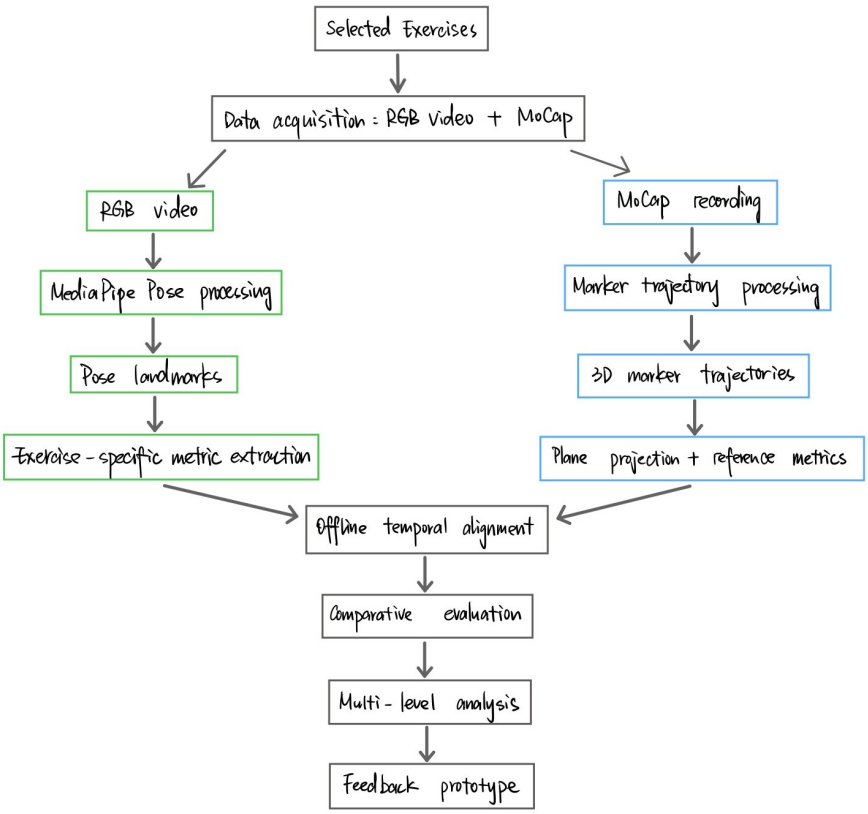


Figure 3.1: Overall methodological workflow of the study.

As shown in Figure 3.1, each exercise trial produced two parallel data streams. The RGB video recordings were processed with MediaPipe Pose to obtain body landmarks, which were then converted into exercise-specific metrics. In parallel, the MoCap marker trajectories were processed as the reference data. Since the MediaPipe measurements were derived from the camera image plane, the 3D MoCap trajectories were projected onto the corresponding frontal or sagittal analysis plane

before applying matched metric definitions.

After the two data streams had been processed, offline temporal alignment was applied because the RGB camera and MoCap system were not hardware-synchronized during recording. The aligned MediaPipe-derived and MoCap-derived metrics were then compared at different evaluation levels, including exercise-level accuracy, subject-level variability, robustness across recording conditions, and repeatability. Finally, selected pose-derived metrics were used in a rule-based feedback prototype to demonstrate how the technical measurements could be translated into patient-facing feedback and physiotherapist-facing session summaries.

3.2 Rehabilitation Exercises and Assessment Criteria

A limited set of rehabilitation-oriented exercises was selected to cover different movement characteristics relevant to home monitoring. The exercise set included SLS, STS, and MINI, representing balance control, functional transition, and lower-limb movement strategy, respectively. For each exercise, a small set of exercise-specific metrics was defined according to the dominant movement plane, the expected rehabilitation-relevant movement characteristics, and the visibility constraints of the single-camera setup [9, 15, 29].

Table 3.1: Exercise-specific metrics used in the evaluation pipeline.

Exercise	Main movement focus	Metrics used
SLS	Balance control and frontal-plane posture	Pelvis horizontal displacement normalized by trunk length, trunk side-tilt angle, hold duration
STS	Functional transition and trunk movement strategy	Pelvis vertical displacement normalized by trunk length, trunk forward-lean angle, repetition duration
MINI	Squat depth, trunk strategy, and knee motion	Squat depth normalized by trunk length, trunk forward-lean angle, knee flexion on the camera-facing side

Table 3.1 summarizes the metrics used in the evaluation pipeline. SLS was mainly evaluated using frontal-plane balance and posture metrics, while STS and MINI were mainly evaluated using sagittal-plane metrics. Continuous metrics were evaluated both over the full trial and, where relevant, over the detected active movement segment. This distinction was used to separate global trial behaviour from movement-specific execution quality.

The metric set was used both for the quantitative comparison and for the feedback prototype. The same variables were compared between MediaPipe Pose and the MoCap reference, and selected measurements were later translated into simple patient feedback and physiotherapist summaries.

3.3 Experimental Setup and Data Collection

Data collection was carried out in a controlled laboratory environment using RGB video recording and MoCap. The RGB videos were recorded with the rear camera of an iPhone 13 at 30 frames per second, while the reference MoCap data were acquired with a Qualisys MoCap system at 60 frames per second. Since the two systems were not hardware-synchronized during acquisition, temporal alignment was performed offline before metric comparison.

The final analysed dataset consisted of paired RGB-video and Qualisys MoCap recordings from five healthy participants. It included 132 trials across SLS, STS, and MINI. For each exercise, recordings were collected under a 3 m camera-to-participant distance with normal room lighting. Additional recordings were then collected with selected variations in camera distance and lighting. These variations were included to examine how practical recording conditions may affect the MediaPipe-derived measurements. Because the number of available trials differed between participants and recording conditions, the dataset was treated as an unbalanced repeated-measures dataset rather than as a fully balanced design.

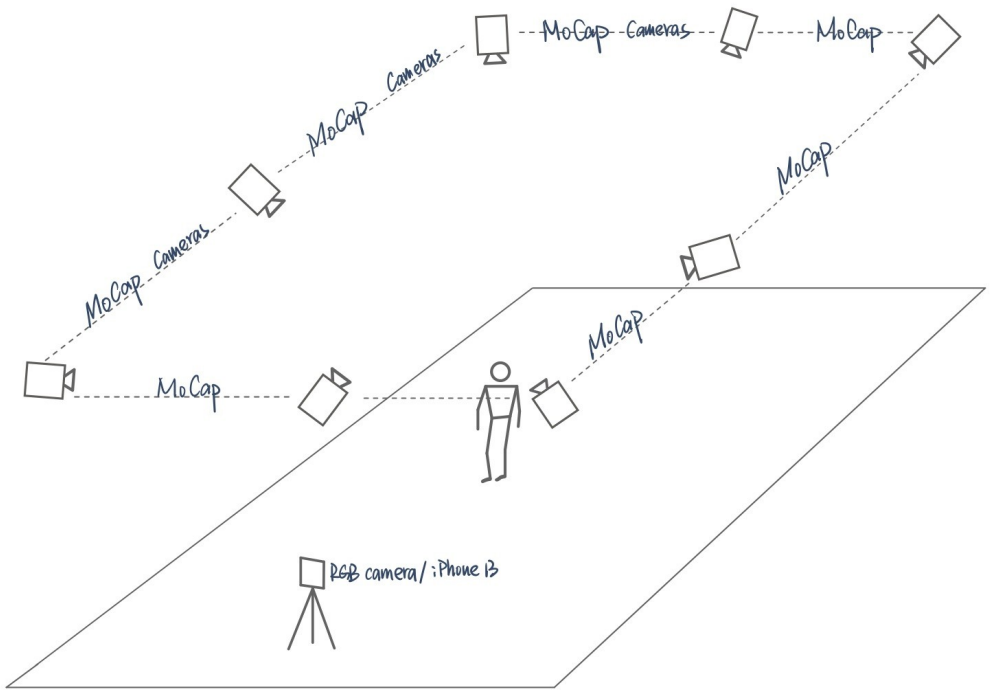


Figure 3.2: (a) Laboratory recording setup.

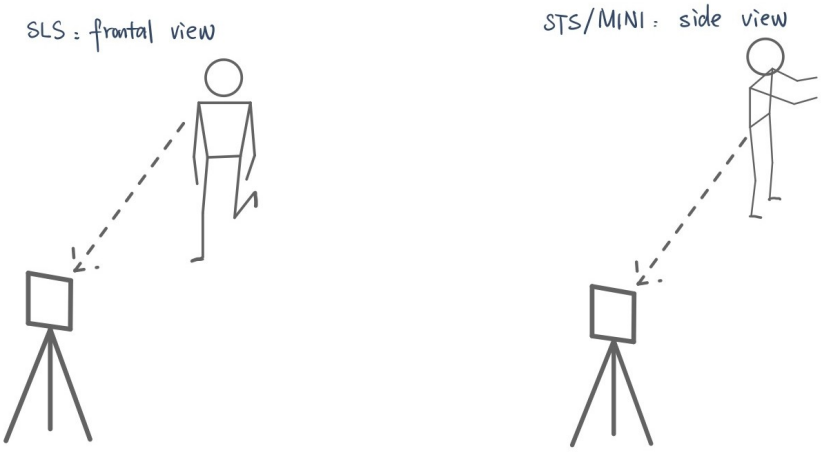


Figure 3.2: (b) Exercise-specific camera views.

Figure 3.2 - Figure 3.2 illustrate the experimental recording setup and the exercise-specific camera views. The recording view was selected according to the dominant movement plane of each exercise, following prior evidence that camera placement can affect markerless kinematic accuracy [8]. SLS was recorded from a frontal view because its main metrics involved frontal-plane trunk side tilt and pelvis sway. STS and MINI were recorded from a side view because trunk forward lean, vertical displacement, and knee flexion are mainly interpreted in the sagittal plane. Camera viewpoint was therefore treated as part of the exercise-specific setup rather than as an independent robustness condition.
The tested recording-condition variations included camera distance and lighting. In the analysed dataset, distance variations were represented by 2 m and 4 m recordings, while lighting variations were represented by low-light and no-light recordings. These variations were intended to approximate selected challenges that may occur in home-based exercise recording, but the study was still conducted in a laboratory setting. No furniture or additional household objects were introduced, so the setup should be understood as a controlled approximation of selected home-like recording conditions rather than a full simulation of real home use.

3.4 MediaPipe-Based Pose Processing and Metric Extraction

The recorded RGB videos were read frame-by-frame using OpenCV's video capture interface and processed with Google's MediaPipe Pose framework, which estimates 33 body landmarks from RGB image frames [1, 32]. For each frame with a detected pose, the 33 MediaPipe landmarks were stored in long-form CSV format, including frame index, timestamp, landmark identity, normalized `x`, `y`, `z`, and visibility. The exported landmark files were then processed trial by trial for synchronization and metric computation.
The normalized landmark trajectories were converted into structured time series

before comparison with MoCap. Because the RGB videos and MoCap recordings had different sampling frequencies and were not hardware-synchronized, the final comparison was performed on temporally aligned signals over a shared timeline rather than directly on raw frame indices. Basic smoothing and low-pass filtering were applied before exercise-specific metrics were computed.

For MediaPipe Pose, each detected landmark was represented as a time-dependent position vector. In this section, $\mathbf{p}_q(t)$ denotes the 2D image-plane position of landmark q at time t, where t corresponds to the video timestamp or frame time. Unless otherwise stated, the MediaPipe landmark coordinates are given in normalized image coordinates, with x and y ranging approximately from 0 to 1 in the image plane. Bold symbols therefore denote 2D position vectors, while scalar symbols denote distances, angles, or metric values.

The pelvis center and shoulder center were defined as the midpoints of the left and right hip landmarks and the left and right shoulder landmarks:

$$\mathbf{p}_{\text{pelvis}}(t) = \frac{\mathbf{p}_{\text{LHip}}(t) + \mathbf{p}_{\text{RHip}}(t)}{2} \tag{3.1}$$

$$\mathbf{p}_{\text{shoulder}}(t) = \frac{\mathbf{p}_{\text{LShoulder}}(t) + \mathbf{p}_{\text{RShoulder}}(t)}{2} \tag{3.2}$$

where $\mathbf{p}_{\text{LHip}}(t)$, $\mathbf{p}_{\text{RHip}}(t)$, $\mathbf{p}_{\text{LShoulder}}(t)$, and $\mathbf{p}_{\text{RShoulder}}(t)$ are the MediaPipe landmark positions for the left hip, right hip, left shoulder, and right shoulder, respectively.

Since MediaPipe provides normalized image coordinates, the landmark positions were converted to pixel-scaled image-plane coordinates before geometric metric extraction:

$$\mathbf{p}_q^{\text{img}}(t) = \begin{bmatrix} x_q(t)W \\ y_q(t)H \end{bmatrix} \tag{3.3}$$

where q denotes a MediaPipe landmark, $x_q(t)$ and $y_q(t)$ are the normalized image coordinates of this landmark at time t, and W and H are the image width and height in pixels. The pixel-scaled position $\mathbf{p}_q^{\text{img}}(t)$ was used for geometric metric extraction so that horizontal and vertical image coordinates were represented on the same pixel scale. This was important for trunk-orientation angles, knee-flexion angles, and trunk-length-normalized displacement metrics.

Displacement-based metrics were expressed relative to non-active reference and normalized by trunk length. The frame-wise trunk length was defined as

$$L_{\text{trunk}}(t) = \|\mathbf{p}_{\text{shoulder}}(t) - \mathbf{p}_{\text{pelvis}}(t)\| \tag{3.4}$$

where $L_{\text{trunk}}(t)$ is the Euclidean distance between the shoulder center and pelvis center at time t. For each trial, the reference trunk length was obtained as the median trunk length during the non-active reference segment. This reference value was used to reduce the influence of participant size and camera scale on displacement-based metrics.

For a displacement signal $d(t)$, the normalized displacement was computed as

$$d_{\text{norm}}(t) = \frac{d(t) - d_{\text{ref}}}{L_{\text{trunk,ref}}} \tag{3.5}$$

where d_{ref} is the median displacement value during the same non-active reference segment, and $L_{\text{trunk,ref}}$ is the trial-specific reference trunk length. In this way, the displacement metric represents movement relative to the participant's initial posture and is expressed in units of trunk length.
Segment-orientation metrics were computed from the image-plane vector between the pelvis center and shoulder center. A generic trunk-orientation angle was computed as

$$\theta_{\text{trunk}}(t) = \text{atan2}\left(x_{\text{shoulder,img}}(t) - x_{\text{pelvis,img}}(t), -\left(y_{\text{shoulder,img}}(t) - y_{\text{pelvis,img}}(t)\right)\right) \tag{3.6}$$

where $x_{\text{shoulder,img}}(t)$ and $y_{\text{shoulder,img}}(t)$ are the pixel-scaled image coordinates of the shoulder center, and $x_{\text{pelvis,img}}(t)$ and $y_{\text{pelvis,img}}(t)$ are the corresponding coordinates of the pelvis center. For SLS, this angle was interpreted as trunk side tilt in the frontal-view recording. For STS and MINI, the same image-plane angle convention was used to describe trunk forward lean in the side-view recording. All orientation metrics were expressed in degrees.
For MINI, knee flexion was computed as a 2D joint angle from the hip, knee, and ankle landmarks in the image plane. For a side s, the knee angle was computed from the angle between the thigh vector and the shank vector

$$\theta_{\text{knee},s}(t) = \cos^{-1}\left(\frac{\mathbf{v}_{\text{thigh},s}(t) \cdot \mathbf{v}_{\text{shank},s}(t)}{\|\mathbf{v}_{\text{thigh},s}(t)\| \, \|\mathbf{v}_{\text{shank},s}(t)\|}\right) \tag{3.7}$$

where $\mathbf{v}_{\text{thigh},s}(t)$ is the vector from the knee landmark to the hip landmark, and $\mathbf{v}_{\text{shank},s}(t)$ is the vector from the knee landmark to the ankle landmark. In the side-view MINI recordings, the visible side was used for the knee-flexion metric because the occluded side was more sensitive to landmark uncertainty.

Table 3.2: Summary of MediaPipe-derived metric computation.

Exercise	Metric	Computation principle
SLS	Pelvis horizontal displacement; trunk side tilt; hold duration	Baseline-relative pelvis displacement normalized by trunk length; trunk orientation from shoulder-pelvis vector; hold phase from foot-index separation.
STS	Pelvis vertical displacement; trunk forward lean; repetition duration	Baseline-relative pelvis displacement normalized by trunk length; trunk orientation in the side-view plane; duration from trunk-based transition segmentation.
MINI	Squat depth; trunk forward lean; visible-side average knee flexion	Baseline-relative squat-depth signal normalized by trunk length; trunk orientation in the side-view plane; 2D knee angle.

Exercise-specific temporal measures were obtained through event-based segmentation. For SLS, the active hold segment was estimated from the relative vertical

separation between the left and right foot-index landmarks. For STS, repetition duration was derived from trunk-based segmentation of the sit-to-stand transition. For MINI, the continuous metrics were evaluated over both the full trial and the active movement segment. The full-trial and active-segment distinction was kept in the final comparison to separate overall recording behaviour from movement-specific execution.

3.5 MoCap Processing and Reference Metric Computation

The Qualisys MoCap data were used as the reference measurement for the comparative evaluation. Since the MediaPipe-based metrics were derived mainly from 2D image-plane landmarks, the MoCap data were not compared directly in their original 3D coordinate form. Instead, the 3D marker trajectories were projected onto frontal or sagittal analysis planes corresponding to the effective camera view used for each exercise. Reference metrics were then computed from the projected MoCap points using the same main metric definitions as the MediaPipe-derived metrics.
The MoCap trajectories were reconstructed offline and exported as marker-coordinate time series. In this section, $\mathbf{p}_q^{\mathrm{ref},3D}(t)$ denotes the original 3D MoCap position of marker or reference point q at time t:

$$\mathbf{p}_q^{\mathrm{ref},3D}(t) = \begin{bmatrix} X_q(t) \\ Y_q(t) \\ Z_q(t) \end{bmatrix} \tag{3.8}$$

where $X_q(t)$, $Y_q(t)$, and $Z_q(t)$ are the MoCap coordinates. The vertical coordinate was represented by $Z(t)$, while the horizontal coordinate used for comparison depended on the selected exercise and camera view.
Reference body points were defined from marker combinations that approximated the corresponding MediaPipe landmark-based body points. In general, the pelvis center was represented using waist markers, and the shoulder center was represented using the left and right shoulder markers. For the MINI knee-angle reference, projected hip-side, knee, and ankle points were selected to correspond as closely as possible to the MediaPipe hip, knee, and ankle landmarks. These definitions were used to make the MoCap reference metrics comparable to the MediaPipe-derived metrics, while recognizing that marker positions and MediaPipe landmarks are not anatomically identical.
To make the MoCap data comparable with the camera-based measurements, the 3D reference points were projected onto a 2D analysis plane. A projected MoCap point was represented as

$$\mathbf{p}_q^{\mathrm{ref},\Pi}(t) = \begin{bmatrix} u_q^{\Pi}(t) \\ Z_q(t) \end{bmatrix} \tag{3.9}$$

where Π denotes the selected analysis plane, $u_q^{\Pi}(t)$ is the selected horizontal coordinate in that plane, and $Z_q(t)$ is the vertical coordinate. For SLS, a frontal-plane projection was used because the main metrics were lateral pelvis displacement and

trunk side tilt. For STS and MINI, a sagittal-plane projection was used because pelvis rise, squat depth, trunk forward lean, and knee flexion were interpreted from the side-view movement plane.
After projection, displacement-based reference metrics were expressed relative to non-active reference segment within each trial and normalized by projected trunk length. The projected trunk length was defined as

$$L_{\text{trunk}}^{\text{ref}}(t) = \left\| \mathbf{p}_{\text{shoulder}}^{\text{ref},\Pi}(t) - \mathbf{p}_{\text{pelvis}}^{\text{ref},\Pi}(t) \right\| \tag{3.10}$$

where $\mathbf{p}_{\text{shoulder}}^{\text{ref},\Pi}(t)$ and $\mathbf{p}_{\text{pelvis}}^{\text{ref},\Pi}(t)$ are the projected shoulder and pelvis centers in the selected analysis plane. For each trial, the reference trunk length was obtained as the median projected trunk length during the non-active reference segment. This value was used to reduce the influence of participant size and scale on the projected displacement metrics.
For a reference displacement signal $d^{\text{ref}}(t)$, the normalized reference displacement was computed as

$$d_{\text{norm}}^{\text{ref}}(t) = \frac{d^{\text{ref}}(t) - d_{\text{ref}}^{\text{ref}}}{L_{\text{trunk,ref}}^{\text{ref}}} \tag{3.11}$$

where $d_{\text{ref}}^{\text{ref}}$ is the median displacement value during the same non-active reference segment, and $L_{\text{trunk,ref}}^{\text{ref}}$ is the trial-specific projected reference trunk length. Here, $d^{\text{ref}}(t)$ denotes the projected MoCap displacement signal corresponding to the MediaPipe-derived displacement metric, such as SLS pelvis horizontal displacement, STS pelvis vertical displacement, or MINI squat depth.
For segment-orientation reference metrics, the trunk angle was computed from the projected vector between the pelvis center and shoulder center. Using $(u(t), v(t))$ to denote the coordinate pair in the projected plane, a generic trunk-orientation angle was computed as

$$\theta_{\text{trunk}}^{\text{ref}}(t) = \operatorname{atan2}\left(u_{\text{shoulder}}(t) - u_{\text{pelvis}}(t), -\left(v_{\text{shoulder}}(t) - v_{\text{pelvis}}(t)\right)\right) \tag{3.12}$$

where $u(t)$ represents the selected horizontal coordinate and $v(t)$ represents the vertical coordinate in the projected plane. For frontal-plane SLS analysis, this angle was interpreted as the reference trunk side tilt. For sagittal-plane STS and MINI analysis, the same orientation definition was interpreted as the reference trunk forward lean. All orientation metrics were expressed in degrees.
For MINI, the MoCap knee-angle reference was computed as a 2D joint angle after sagittal-plane projection. For a side s, the knee angle was computed from the angle between the projected thigh vector and the projected shank vector:

$$\theta_{\text{knee},s}^{\text{ref}}(t) = \cos^{-1}\left(\frac{\mathbf{v}_{\text{thigh},s}^{\text{ref}}(t) \cdot \mathbf{v}_{\text{shank},s}^{\text{ref}}(t)}{\left\|\mathbf{v}_{\text{thigh},s}^{\text{ref}}(t)\right\| \left\|\mathbf{v}_{\text{shank},s}^{\text{ref}}(t)\right\|} \right) \tag{3.13}$$

where $\mathbf{v}_{\text{thigh},s}^{\text{ref}}(t)$ is the projected vector from the knee point to the hip-side point, and $\mathbf{v}_{\text{shank},s}^{\text{ref}}(t)$ is the projected vector from the knee point to the ankle point. This definition matched the MediaPipe knee-flexion metric at the level of a planar hip-knee-ankle angle. However, because the MoCap markers and MediaPipe landmarks

do not represent exactly the same anatomical points, the knee-angle comparison should be interpreted as a matched technical reference rather than a perfect anatomical equivalence.
The same full-trial and active-segment definitions used in the MediaPipe evaluation were applied to the projected MoCap reference signals after temporal alignment. This allowed the comparison to be performed over matched time intervals, both for the complete recordings and for the movement-relevant portions of each trial.

3.6 Synchronization and Comparative Evaluation

3.6.1 Synchronization

The MediaPipe-derived and MoCap-derived metric signals were synchronized offline for each trial before comparison. Before alignment, the MoCap signals were low-pass filtered with a cutoff frequency of 2.5 Hz and downsampled to the estimated MediaPipe sampling rate. The main alignment method was signal-based cross-correlation with a maximum searched lag of 12 s. Exercise-specific candidate signals were used for alignment: toe-gap and pelvis-horizontal motion were prioritized for SLS, while pelvis-vertical motion was prioritized for STS and MINI. If the cross-correlation score was below 0.5, a peak-based fallback shift was used when available. After alignment, the MediaPipe and MoCap signals were interpolated to a common analysis timeline.

3.6.2 Comparative Evaluation

The comparative evaluation was organized at four levels. First, the exercise-level analysis compared the selected metrics across SLS, STS, and MINI under the 3 m camera-to-participant distance and normal room lighting setup, defined as Condition 1. This analysis was used to identify which exercise types and metric families were more suitable for the MediaPipe-based pipeline. Second, the subject-level analysis examined the Condition 1 trials separately across participants to describe inter-subject variability. Third, the condition-level analysis compared selected distance and lighting variations with matched Condition 1 trials. Finally, the repeatability analysis examined whether the MediaPipe-MoCap error remained stable across repeated recordings within the same session and, where available, across different days.
For the condition-level robustness analysis, the recording conditions were defined according to the practical setup variations introduced during data collection. Condition 1 was the standard setup, with a 3 m camera-to-participant distance and normal room lighting. Condition 2 used a shorter camera distance of 2 m, while Condition 3 used a longer camera distance of 4 m. Condition 4 used low-light recording, and Condition 5 used no-light recording.
The comparison used the error metrics defined in Section 2.6. For each metric, the MediaPipe-derived signal was compared with the corresponding projected MoCap reference signal after temporal alignment. Continuous metrics were evaluated over

the full trial and, where relevant, over the active movement segment. The main reported measures were MAE and RMSE.
For the condition-level robustness analysis, each varied recording condition was compared with the matched Condition 1 trials from the same exercise setting. The change in MAE was computed as

$$\Delta\mathrm{MAE}_{C_i} = \mathrm{MAE}_{C_i} - \mathrm{MAE}_{C_1}, \qquad i \in \{2, 3, 4, 5\} \tag{3.14}$$

where C_1 denotes Condition 1 and C_i denotes one of the varied recording conditions. A positive $\Delta\mathrm{MAE}_{C_i}$ means that the MediaPipe-MoCap error increased under the tested recording condition compared with the matched Condition 1 trials, while a negative value means that the error decreased.
Repeatability was analysed from the error perspective. The purpose was not to estimate the internal repeatability of MediaPipe alone, but to examine whether the difference between MediaPipe and MoCap stayed consistent across repeated recordings. A matched repeat group was defined as a set of repeated recordings with the same participant, exercise, metric, and recording setup. For each recording r in a group, a trial-level signed error was computed as

$$e_r = x_{\mathrm{MP},r} - x_{\mathrm{MoCap},r} \tag{3.15}$$

where $x_{\mathrm{MP},r}$ and $x_{\mathrm{MoCap},r}$ are the trial-level MediaPipe-derived and MoCap-derived metric values from the same repeated recording.
For a matched repeat group with m recordings, the repeatability error change was computed as the mean pairwise absolute difference between signed errors:

$$\Delta e_{\mathrm{repeat}} = \frac{2}{m(m-1)} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} |e_j - e_k| \tag{3.16}$$

This calculation first averages all pairwise error changes within the same matched repeat group. The final repeatability results were then summarized as mean ± standard deviation across matched repeat groups. This aggregation was used so that groups with more repeated recordings did not automatically receive a larger weight in the final summary. For the within-session analysis, a matched repeat group consisted of repeated recordings from the same participant, exercise, session, and setup. For the across-day analysis, it consisted of recordings from the same participant, exercise, and setup collected on different days.

3.7 Feedback Prototype Design

The feedback prototype was implemented as a rule-based monitoring layer on top of the pose-processing pipeline. It used selected pose-derived metrics and exercise-specific thresholds to generate simple feedback messages during exercise execution. The purpose of the prototype was not to provide clinical diagnosis, but to demonstrate how the extracted movement metrics could be translated into understandable feedback for home rehabilitation monitoring. Table 3.3 summarizes the main rule-based feedback components implemented for each exercise.

Table 3.3: Rule-based feedback components implemented in the prototype.

Exercise	Main detection logic	Feedback focus
SLS	Active hold detection from relative foot-index separation	Trunk lateral lean, arm compensation, and excessive pelvis sway.
STS	Finite-state transition from seated posture to standing	Use of arms, excessive trunk momentum, and incomplete extension.
MINI	Side-view state sequence for repetition counting	Insufficient depth, excessive forward lean, excessive hip flexion, knee-over-toe movement, and hands-on-thigh compensation.

Two output perspectives were considered in the prototype. The patient-facing feedback was designed to be short and action-oriented, for example indicating whether posture should be corrected, whether balance should be improved, or whether the movement should be adjusted. The physiotherapist-facing output was designed as a more structured session summary, including repeated compensations, movement trends, and overall exercise-quality information across trials.

4

Results

This chapter presents the evaluation results of the MediaPipe-based metric pipeline. The first part reports the metric errors under Condition 1 setup. The following sections then examine how the errors varied between participants, how they changed under the tested distance and lighting conditions, how stable the MediaPipe-MoCap error was across repeated recordings, and how the same metric pipeline was used in the feedback prototype.

4.1 Exercise-Level Accuracy Under the Condition 1 Recording Setup

Table 4.1 and Table 4.2 report the exercise-level errors for the selected metrics under the Condition 1 recording setup. The full-trial evaluation used the entire recording, including preparation, repetitions, and static intervals. The active-segment evaluation used only the movement-relevant part of each trial. In the following interpretation, the active-segment results are emphasized because they better represent the actual exercise execution, while the full-trial results show how much the average error changes when non-active periods are included.

Table 4.1: Exercise-level results for full-trial evaluation under Condition 1.

Exercise	Metric	Unit	Trials	MAE (mean ± SD)	RMSE (mean ± SD)
SLS	Pelvis horizontal displacement / trunk length	unitless	15	0.0275 ± 0.0148	0.0369 ± 0.0205
	Trunk side tilt	degree	15	1.0474 ± 0.6020	1.2012 ± 0.5694
STS	Trunk forward lean	degree	15	3.1644 ± 1.0778	3.8824 ± 1.4205
	Pelvis vertical displacement / trunk length	unitless	15	0.0516 ± 0.0232	0.0670 ± 0.0311
MINI	Squat depth / trunk length	unitless	15	0.0236 ± 0.0127	0.0319 ± 0.0169
	Knee flexion (visible side)	degree	15	4.4693 ± 1.6302	5.5963 ± 1.7198
	Trunk forward lean	degree	15	2.2030 ± 0.9436	2.4003 ± 0.8831

The active-segment evaluation generally produced higher errors than the full-trial evaluation. This was expected to some extent, since the active segment contains the main movement phase, where landmark errors and timing differences have a stronger influence on the computed metrics. In contrast, the full-trial evaluation also includes preparation and static periods, which can reduce the average error and make the metric appear more stable than it is during the actual movement.

Table 4.2: Exercise-level results for active-segment evaluation under Condition 1.

Exercise	Metric	Unit	Trials	MAE (mean ± SD)	RMSE (mean ± SD)
SLS	Pelvis horizontal displacement / trunk length	unitless	15	0.0392 ± 0.0397	0.0414 ± 0.0395
	Trunk side tilt	degree	15	1.1091 ± 0.8177	1.2190 ± 0.7856
STS	Trunk forward lean	degree	15	3.2688 ± 1.3354	3.8772 ± 1.4558
	Pelvis vertical displacement / trunk length	unitless	15	0.0604 ± 0.0310	0.0711 ± 0.0344
MINI	Squat depth / trunk length	unitless	15	0.0453 ± 0.0243	0.0480 ± 0.0241
	Knee flexion (visible side)	degree	15	8.6963 ± 2.4262	8.9577 ± 2.4092
	Trunk forward lean	degree	15	1.8821 ± 1.0911	1.9689 ± 1.0448

This difference was most visible for MINI knee flexion on the camera-facing side. The MAE increased from 4.4693° in the full-trial evaluation to 8.6963° in the active-segment evaluation, making this the most challenging metric under Condition 1. A likely reason is that the 2D knee-angle estimate depends on the relative positions of the hip, knee, and ankle landmarks. Small landmark errors, side-view visibility issues, or partial occlusion can therefore have a larger effect on this metric than on trunk-orientation or displacement measures.

The results also suggest that the performance of the pipeline was metric-dependent. Trunk-orientation metrics and normalized displacement metrics were generally more stable than the 2D knee-flexion metric. However, active-segment selection did not affect all metrics in the same way. For example, MINI trunk forward lean did not show the same increase in error as MINI knee flexion. This shows that the evaluation result depends not only on the exercise itself, but also on the metric definition and on which part of the trial is analysed.

The Bland–Altman plots in Figure 4.1 provide an additional view of the agreement pattern under Condition 1. In these plots, the horizontal axis represents the average of the MediaPipe-derived and MoCap-derived value for each trial, which indicates the approximate magnitude of the metric in that trial. The vertical axis represents the signed difference between the two methods, defined as MediaPipe minus MoCap. Therefore, points above zero indicate overestimation by MediaPipe, while points below zero indicate underestimation. Unlike Table 4.1 and Table 4.2, which summarize absolute error using MAE and RMSE, the Bland–Altman plots show the direction and spread of the signed differences between MediaPipe and MoCap. The trunk-orientation metrics showed relatively narrow limits of agreement compared with the more challenging knee-flexion metric. STS trunk forward lean showed a positive bias, indicating that MediaPipe tended to estimate larger forward-lean values than MoCap in these trials. In contrast, MINI knee flexion on the camera-facing side showed a negative bias, indicating underestimation by MediaPipe.

The displacement-based ratio metrics also showed negative mean bias values, suggesting that MediaPipe tended to underestimate the corresponding MoCap-derived displacement measures under Condition 1. For SLS pelvis horizontal displacement, the plot suggested a possible proportional pattern, where larger displacement values were associated with stronger underestimation. Since each plot was based on 15 balanced trials from Condition 1, these agreement patterns should be interpreted

descriptively. The numerical Bland–Altman summary values are provided in Appendix C.1.

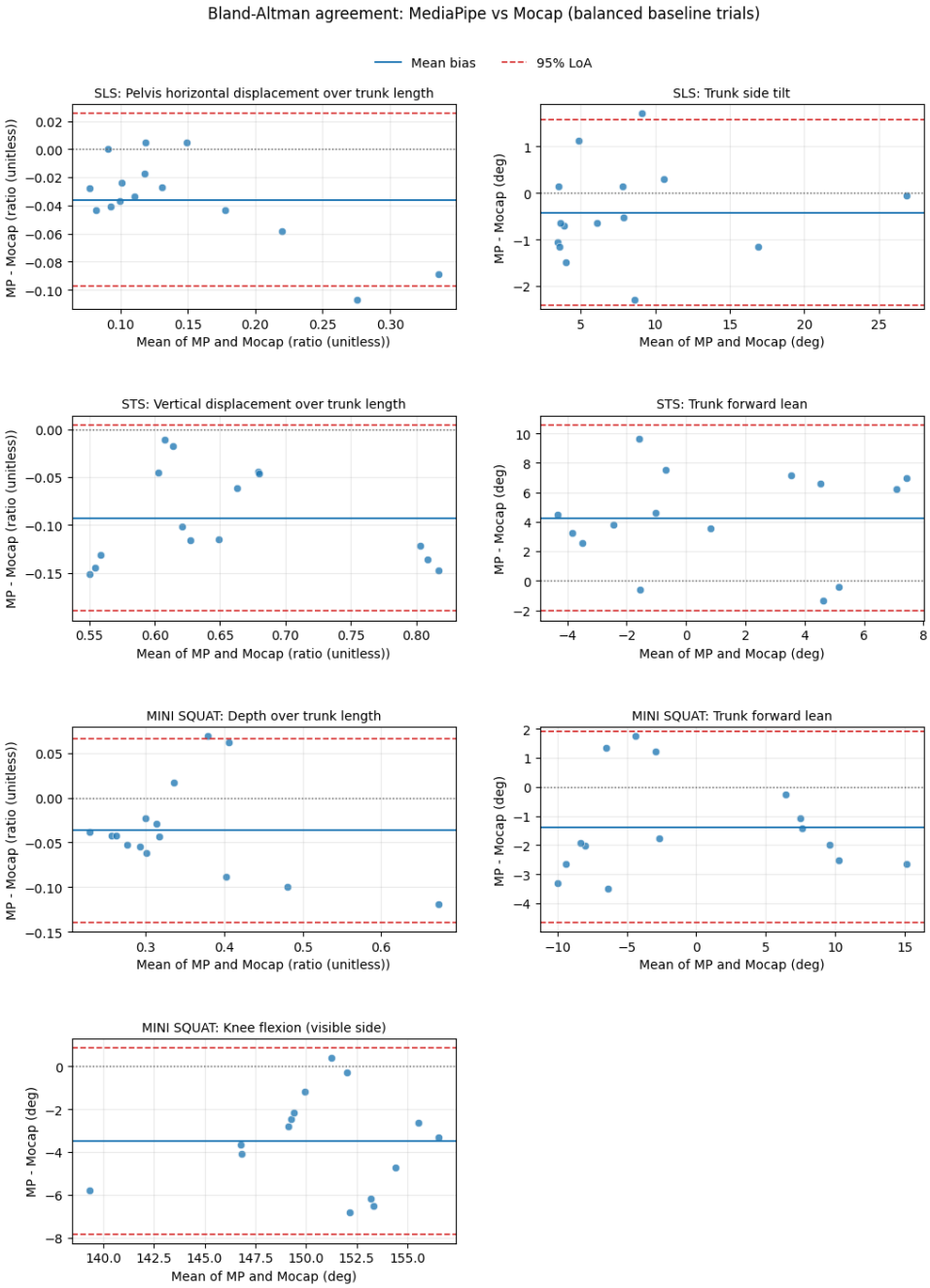


Figure 4.1: Bland-Altman agreement plots for the balanced Condition 1 trials. The horizontal axis shows the average of the MediaPipe-derived and MoCap-derived metric values for each trial, and the vertical axis shows the signed difference, defined as MediaPipe minus MoCap. The blue line indicates the mean bias, and the red dashed lines indicate the 95% limits of agreement.

4.2 Subject-Level Variability

Table 4.3: Subject-level variability in active-segment MAE for each subject.

Exercise	Metric	Unit	S01 (mean ± SD)	S02 (mean ± SD)	S03 (mean ± SD)	S04 (mean ± SD)	S05 (mean ± SD)
SLS	Pelvis horizontal displacement / trunk length	unitless	0.0076 ± 0.0018	0.0094 ± 0.0026	0.0767 ± 0.0089	0.0928 ± 0.0186	0.0094 ± 0.0039
	Trunk side tilt	degree	2.4462 ± 0.4314	1.2270 ± 0.6139	0.4184 ± 0.1473	0.9179 ± 0.3388	0.5361 ± 0.1226
STS	Trunk forward lean	degree	2.7336 ± 0.5472	2.3405 ± 0.3836	3.1370 ± 0.0381	5.5519 ± 1.3209	2.5811 ± 0.1153
	Pelvis vertical displacement / trunk length	unitless	0.0413 ± 0.0133	0.0515 ± 0.0093	0.0981 ± 0.0055	0.0891 ± 0.0141	0.0221 ± 0.0037
MINI	Squat depth / trunk length	unitless	0.0239 ± 0.0039	0.0169 ± 0.0074	0.0740 ± 0.0177	0.0499 ± 0.0065	0.0621 ± 0.0108
	Knee flexion (visible side)	degree	8.2653 ± 0.7999	11.9473 ± 0.4750	8.3921 ± 0.6863	9.5757 ± 1.1953	5.3031 ± 1.9352
	Trunk forward lean	degree	3.1686 ± 0.2366	1.8547 ± 0.2452	2.9186 ± 0.1752	0.8920 ± 0.2283	0.5766 ± 0.1129

The subject-level analysis compared the active-segment MAE values across the five participants. Table 4.3 reports the mean and standard deviation of the MAE for each subject and metric.

The size of the subject-level variation depended on the metric. Among the normalized displacement metrics, SLS pelvis horizontal displacement had the largest best-worst subject gap, followed by STS pelvis vertical displacement and MINI squat depth. Among the angular metrics, MINI visible side knee flexion showed the largest gap. The subjects with the highest errors were not the same for every metric, so the variability was not simply a case of one participant being consistently easier or harder to track.

These differences may come from several sources, including individual movement strategy, apparent body scale in the image, landmark visibility, and markerless tracking behaviour. With only five healthy participants, it is not possible to separate these factors or link the observed differences to demographic variables such as sex, height, or body type. The subject-level results should therefore be read as evidence that metric errors can vary between individuals even under the same baseline recording setup, not as a judgement of participant performance.

4.3 Robustness Across Recording Conditions

The robustness analysis compared active-segment errors under the available recording-condition variations with their matched Condition 1 trials. The tested variations were camera distance and lighting. Several conditions had only a small number of matched pairs, so the focus here is on the observed error patterns rather than on formal statistical conclusions.

4.3.1 Single-Leg Stance

Table 4.4 reports the SLS active-segment errors under the frontal-view setup.

Table 4.4: SLS active-segment condition results. Positive error difference values indicate increased MAE compared with matched Condition 1 trials.

Condition	Pelvis horizontal displacement MAE	Error difference [95% CI]	Trunk side tilt MAE	Error difference [95% CI]	Paired trials
Condition 2	0.0329	+0.0059 [-0.0011, +0.0129]	2.1681°	+0.8779° [-0.7934, +2.6443]	4
Condition 3	0.0412	-0.0016 [-0.0036, +0.0005]	1.3551°	−0.1958° [-0.4615, +0.0699]	2
Condition 4	0.0272	+0.0024 [-0.0021, +0.0050]	1.4660°	+0.4567° [-0.1664, +0.9332]	4
Condition 5	0.0404	+0.0099 [-0.0018, +0.0287]	1.4410°	+0.3171° [-0.3406, +0.8691]	6

For SLS, the two metrics responded differently to the tested conditions. Pelvis horizontal displacement changed only slightly across distance and lighting variations. Trunk side tilt showed larger changes, especially in condition 2 and condition 4. Condition 3 did not show the same increase. Because the matched-pair counts were small and the confidence intervals were wide, these results are best read as early signs of metric-specific sensitivity rather than as a ranking of the recording conditions.

4.3.2 Sit-to-Stand

Table 4.5 reports the STS active-segment errors under the sagittal-view setup.

Table 4.5: STS active-segment condition results. Positive error difference values indicate increased MAE compared with matched Condition 1 trials.

Condition	Trunk forward lean MAE	Error difference [95% CI]	Vertical displacement MAE	Error difference [95% CI]	Paired trials
Condition 2	2.6893°	−0.0862° [-1.2383, +0.7320]	0.0723	+0.0203 [+0.0098, +0.0410]	3
Condition 3	2.8753°	−0.1131° [-0.9250, +0.6987]	0.0823	+0.0204 [+0.0198, +0.0211]	2
Condition 4	2.7832°	+0.1917° [-0.3065, +0.5309]	0.0659	+0.0168 [+0.0004, +0.0390]	4
Condition 5	3.0986°	−0.2097° [-0.7126, +0.3424]	0.0752	+0.0093 [-0.0002, +0.0204]	6

For STS, pelvis vertical displacement was more affected by the tested recording variations than trunk forward lean. The vertical-displacement error increased for both distance changes and both lighting variations. Trunk forward lean did not show the same pattern, and in several conditions its error was close to or lower than the matched condition 1. In this dataset, the displacement-based STS metric was therefore more sensitive to recording changes than the sagittal trunk-orientation metric.

4.3.3 Mini-Squat

For MINI as shown in Table 4.6, the clearest change appeared in condition 2, where squat depth and average knee flexion had higher errors than their matched condition 1. Trunk forward lean changed less across the tested conditions. The lighting results were less consistent. In the available low-light and no-light pairs, knee-flexion

error was lower than in the matched condition 1 trials. This should not be read as an improvement caused by poorer lighting. A more likely explanation is the small and unbalanced set of matched comparisons, together with trial-specific variation.

Table 4.6: MINI active-segment condition results. Positive error difference values indicate increased MAE compared with matched Condition 1 trials.

Condition	Squat Depth MAE	Error difference [95% CI]	Knee flexion (visible side) MAE	Error difference [95% CI]	Trunk lean MAE	Error difference [95% CI]	Paired trials
Condition 2	0.0625	+0.0216 [-0.0155, +0.0602]	9.6579°	+1.1519° [-2.0525, +3.5144]	2.5130°	+0.0334° [-0.6402, +0.7069]	4
Condition 3	0.0583	-0.0029 [-0.0217, +0.0159]	7.1840°	−1.1164° [-3.7787, +1.5456]	3.0723°	−0.0062° [-0.2758, +0.2634]	2
Condition 4	0.0419	-0.0076 [-0.0247, +0.0084]	9.9107°	+1.2241° [-0.9198, +3.3337]	2.0122°	−0.1455° [-0.6000, +0.3091]	4
Condition 5	0.0492	-0.0031 [-0.0156, +0.0088]	8.8330°	+0.0011° [-2.4393, +2.3697]	2.7004°	+0.0828° [-0.8231, +1.1859]	6

Overall, the condition-level results show that robustness differed by exercise and metric. SLS trunk side tilt, STS pelvis vertical displacement, and MINI squat depth or knee flexion showed clearer changes under the tested conditions than the STS and MINI trunk-orientation metrics. Therefore, distance and lighting effects should be interpreted at the metric level, rather than as a single robustness value for the whole MediaPipe-based pipeline.

4.4 Repeatability

Repeatability is interpreted as the stability of the MediaPipe-MoCap error across repeated recordings. The main question here is whether MediaPipe stayed close to the MoCap reference when the same exercise was repeated within the same session or across different days.

Table 4.7: Repeatability of MediaPipe-MoCap error across repeated recordings.

Exercise	Metric	Within-session MAE (mean ± SD)	Within-session bias (mean ± SD)	Within-session error change (mean ± SD)	Across-day MAE (mean ± SD)	Across-day bias (mean ± SD)	Across-day error change (mean ± SD)
SLS	Pelvis horizontal displacement / trunk length (ratio)	0.0432 ± 0.0285	-0.0423 ± 0.0299	0.0250 ± 0.0127	0.0410 ± 0.0238	-0.0410 ± 0.0238	0.0209 ± 0.0079
	Trunk side tilt (degree)	1.10 ± 0.92	-0.63 ± 1.30	0.72 ± 0.57	1.09 ± 0.88	-0.47 ± 1.31	1.29 ± 0.31
STS	Trunk forward lean (degree)	4.65 ± 2.74	4.26 ± 3.34	2.84 ± 2.65	4.10 ± 2.34	2.95 ± 3.64	3.92 ± 2.05
	Pelvis vertical displacement / trunk length (ratio)	0.1050 ± 0.0670	-0.1050 ± 0.0670	0.0302 ± 0.0231	0.0947 ± 0.0712	-0.0947 ± 0.0712	0.0434 ± 0.0234
MINI	Depth / trunk length (ratio)	0.0501 ± 0.0318	-0.0368 ± 0.0472	0.0180 ± 0.0123	0.0430 ± 0.0304	-0.0408 ± 0.0336	0.0310 ± 0.0062
	Knee flexion, visible side (degree)	3.77 ± 2.41	-3.09 ± 3.27	1.71 ± 1.18	4.53 ± 2.97	-4.11 ± 3.57	3.01 ± 1.71
	Trunk forward lean (degree)	2.21 ± 1.61	-0.48 ± 2.73	1.12 ± 1.50	2.21 ± 1.59	0.07 ± 2.76	2.12 ± 2.32

The repeatability results show that the stability of the MediaPipe-MoCap error depended strongly on the selected metric. The SLS metrics, MINI depth, and MINI trunk forward lean showed more stable agreement with MoCap across repeated recordings. In contrast, STS trunk forward lean and MINI knee flexion on the camera-facing side were more sensitive to repeated-recording variation, especially in the across-day comparison.

For MINI, the depth metric was more repeatable than knee flexion. This supports the pattern seen in the condition 1 results, where displacement-based metrics were generally more dependable than knee-angle estimation. Although using the camera-facing leg made the knee-flexion metric more appropriate for the side-view setup, it still remained sensitive to landmark visibility, small changes in body orientation, and the limitations of 2D angle estimation.

4.5 Illustrative Feedback and User-Interface Outputs

The feedback prototype was included to show how the computed movement metrics could be presented after pose processing. Two output formats were implemented: real-time feedback for the exercising user and a session-review dashboard for physiotherapists. The prototype was not clinically validated, but it provides an example of how the metric pipeline could be turned into more interpretable outputs.

4.5.1 Patient-Facing Feedback

The patient-facing interface converted pose-derived measurements into short messages during exercise execution. The messages covered posture correction, balance guidance, repetition counting, compensation warnings, and camera or pose-detection issues. The purpose was to give action-oriented guidance during the exercise, instead of showing raw landmark coordinates or numerical error values.

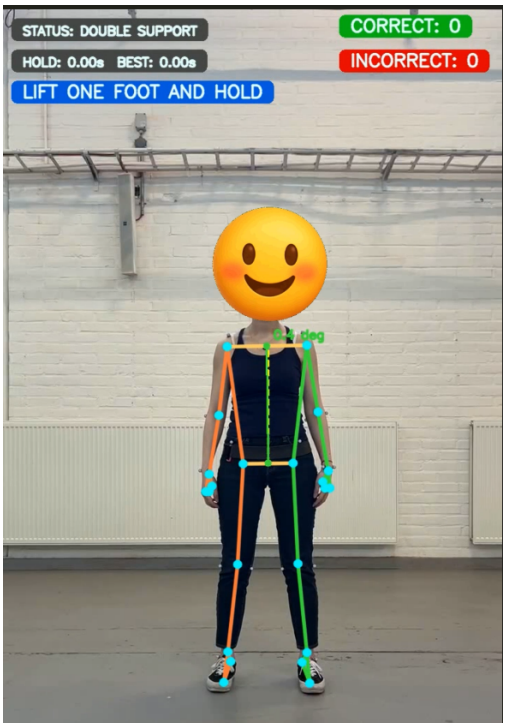


Figure 4.2: Preparation-stage patient-facing interface for SLS.

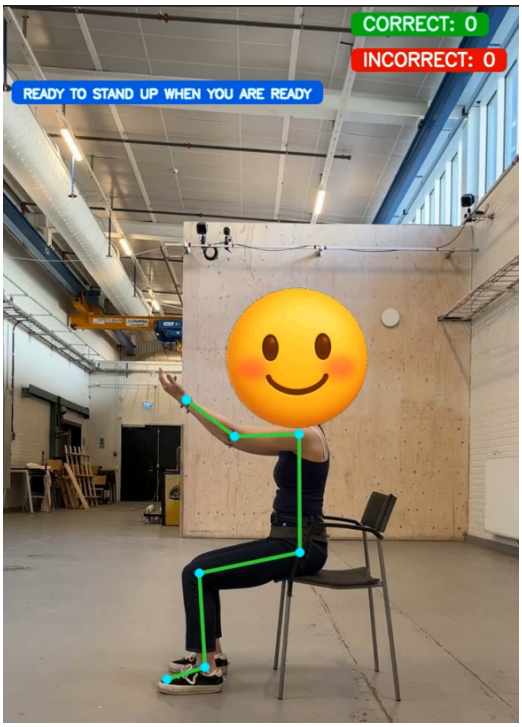


Figure 4.3: Preparation-stage patient-facing interface for STS.

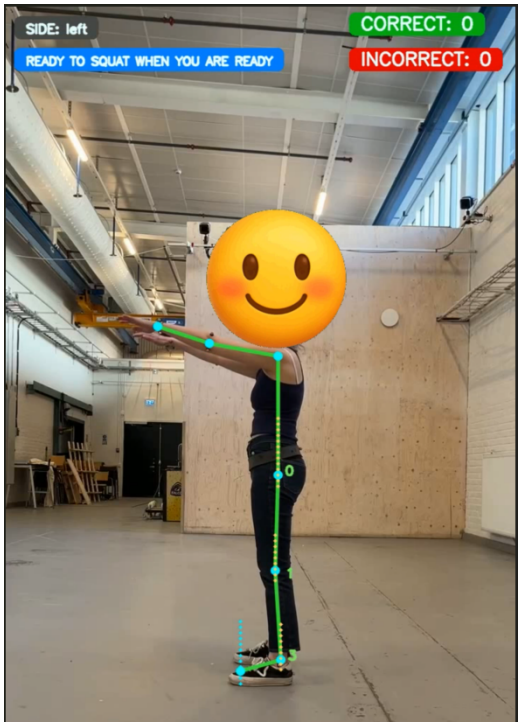


Figure 4.4: Preparation-stage patient-facing interface for MINI.

Tables 4.8–4.10 list the feedback messages implemented for each exercise and the corresponding trigger conditions.

Table 4.8: Subject-interface feedback messages for SLS.

Subject-interface feedback	**Trigger / interpretation**
LIFT ONE FOOT AND HOLD	The participant is in starting position
LIFT YOUR FOOT HIGHER	The foot is lifted, but the height is below the hold threshold
GOOD POSTURE	Trunk lateral lean is within the acceptable range
TRUNK LEANING. RETURN TO CENTER	Trunk lateral lean exceeds the warning threshold
EXCESSIVE TRUNK LATERAL LEAN	Trunk lateral lean exceeds the compensation threshold
GREAT BALANCE	Pelvis sway during the hold is very small
GOOD BALANCE	Pelvis sway remains within the acceptable range
KEEP HIPS STEADY	Pelvis sway exceeds the warning threshold
ARM COMPENSATION	Arm elevation compensation is detected during the hold
CAMERA NOT ALIGNED PROPERLY	The frontal camera view is not sufficiently reliable
NO POSE DETECTED	MediaPipe Pose does not detect a valid body pose

Table 4.9: Subject-interface feedback messages for STS.

Subject-interface feedback	**Trigger / interpretation**
RAISE ARMS IN FRONT	The participant is seated, but the arms are not in the required starting position
READY TO STAND UP WHEN YOU ARE READY	The starting posture is detected as ready
LEAN FORWARD MORE	More forward trunk lean is needed
KEEP GOING. STAND UP FULLY	The participant is extending toward standing
USE OF HANDS	Hand support is detected during the movement
EXCESSIVE TRUNK MOMENTUM	Excessive trunk-driven movement is detected
INCOMPLETE EXTENSION	The final standing posture is not fully extended
COMPLETE! SIT-TO-STAND: x.x s	A sit-to-stand repetition is completed and the repetition duration is displayed
SIT DOWN TO START	Standing posture is detected before exercise session
CAMERA NOT ALIGNED PROPERLY	The sagittal camera view is not sufficiently reliable
NO POSE DETECTED	MediaPipe Pose does not detect a valid body pose

Table 4.10: Subject-interface feedback messages for MINI.

Subject-interface feedback	**Trigger / interpretation**
READY TO SQUAT WHEN YOU ARE READY	The participant is in the standing posture before exercise
GOOD FORM. KEEP GOING	No active fault feedback is currently triggered
LOWER YOUR HIPS	Squat depth is insufficient
SQUAT TOO DEEP	Squat depth exceeds the expected range
BEND BACKWARDS	Trunk forward lean angle exceeds the threshold but below the excessive forward lean threshold
BEND FORWARD	More forward leaning is needed
EXCESSIVE FORWARD LEAN	Trunk forward lean exceeds the compensation threshold
EXCESSIVE HIP FLEXION	Hip flexion exceeds the compensation threshold
KNEE FALLING OVER TOE	The knee translates forward beyond the toe reference
HANDS ON THIGHS	Hands are detected on the thighs during the movement
CAMERA NOT ALIGNED PROPERLY	The sagittal camera view is not sufficiently reliable
NO POSE DETECTED	MediaPipe Pose does not detect a valid body pose

4.5.2 Therapist-Facing Review Dashboard

The therapist-facing dashboard was implemented for asynchronous session review. After each monitored session, the pipeline exported a structured JSON file containing exercise summaries, repetition-level labels, selected frame-level metrics, and review status information. These saved sessions could then be opened in a Streamlit-based dashboard.

The dashboard included a session list as shown in Figure 4.5, a session detail view with within-session time-series plots and a therapist feedback as shown from Figure 4.6a to Figure 4.6c. The detail view presented exercise-specific summaries, including repetition counts, correct and incorrect repetitions, best hold time for SLS, fastest repetition time for STS, and movement-quality indicators for individual repetitions. When frame-level metrics were available, selected time-series signals were plotted to support a more detailed review.

The two interfaces used the same underlying pose-derived information in different ways. The patient-facing view emphasized immediate guidance during exercise execution, while the therapist-facing dashboard emphasized later review, session summaries, and follow-up. This separation was useful for the prototype because the two user groups need different levels of detail from the same monitoring pipeline.

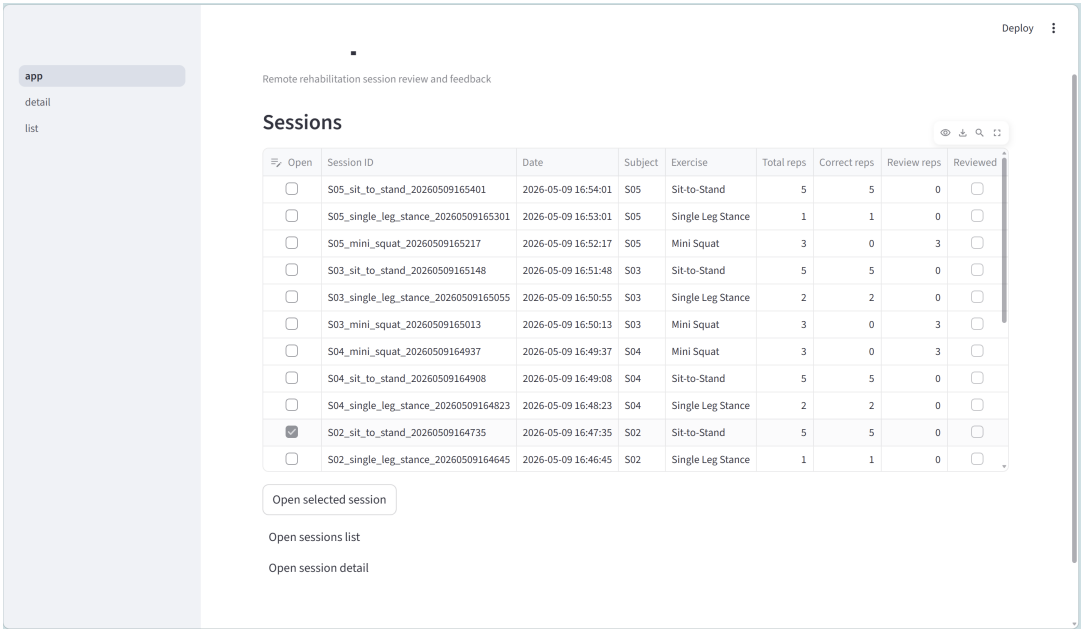
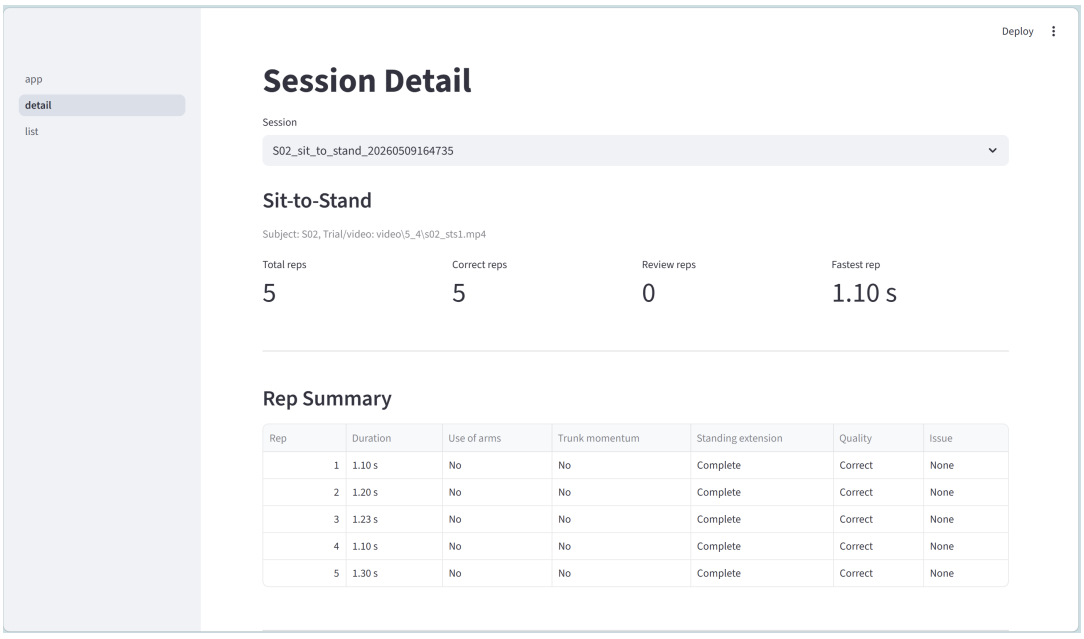


Figure 4.5: Session list view of the therapist-facing dashboard.

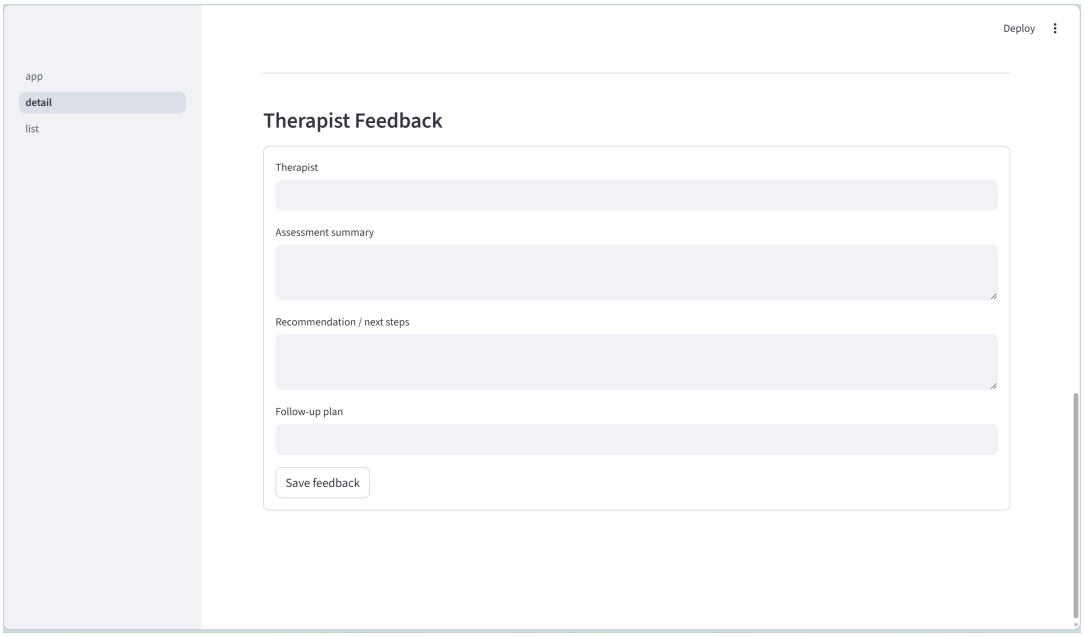


(a) Session detail view

Figure 4.6: Session detail views from the therapist-facing dashboard.

(b) Within-session time-series plots

Figure 4.6: Session detail views from the therapist-facing dashboard continued.



(c) Therapist feedback form

Figure 4.6: Session detail views from the therapist-facing dashboard continued.

5

Discussion

5.1 Main Findings

This thesis evaluated a MediaPipe Pose-based pipeline for monitoring the selected rehabilitation-oriented exercises with a single RGB camera. The main finding is that the pipeline did not perform equally well for all exercises or metrics. Trunk-orientation metrics and some normalized displacement metrics were the most consistent, especially SLS trunk side tilt and MINI trunk forward lean. In contrast, MINI knee flexion on the camera-facing side produced the largest active-segment error, showing the difficulty of estimating lower-limb joint angles from a 2D camera view in this setup.

The distinction between full-trial and active-segment evaluation was important. Full-trial errors were often lower because the recordings included preparation phases, static periods, and other parts of the trial where little movement occurred. Active-segment evaluation gave a stricter view of performance during the actual exercise movement, and was therefore more relevant for judging exercise execution quality.

The results also showed clear variation between participants and recording conditions. Some metrics were more affected by subject-specific movement patterns, while others changed more under distance or lighting variations. This means that the pipeline should be judged by individual metric and exercise, rather than by treating MediaPipe Pose as equally accurate across all movement tasks.

The repeatability analysis focused on the stability of the MediaPipe-MoCap error across repeated recordings. The results showed that the error was generally more stable within the same session than across different days, although the pattern depended on the metric. SLS trunk side tilt, SLS pelvis horizontal displacement, MINI depth, and MINI trunk forward lean showed relatively stable errors. In contrast, STS trunk forward lean and MINI visible-side knee flexion were more sensitive to repeated recording and day-to-day variation. This indicates that repeatability in this study should be understood as the reproducibility of the agreement with MoCap, rather than only as the internal consistency of MediaPipe-derived measurements.

5.2 Limitations

The results should be interpreted in relation to the size and composition of the dataset. Only five healthy participants were included, and the exercises were performed under structured laboratory conditions. The subject-level results are therefore useful for identifying possible variation between participants, but they are not

sufficient to describe how the system would perform in a wider clinical population. Patients with pain, reduced mobility, neurological impairments, or stronger compensatory strategies may produce different movement patterns from those observed in this study.
The laboratory setup also limited how far the findings can be transferred to real home use. Although distance and lighting were varied, the recordings did not include the full range of home-environment challenges, such as furniture, background clutter, loose clothing, partial occlusion, or inconsistent camera placement. For this reason, the robustness results should be read as an evaluation of selected recording changes, not as a complete test of home deployment.
The use of a single RGB camera made the system simple and practical, but it also constrained the types of measurements that could be estimated reliably. Metrics that were mostly visible in the camera plane, such as trunk orientation and normalized displacement, were generally more stable. In contrast, depth-related or out-of-plane movement was harder to assess, which helps explain the larger errors observed for lower-limb joint-angle metrics such as MINI knee flexion.
Another limitation was that the RGB video and MoCap systems were not hardware-synchronized. Offline signal-based alignment enabled frame-wise comparison, but it may still introduce uncertainty when the alignment signal is weak or when the movement contains only small dynamic changes. This uncertainty should be considered when interpreting time-dependent errors and active-segment comparisons.
Finally, the feedback prototype was rule-based and exploratory. The thresholds were used to demonstrate how pose-derived metrics could be translated into patient-facing and physiotherapist-facing outputs, but they were not clinically validated. The interface was also not tested with patients or physiotherapists, so no conclusion can be made about usability, clinical usefulness, or user acceptance.

5.3 Practical Challenges for Real-World Use

Although the results show that some MediaPipe-derived metrics can be useful under controlled conditions, using this type of system in real rehabilitation practice would involve several additional challenges. A camera-based system depends not only on the pose-estimation model, but also on how consistently the recording setup is used. Patients would need to place the camera at a suitable distance and angle, keep the full body visible, avoid occlusion, and follow exercise instructions without direct supervision. Small changes in camera position, lighting, clothing, room layout, or body orientation may affect landmark quality and therefore change the computed metrics.
Another practical challenge is interpretation. The system can produce numerical metrics and rule-based feedback, but these outputs still need to be meaningful for both patients and physiotherapists. Feedback that is too sensitive may confuse or discourage patients, while feedback that is too general may not be useful for correcting movement quality. For physiotherapists, the system should summarize the most relevant information without creating too much extra review work.
Real-world use would also require attention to privacy and data handling. Video-based rehabilitation monitoring may capture identifiable body images and parts of

the home environment. Therefore, future implementations would need clear consent procedures, secure data storage, and careful decisions about whether to store raw videos, pose landmarks, or only summary metrics. These issues are outside the technical evaluation in this thesis, but they are important before a camera-based rehabilitation system can be used in practice.

5.4 Future Work

Future work should include a larger and more diverse dataset. This should include more participants, a wider range of body types and clothing conditions, and eventually clinical participants performing rehabilitation exercises. A larger dataset would make it possible to examine subject-level effects more systematically and to test whether the observed metric-level patterns remain stable beyond a small healthy-participant sample.

The system should also be evaluated in more realistic home environments. Future recordings should include uncontrolled backgrounds, furniture, partial occlusion, different lighting conditions, and more varied camera placements. Clear camera-placement guidance will be important, since the results showed that metric interpretation depends on matching the camera view to the dominant movement plane.

Further refinement is also needed in the processing pipeline and metric definitions. Lower-limb joint-angle estimation, especially MINI knee flexion, should be improved before it can be used with confidence for rehabilitation monitoring. Future studies could compare different pose-estimation models, add camera calibration, use hardware synchronization during validation, or combine RGB pose estimation with additional sensors when higher accuracy is required.

The feedback prototype should be evaluated with physiotherapists and potential users. Clinical input would help refine the compensation rules, threshold values, and feedback messages. Usability testing would also be needed to determine whether patients can understand and act on the feedback, and whether physiotherapists find the session summaries useful for remote follow-up.

6
Conclusion

The implemented pipeline shows that single-camera pose estimation can provide useful information for selected rehabilitation monitoring tasks when the exercise, camera view, and metric definition are carefully chosen. The strongest results were obtained for trunk-orientation and normalized displacement metrics, while lower-limb joint-angle estimation remained more limited. MediaPipe Pose should therefore be viewed as a practical support tool for structured home exercise monitoring, rather than as a direct substitute for laboratory motion capture. Further testing with more participants, clinical users, and real home environments is needed before such a system can be considered for practical rehabilitation use.

Declaration of AI Tool Usage

AI tools were used as supporting tools during the preparation of this thesis. They were used to support literature reading, online information searching, clarification of technical concepts and formulas, and brainstorming of ideas and interpretations. AI tools were also used to discuss alternative viewpoints, improve the clarity of written explanations, and assist with LaTeX formatting of paragraphs, tables, and equations.
AI tools were also used as support during the coding and analysis work. For example, they helped me understand parts of the code structure, discuss possible implementation approaches, and troubleshoot errors. However, the implementation, analysis workflow, numerical results were checked by me against the project code, data outputs, and thesis scope before they were included in the thesis.

Bibliography

[1] Google AI Edge. (2024). *MediaPipe Pose documentation*. Available at: `https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/pose.md` Accessed: 2026-05-14.

[2] Scott, B., Seyres, M., Philp, F., Chadwick, E. K., and Blana, D. (2022). *Healthcare applications of single camera markerless motion capture: a scoping review*. PeerJ, 10, e13517.

[3] Lam, W. W. T., Tang, Y. M., and Fong, K. N. K. (2023). *A systematic review of the applications of markerless motion capture technology for clinical measurement in rehabilitation*. Journal of NeuroEngineering and Rehabilitation, 20, 57.

[4] Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., and Grundmann, M. (2020). *BlazePose: On-device Real-time Body Pose Tracking*. CoRR, abs/2006.10204.

[5] Grishchenko, I., Bazarevsky, V., Zanfir, A., Bazavan, E. G., Zanfir, M., Yee, R., Raveendran, K., Zhdanovich, M., Grundmann, M., and Sminchisescu, C. (2022). *BlazePose GHUM Holistic: Real-time 3D Human Landmarks and Pose Estimation*. CoRR, abs/2206.11678.

[6] Pardell, M., Dolgoy, N. D., Bernard, S., Bayless, K., Hirsche, R., Dennett, L., and Tandon, P. (2024). *Movement Outcomes Acquired via Markerless Motion Capture Systems Compared with Marker-Based Systems for Adult Patient Populations: A Scoping Review*. *Biomechanics*, 4(4), 618–632.

[7] Hii, C. S. T., Gan, K. B., Zainal, N., Mohamed Ibrahim, N., Azmin, S., Mat Desa, S. H., van de Warrenburg, B., and You, H. W. (2023). Automated gait analysis based on a marker-free pose estimation model. *Sensors*, 23(14), 6367.

[8] Yang, J. and Park, K. (2024). Improving gait analysis techniques with markerless pose estimation based on smartphone location. *Bioengineering*, 11(2), 174.

[9] Ferraris, C., Amprimo, G., Cerfoglio, S., Vismara, L., and Cimolin, V. (2025). *A deep dive into MediaPipe Pose for postural assessment: a comparative investigation*. *IEEE Access*, 13, 211055–211074.

[10] Dill, S., Rösch, A., Rohr, M., Güney, G., Witte, L., Schwartz, E., and Hoog Antink, C. (2023). Accuracy evaluation of 3D pose estimation with MediaPipe Pose for physical exercises. *Current Directions in Biomedical Engineering*, 9(1), 563–566.

[11] Wagh, V., Scott, M. W., and Kraeutner, S. N. (2024). *Quantifying Similarities Between MediaPipe and a Known Standard to Address Issues in Tracking 2D Upper Limb Trajectories: Proof of Concept Study*. *JMIR Formative Research*, 8, e56682.

[12] Wagh, V., Scott, M. W., Andrushko, J. W., Jones, C. B., Larssen, B. C., Boyd, L. A., and Kraeutner, S. N. (2025). *Using MediaPipe to track upper-limb reaching movements after stroke: a proof-of-principle study. Journal of NeuroEngineering and Rehabilitation*, 22(1), 268.

[13] QuickPose AI. (2026). *QuickPose AI: AI pose estimation solutions and SDK documentation.* Available at: `https://quickpose.ai/` and `https://docs.quickpose.ai/` Accessed: 2026-05-04.

[14] UCSF Department of Orthopaedic Surgery. (2026). *OrthoCAP.* Available at: `https://orthosurgery.ucsf.edu/research/resources/orthocap-app` Accessed: 2026-05-04.

[15] Roggio, F., Trovato, B., Sortino, M., and Musumeci, G. (2024). *A comprehensive analysis of the machine learning pose estimation models used in human movement and posture analyses: A narrative review. Heliyon*, 10(21), e39977.

[16] Das, K., de Paula Oliveira, T., and Newell, J. (2023). *Comparison of markerless and marker-based motion capture systems using 95% functional limits of agreement in a linear mixed-effects modelling framework. Scientific Reports*, 13, 22880.

[17] Scataglini, S., Abts, E., Van Bocxlaer, C., Van den Bussche, M., Meletani, S., and Truijen, S. (2024). *Accuracy, Validity, and Reliability of Markerless Camera-Based 3D Motion Capture Systems versus Marker-Based 3D Motion Capture Systems in Gait Analysis: A Systematic Review and Meta-Analysis. Sensors*, 24(11), 3686.

[18] Seron, P., Oliveros, M.-J., Gutierrez-Arias, R., Fuentes-Aspe, R., Torres-Castro, R. C., Merino-Osorio, C., Nahuelhual, P., Inostroza, J., Jalil, Y., Solano, R., Marzuca-Nassr, G. N., Aguilera-Eguía, R., Lavados-Romo, P., Soto-Rodríguez, F. J., Sabelle, C., Villarroel-Silva, G., Gomolán, P., Huaiquilaf, S., and Sanchez, P. (2021). *Effectiveness of Telerehabilitation in Physical Therapy: A Rapid Overview. Physical Therapy*, 101(6), pzab053.

[19] Simmich, J., Ross, M. H., and Russell, T. (2024). *Real-time video telerehabilitation shows comparable satisfaction and similar or better attendance and adherence compared with in-person physiotherapy: a systematic review. Journal of Physiotherapy*, 70(3), 181–192.

[20] Amin, J., Ahmad, B., Amin, S., Siddiqui, A. A., and Alam, M. K. (2022). *Rehabilitation Professional and Patient Satisfaction with Telerehabilitation of Musculoskeletal Disorders: A Systematic Review. BioMed Research International*, 2022, 7366063.

[21] Nordin, N., Xie, S. Q., and Wünsche, B. (2014). *Assessment of movement quality in robot-assisted upper limb rehabilitation after stroke: a review. Journal of NeuroEngineering and Rehabilitation*, 11, 137.

[22] Sellmann, A., Wagner, D., Holtz, L., Eschweiler, J., Diers, C., Williams, S., and Disselhorst-Klug, C. (2022). *Detection of Typical Compensatory Movements during Autonomously Performed Exercises Preventing Low Back Pain (LBP). Sensors*, 22(1), 111.

[23] Gogu, S. and Gandbhir, V. N. (2022). *Trendelenburg Sign.* In: StatPearls. Treasure Island (FL): StatPearls Publishing.

[24] Prior, S., Mitchell, T., Whiteley, R., O'Sullivan, P., Williams, B. K., Racinais, S., and Farooq, A. (2014). *The influence of changes in trunk and pelvic posture during single leg standing on hip and thigh muscle activation in a pain free population. BMC Sports Science, Medicine and Rehabilitation*, 6, 13.

[25] Janssen, W. G. M., Bussmann, H. B. J., and Stam, H. J. (2002). *Determinants of the sit-to-stand movement: A review. Physical Therapy*, 82(9), 866–879.

[26] van der Kruk, E., Strutton, P., Koizia, L. J., Fertleman, M., Reilly, P., and Bull, A. M. J. (2022). *Why do older adults stand-up differently to young adults?: investigation of compensatory movement strategies in sit-to-walk. npj Aging*, 8, 13.

[27] Ageberg, E., Bennell, K. L., Hunt, M. A., Simic, M., Roos, E. M., and Creaby, M. W. (2010). *Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat. BMC Musculoskeletal Disorders*, 11, 265.

[28] Zawadka, M., Smolka, J., Skublewska-Paszkowska, M., Lukasik, E., and Gawda, P. (2025). *Effect of Limited Ankle Dorsiflexion on Lower Limbs and Trunk Kinematics During Squatting. Journal of Manipulative and Physiological Therapeutics.*

[29] Millor, N., Lecumberri, P., Gomez, M., Martinez-Ramirez, A., and Izquierdo, M. (2014). *Kinematic parameters to evaluate functional performance of sit-to-stand and stand-to-sit transitions using motion sensor devices: a systematic review. IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(5), 926–936.

[30] OpenStax. (2022). *Anatomy and Physiology 2e, Section 1.6: Anatomical Terminology.* Rice University. Available at: `https://openstax.org/books/anatomy-and-physiology-2e/pages/1-6-anatomical-terminology` Accessed: 2026-05-09.

[31] Bland, J. M. and Altman, D. G. (1986). *Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet*, 327(8476), 307–310.

[32] OpenCV. *cv::VideoCapture Class Reference.* Available at: `https://docs.opencv.org/4.x/d8/dfe/classcv_1_1VideoCapture.html` (accessed 19 May 2026).

A

Appendix A - Source Code for Video-Based Landmark Extraction

The following Python script was used to extract pose landmarks from recorded videos using MediaPipe Pose and save them as long-format CSV files.

Listing A.1: Python script for extracting MediaPipe Pose landmarks from video recordings.

```
from __future__ import annotations

import csv
from datetime import datetime
from pathlib import Path
from typing import Sequence

import cv2
import mediapipe as mp


INPUT_PATH = Path(r"video\5_5")
OUTPUT_DIR = Path(r"Outputs\5_5")
RECURSIVE = False

VIDEO_EXTENSIONS = {".mp4", ".mov", ".avi", ".mkv", ".m4v", ".wmv"}

MIN_DETECTION_CONFIDENCE = 0.5
MIN_TRACKING_CONFIDENCE = 0.5
MODEL_COMPLEXITY = 1
SMOOTH_LANDMARKS = True

LONGFORM_FIELDNAMES = [
    "frame",
    "timestamp_s",
    "input_mode",
    "landmark_id",
    "landmark_name",
    "x",
    "y",
    "z",
    "visibility",
]


def list_videos(input_dir: Path, recursive: bool = False) -> list[Path]:

    candidates = input_dir.rglob("*") if recursive else input_dir.glob("*")
    videos = [
        path
        for path in candidates
        if path.is_file() and path.suffix.lower() in VIDEO_EXTENSIONS
    ]
    return sorted(videos)


```

```
47 def unique_path(path: Path) -> Path:
48
49     if not path.exists():
50         return path
51
52     index = 2
53     while True:
54         candidate = path.with_name(f"{path.stem}_{index}{path.suffix}")
55         if not candidate.exists():
56             return candidate
57         index += 1
58
59
60 def write_longform_landmarks(
61     writer: csv.DictWriter,
62     frame_idx: int,
63     timestamp_s: float,
64     landmarks: Sequence,
65     landmark_names: Sequence[str],
66 ) -> None:
67
68     for landmark_id, landmark in enumerate(landmarks):
69         writer.writerow(
70             {
71                 "frame": frame_idx,
72                 "timestamp_s": f"{timestamp_s:.4f}",
73                 "input_mode": "video",
74                 "landmark_id": landmark_id,
75                 "landmark_name": landmark_names[landmark_id],
76                 "x": f"{landmark.x:.6f}",
77                 "y": f"{landmark.y:.6f}",
78                 "z": f"{landmark.z:.6f}",
79                 "visibility": f"{landmark.visibility:.6f}",
80             }
81         )
82
83
84 def extract_video_landmarks(
85     video_path: Path,
86     output_dir: Path,
87     pose: object,
88     landmark_names: list[str],
89 ) -> dict[str, object]:
90
91     cap = cv2.VideoCapture(str(video_path))
92     if not cap.isOpened():
93         return {
94             "video": str(video_path),
95             "status": "error",
96             "message": "Cannot␣open␣video",
97             "frames": 0,
98             "detected_frames": 0,
99             "csv_path": None,
100         }
101
102     fps = cap.get(cv2.CAP_PROP_FPS)
103     if fps is None or fps <= 0:
104         fps = 30.0
105
106     run_date = datetime.now().strftime("%Y%m%d")
107     csv_path = unique_path(output_dir / f"{video_path.stem}_landmarks_{run_date}.
        csv")
108
109     frame_idx = 0
110     detected_frames = 0
111
112     with csv_path.open("w", newline="", encoding="utf-8") as csv_file:
113         writer = csv.DictWriter(csv_file, fieldnames=LONGFORM_FIELDNAMES)
114         writer.writeheader()
115
```

```
        while cap.isOpened():
            success, frame = cap.read()
            if not success:
                break

            frame_idx += 1
            timestamp_s = frame_idx / float(fps)

            image_rgb = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)
            image_rgb.flags.writeable = False
            results = pose.process(image_rgb)
            image_rgb.flags.writeable = True

            if results.pose_landmarks:
                detected_frames += 1
                write_longform_landmarks(
                    writer,
                    frame_idx,
                    timestamp_s,
                    results.pose_landmarks.landmark,
                    landmark_names,
                )

    cap.release()

    return {
        "video": str(video_path),
        "status": "ok",
        "message": "",
        "frames": frame_idx,
        "detected_frames": detected_frames,
        "csv_path": str(csv_path),
    }


def main() -> None:
    OUTPUT_DIR.mkdir(parents=True, exist_ok=True)

    if INPUT_PATH.is_file():
        if INPUT_PATH.suffix.lower() not in VIDEO_EXTENSIONS:
            raise ValueError(f"Unsupported video extension: {INPUT_PATH.suffix}")
        videos = [INPUT_PATH]
    elif INPUT_PATH.is_dir():
        videos = list_videos(INPUT_PATH, recursive=RECURSIVE)
    else:
        raise FileNotFoundError(f"Input path does not exist: {INPUT_PATH.resolve()}
            ")

    if not videos:
        raise FileNotFoundError(f"No videos found in: {INPUT_PATH.resolve()}")

    mp_pose = mp.solutions.pose
    landmark_names = [landmark.name for landmark in mp_pose.PoseLandmark]

    results = []
    with mp_pose.Pose(
        model_complexity=MODEL_COMPLEXITY,
        smooth_landmarks=SMOOTH_LANDMARKS,
        min_detection_confidence=MIN_DETECTION_CONFIDENCE,
        min_tracking_confidence=MIN_TRACKING_CONFIDENCE,
    ) as pose:
        total = len(videos)
        for index, video_path in enumerate(videos, start=1):
            result = extract_video_landmarks(
                video_path,
                OUTPUT_DIR,
                pose,
                landmark_names,
            )
            results.append(result)
```

```
185             print(
186                 f"[{index}/{total}] {video_path.name} | "
187                 f"status={result['status']} | "
188                 f"frames={result['frames']} | "
189                 f"detected={result['detected_frames']}"
190             )
191 
192     ok_count = sum(1 for result in results if result["status"] == "ok")
193     error_count = len(results) - ok_count
194     print(f"Done. videos={len(results)}, ok={ok_count}, error={error_count}")
195 
196 
197 if __name__ == "__main__":
198     main()
```

B

Appendix B - Additional Full-Trial Evaluation Tables

The main text presents the active-segment comparison tables for the subject-level and condition-level analyses. The corresponding full tables are provided here for completeness.

Table B.1: Subject-level variability in full-trial MAE for each subject.

Exercise	Metric	Unit	S01 (mean ± SD)	S02 (mean ± SD)	S03 (mean ± SD)	S04 (mean ± SD)	S05 (mean ± SD)
SLS	Pelvis horizontal displacement / trunk length	unitless	0.0125 ± 0.0043	0.0311 ± 0.0079	0.0340 ± 0.0069	0.0466 ± 0.0106	0.0133 ± 0.0052
	Trunk side tilt	degree	2.1366 ± 0.2470	0.9532 ± 0.3379	0.8666 ± 0.0905	0.6831 ± 0.0510	0.5972 ± 0.0784
STS	Trunk forward lean	degree	3.0262 ± 0.1832	3.2288 ± 0.3358	2.3192 ± 0.0752	5.0440 ± 0.4219	2.2036 ± 0.1022
	Pelvis vertical displacement / trunk length	unitless	0.0400 ± 0.0039	0.0473 ± 0.0134	0.0712 ± 0.0086	0.0795 ± 0.0007	0.0201 ± 0.0050
MINI	Squat depth / trunk length	unitless	0.0110 ± 0.0004	0.0097 ± 0.0027	0.0339 ± 0.0098	0.0251 ± 0.0022	0.0381 ± 0.0048
	Knee flexion (visible side)	degree	4.1300 ± 0.0796	4.9620 ± 0.7124	3.0930 ± 0.4977	4.2050 ± 0.8161	2.3678 ± 0.3092
	Trunk forward lean	degree	3.5478 ± 0.1540	1.5611 ± 0.1727	3.0125 ± 0.1383	1.4536 ± 0.3433	1.4402 ± 0.1166

Table B.2: SLS full-segment condition results. Positive error difference values indicate increased MAE compared with matched condition 1 trials.

Condition	Pelvis horizontal displacement MAE	Error difference [95% CI]	Trunk side tilt MAE	Error difference [95% CI]	Paired trials
Condition 2	0.0279	+0.0107 [+0.0006, +0.0208]	2.3024°	+1.0247° [-0.5628, +3.0767]	4
Condition 3	0.0285	+0.0053 [+0.0043, +0.0062]	1.3355°	−0.3031° [-0.4869, -0.1192]	2
Condition 4	0.0249	+0.0052 [-0.0002, +0.0108]	1.4768°	+0.3878° [+0.0196, +0.6570]	4
Condition 5	0.0322	+0.0103 [-0.0027, +0.0295]	1.2829°	+0.1161° [-0.2807, +0.4772]	6

Table B.3: STS full-segment condition results. Positive error difference values indicate increased MAE compared with matched condition 1 trials.

Condition	Trunk forward lean MAE	Error difference [95% CI]	Pelvis vertical displacement MAE	Error difference [95% CI]	Paired trials
Condition 2	2.7563°	+0.1757° [-0.1834, +0.6394]	0.0592	+0.0189 [-0.0027, +0.0354]	3
Condition 3	2.8687°	+0.2393° [+0.2346, +0.2440]	0.0592	+0.0103 [+0.0009, +0.0198]	2
Condition 4	2.8365°	+0.0672° [-0.3074, +0.4010]	0.0511	+0.0073 [-0.0053, +0.0224]	4
Condition 5	3.0613°	−0.1018° [-0.7164, +0.4721]	0.0603	+0.0017 [-0.0090, +0.0133]	6

Table B.4: Mini-squat full-segment condition results. Positive error difference values indicate increased MAE compared with matched condition 1 trials.

Condition	Squat depth MAE	Error difference [95% CI]	Knee flexion MAE	Error difference [95% CI]	Trunk forward lean MAE	Error difference [95% CI]	Paired trials
Condition 2	0.0294	+0.0065 [-0.0123, +0.0282]	5.0595°	−0.2386° [-1.4592, +0.7390]	2.1997°	−0.5182° [-0.9911, -0.0454]	4
Condition 3	0.0225	-0.0057 [-0.0139, +0.0025]	3.5695°	−0.2950° [-1.3443, +0.7543]	2.6698°	−0.5859° [-1.1895, +0.0177]	2
Condition 4	0.0210	-0.0038 [-0.0136, +0.0028]	3.4117°	−0.2918° [-1.1002, +0.3136]	2.1206°	−0.2490° [-1.1321, +0.4925]	4
Condition 5	0.0246	-0.0024 [-0.0089, +0.0035]	3.2148°	−0.3136° [-0.9204, +0.4257]	2.5931°	+0.1004° [-0.7982, +0.9503]	6

C

Appendix C - Supplementary Results

C.1 Bland–Altman Summary Values

Table C.1 reports the numerical Bland–Altman summary values for the balanced condition 1 trials. The signed difference was defined as MediaPipe minus MoCap.

Table C.1: Bland–Altman summary values for balanced condition 1 trials.

Exercise	Metric	Unit	n	MP mean	MoCap mean	Mean bias (MP-MoCap)	SD difference	95% LoA lower	95% LoA upper
SLS	Pelvis horizontal displacement / trunk length	unitless	15	0.1272	0.1632	-0.0359	0.0313	-0.0972	0.0254
	Trunk side tilt	degree	15	7.8448	8.2626	−0.4179°	1.0199°	−2.4170°	1.5812°
STS	Trunk forward lean	degree	15	3.0849	-1.1750	4.2599°	3.2247°	−2.0606°	10.5803°
	Vertical displacement / trunk length	unitless	15	0.6093	0.7020	-0.0927	0.0496	-0.1899	0.0045
MINI	Squat depth / trunk length	unitless	15	0.3296	0.3662	-0.0366	0.0525	-0.1396	0.0664
	Knee flexion (visible side)	degree	15	148.8647	152.3402	−3.4755°	2.2186°	−7.8239°	0.8729°
	Trunk forward lean	degree	15	-0.8344	0.5444	−1.3789°	1.6731°	−4.6581°	1.9004°

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY