SCIENTIFIC
JOURNAL OF
SPORT AND
PERFORMANCE

ORIGINAL ARTICLE

Comparison of different pose estimation models for lower-body kinematics: A validation study

Takashi Fukushima ✉ . Chair of Performance Analysis and Sports Informatics. Technical University of Munich. Munich, Germany.
Patrick Blauberger. Chair of Performance Analysis and Sports Informatics. Technical University of Munich. Munich, Germany.
Tiago Guedes Russomanno. Chair of Performance Analysis and Sports Informatics. Technical University of Munich. Munich, Germany.
University of Brasilia. Campus Universitário Darcy Ribeiro, Brasília, Brazil.
Martin Lames. Chair of Performance Analysis and Sports Informatics. Technical University of Munich. Munich, Germany.

ABSTRACT

As pose estimation has garnered considerable attention for kinematic analysis, numerous pose estimation models have been developed in recent times. A pose estimation model is a trained neural network that predicts human body landmarks from an image. Each model contains different strong and weak points, which make it difficult for users to decide which model to use for kinematic analysis. The accuracy of the model can be one big factor for model selection, but there are not many studies investigating this critical point. Therefore, this study aims to investigate the accuracy of different models and variants by comparing the measurements from the models and variants against reference measurements. Five male participants were invited to this study. Each participant was asked to perform five exercises: squat, squat jump, counter movement jump, walk, and jog while being recorded by twelve normal RGB cameras (Contemplas) and ten marker-based tracking cameras (VICON). The video recordings from the Contemplas were processed by six different pose estimation models and variants: Mediapipe, MeTRAbs Small, MeTRAbs X Large, YOLO, MoveNet Lightning, and MoveNet Thunder to detect joint positions. From the detected joint positions, four joint angles, left hip, right hip, left knee, and right knee, were calculated. Three-way repeated measures ANOVA and Tukey HSD post-hoc analysis were applied to compare the pose estimation models with VICON measurements. The ANOVA result showed that exercise and model factors had a significant impact on the measurement errors although angle factor did not. In the post-hoc analysis, knee joint angle errors from YOLO, MoveNet Lightning, and MoveNet Thunder in jog and walk were significantly higher than those from Mediapipe, MeTRAbs Small, and MeTRAbs X Large. In conclusion, differentiated recommendations can be given for optimum model and variant choice in different conditions in kinematic analyses.

Keywords: Performance analysis, Validation, Pose estimation, Computer vision, Kinematics, Sports informatics.

Cite this article as:
Fukushima, T., Blauberger, P., Guedes Russomanno, T., & Lames, M. (2026). Comparison of different pose estimation models for lower-body kinematics: A validation study. *Scientific Journal of Sport and Performance*, *5*(2), 253-268. https://doi.org/10.55860/XWJL7156

✉ **Corresponding author.** *Technical University of Munich, Chair of Performance Analysis and Sports Informatics, Georg-Brauchle-Ring 60/62, 80992 München, Germany.*
E-mail: takashi.fukushima@tum.de

Submitted for publication September 27, 2025.
Accepted for publication November 11, 2025.
Published December 09, 2025.
Scientific Journal of Sport and Performance. ISSN 2794-0586.

doi: https://doi.org/10.55860/XWJL7156

INTRODUCTION

Motion capture

Motion capture is a widespread method to analyse human kinematics. There are mainly three types of motion capture systems: optoelectrical systems, inertial measurement unit (IMU) systems, and video-based systems.

The optoelectrical systems use reflection markers and infrared cameras to capture human movements. An example is VICON (VICON Motion Systems Ltd., Oxford, the UK). The system is highly accurate and considered a gold standard measurement for kinematic analysis. Indeed, many studies use this system as a reference to validate measurements of new devices (Fukushima et al., 2024; Jeong et al., 2024; Merker et al., 2023; Trowell et al., 2024). The main concern of using this system is accessibility. The system is expensive and needs a proper laboratory environment. For example, since it captures the reflection of infrared markers, other infrared sources such as sunlight must be shut down. In addition, there are some risks of errors. One is the misplacement of markers. A previous paper simulated the marker misplacement effects with random anterior–posterior displacements of lateral thigh (THI), lateral femoral epicondyle (KNE), and tibial (TIB) markers on the assessment of inter-limb differences during a 90° change of direction (CoD) (McFadden et al., 2021). Using Gaussian distributions with standard deviations of 5.8 mm (THI), 3.9 mm (KNE), and 3.4 mm (TIB), the authors quantified the resultant kinematic variability. The 95% confidence intervals attributable solely to marker misplacement ranged from ±2.5° (hip and ankle flexion) to ±13.4° (knee rotation) and ±12.5° (ankle rotation). Another concern is marker occlusion. A study investigated how marker occlusions of marker-based motion capture systems influence marker displacements and distort inter-marker distances (IMD) in static and dynamic settings (Conconi et al., 2021). They found that, in static tests, partial occlusions produced maximum marker displacements of around 1.8 mm, while total occlusions produced smaller displacements of around 0.45 mm. In dynamic tests replicating gait-like motion, maximum IMD variation due to occlusion reached 7.20 mm. Moreover, marker movement on the skin, called soft tissue artefact, can affect motion capture accuracy. A study quantified the artefact of a marker-based motion capture system using segment length variation and apparent joint dislocations during walking, jumping, and sit-to-stand transitions (Bousigues et al., 2025). They found that the segment length variation reached up to 7%, with the thigh segment exhibiting the greatest variability across tasks. Also, apparent joint dislocations were considerable, with mean values of 40.8 mm at the hip (range: 1.1–159.7 mm), 10.9 mm at the knee (range: 0.3–52.8 mm), and 3.4 mm at the ankle (range: 0.1–20.6 mm).

The IMU consists of an accelerometer, gyroscope, and magnetometer. By combining several IMUs, an IMU system can capture human motions. Xsens motion capture suit (Xsens Technologies B.V., Enschede, Netherlands) is one of the examples of IMU systems. These systems are also often used in validation studies (English et al., 2023; Islam et al., 2020; Leung et al., 2024). Compared to optoelectronic systems, the capture volume is wider in IMU systems, meaning that motion data are collected as long as the IMU sensors are attached to the body. However, data can be negatively affected when magnetic forces are in the environment. Also, due to the characteristics of integration to measure positions of the human body, a drift error can cause gradual shifts of the actual position data, making it hard to get speed data or absolute positions, for example.

The video-based systems detect either objects or human body landmarks from images or videos. Traditionally, objects or human body landmarks were manually annotated and used as a gold standard measurement (Menychtas et al., 2023; Reilly et al., 2021). Due to the advancement of computer vision and machine learning research fields, artificial intelligence (AI) can perform the latter task, called pose estimation. Compared to the optoelectronic and IMU-based systems, these systems are more accessible and affordable since they only require running AI, i.e., a pose estimation model, on images or videos. However, these systems have not been well validated for kinematic analysis although there are some studies to validate the system for kinematic analysis in certain movements (Aleksic et al., 2024; D'Haene et al., 2024; Fukushima et al., 2024).

Pose estimation model

A pose estimation model is a trained neural network or algorithm that takes input images or videos and predicts human body landmarks or body poses. There are several pose estimation models and variants that are publicly available, for example, Mediapipe pose landmark detection (Grishchenko et al., 2022), YOLO-pose (Maji et al., 2022), MoveNet (TensorFlow, n.d.-a; TensorFlow, n.d.-b), and MeTRAbs (Sarandi et al., 2021). Each model and variant contains distinct characteristics, ranging from mobile application integration to absolute three-dimensional (3D) estimation.

The Mediapipe pose landmark detection is based on a model called BlazePose (Grishchenko et al., 2022). According to the paper, the model follows a detector–tracker pipeline that first identifies a region of interest containing the human subject and then performs inference to extract 33 two-dimensional (2D) and three-dimensional (3D) landmarks. The core of the model leverages a statistical 3D human mesh representation (GHUM) that captures realistic body shapes and articulations. Ground-truth 3D annotations were generated by fitting the GHUM model to a diverse dataset of 2D human pose annotations spanning multiple physical activities (e.g., yoga, fitness, dance). This fitting process employed additional depth-ordering annotations to mitigate the inherent ambiguity in monocular 3D reconstruction. There are three model types: Lite, Full, and Heavy. On a desktop system equipped with an Intel i9-10900K CPU and NVIDIA GTX 1070 GPU, the Lite, Full, and Heavy variants achieve average inference latencies of 7 ms, 8 ms, and 10 ms per frame, respectively. For mobile deployment, tests conducted on a Google Pixel 4 indicate that the model executes in 25–147 ms on the CPU and 8–22 ms on the GPU, depending on the variant. In-browser inference on a 15-inch MacBook Pro (2017) yields frame processing times of 13 ms (Lite), 15 ms (Full), and 29 ms (Heavy). The Lite model is the fastest but least accurate; in contrast, the Heavy model is the slowest but most accurate. It is available for mobile and web application development through the Mediapipe or ML Kit application programming interfaces (APIs).

The YOLO-pose model regresses 2D landmarks associated with each detected person, thereby eliminating the need for post-hoc landmark grouping or non-differentiable post-processing steps. Its backbone extracts multi-scale visual features from the input image; these features are then aggregated using a Path Aggregation Network (PANet), which facilitates feature fusion across different resolutions to improve localization accuracy for objects and landmarks at various scales. The YOLO-pose model is trained on pose landmarks from the COCO dataset (Lin et al., 2014). It can detect 17 landmarks in 2D. The YOLO-pose model is accessible through Ultralytics (Jocher et al., 2023) and can easily be trained with custom datasets for transfer learning and fine-tuning within its framework, allowing users to customize pose estimation for specific applications. Depending on the model variant, different sizes are available; smaller models offer faster inference but lower accuracy, while larger models provide higher accuracy at the cost of slower processing.

MoveNet uses MobileNetV2 (Sandler et al., 2018) as its feature extractor, which is well-suited for mobile inference due to its depth-wise separable convolutions and low parameter count. This backbone is augmented with a Feature Pyramid Network (FPN; Lin et al., 2016) to enable multi-scale feature aggregation, enhancing the model's robustness to scale variations and occlusions commonly encountered in unconstrained environments. MoveNet is designed for speed, operating in a single forward pass rather than relying on multi-stage refinement or post-hoc grouping of detected landmarks. MoveNet offers two models:

Thunder and Lightning. The Thunder model is slower but more accurate than the Lightning model. Both models can detect 17 2D landmarks. TensorFlow provides a platform that enables users to run MoveNet on web, desktop, and mobile applications (Abadi et al., 2015).

The MeTRAbs model uses a backbone such as ResNet (He et al., 2015), EfficientNet (Tan & Le, 2019), or MobileNet (Lin et al., 2016) as a fully convolutional network (FCN) without additional decoder modules to extract features from an input image. The output is a set of volumetric heatmaps—one for each joint—where each heatmap voxel corresponds to a fixed physical location in 3D metric space. The model applies a soft-argmax operation across the 3D heatmap volume to regress continuous joint coordinates in millimetres. MeTRAbs is trained on both 2D and 3D landmarks from publicly available datasets. Several model variants exist, distinguished by backbone type, model size, and dataset configurations. Unlike the Mediapipe pose landmark detection, which estimates depth relative to the middle-hip point, MeTRAbs can detect absolute 3D coordinates. It can also detect landmarks using different topologies, such as the COCO topology, the SMPL topology (Loper et al., 2015), or the Kinect topology (Microsoft, Redmond, WA).

State of the art of pose estimation validation

There are several studies that have investigated the validity of pose estimation models. One study found a mean absolute error of up to 9.9 degrees during gait using OpenPose (D'Antonio et al., 2020). A previous validation study also used OpenPose for athletic and sports movements and reported errors of 9.7 ± 4.7 degrees and 9.0 ± 3.3 degrees, respectively (Fukushima et al., 2024). Another study examined the validity of OpenCap for lower-body movements and found errors between 11.6 and 14.7 degrees (Lima et al., 2023), although a similar study reported lower errors between 1.91 and 6.87 degrees (Turner et al., 2024). One study compared the accuracy of four different models and variants—OpenPose, MoveNet Lightning, MoveNet Thunder, and DeepLabCut—during gait using a marker-based motion capture system (Washabaugh et al., 2022). They found errors of 3.7 ± 1.3°, 4.6 ± 1.8°, 6.8 ± 1.6°, and 5.9 ± 3.6° for OpenPose, MoveNet Thunder, DeepLabCut, and MoveNet Lightning, respectively, at the hip joint angle. For knee joint angles, the errors were 5.1 ± 2.5°, 7.5 ± 2.5°, 9.4 ± 2.4°, and 9.1 ± 3.0° for OpenPose, MoveNet Thunder, DeepLabCut, and MoveNet Lightning, respectively. Another study investigated the accuracy of OpenPose, AlphaPose, and DeepLabCut during walking, running, and jump tasks using a marker-based motion capture system (Needham et al., 2021). They found errors of approximately 30–50 mm at the knee and hip points, whereas ankle point errors ranged from 1 to 15 mm.

Aim

Considering the challenges of using a gold standard measurement for kinematic analysis, pose estimation models can potentially be a replacement for optoelectronic systems and IMU systems. However, the pose estimation models need to be validated before their use for kinematic analysis. Also, each available model and variant should be compared to find the best model and variant for kinematic analysis. Therefore, this study aims to validate pose estimation models and variants by using a gold standard measurement for lower limbs and to compare the accuracy of each model and variant.

METHODS

Participants

In total, five male participants (Age (mean ± standard deviation): 30.2 ± 6.6 years old, Height: 176.2 ± 6.7 cm, Body mass: 74.2 ± 9.1 kg) participated in this study.

Test exercises

Each participant was asked to perform five exercises: a squat (Figure 1), squat jump (Figure 2), counter movement jump (Figure 3), walk (Figure 4), and jog (Figure 5) two times. The exercises were selected because they were investigated in previous validation studies (D'Antonio et al., 2020; Fukushima et al., 2024; Lima et al., 2023; Turner et al., 2024). They are also commonly used for gait and performance analysis.

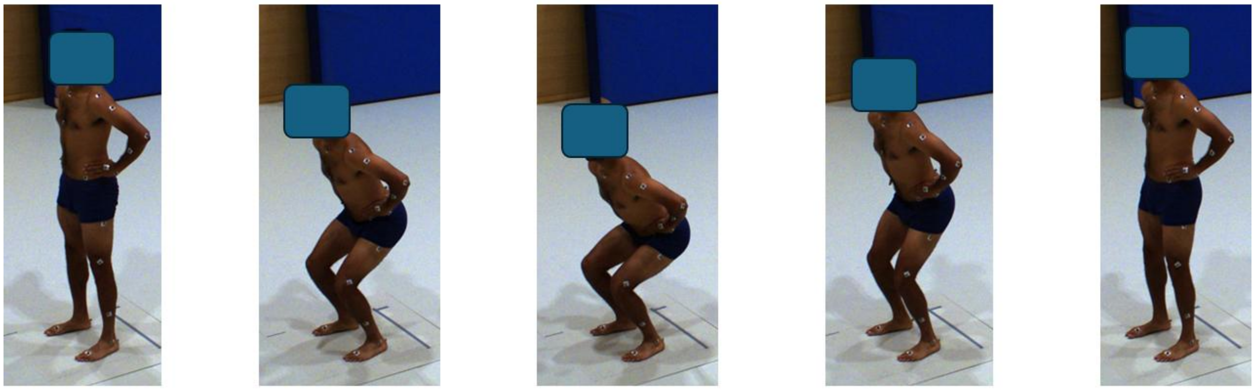

Figure 1. Squat.

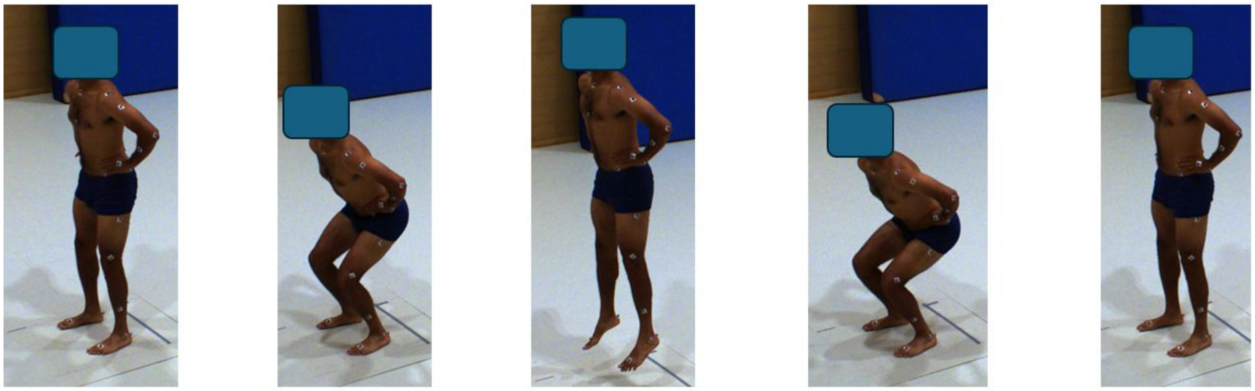

Figure 2. Squat jump.

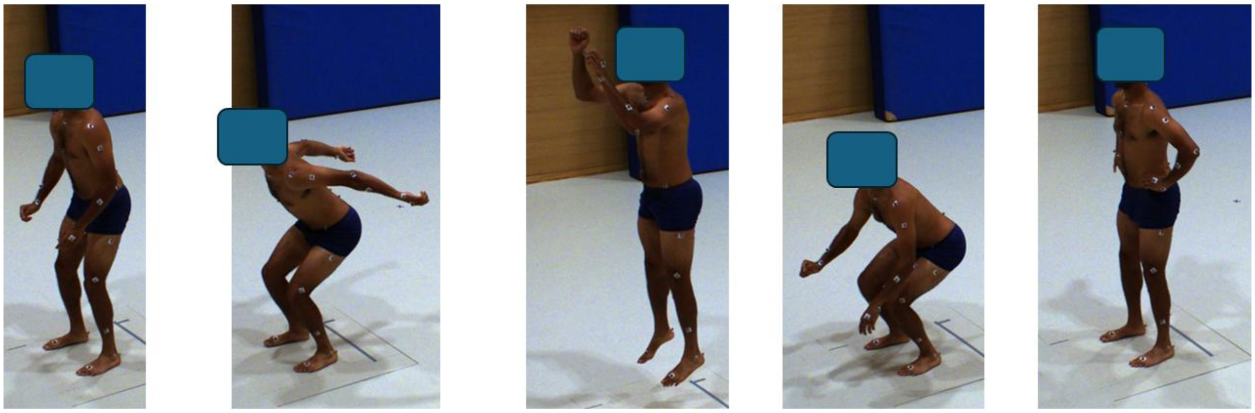

Figure 3. Counter movement jump.

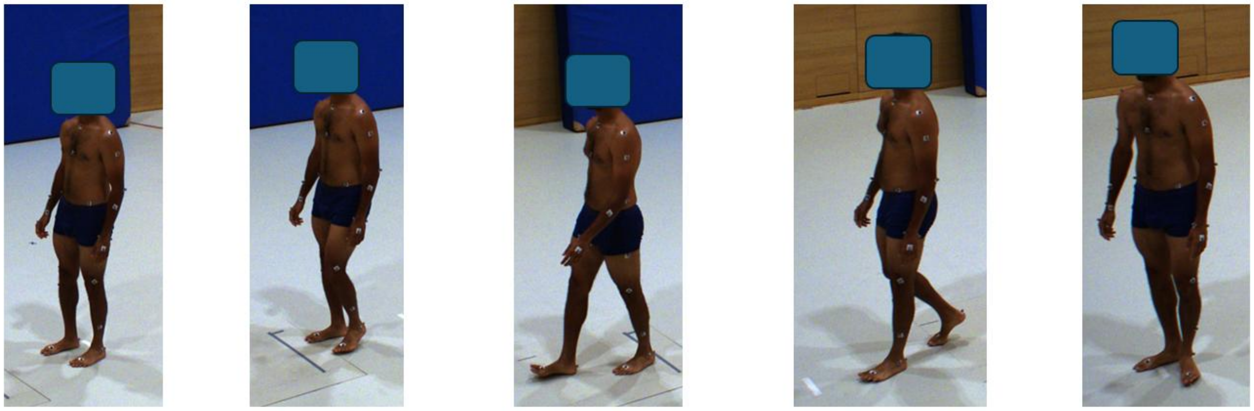

Figure 4. Walk.

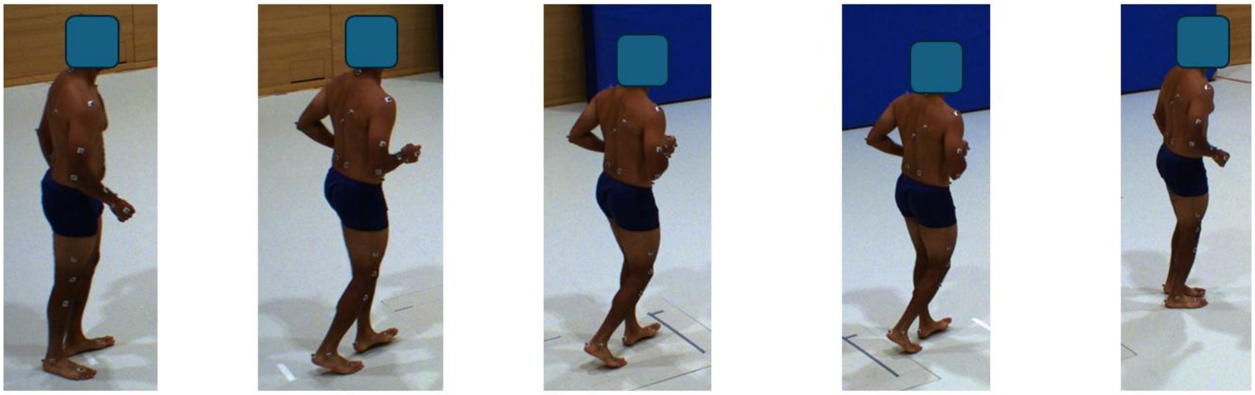

Figure 5. Jog.

Set-ups

To validate the measurements from pose estimation models and variants, two different arrays of cameras were prepared. One is Contemplas cameras "*ab Baumer VLXT-31C*" (CONTEMPLAS GmbH., Kempten, Germany), and the other is VICON cameras. Figure 6 shows the camera setups.

For the VICON system, used as a gold standard, reflecting markers were placed on the body based on the VICON Plug-in Gait model ("*Full body modelling with Plug-in Gait*," n.d.). In total, ten VICON infrared cameras were calibrated using a VICON calibration wand and VICON NEXUS software. Two cameras had a sensor size of 1024 × 1024, two others had a sensor size of 2048 × 1088, and six cameras had a sensor size of 2432 × 2048. All cameras had a sampling rate of 100 Hz. After data collection using the VICON system, the 3D central points of the left and right wrists, elbows, shoulders, hips, knees, and ankles were estimated from detected marker points and extracted using the VICON NEXUS software.

For the Contemplas system, used as input for the pose estimation models, twelve Contemplas cameras with a resolution of 2048 × 1536 pixels and a frame rate of 100 Hz were calibrated with a calibration cage consisting of twelve markers as 3D reference points. Three reflective markers were affixed to each vertical pole of the calibration cage at equal intervals, spanning from ground level up to a height of 100 cm. The horizontal spacing between adjacent markers was set at 100 cm. For each marker with known 3D coordinates, the corresponding 2D image coordinates were manually annotated in each camera view. A projection matrix for each camera was initially estimated using the Direct Linear Transformation (DLT) method (Molnár, 2010) and subsequently refined through bundle adjustment optimization (Triggs et al., 2000) to improve the accuracy of the camera calibration.

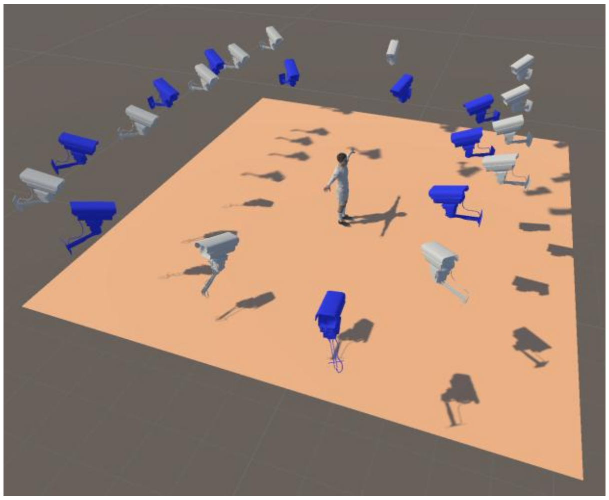

Figure 6. Camera setup. White cameras: Contemplas cameras, Blue cameras: VICON infrared cameras.

Each 3D point was matched with a 2D point observed from each Contemplas camera. The matched point pairs were processed using the Direct Linear Transformation (DLT) method (Molnár, 2010) to compute a projection matrix. Six pose estimation models and variants—MeTRAbs EfficientNet backbone X Large (MeTRAbs X Large) and Small models (MeTRAbs Small), MoveNet Thunder and Lightning models, the YOLOv8-pose model (YOLO), and the Mediapipe full pose landmark detection model (Mediapipe)—were run on the videos recorded from the Contemplas cameras to detect landmarks. Missing landmarks were linearly interpolated from the same point in adjacent frames. Figure 7 shows the skeleton overlay on a video frame and the stick figure derived from VICON data. The detected landmarks were then triangulated using the computed projection matrix to reconstruct the landmarks in 3D (Hartley & Sturm, 1997).

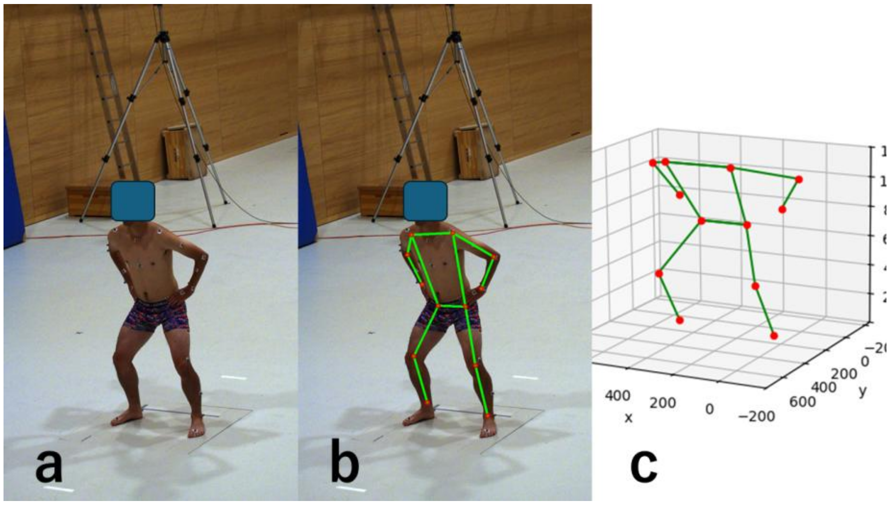

Figure 7 Visual comparison between pose estimation and VICON data. a: Raw video frame, b: Skeleton overlay on video frame, c: Stick figure of VICON data.
The reconstructed 3D landmarks and VICON 3D data points were used to calculate right and left knee and hip joint angles using Equation (1).

$$\theta = a \tan 2\, d\big(||V_1 \times V_2||, (V_1 \cdot V_2)\big) \qquad (1)$$

where V_1 is a 3D vector of the 3D joint position of interest and its proximal adjacent 3D joint position. V_2 is a 3D vector of the 3D joint position of interest and its distal adjacent 3D joint position.

The recording time of measurements was synchronized by dropping a reflection ball and finding the frame in which the ball hit the ground. The study investigated only knee and hip angles since these angles are key to the selected exercises.

Statistics

For each joint angle, the mean error and mean absolute error for the VICON angle were calculated over the course of each exercise. Thus, measurement errors for six models and variants, five exercises, and four joint angles were analysed. The two trials were averaged to get a single mean value. A three-way (6 x 5 x 4) repeated measures ANOVA was run to find significant factors and interactions using the mean absolute error. A post-hoc Tukey HSD was conducted to obtain specific differences between single cells. Bland-Altman plots were used to quantitatively and qualitatively evaluate the overall errors and biases of each model and variant measurement against the VICON measurement. Python 3.10 was used to compute all data analysis and statistics.

Ethics

This study was conducted by following the ethical guidelines of the Technical University of Munich and the Declaration of Helsinki. All participants signed a consent form to participate in this research.

RESULTS

Table 1 shows the results of the Three-way Repeated Measures ANOVA. The main factor exercise and model were highly significant (p < .001), but the main factor angle was not (p = .301). Regarding the exercise factor, the errors were especially different between ipsilateral exercises, which are squat, squat jump, and counter movement jump, and contralateral exercises, which are jog and walk. The contralateral exercises showed higher errors than the ipsilateral exercises overall. The marginal mean error of the exercise factor was 9.2, 9.5, 7.7, 10.8, and 11.7 degrees in counter movement jump, jog, squat, squat jump, and walk, respectively.

In the model factor, the error was higher in YOLO, MoveNet Lightning, and MoveNet Thunder in the ipsilateral exercises compared to those in the contralateral exercises, but the opposite was mostly true in the other models and variants. The marginal mean error of the model factor was 7.7, 7.3, 7.4, 11.6, 11.1, and 13.7 degrees in Mediapipe, MeTRAbs Small, MeTRAbs X Large, MoveNet Lightning, MoveNet Thunder, and YOLO.

Figure 8 shows the marginal mean errors for the interactions between models and exercises, as well as models and joint angles. MoveNet Thunder was erroneous as YOLO and MoveNet Lightning in walk and jog, but it was as accurate as Mediapipe, MeTRAbs X Large, and MeTRAbs Small in other exercises. Also, all models except for YOLO showed a similar error in the counter-movement jump.

In the marginal mean error by models and joint angles, Mediapipe, MeTRAbs X Large, and MeTRAbs Small showed a similar error pattern. In all cases, the YOLO model showed the highest error. Whereas the marginal means of the errors in knee angles were higher than in hip angles for MoveNet Lightning, MoveNet Thunder, and YOLO, they were lower for Mediapipe, MeTRAbs Small, MeTRAbs X Large thus accounting for the highly significant interaction between factors model and exercise.

Table 1. Three-way Repeated Measures ANOVA result.

	F	DF	*p*
Model	35.0	5	<.001
Exercise	10.8	4	<.001
Angle	1.4	3	.301
Model x exercise	24.8	20	<.001
Model x angle	5.2	15	<.001
Exercise x angle	26.9	12	<.001
Model x exercise x angle	10.5	60	<.001

Note. F = F-statistic, DF = Degree of freedom, p = p-value.

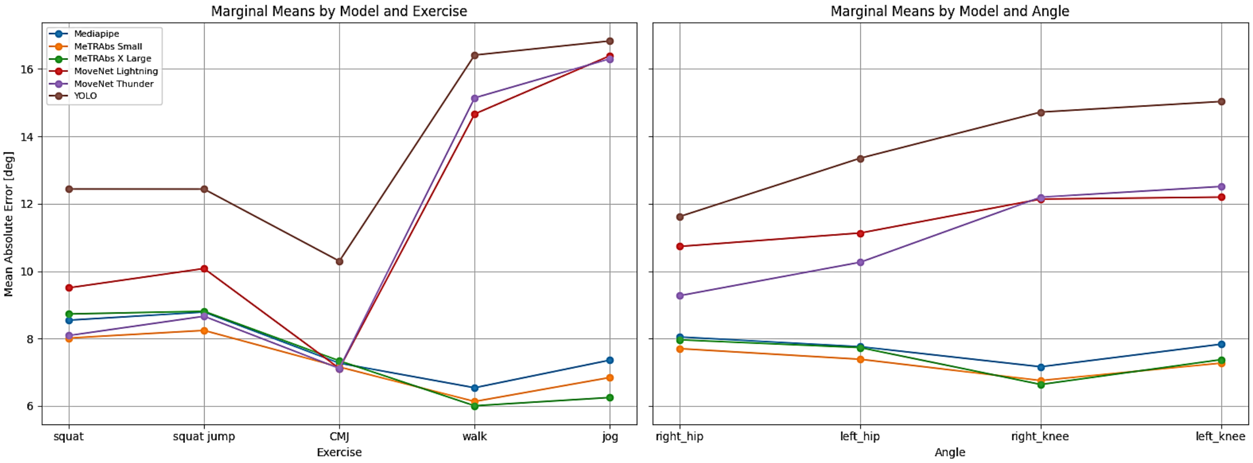

Figure 8. Marginal mean error by Model and Exercise, as well as Model and Angle. CMJ = Counter movement jump.

Table 2 shows selected post-hoc comparisons (total number of post-hoc comparisons = 7140). The comparison pair comes from 120 unique combinations of 6 models and variants, 5 exercises, and 4 joint angles. Then, the possible pair of the combination is (120×119) / 2 = 7140. The analyses compare mean absolute errors between Mediapipe, MeTRAbs Small, MeTRAbs X Large and YOLO, MoveNet Lightning, MoveNet Thunder in jog. All the pairs between Mediapipe, MeTRAbs Small, MeTRAbs X Large and YOLO, MoveNet Lightning, MoveNet Thunder in jog at knee angle were statistically significantly different although this was not the case for hip angles. The same scheme of results was found in the walk (not reported here). However, no statistically significant differences were found in squat, squat jump, and counter movement jump within the model factor.

Table 2. Post-hoc analysis between model 1,2,3 vs model 4,5,6 in jog.

Group1	Group2	Mean Diff (deg)	*p*-adj	Lower CL (deg)	Upper CL (deg)
Mod1Exc5Ang3	Mod4Exc5Ang3	12.9	<.001	4.8	21.0
Mod1Exc5Ang3	Mod5Exc5Ang3	13.7	<.001	5.6	21.8
Mod1Exc5Ang3	Mod6Exc5Ang3	14.1	<.001	6.0	22.2
Mod2Exc5Ang3	Mod4Exc5Ang3	13.5	<.001	5.4	21.5
Mod2Exc5Ang3	Mod5Exc5Ang3	14.2	<.001	6.2	22.3
Mod2Exc5Ang3	Mod6Exc5Ang3	14.6	<.001	6.6	22.7
Mod3Exc5Ang3	Mod4Exc5Ang3	14.4	<.001	6.3	22.5
Mod3Exc5Ang3	Mod5Exc5Ang3	15.2	<.001	7.1	23.2
Mod3Exc5Ang3	Mod6Exc5Ang3	15.6	<.001	7.5	23.6
Mod1Exc5Ang1	Mod4Exc5Ang1	6.9	.372	-1.2	14.9
Mod1Exc5Ang1	Mod5Exc5Ang1	5.7	.885	-2.3	13.8
Mod1Exc5Ang1	Mod6Exc5Ang1	4.5	1.000	-3.5	12.6
Mod2Exc5Ang1	Mod4Exc5Ang1	7.5	.135	-0.5	15.6
Mod2Exc5Ang1	Mod5Exc5Ang1	6.4	.587	-1.6	14.5
Mod2Exc5Ang1	Mod6Exc5Ang1	5.2	.979	-2.8	13.3
Mod3Exc5Ang1	Mod4Exc5Ang1	8.0	.053	0.0	16.1
Mod3Exc5Ang1	Mod5Exc5Ang1	6.9	.340	-1.1	15.0
Mod3Exc5Ang1	Mod6Exc5Ang1	5.7	.887	-2.3	13.8

Note. Mod1 = Mediapipe, Mod2 = MeTRAbs Small, Mod3 = MeTRAbs X Large, Mod4 = MoveNet Lightning, Mod5 = MoveNet Thunder, Mod6 = YOLOpose*, Exc5 = jog, Ang3 = Right Knee Angle, Ang1 = Right Hip Angle, p-adj = adjusted p-value, CL = Confidence intervals, Mean Diff = Mean difference.*

Figure 9 visually summarizes the mean absolute errors across all the exercises, joint angles, and models. The mean absolute differences between Mediapipe, MeTRAbs Small, MeTRAbs X Large and YOLO, MoveNet Lightning, MoveNet Thunder in walk and jog were more than 10 degrees.

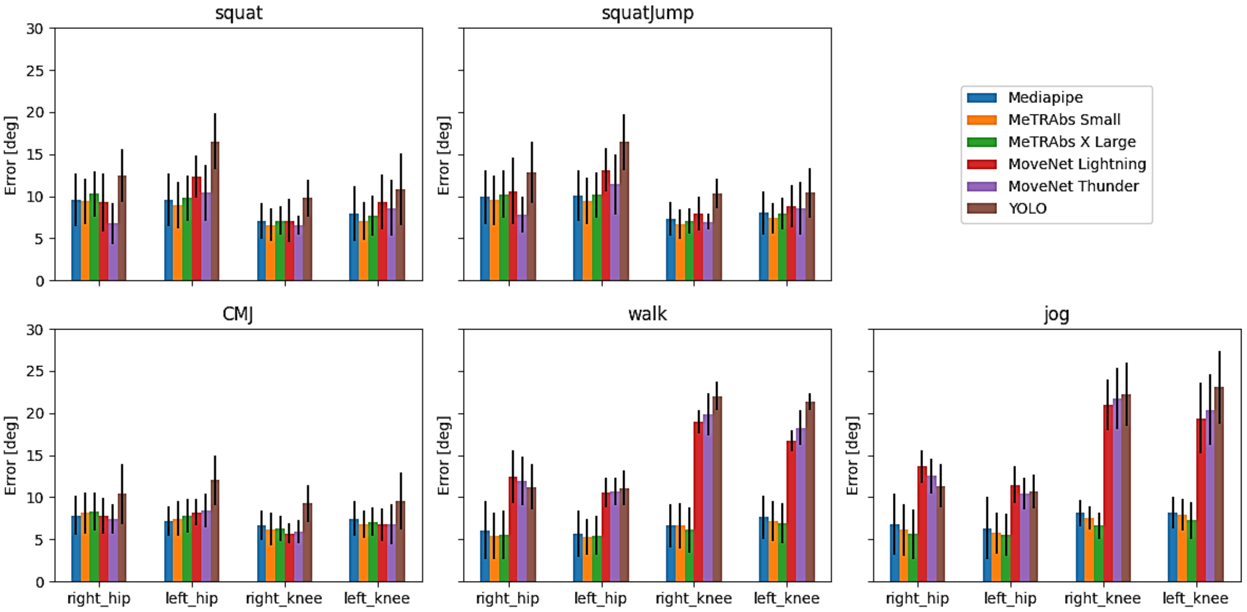

Figure 9. Mean absolute error of all the exercises, joint angles, and models.

Figure 10 shows box plots of mean errors in addition to mean absolute errors reported so far across all the exercises, joint angles, and models. Most of the mean errors were slightly negative. The whisker lines and boxes in YOLO, MoveNet Lightning, and MoveNet Thunder in jog and walk were relatively long compared to others. The mean and standard deviation of the whisker lower and upper values for those models and exercises were –48.2 ± 12.9 and 39.7 ± 11.6 degrees, respectively. The mean and standard deviation of the whisker lower and upper values for other models and exercises were –21.0 ± 4.8 and 5.8 ± 4.1 degrees, respectively.

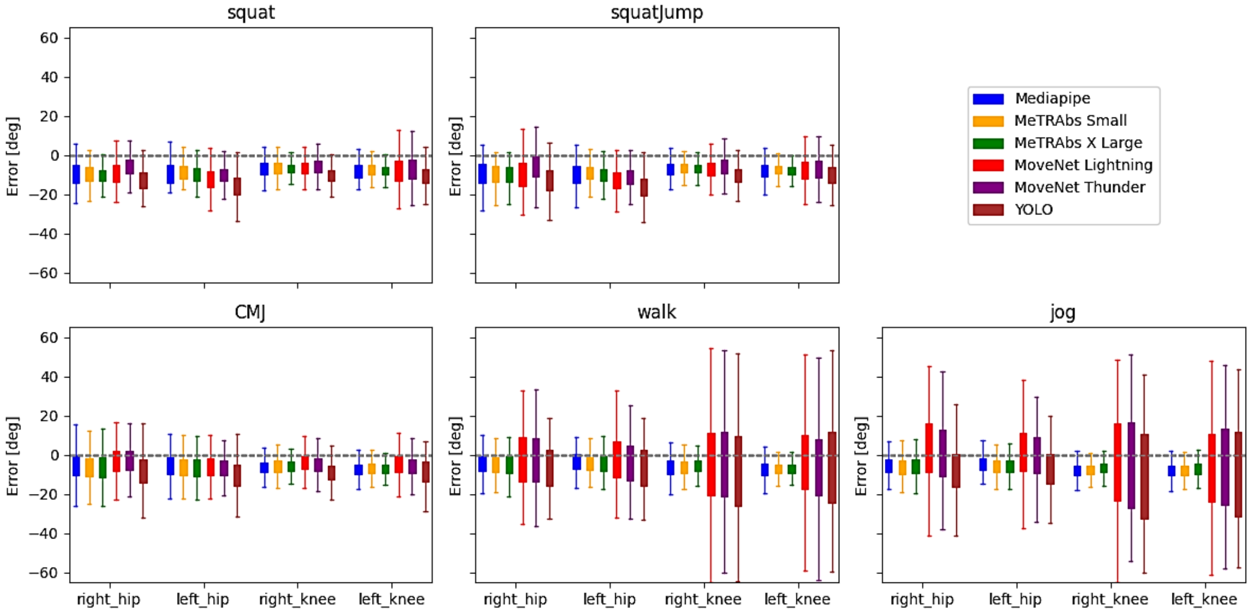

Figure 10. Box plot of mean error ± SD of all the exercises, joint angles, and models. The box represents the Interquartile range. The whisker represents the maximum and minimum range without outliers.

Figure 11 represents the Bland-Altman plot for each model and variant measurements against VICON measurements. The results revealed a consistent positive bias across all models, indicating that predicted joint angles tended to be overestimated relative to the VICON reference. The mean differences (biases) ranged from +4.51° (MoveNet Thunder) to +6.61° (YOLO), with all models exhibiting limits of agreement exceeding ±50°, suggesting substantial variability in joint angle estimation. Notably, MeTRAbs Small and MeTRAbs XL demonstrated the narrowest limits of agreement (~±55–60°) and relatively consistent bias across the range of motion, indicating more stable agreement with the gold standard. In contrast, YOLO and MoveNet Lightning showed wider limits of agreement (up to ±75°), reflecting greater inconsistency and variability in joint angle predictions. Across most models, the Bland–Altman plots exhibited a diamond-shaped distribution, with larger errors observed at mid-range joint angles (~90–110°) and smaller errors at the extremes. This pattern suggests the presence of non-linear systematic bias, likely due to increased joint occlusion and model instability during high-flexion movements. While proportional bias was visually evident in some models (e.g., YOLO), formal regression analysis would be required to statistically confirm such trends.

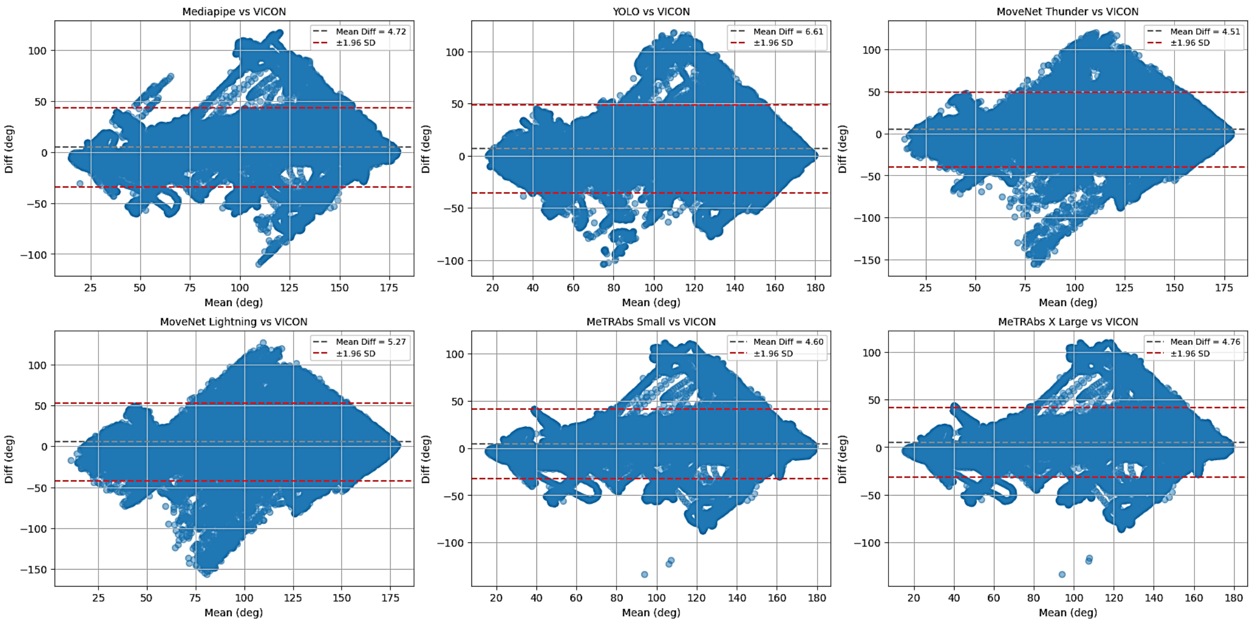

Figure 11. Bland-Altman plot for each model and variant vs VICON measurements.

DISCUSSION

The errors from the Mediapipe, MeTRAbs Small, and MeTRAbs X Large models were within the range reported in previous studies, which was around 9 degrees (D'Antonio et al., 2020; Fukushima et al., 2024). However, YOLO, MoveNet Lightning, and MoveNet Thunder showed more than 10 degrees of error in walking and jogging. Another study reported errors of only 4.6 ± 1.8° and 5.9 ± 3.6° for hip joint angles in MoveNet Thunder and Lightning, respectively (Washabaugh et al., 2022). The same study also found 7.5 ± 2.5° and 9.1 ± 3.0° errors for knee joint angles in MoveNet Thunder and Lightning, respectively, which is lower than the findings of the present study. YOLO was the most erroneous among all models across most exercises and joint angles.

In the three-way repeated measures ANOVA, the model and exercise factors were statistically significant, but the angle factor was not (Table 1). This indicates that the model and exercise have a significant impact on measurement errors, whereas the errors do not depend on the specific joint angle evaluated.

The marginal mean errors of the exercise factor showed that walk was the most erroneous, while squat was the least erroneous (Figure 8). YOLO, MoveNet Thunder, and MoveNet Lightning showed the highest and lowest errors in walk and squat, respectively, whereas the other models showed relatively similar errors across exercises. This suggests that all models were likely well trained and appropriately designed for squatting movements, or that squats are inherently easier for the models to detect. The marginal mean errors by model and angle indicate that MeTRAbs X Large, MeTRAbs Small, and Mediapipe may be suitable for practical applications requiring knee and hip angle analysis, compared to the other models. Mediapipe followed almost the same pattern as MeTRAbs X Large and MeTRAbs Small, suggesting their predictions may be similar under certain conditions.

Mediapipe, MeTRAbs Small, and MeTRAbs X Large showed smaller knee angle errors compared to YOLO, MoveNet Lightning, and MoveNet Thunder in jog and walk (Figures 9 and 10). YOLO, MoveNet Lightning, and MoveNet Thunder showed difficulties detecting landmarks in ipsilateral exercises (Figures 9 and 10). Model size and architecture may have contributed to these issues. The MoveNet model was developed for mobile deployment; therefore, accuracy may not have been the top priority (TensorFlow, n.d.-a; TensorFlow, n.d.-b). YOLO originated as an object detector, so it may not yet be fully optimized for pose estimation. However, YOLO provides a platform for fine-tuning (Jocher et al., 2023), meaning that model accuracy for walking and jogging could be improved if task-specific datasets are used. A previous study employed a similar fine-tuning approach with another model and found improvements (He et al., 2015). This could be a direction for future research.

Regarding the mean error shown in the box plots, negative and positive values indicate that the pose estimator overestimated and underestimated joint angle measurements relative to the reference, respectively (Figure 10). Both overestimation and underestimation were greater during walk and jog for YOLO, MoveNet Lightning, and MoveNet Thunder, suggesting that the errors may not have been systematic.

The observed positive biases across all models suggest a systematic tendency for marker less pose estimation methods to overestimate joint angles relative to the VICON system (Figure 11). While small biases (approximately +4–5°) were found for MeTRAbs and MoveNet Thunder, larger deviations in YOLO and MoveNet Lightning highlight limitations in predictive accuracy, particularly during complex joint movements. The width of the limits of agreement further reflects model consistency, with narrower ranges in MeTRAbs models indicating greater reliability. The diamond-shaped pattern evident in the Bland–Altman plots suggests a non-linear error distribution, with larger errors occurring at mid-range joint angles (approximately 90–110°) and smaller errors at the extremes. This may indicate that models struggle when joints are moderately flexed, possibly due to self-occlusion, limb overlap, or underrepresentation of such poses in training data. The absence of constant error across all joint angles suggests that the inaccuracies reflect systematic rather than random bias. These findings highlight the need to consider not just average accuracy, but also variability and error distribution across different joint ranges. For applications requiring precise kinematic analysis—especially during deep flexion or dynamic movements—MeTRAbs models appear to provide more stable performance than lightweight alternatives such as YOLO or MoveNet Lightning.

Considering these results, Mediapipe, MeTRAbs Small, and MeTRAbs X Large show potential for kinematic analysis with acceptable accuracy. Mediapipe, in particular, is user-friendly for mobile application development, making it suitable for practical kinematic analysis applications. The other models may be usable for contralateral exercises, but they appear less reliable for ipsilateral movements. However, YOLO could be improved by fine-tuning using task-specific datasets.

For practical use cases, each model requires further accuracy improvements, particularly YOLO. Among the models examined, Mediapipe appears to be the most appropriate choice for many applications due to its accuracy, developer-friendliness, and inference speed. Surprisingly, Mediapipe's accuracy was similar to that of the MeTRAbs models, even though MeTRAbs models are computationally heavier. For detecting counter-movement jumps, squats, and jogs, MoveNet Thunder can also be a reasonable option alongside Mediapipe. These movements are often analysed in performance research. For squat analysis using pose estimation, all models except YOLO appear usable.

Finally, typical average absolute errors of approximately 9–10 degrees should be considered acceptable for current state-of-the-art pose estimation. When higher accuracy is required, such as in clinical gait analysis, pose estimation should not be recommended. Nevertheless, improvements may be anticipated in the near future.

Although this study found meaningful results, the small sample size must be acknowledged. More participants are needed to draw stronger conclusions. Additionally, this study analysed only hip and knee joint angles; therefore, the findings may not generalize to other joints, including those in the lower extremities such as the ankle. Future research should evaluate a wider range of joint angles.

CONCLUSION

This study compared six different models and variants by joint angle measurement errors against VICON measurements. Three-way repeated measurement ANOVA and post-hoc measurements revealed that Mediapipe, MeTRAbs Small, and MeTRAbs X Large outperformed YOLO, MoveNet Lightning, and MoveNet Thunder overall. Considering application development friendliness, Mediapipe would be a proper choice among the models and variants investigated in this study. However, YOLO contains the potential to improve accuracy by fine-tuning. Although statistical analysis resulted in some significant differences in factor and interaction level, post-hoc comparisons were mostly insignificant due to low power resulting from the low number of only 5 participants.

AUTHOR CONTRIBUTIONS

Data collection was done by Takashi Fukushima, Patrick Blauberger, and Tiago Guedes Russomanno. Data analysis and paper writing were done by Takashi Fukushima. This study was supervised by Martin Lames.

SUPPORTING AGENCIES

No funding agencies were reported by the authors.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., ... Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems [Computer software]. Retrieved from [Accessed 2025, 20 November]: https://www.tensorflow.org

Aleksic, J., Kanevsky, D., Mesaroš, D., Knezevic, O. M., Cabarkapa, D., Bozovic, B., & Mirkov, D. M. (2024). Validation of automated countermovement vertical jump analysis: Markerless pose estimation vs. 3D marker-based motion capture system. Sensors, 24(20). https://doi.org/10.3390/s24206624

Bousigues, S., Naaim, A., Robert, T., Muller, A., & Dumas, R. (2025). The effects of markerless inconsistencies are at least as large as the effects of the marker-based soft tissue artefact. Journal of Biomechanics, 182, 112566. https://doi.org/10.1016/j.jbiomech.2025.112566

Conconi, M., Pompili, A., Sancisi, N., & Parenti-Castelli, V. (2021). Quantification of the errors associated with marker occlusion in stereophotogrammetric systems and implications on gait analysis. Journal of Biomechanics, 114, 110162. https://doi.org/10.1016/j.jbiomech.2020.110162

D'Antonio, E., Taborri, J., Palermo, E., Rossi, S., & Patane, F. (2020). A markerless system for gait analysis based on OpenPose library. In 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) (pp. 1-6). https://doi.org/10.1109/I2MTC43012.2020.9128918

D'Haene, M., Chorin, F., Colson, S. S., Guérin, O., Zory, R., & Piche, E. (2024). Validation of a 3D markerless motion capture tool using multiple pose and depth estimations for quantitative gait analysis. Sensors, 24(22). https://doi.org/10.3390/s24227105

English, D. J., Weerakkody, N., Zacharias, A., Green, R. A., Hocking, C., & Bini, R. R. (2023). The validity of a single inertial sensor to assess cervical active range of motion. Journal of Biomechanics, 159, 111781. https://doi.org/10.1016/j.jbiomech.2023.111781

Fukushima, T., Blauberger, P., Guedes Russomanno, T., & Lames, M. (2024). The potential of human pose estimation for motion capture in sports: A validation study. Sports Engineering, 27(1). https://doi.org/10.1007/s12283-024-00460-w

Full body modeling with Plug-in Gait. (n.d.). VICON Documentation. Retrieved from [Accessed 2025, 20 November]: https://docs.vicon.com/display/Nexus212/Full+body+modeling+with+Plug-in+Gait

Grishchenko, I., Bazarevsky, V., Zanfir, A., Bazavan, E. G., Zanfir, M., Yee, R., ... Sminchisescu, C. (2022). BlazePose GHUM Holistic: Real-time 3D human landmarks and pose estimation. https://doi.org/10.48550/arXiv.2206.11678

Hartley, R. I., & Sturm, P. (1997). Triangulation. Computer Vision and Image Understanding, 68(2), 146-157. https://doi.org/10.1006/cviu.1997.0547

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. https://doi.org/10.1109/CVPR.2016.90

Islam, R., Bennasar, M., Nicholas, K., Button, K., Holland, S., Mulholland, P., ... Al-Amri, M. (2020). A nonproprietary movement analysis system (MoJoXlab) based on wearable inertial measurement units: Validation study. JMIR mHealth and uHealth, 8(6), e17872. https://doi.org/10.2196/17872

Jeong, M. G., Kim, J., Lee, Y., & Kim, K. T. (2024). Validation of a newly developed low-cost, high-accuracy, camera-based gait analysis system. Gait & Posture, 114, 8-13. https://doi.org/10.1016/j.gaitpost.2024.08.077

Jocher, G., Qiu, J., & Chaurasia, A. (2023, January 10). Ultralytics YOLO (Version 8.0.0) [Computer software]. Ultralytics. Retrieved from [Accessed 2025, 20 November]: https://github.com/ultralytics/ultralytics

Kitamura, T., Teshima, H., Thomas, D., & Kawasaki, H. (2022). Refining OpenPose with a new sports dataset for robust 2D pose estimation. In 2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW). IEEE. https://doi.org/10.1109/WACVW54805.2022.00074

Leung, K. L., Li, Z., Huang, C., Huang, X., & Fu, S. N. (2024). Validity and reliability of gait speed and knee flexion estimated by a vision-based smartphone application. Sensors, 24(23). https://doi.org/10.3390/s24237625

Lin, T.-Y., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., ... Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. arXiv:1405.0312. Retrieved from [Accessed 2025, 20 November]: http://arxiv.org/abs/1405.0312

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2016). Feature pyramid networks for object detection. https://doi.org/10.48550/arXiv.1612.03144

Lima, Y., Collings, T., Hall, M., Bourne, M., & Diamond, L. (2023). Assessing lower-limb kinematics via OpenCap during dynamic tasks: A validity study. Journal of Science and Medicine in Sport, 26(Suppl.), S105. https://doi.org/10.1016/j.jsams.2023.08.123

Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: A skinned multi-person linear model. ACM Transactions on Graphics, 34(6), 248. https://doi.org/10.1145/2816795.2818013

Maji, D., Nagori, S., Mathew, M., & Poddar, D. (2022). YOLO-pose: Enhancing YOLO for multi-person pose estimation using object keypoint similarity loss. https://doi.org/10.48550/arXiv.2204.06806

McFadden, C., Daniels, K., & Strike, S. (2021). The effect of simulated marker misplacement on inter-limb differences during a change of direction task. Journal of Biomechanics, 116, 110184. https://doi.org/10.1016/j.jbiomech.2020.110184

Menychtas, D., Petrou, N., Kansizoglou, I., Giannakou, E., Grekidis, A., Gasteratos, A., ... Aggelousis, N. (2023). Gait analysis comparison between manual marking, 2D pose estimation algorithms, and a 3D marker-based system. Frontiers in Rehabilitation Sciences, 4, 1238134. https://doi.org/10.3389/fresc.2023.1238134

Merker, S., Pastel, S., Bürger, D., Schwadtke, A., & Witte, K. (2023). Measurement accuracy of the HTC VIVE Tracker 3.0 compared to Vicon system. Sensors, 23(17), 7371. https://doi.org/10.3390/s23177371

Molnár, B. (2010). Direct linear transformation based photogrammetry software on the web. ISPRS Commission, 38, 5-8. Retrieved from [Accessed 2025, 20 November]: http://www.isprs.org/proceedings/XXXVIII/part5/papers/130.pdf

Needham, L., Evans, M., Cosker, D. P., Wade, L., McGuigan, P. M., Bilzon, J. L., & Colyer, S. L. (2021). The accuracy of several pose estimation methods for 3D joint centre localisation. Scientific Reports, 11, 20673. https://doi.org/10.1038/s41598-021-00212-x

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. https://doi.org/10.48550/arXiv.1801.04381

Sarandi, I., Linder, T., Arras, K. O., & Leibe, B. (2021). MeTRAbs: Metric-scale truncation-robust heatmaps for absolute 3D human pose estimation. IEEE Transactions on Biometrics, Behavior, and Identity Science, 3(1), 16-30. https://doi.org/10.1109/TBIOM.2020.3037257

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Retrieved from [Accessed 2025, 20 November]: https://doi.org/10.48550/arXiv.1905.11946

TensorFlow. (n.d.-a). MoveNet: Ultra fast and accurate pose detection model. Retrieved from [Accessed 2025, 20 November]: https://www.tensorflow.org/hub/tutorials/movenet

TensorFlow. (n.d.-b). Pose estimation and classification on edge devices with MoveNet and TensorFlow Lite. Retrieved from [Accessed 2025, 20 November]: https://blog.tensorflow.org/2021/08/pose-estimationand-classification-on-edge-devices-with-MoveNet-andTensorFlow-Lite.html

Triggs, B., McLauchlan, P. F., Hartley, R. I., & Fitzgibbon, A. W. (2000). Bundle adjustment: A modern synthesis. In Vision Algorithms: Theory and Practice (pp. 298-372). Springer. https://doi.org/10.1007/3-540-44480-7_21

Trowell, D. A., Carruthers Collins, A. G., Hendy, A. M., Drinkwater, E. J., & Kenneally-Dabrowski, C. (2024). Validation of a commercially available mobile application for velocity-based resistance training. PeerJ, 12, e17789. https://doi.org/10.7717/peerj.17789

Turner, J. A., Chaaban, C. R., & Padua, D. A. (2024). Validation of OpenCap: A low-cost markerless motion capture system for lower-extremity kinematics during return-to-sport tasks. Journal of Biomechanics, 171, 112200. https://doi.org/10.1016/j.jbiomech.2024.112200

Washabaugh, E. P., Shanmugam, T. A., Ranganathan, R., & Krishnan, C. (2022). Comparing the accuracy of open-source pose estimation methods for measuring gait kinematics. Gait & Posture, 97, 188-195. https://doi.org/10.1016/j.gaitpost.2022.08.008