Article

Movement Recognition and Knee Angle Estimation for Home-Based Lower Limb Rehabilitation Using an RGB-D Camera

GRUPPO 19. Vincenza Claudia Dicuonzo [1], Lorenzo Ricci [1], Marco Sportelli [1], Francesca Ursi [1], Fabiano Vaglio [1]

[1] Politecnico di Torino, 10129 Torino, Italy;

Abstract

Home-based rehabilitation is a promising application of Human Activity Recognition, which has emerged as a solution to improve patient adherence and reduce healthcare costs. Nevertheless, ensuring correct execution of exercises outside clinical supervision remains a significant challenge. This study presents an integrated system for the classification of lower limb motor gestures and knee angle estimation, designed to assist patients during autonomous rehabilitation sessions at home.
RGB-D data from an Intel RealSense camera were processed with MediaPipe to extract 3D joint positions. These were used to train a Transformer-based network coupled with a Multi Layer Perceptron (MLP) classifier for movement recognition. The model demonstrated high accuracy in classifying four target exercises and provided reliable estimates of knee joint angles during the entire movement. A Unity-based grafical user interface enables users to record movements, send it to the algorithm, and visualise execution correctness. The system includes a 3D humanoid avatar displaying the correct execution of each gesture and a "history" module to track user progress. Preliminary results indicate that the proposed solution can enhance the accuracy and autonomy of rehabilitation exercises, offering a sensorless sistem capable of identifying movement patterns. This approach could benefit both patients and clinicians.

Keywords: 3D Model; Gesture Recognition; Joints Estimation; Lower Limb Rehabilitation; RGB-D Camera; Transformer-based Neural Network; Unity Graphical User Interface

1. Introduction

Human Activity Recognition (HAR) has emerged as an increasingly important approach, with a wide range of applications across domains such as healthcare, robotics, virtual reality, and human–computer interaction. In particular, in the clinical field, HAR technologies are playing an expanding role by enabling the analysis of joint angles and movement trajectories to detect and classify physical gestures [1]. In this context, machine learning models are employed to interpret data from depth cameras, inertial measurement units (IMUs), and RGB videos, thereby enhancing the accuracy and generalisability of gesture recognition systems [2]. The integration of smart sensing technologies with advanced learning algorithms has proven especially valuable in remote rehabilitation, providing real-time feedback to patients, and actionable performance metrics to clinicians [3]. One of the most promising applications of HAR is home-based rehabilitation, which offers a more accessible, flexible, and cost-effective alternative to conventional in-clinic therapy. It is particularly beneficial for individuals recovering from injury, surgery, or neuromuscular conditions. Nevertheless, home-based therapy presents significant challenges, notably the absence of continuous professional supervision, which can lead to incorrect exercise execution and reduced adherence to treatment protocols. To address these limitations, recent studies have investigated the use of RGB-D cameras, wearable sensors, and neural networks to enable automated monitoring and quantitative movement assessment [3].

In response to the need for effective remote monitoring in home-based rehabilitation, this work presents an interactive application that employs an RGB-D camera in conjunction with neural network algorithms to support lower limb recovery. Using an Intel RealSense depth camera, the system captures patients' leg movements and processes them through a Transformer-based neural network [4] trained to recognize four specific rehabilitation exercises: backward knee flexion, standing hip flexion, seated leg extension and squat. This approach enables offline classification of recorded gestures and estimation of knee joint angles, allowing the system to deliver objective feedback following each exercise session.

The application is designed with a dual target audience in mind: patients and clinicians. The system is primarily intended for patients, monitoring their progress through knee angle measurements over time and providing them with visual insight into the correctness of their movements. To support this, the system features a humanoid 3D model that demonstrates the correct execution of each exercise, allowing patients to compare their own performance against an accurate reference. The secondary target is the physiotherapist, who can utilize the application as a remote monitoring tool to analyse the patient's rehabilitation progress, evaluating both movement execution and changes in the knee joint's range of motion (ROM) through quantitative assessment of joint angles.

This dual focus aims to empower patient autonomy while equipping clinicians with meaningful data to support treatment planning.

1.1. Related work

In the context of lower limb rehabilitation, RGB-D cameras have demonstrated effectiveness in accurately measuring hip and knee joint angles during unsupervised, home-based exercise sessions, demonstrating the potential of depth-sensing technologies to support patient monitoring in non-clinical environments [3]. Sensor fusion approaches, such as the integration of IMU data with depth camera inputs, as illustrated in DeepFuse by Huang et al., have further improved real-time 3D human pose estimation [5]. More recent developments include the combination of surface electromyography (sEMG) signals with RGB-D data to enhance the estimation of lower limb joint angles, emphasizing the potential of multimodal sensor integration for improved movement analysis in rehabilitation settings [1]. Kumar et al. developed a sensor-free, vision-based system that estimates lower limb joint kinematics during daily exercises using a combination of convolutional and recurrent neural networks, highlighting its suitability for home-based rehabilitation monitoring [6].

The structure of this paper is organized as follows. Section 2 describes data acquisition and processing, the neural network model and the Unity-based application. Section 3 presents and discusses the results of gesture classification and joint angle estimation, evaluating the model's performance, limitations, and knee angle trends in comparison with findings from existing literature. Finally, Section 4 concludes the study and outlines directions for future work.

2. Materials and Methods

The present study proposes an application aimed at recognising and classifying specific lower limb movements through the analysis of RGB-D video data, employing a Transformer-based neural network. Figure 1 provides a summarised pipeline of the various stages of the work carried out.

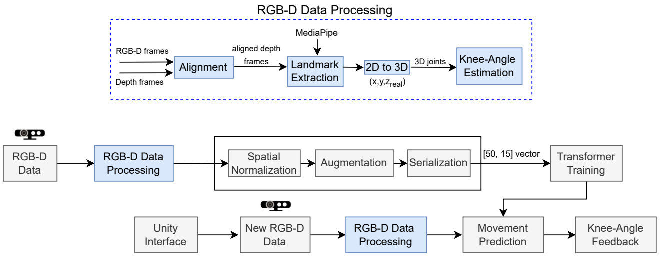


Figure 1. *Overview of the proposed pipeline for lower limb movement recognition.*

2.1. Data Acquisition

As a preliminary step of the study, data acquisition was carried out through video recordings using an Intel RealSense D435i depth camera. An RGB-D camera generates two concurrent data streams: colour (RGB) video frames and associated depth frames. The depth frames encode spatial information by representing the distance of each pixel from the sensor as a two-dimensional grayscale image, where pixel intensity corresponds to depth values [3]. The depth video stream is visualised by overlaying a pseudocolour map to enhance visual contrast resulting from differences in distance from the camera. The RealSense Viewer software was employed to configure the acquisition parameters: both the colour and depth video streams were set to a resolution of 640×480 pixels with a frame rate of 30 frames per second (FPS), and a depth format of Z16. Recordings were saved in `.bag` format. The experimental setup consisted of a fixed camera mounted on a tripod. Subjects were instructed to stand in a lateral position, both left and right profile, so that the moving leg appeared as parallel as possible to the camera image plane. This orientation ensured that leg movements could be captured in a predominantly two-dimensional plane. Additionally, each subject was positioned at a suitable distance from the camera to ensure that the field of view included the anatomical region extending from the pelvis to the toes, thereby enabling accurate extraction of landmarks associated with the lower limb. A total of 19 subjects participated in the data collection. The protocol involved the execution of four specific lower limb rehabilitation exercises:

- backward knee flexion
- standing hip flexion
- seated leg extension
- squat.

For each subject, a single exercise was recorded at a time. Each movement was performed twice in succession to maximise the chance of capturing at least one complete and artifact-free execution and, furthermore, recordings were conducted for both the right and left leg, to account for bilateral symmetry and inter-limb variability.

2.2. Landmark Extraction

Once the video acquisition phase was completed, the recorded data were processed to extract the anatomical landmarks of the lower limb. Pose estimation was performed using MediaPipe Pose [7], a machine learning framework for body tracking. For image processing tasks such as frame manipulation and visualization, the Open Source Computer Vision Library (OpenCV) [8] was used. MediaPipe's Pose Landmarker model detects a total of 33 anatomical keypoints, including those related to the lower limbs, providing normalised 3D coordinates and a Z-scale relative to the pelvis-centred origin.

Specifically, landmarks with indices 23–32 correspond to five key reference points essential for monitoring leg motion. These include the hip, knee, ankle, heel, and foot index (toe), each playing a crucial role in evaluating joint angles, limb orientation, and foot-ground interactions. [7] Table 1 summarises the corresponding landmark indices for the left and right lower limbs as defined in the MediaPipe Pose model.

Table 1. *MediaPipe Pose indices for key lower limb landmarks.*

Landmark	**Left leg index**	**Right leg index**
Hip	23	24
Knee	25	26
Ankle	27	28
Heel	29	30
Foot Index (Toe)	31	32

The combination of hip, knee, and ankle landmarks allows for the estimation of the orientation of the thigh and lower leg segments, thus enabling the calculation of the knee joint angle in the sagittal plane, an essential parameter for assessing the joint's ROM during rehabilitation exercises.

Assuming that knee rotation is limited to the plane defined by the two adjacent long bones, the knee rotation angle θ_{knee} can be estimated during different exercises using the equation 1:

$$\theta_{\text{knee}} = \arccos\left(\frac{(\mathbf{h} - \mathbf{k}) \cdot (\mathbf{a} - \mathbf{k})}{\|\mathbf{h} - \mathbf{k}\| \, \|\mathbf{a} - \mathbf{k}\|} \right) \tag{1}$$

where **h**, **k**, and **a** denote the three-dimensional coordinates of the hip, knee, and ankle joints, respectively, of either the left or right limb [3]. This formula computes the internal angle between the femur and tibia segments, effectively measuring the flexion or extension angle of the knee in the sagittal plane. The calculated angle θ_{knee} corresponds to the measured angle illustrated in Figure 2, where the anatomical landmarks, estimated using MediaPipe, are visualised along with the resulting joint angles for the four analysed movements.

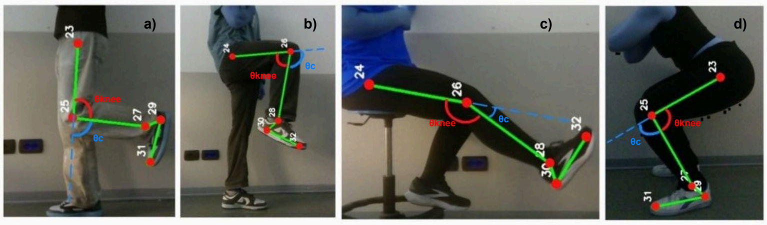


Figure 2. *Anatomical landmarks of the hip, knee, and ankle detected using MediaPipe are shown along with the knee angle θ_{knee} and its complementary angle θ_c in the sagittal plane during four exercises: (a) Backward Knee Flexion, (b) Standing Hip Flexion, (c) Seated Leg Extension, and (d) Squat.*

The implemented Python script processes the recorded `.bag` file by extracting a fixed-length sequence of 50 frames, each containing the estimated positions of the five landmarks. After aligning the depth and color streams to ensure pixel-wise correspondence, a depth mask is applied to remove distant background regions. While MediaPipe provides an estimated z-coordinate relative to a pelvis-centered reference frame, this information is enhanced by replacing it with the actual depth values obtained from the RealSense sensor. For each landmark, the real-world depth is converted into metric units, enabling a more accurate spatial representation of lower limb motion.

The resulting landmark positions per frame are stored as $[x, y, z]$ triplets, where x and y denote the horizontal and vertical positions normalized by the frame width and height, respectively, and z corresponds to the real depth in meters. All data are compiled into a NumPy array of shape $[50, 5, 3]$, corresponding to the number of frames, landmarks, and spatial coordinates, respectively.

2.3. Data Pre-processing

Each 50-frame sequence, consisting of five 3D landmarks of the lower limb, was used to classify the type of movement performed through a Transformer-based neural network trained on preprocessed and normalized input data. Initially, to ensure spatial consistency across sequences, a body alignment procedure was applied. The first frame was used to compute the direction from the hip to the foot, and the entire sequence was rotated such that this principal body axis pointed vertically downwards. This step reduces inter-subject variability caused by differing orientations during recording. Subsequently, sequences were rescaled using the hip-to-ankle distance from the first frame, ensuring independence from the subject's body height. Each normalised sequence was then standardised using its own mean and standard deviation, to further reduce inter-sample variability.

To improve the model's generalization capability, data augmentation techniques were applied, exclusively during the training phase. These techniques include:

- Rotation of the entire sequence around the Z-axis by a random angle (±60°) to simulate changes in the subject's lateral orientation;
- Random scaling to simulate apparent size variations;
- Random spatial translation to increase robustness to figure displacements;

- Addition of Gaussian noise with a standard deviation of 0.01;
- Random temporal shifting up to ±5 frames to introduce temporal variability;
- Spatial dropout on individual landmarks with a probability of 0.5, simulating partial occlusions or acquisition errors.

These transformations are applied exclusively to training samples, while validation and test sets remain unchanged to ensure unbiased evaluation.

At the end of preprocessing and augmentation steps, each frame is serialized into a 1D vector by concatenating the 3D coordinates of all landmarks in an input tensor of shape $[50, 15]$.

2.4. Transformer-based Model Architecture for Movement Classification

For the task of movement classification, we implemented a neural network architecture based on the *Transformer Encoder*, specifically tailored to learn temporal dependencies from sequences of 3D joint positions acquired over time. The aim was to enable the model to classify motor gestures by analysing the dynamic evolution of body landmarks throughout each sequence. The architecture processes input data structured as a tensor of shape $[\text{batch\_size}, \text{sequence\_length}, \text{input\_size}]$, where `input_size` represents the number of features per frame. A linear embedding layer projects the input features into a higher-dimensional space, generating dense feature vectors more suitable for subsequent Transformer processing.

To retain information about the temporal order of the input sequence a positional encoding block is applied, following the sinusoidal formulation proposed by Vaswani et al. [4]. This step allows the model to distinguish the relative and absolute positions of frames within a sequence.

The core of the network comprises a stack of Transformer encoder layers. Each layer includes a multi-head self-attention mechanism and a feedforward sub-network, enabling the model to capture complex and long-range temporal relationships across the entire sequence. This capability is particularly valuable for modelling the spatio-temporal progression of human movement.

After passing through the encoder stack, temporal features are aggregated via *mean pooling* across the sequence dimension, resulting in a single compact representation per gesture. This vector is then processed by a classification head, which includes a `LayerNorm` block (32 units), a fully connected layer with `ReLU` activation, `Dropout` regularisation, and a final linear layer that outputs the class logits corresponding to the number of gesture classes.

Model training was performed in a supervised learning framework. To mitigate potential class imbalance, we employed a weighted `CrossEntropyLoss`, where class weights were computed based on their frequency in the training set.

The preprocessed dataset, described in detail in the previous section, was partitioned into training, validation, and test subsets using a stratified shuffle split, ensuring balanced class distributions across all sets.

Model optimisation was carried out using the `Adam` optimiser with weight decay regularisation and an initial learning rate of 5×10^{-4}. The learning rate was adaptively adjusted during training using a `ReduceLROnPlateau` scheduler, which automatically reduced the learning rate when the validation loss failed to improve over a set number of consecutive epochs.
The hyperparameters were set as follows:

- `batch_size` = 8
- `num_epochs` = 100
- `weight_decay` = 5×10^{-4}

To prevent overfitting, an *early stopping* mechanism was employed, saving the model checkpoint corresponding to the lowest validation loss with a patience of ten consecutive epochs.

Loss and accuracy curves were plotted to monitor training dynamics and to detect potential overfitting or underfitting. Final evaluation metrics, including overall accuracy and gesture recognition rate, were computed on the test set.

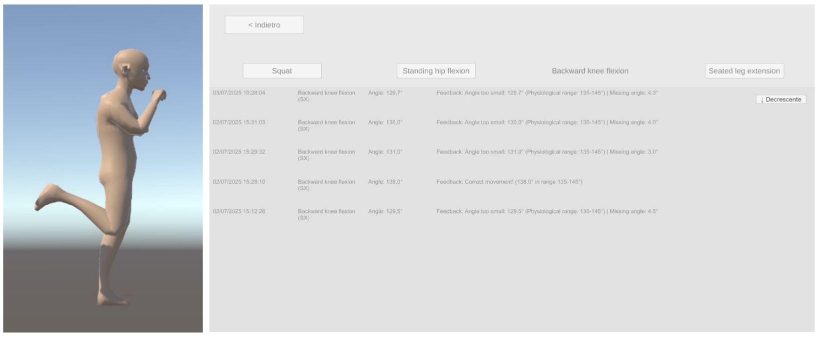
2.5. User Interface for Visualization and Feedback

Unity 3D is a cross-platform engine developed by Unity Technologies, widely adopted for the creation of interactive applications. In this project, the Unity-based interface was specifically designed to allow patients to independently perform motor rehabilitation exercises.

Upon launching the interface, the user is presented with an initial screen featuring a button to start the recording and a button to open the "history" section. Once pressed "Recording", a countdown timer is activated, granting the subject a few seconds to assume the correct position: the distance from the camera must ensure full capture of the entire lower limb involved, whereas the orientation must allow a lateral view of the active leg. At the end of the countdown, recording begins automatically and lasts approximately seven seconds. During the movement execution, the patient can monitor their performance in real time through visual feedback provided by the camera, which displays both the colour video stream and the corresponding depth map. Once the movement is completed, the video is saved and the interface offers the user the option to either retain or repeat the recording. If the user chooses to save the video, it is sent to the back-end server, where an automated processing pipeline is executed. Specifically, the system extracts the landmarks of the lower limb associated with the moving leg, processes the data using a Transformer-based model to classify the type of motor gesture, and calculates, for each frame, the angle formed by the hip-knee-ankle landmarks. Upon completion of processing, the interface returns the recognised gesture classification, the maximum angle achieved during execution and a personalised feedback message.

Furthermore, a 3D humanoid model is displayed to illustrate the correct execution of the movement, providing a useful visual reference. An `.fbx` format humanoid model was employed [9], to which animations corresponding to the four target movements were applied. These animations were initially created using a base model downloaded from Mixamo [10] and were subsequently manually refined in Blender, adjusting joint segment rotations frame by frame to accurately replicate each gesture. If the calculated joint angle falls within the expected physiological range for that specific exercise, the interface confirms correct execution. Otherwise, the system notifies the user of how many degrees are lacking to reach the maximum joint angle permitted by physiological limits.

Additionally, a "history" section was implemented to store recorded movements. For each acquisition, the system logs the date, time, and the feedback provided by the neural model regarding the correctness of the performed gesture. Figure 3 illustrates the reference model during the backward knee flexion exercise, alongside an example of the corresponding "history" section obtained after multiple executions of the movement.



03/07/2025 10:28:04	Backward knee flexion (SX)	Angle: 129,7°	Feedback: Angle too small: 129.7° (Physiological range: 135-145°) | Missing angle: 4.3°
02/07/2025 15:31:03	Backward knee flexion (SX)	Angle: 130,0°	Feedback: Angle too small: 130.0° (Physiological range: 135-145°) | Missing angle: 4.0°
02/07/2025 15:29:32	Backward knee flexion (SX)	Angle: 131,0°	Feedback: Angle too small: 131.0° (Physiological range: 135-145°) | Missing angle: 3.0°
02/07/2025 15:28:10	Backward knee flexion (SX)	Angle: 138,0°	Feedback: Correct movement! (138.0° in range 135-145°)
02/07/2025 15:12:26	Backward knee flexion (SX)	Angle: 129,5°	Feedback: Angle too small: 129.5° (Physiological range: 135-145°) | Missing angle: 4.5°

Figure 3. *Example of the humanoid model and "history" section for backward knee flexion.*

3. Results and Discussions

3.1. Performance Evaluation on Training, Validation, and Test Sets

The learning behaviour of the developed model was analysed by monitoring the evolution of the loss and accuracy curves over the 100 training epochs, for both the training and validation sets. The results are reported in Figure 4.

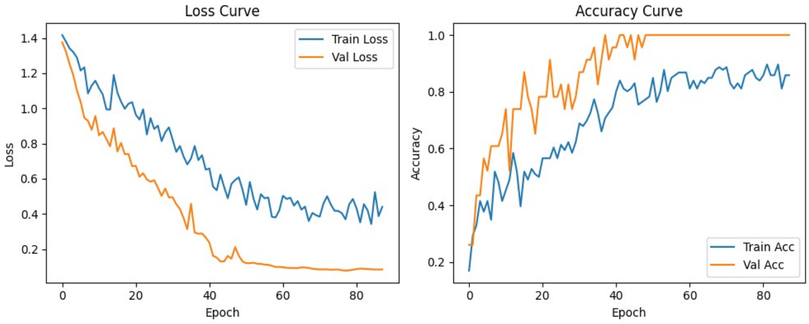


Figure 4. *Training and validation curves for loss and accuracy over 100 epochs.*

The analysis of the loss and accuracy curves indicates an effective and stable learning process. The training loss exhibits a steady decline throughout the epochs, with minor residual fluctuations attributable to the stochastic nature of the Adam optimiser, eventually stabilising around 0.4. The validation loss decreases more rapidly and consistently, reaching a plateau near 0.1 from approximately the 40th epoch, suggesting good generalisation capability.

Similarly, the training accuracy increases progressively, with a plateau around 85%. In contrast, the validation accuracy rises sharply, exceeding 90% before the 40th epoch and remaining close to 100% from this epoch onwards. This pattern reflects that the model achieves excellent performance on the validation set. However, the fact that validation loss is significantly lower and validation accuracy markedly higher than training performance could indicate that the validation set may not fully represent the variability present in the training data, or that the data split led to a validation set that is easier to classify.

To counteract overfitting, early stopping was applied when the validation loss ceased to improve, and a ReduceLROnPlateau scheduler was used to dynamically reduce the learning rate. These strategies contributed to stable training, as reflected in the smooth convergence of the curves.

Despite these measures, the persistent gap between training and validation metrics highlights the need for further refinement. Future efforts should focus on increasing validation data diversity and adopting stronger regularisation techniques to enhance generalisation.

At the end of training, the model was evaluated on an independent test set comprising data from 5 subjects, to assess its generalisation capability in gesture classification. In this context, accuracy reflects the recognition rate achieved by the model. The model achieved an Accuracy of 91% .

Although these results appear excellent, it is important to emphasise that the small size of the test set and the similarity between training, validation, and test samples may have facilitated classification. Therefore, these results should be interpreted with caution, as more heterogeneous test data might reveal limitations in the model's generalisation ability.

3.2. Knee Angle Dynamics and Variability Analysis

A further evaluation was conducted to analyse the temporal evolution of the knee angle across test set subjects, aiming to assess inter-subject variability. The plots (Figure 5) display unstandardised single-movement sequences, as no constraints were imposed on execution time or speed during data acquisition. Moreover, no explicit instructions were provided regarding movement depth, which led to increased variability and, in some cases, outlier values beyond expected physiological ranges.

The angles evaluated were the external angle, θ_c, for the Backward Knee Flexion, Standing Hip Flexion, and Squat movements, and the internal angle, θ_{knee}, for the Seated Leg Extension. This approach was adopted to maintain consistency with data reported in the literature. Individual executions of each movement are reported.

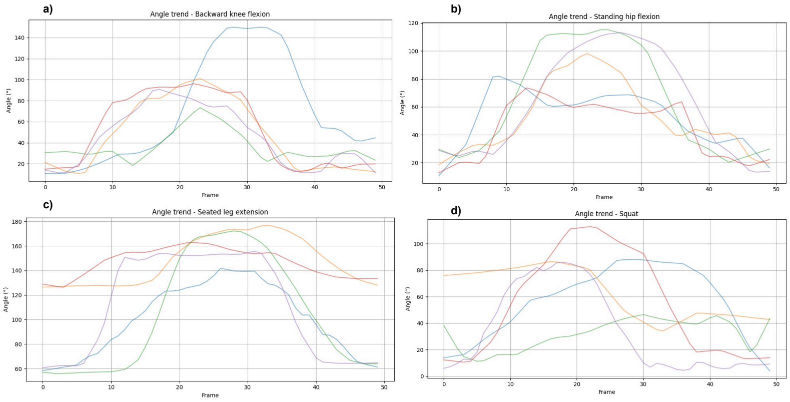


Figure 5. *Angular evolution during one execution of the 4 movements: (a) Backward Knee Flexion, (b) Standing Hip Flexion, (c) Seated Leg Extension, (d) Squat.*

For a given gesture, the temporal profiles among different subjects were qualitatively similar, indicating that the MediaPipe framework successfully captured the movements without frame loss or incoherent values. Generally, all movements exhibit a progressive increase in joint angle, followed by a return to initial values, consistent with physiologically expected dynamics.

In posterior knee flexion, standing hip flexion, and squat movements, the initial and final angles are typically close to zero, except for isolated cases, such as the subject shown in yellow in Figure 5d during the squat movement.

For the seated leg extension movement, two initial configurations were observed: three subjects started from 60°, while the remaining two started from 120°. This discrepancy is likely due to differences in positioning relative to the camera; however, the qualitative trends are preserved.

3.3. Evaluation of Maximum Joint Angles

In view of potential rehabilitation applications, the maximum joint angles reached by each subject were deemed clinically relevant and are presented in the user interface history, as they allow for easier differentiation between pathological and healthy movement patterns. All subjects analysed were healthy, and the measured angles fell within or near the expected physiological ranges (Table 2).

- **Backward knee flexion:** maximum angles recorded ranged from 70° to 150°, partially below and partially above the typical range reported in the literature for healthy individuals, which is generally between 135° and 145° [11].
- **Standing hip flexion:** in this movement, where subjects were instructed to lift one leg by flexing the hip, two patterns emerged: some patients reached a maximum knee angle of approximately 60°, while others achieved around 120°. The physiological reference maximum angle is higher (up to 140°) [12], so these values fall within the expected range.
- **Seated leg extension:** this exercise is typically used to strengthen the quadriceps and promote full joint extension. It starts from a 90° flexed position and proceeds to full extension; the measured maximum angles ranged from 140° to 180°, which is consistent with expected physiological values.
- **Squat:** maximum angles were just below 120°, which falls within the expected range for this exercise. The classification did not account for different squat depths (e.g., partial, half, deep) or segmental inclinations required for proper execution [13].

Table 2. *Comparison between measured joint angles and physiological reference values.*

Movement	Measured maximum angle	Physiological reference values
Backward knee flexion	70°–150°	135°–145°
Standing hip flexion	60°–120°	Up to 140°
Seated leg extension	140°–180°	Up to 180° (full extension)
Squat	∼120°	Partial/shallow squat (90°) Half squat (90–110°) Deep squat (110–135°)

All deviations are physiologically plausible and probably caused by the unconstrained nature of acquisition.

3.4. Limits and Future Perspectives

The proposed approach, based on a Transformer Encoder neural network, represents an initial step towards an automatic system capable of recognising pathological movement patterns. This architecture, still relatively underused in rehabilitation contexts compared to conventional solutions such as CNNs [1] or LSTMs [14], proved particularly effective for temporal sequence analysis, thanks to its ability to process all frames simultaneously.

Its integration with the markerless MediaPipe framework, which is simple to deploy and economically accessible, renders the model suitable for real-world clinical applications.

In the future, the analysis could be extended to a larger set of motor classes and a more comprehensive evaluation of lower limb joint angles, conducted on a broader sample and using standardised acquisition protocols. This would help prevent the overfitting observed in the model, thereby enabling the development of models capable of classifying the degree of functional impairment using clinically meaningful severity scales. Furthermore, the integration of additional sources of information, such as sEMG data, could further enhance the system's discriminative capabilities [1]. In this context, automatic movement recognition has the potential to support clinical assessment, contributing to the analysis of patients' functional status and the formulation of more targeted and personalised rehabilitation pathways.

4. Conclusions

This study proposed an innovative system for gesture recognition and joint angle analysis aimed at supporting lower limb rehabilitation in home-based contexts. The developed pipeline integrates RGB-D data acquisition, markerless pose estimation via MediaPipe, and gesture classification using a Transformer-based neural network. The architecture effectively models the temporal structure of human motion through linear embedding, sinusoidal positional encoding, and multi-head self-attention mechanisms. The classification head, implemented via a multi-layer perceptron (MLP), combined with the Adam optimiser and a weighted CrossEntropy loss function, enabled robust learning and accurate gesture identification. In parallel, a Unity-based graphical interface was developed to deliver an intuitive user experience. This includes real-time visual feedback, guidance through animated 3D avatars, and longitudinal tracking of performance metrics. Such integration supports user engagement and promotes correct execution of rehabilitation exercises.

The experimental results are promising: the model showed high levels of accuracy on the test set, while the analysis of knee angle trajectories confirmed that all movements remained within or near expected physiological ranges. These findings indicate the system's potential to support both quantitative motor assessment and early detection of abnormal patterns. Nonetheless, current limitations include the relatively small and homogeneous dataset, which may restrict the model's generalisability. Future work will address this by incorporating a broader and more diverse sample, introducing stricter acquisition protocols, and exploring the integration of multimodal data. These improvements may enable the system to detect subtle deviations in movement patterns and facilitate the classification of functional impairments along clinically relevant severity scales. Overall, the proposed approach represents a step towards the development of an accessible, cost-effective, and clinically useful tool for remote rehabilitation, capable of enhancing both patient autonomy and clinical decision-making.

References

1. Du, G.; Ding, Z.; Guo, H.; Song, M.; Jiang, F. Estimation of Lower Limb Joint Angles Using sEMG Signals and RGB-D Camera. *Bioengineering* **2024**, *11*, 1026. https://doi.org/10.3390/bioengineering11101026.
2. Seel, T.; Raisch, J.; Schauer, T. IMU-Based Joint Angle Measurement for Gait Analysis. *Sensors* **2014**, *14*, 6891–6909. https://doi.org/10.3390/s140406891.
3. Uccheddu, F.; Furferi, R.; Governi, L.; Carfagni, M. RGB-D-Based Method for Measuring the Angular Range of Hip and Knee Joints during Home Care Rehabilitation. *Sensors* **2021**, *22*(1), 184. https://doi.org/10.3390/s22010184.
4. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, *arXiv:1706.03762*, 1–15. https://doi.org/10.48550/arXiv.1706.03762.
5. Huang, F.; Zeng, A.; Liu, M.; Lai, Q.; Xu, Q. DeepFuse: An IMU-Aware Network for Real-Time 3D Human Pose Estimation from Multi-View Image. *CoRR* **2019**, *abs/1912.04071*. http://arxiv.org/abs/1912.04071.
6. Kumar, K. S.; Jamsrandorj, A.; Kim, J.; Mun, K. R. Prediction of Lower Limb Kinematics from Vision-Based System Using Deep Learning Approaches. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2022**, *2022*, 177–181, DOI: 10.1109/EMBC48229.2022.9871577. PMID: 36086538.
7. NizdarLaila. *Pose Estimation using MediaPipe*. Kaggle Notebooks, Kaggle, 2022. https://www.kaggle.com/code/nizdarlaila/pose-estimation-using-mediapipe (accessed June 30, 2025).
8. OpenCV Org. *OpenCV – Open Computer Vision Library*; OpenCV Foundation: 2000–2025. https://opencv.org/ (accessed June 30, 2025).
9. Free3D. *Male Base 3D Model*, 2025. Retrieved June 30, 2025, from https://free3d.com/it/3d-model/male-base-88907.html.
10. Mixamo. *Mixamo—3D character animation*, 2025. Retrieved June 30, 2025, from https://www.mixamo.com/.
11. Samarpan Physio Clinic. *Knee Flexion and Extension — Movement, ROM, Function, Exercise*; Samarpan Physio Clinic, 2023. https://samarpanphysioclinic.com/knee-flexion-and-extension/ (accessed June 30, 2025).
12. Tsuda, E. Range of Motion (ROM) and Gait Function. In *Advances in Total Knee Arthroplasty*; Matsuda, S., Ed.; Springer Nature Singapore: Singapore, 2024; pp 399–407. DOI: https://doi.org/10.1007/978-981-97-4920-1_67.
13. Straub, R.K.; Powers, C.M. A Biomechanical Review of the Squat Exercise: Implications for Clinical Practice. *Int. J. Sports Phys. Ther.* **2024**, *19*(4), 490–501. https://doi.org/10.26603/001c.94600.
14. Toro-Ossaba, A.; Jaramillo-Tigreros, J.; Tejada, J. C.; Peña, A.; López-González, A.; Castanho, R. A. LSTM Recurrent Neural Network for Hand Gesture Recognition Using EMG Signals. *Appl. Sci.* **2022**, *12*(19), 9700. https://doi.org/10.3390/app12199700.