Журнал Белорусского государственного университета. Математика. Информатика. 2026;1:130–139
Journal of the Belarusian State University. Mathematics and Informatics. 2026;1:130–139

*УДК 004.89, 004.021*

# СИСТЕМА РАСПОЗНАВАНИЯ ПОХОДКИ В РЕАЛЬНОМ ВРЕМЕНИ НА ОСНОВЕ ЛЕГКОВЕСНОЙ ДВУХМОДАЛЬНОЙ ОЦЕНКИ ПОЛОЖЕНИЯ

***ДИН АОДИ[1)], ЙЕ ШИПИНГ[2)], А. М. НЕДЗЬВЕДЬ[1), 3)], В. С. АНОСОВ[4)]***

*[1)]Белорусский государственный университет, пр. Независимости, 4, 220030, г. Минск, Беларусь*
*[2)]Чжэцзян-Шуренский университет, ул. Шурен, 8, 310015, г. Ханчжоу, Китай*
*[3)]Объединенный институт проблем информатики НАН Беларуси, ул. Сурганова, 6, 220012, г. Минск, Беларусь*
*[4)]6-я городская клиническая больница, ул. Уральская, 5, 220037, г. Минск, Беларусь*

***Аннотация.*** Предлагается облегченный подход к совместному анализу походки в 2D- и 3D-пространствах. Реализуется двухмодальная система распознавания походки, обеспечивающая ее высокоточную оценку в реальном времени за счет архитектуры совместной 2D- и 3D-реконструкции. В рамках системы для детекции ключевых точек скелета в 2D-пространстве и извлечения биометрических метрик походки применяется библиотека MediaPipe. Для трехмерной оценки позы используется модифицированная облегченная архитектура MTSA-former, которая содержит 2,5 млн параметров. Данная модель реализует каскадную обработку, включающую отдельные временные и пространственные модули, а также матрицу смежности, которая построена на основе топологической структуры скелета, что позволяет эффективно моделировать физические ограничения между суставами. Для достижения высокой производительности при сохранении вычислительной легкости предлагается гибридная локально-глобальная стратегия пространственного моделирования, интегрирующая графовые сверточные сети. Такой подход обеспечивает точную и эффективную реконструкцию 3D-позы в условиях реального времени.

***Ключевые слова:*** анализ походки; оценка 3D-позы; облегченная модель; двухмодальное распознавание.

***Благодарность.*** Работа была выполнена при финансовой поддержке Китайского стипендиального совета и Министерства трудовых ресурсов и социального обеспечения Китая (гранты № H20250551 и H20240330).

**Образец цитирования:**
Дин Аоди, Йе Шипинг, Недзьведь АМ, Аносов ВС. Система распознавания походки в реальном времени на основе легковесной двухмодальной оценки положения. *Журнал Белорусского государственного университета. Математика. Информатика.* 2026;1:130–139 (на англ.).
EDN: XDQZRA

**For citation:**
Din Aodi, Ye Shiping, Nedzved AM, Anosov VS. Real-time gait recognition system based on lightweight dual-modal pose estimation. *Journal of the Belarusian State University. Mathematics and Informatics.* 2026;1:130–139.
EDN: XDQZRA

**Авторы:**
***Дин Аоди*** – аспирант кафедры информационных систем управления факультета прикладной математики и информатики. Научный руководитель – А. М. Недзьведь.
***Йе Шипинг*** – проректор.
***Александр Михайлович Недзьведь*** – доктор технических наук, доцент; заведующий кафедрой информационных систем управления факультета прикладной математики и информатики[1)], главный научный сотрудник отдела интеллектуальных информационных систем[3)].
***Виктор Сергеевич Аносов*** – главный врач.

**Authors:**
***Din Aodi***, postgraduate student at the department of information management systems, faculty of applied mathematics and computer science.
*aodiding541@gmail.com*
***Ye Shiping***, vice-rector.
*zjsruysp@163.com*
***Alexander M. Nedzved***, doctor of science (engineering), docent; head of the department of information management systems, faculty of applied mathematics and computer science[a], and chief researcher at the department of intellectual information systems[c].
*nedzveda@tut.by*
*https://orcid.org/0000-0001-6367-5900*
***Victor S. Anosov***, chief physician.
*avictor8@gmail.com*

# REAL-TIME GAIT RECOGNITION SYSTEM BASED ON LIGHTWEIGHT DUAL-MODAL POSE ESTIMATION

***DIN AODI*[a], *YE SHIPING*[b], *A. M. NEDZVED*[a, c], *V. S. ANOSOV*[d]**

[a]*Belarusian State University, 4 Niezaliezhnasci Avenue, Minsk 220030, Belarus*
[b]*Zhejiang Shuren University, 8 Shuren Street, Hangzhou 310015, China*
[c]*United Institute of Informatics Problems, National Academy of Sciences of Belarus, 6 Surganava Street, Minsk 220012, Belarus*
[d]*6th City Clinical Hospital, 5 Uralskaja Street, Minsk 220037, Belarus*

*Corresponding author: A. M. Nedzved (nedzveda@tut.by)*

***Abstract.*** A lightweight approach for joint 2D and 3D gait analysis is proposed. A dual-modal gait recognition system is adopted, achieving high-precision real-time analysis through a 2D and 3D collaborative estimation architecture. The system employs MediaPipe for real-time 2D keypoint detection and gait metric extraction. An improved lightweight architecture MTSA-former with 2.5 mln parameters is used for 3D pose estimation. This model utilises cascaded processing incorporating temporal and spatial modules, along with an adjacency matrix based on skeletal topological structure to model physical constraints between joints. A local-global hybrid spatial modelling strategy fusing graph convolutional networks is implemented to achieve efficient 3D pose estimation, attaining high performance while maintaining lightweight characteristics. This approach ensures accurate and efficient 3D pose reconstruction in real time.

***Keywords:*** gait analysis; 3D pose estimation; lightweight model; dual-modal recognition.

***Acknowledgements.*** This work was carried out with the financial support of the China Scholarship Council and the Ministry of Human Resources and Social Security of China (grants No. H20250551 and H20240330).

## Introduction

Gait analysis, as a non-invasive method for motor function assessment and clinical diagnosis, holds significant application value in early screening. For example, Parkinson's disease patients exhibit typical characteristics such as shortened stride length, reduced gait speed and shuffling gait, while stroke patients present hemiplegic gait and limited joint range of motion [1]. These abnormal gait patterns can often be captured through quantitative analysis in the early stages of disease [2].

In fig. 1 the lightweight dual-modal gait recognition system proposed in this study, which balances real-time performance and accuracy through a 2D and 3D collaborative analysis architecture, is illustrated. MediaPipe first extracts 2D keypoint sequences from video in real-time, which are then input into the multi-scale temporal self-attention transformer model (here and further – MTSA-former model). This model captures gait cycles through multi-scale temporal convolution, reinforces skeletal constraints via spatial graph convolutional network (GCN) and ensures motion continuity through velocity and acceleration loss functions, thereby lifting 2D coordinates into high-precision 3D poses. Finally, based on the reconstructed 3D joint trajectories (particularly the lower limb hip, knee and ankle joints), end-to-end real-time gait analysis is achieved. Multi-dimensional gait features, including joint angle temporal curves, gait symmetry indices and time ratios of stance phase and swing phase, are extracted [3]. The lightweight design and dual-modal collaborative mechanism of this study enhance gait parameter measurement accuracy while maintaining computational efficiency [4]. Providing a novel technical approach for early warning of neurodegenerative diseases and quantitative assessment of rehabilitation outcomes.

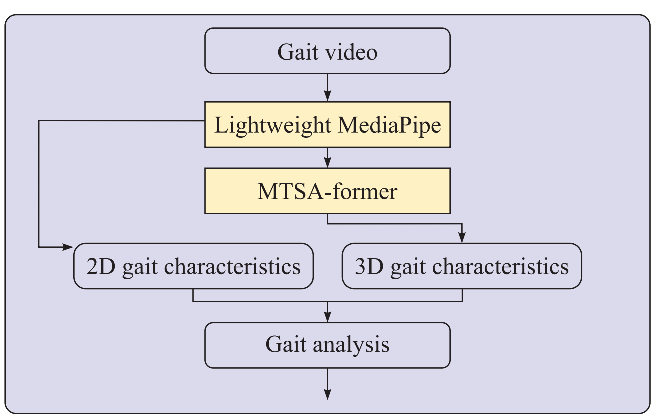


*Fig. 1.* Dual-modal gait recognition framework

## Methods and principles

**2D pose estimation method.** MediaPipe utilises a lightweight detector to rapidly localise human body regions in images. During the pose keypoint localisation stage, the region of interest is cropped and normalised, then input into the pose estimation network to extract 2D coordinates of 22 human body keypoints. MediaPipe Lite model has approximately 3 mln parameters, achieving inference speeds of 30–60 frames per second on mobile central processing units and over 100 frames per second on graphics processing units [5]. With a model size of approximately 12 MB, it is suitable for deployment on resource-constrained edge devices. MediaPipe Lite significantly reduces computational cost and storage requirements while maintaining acceptable accuracy, providing an efficient 2D pose estimation foundation for the real-time dual-modal gait analysis system in this study [6].

From fig. 2, it can be seen that 11 skeletal nodes were removed from MediaPipe's original 33 keypoints, reducing the data dimensionality by approximately 33 %. The 22 core keypoints serve as the foundational skeletal model for gait analysis, encompassing the major movement joints of the human body, including key positions of the head, trunk, upper limbs and lower limbs. The most critical joints for gait analysis (lower limb joints (hip, knee and ankle joints) and trunk joints (shoulder and hip joints)) enable the model to more effectively capture key gait parameters such as periodic movement patterns during walking, joint angle variations and body center of gravity shifts.

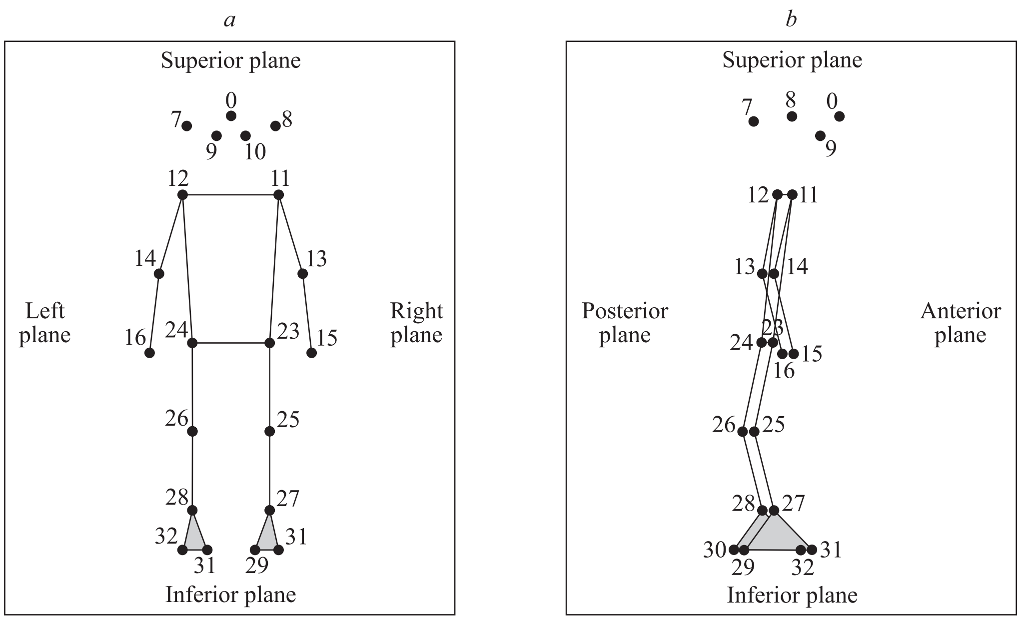


*Fig. 2.* MediaPipe key skeletal structure:
*a* – frontal view; *b* – sagittal view (left lateral)

**3D pose estimation method.** Traditional 3D human pose estimation models typically require large-scale annotated data, which poses challenges in clinical gait analysis scenarios due to high data collection costs for patients with gait disorders and limited sample sizes. This paper proposes MTSA-former model – a lightweight, highly efficient and rapidly convergent gait analysis-specific model trained on small samples from Human3.6M dataset, achieving fast training and accurate prediction under few-shot settings (fig. 3) [7]. The model employs depth-wise separable convolution, linear complexity attention mechanisms and parameter sharing strategies to significantly reduce model size while maintaining performance. The core of the transformer uses two main parts: attention and feed-forward network (FFN). The standard GELU function is used as the activation function.

The model first performs linear projection and positional encoding on the input sequence to obtain spatiotemporal feature embeddings. Subsequently, within multiple encoding blocks, temporal modelling (utilising multi-scale convolution to capture short-term and long-term motion variations) and spatial modelling (learning structural relationships between joints through graph convolution and attention mechanisms) are conducted simultaneously. Finally, multi-scale pyramid pooling fuses information from different temporal ranges, and the regression head outputs the 3D human pose. This design effectively enhances the model's spatiotemporal perception capabilities and prediction accuracy, demonstrating excellent performance in motion continuity and smoothness.

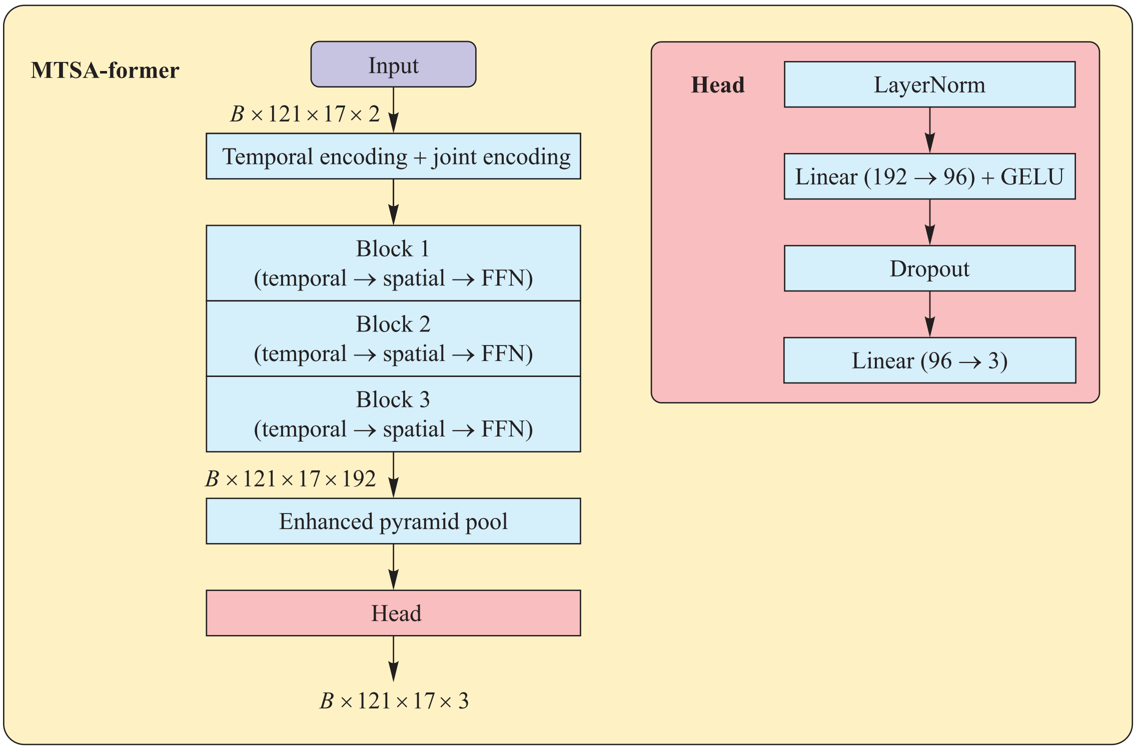


*Fig. 3.* 3D pose estimation framework using the MTSA-former model

**Human skeletal spatial graph convolution.** The human skeletal structure can be abstracted as a graph, where joint points are the nodes of the graph and bones are the edges connecting the nodes. GCN can explicitly model the physical connections and kinematic constraints between joint points by propagating and aggregating information across the graph [8]. Each GCN layer updates the feature representation of each node, where the new feature is a weighted sum of its own features and those of neighbouring nodes. By stacking GCN layers, the model can learn high-order inter-joint dependencies that conform to human structural priors, thereby effectively enhancing the physical realism and structural integrity of the generated poses. The composite loss function formula is as follows:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right).$$

Describes the feature propagation and transformation processes in a GCN from layer $l$ to layer $l+1$. Here $H^{(l)}$ denotes the node feature matrix at layer $l$; $\tilde{A}$ represents the adjacency matrix; $\tilde{D}$ is the corresponding degree matrix; $W^{(l)}$ represents the trainable weight matrix at layer $l$.

The model is guided to learn a more comprehensive mapping relationship. It not only learns how to accurately localise the keypoints but also acquires prior knowledge of human body structure and the dynamic patterns of motion, ultimately generating highly realistic 3D human poses across spatial, structural and temporal dimensions. In formula

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MPJPE}} + \alpha\mathcal{L}_{\text{bone}} + \gamma\mathcal{L}_{\text{velocity}} + \delta\mathcal{L}_{\text{acceleration}}$$

$\mathcal{L}_{\text{MPJPE}}$ measures the average Euclidean distance between the predicted 3D joint positions and the ground-truth joint positions; $\mathcal{L}_{\text{bone}}$ constrains the predicted pose to maintain bone lengths consistent with those of the real human body, preventing unrealistic stretching or compression; $\mathcal{L}_{\text{velocity}}$ is used to smooth the motion of joints between consecutive frames by penalising abrupt temporal changes, thereby ensuring smooth and continuous motion.

**The 3D human pose estimation results of the MTSA-former model.** As shown in fig. 4, the image on the left depicts the input 2D human keypoint detection, while the image on the right illustrates the reconstructed 3D human skeleton predicted by the model. This method can maintain high 3D reconstruction accuracy in gait detection scenarios, generating pose estimation results with strong structural consistency and smooth motion. It demonstrates the effectiveness of the improved model in multi-scale spatiotemporal feature fusion.

*a* *b*

*Fig. 4.* 3D pose estimation using the MTSA-former model:
*a* – input 2D human keypoint detection;
*b* – reconstructed 3D human skeleton predicted by the model

**Gait analysis techniques.** Both 2D and 3D gait analysis are essential tools for evaluating gait and identifying deviations or abnormalities in movement. Gait analysis plays a vital role in the diagnosis and treatment of various musculoskeletal and neurological disorders. Using the logical framework motion features – normal range and abnormality grading – clinical significance, the angular variation parameters commonly observed during gait are systematically categorised and interpreted [9]. As shown in the table, the maximum ankle dorsiflexion angle, peak knee flexion during the swing phase and pelvic obliquity at midstance represent the core lowerlimb motion patterns in the 2D plane, whereas the knee progression angle and maximum trunk lateral shift reflect integrated manifestations of coordination and balance control in 3D space [10].

**Gait analysis features**

| Feature | Expression dimension | Involved plane | Normal range | Mild abnormality | Moderate abnormality | Severe abnormality | Clinical significance |
|---|---|---|---|---|---|---|---|
| Maximum ankle dorsiflexion in stance | 2D single plane | Sagittal plane | Dorsiflexion by 10–15° | Dorsiflexion by 5–9° or 16–20° | Dorsiflexion by 0–4° or 21–30° | Plantarflexion position or dorsiflexion by >30° | Assess ankle range of motion and gastrocnemius tightness |
| Peak knee flexion in swing | 2D single plane | Sagittal plane | Flexion by 55–65° | Flexion by 45–54° or 66–75° | Flexion by 35–44° or 76–85° | Flexion by <35° or >85° | Assess foot clearance ability and quadriceps strength |
| Pelvic obliquity at mid-stance | 2D single plane | Coronal plane | Drop of stance side by 4–6° | Drop by 2–3° or 7–9° | Drop by 0–1° or 10–12° | Elevation or drop by >12° | Assess hip abductor strength and trunk stability |
| Knee progression angle at mid-stance | 3D composite | Sagittal and transverse planes | Knee forward direction by ±5° | External or internal rotation by 6–10° | External or internal rotation by 11–20° | Rotational deviation by >20° | Assess lower limb rotational alignment and knee joint stress |
| Maximum trunk lateral shift | 3D composite | Coronal and transverse planes | Lateral shift toward stance side by 3–5 cm | 2–3 or 6–7 cm | 1–2 or 8–10 cm | <1 or >10 cm | Assess balance control and energy efficiency |

In gait analysis these indicators use joint angular variations as the core quantitative diagnostic basis. By precisely measuring the angular range of specific joints during the gait cycle and comparing it with standard normal intervals, clinicians can quantitatively assess the severity of issues such as muscle weakness, joint contracture or neuromotor impairment. For example, an ankle dorsiflexion angle of less than 5° during the stance phase often indicates tightness of the gastrocnemius or soleus muscles, while a knee progression angle deviating more than ±20° suggests potential lowerlimb rotational malalignment or deviation of the knee joint's mechanical axis.

In summary, this table establishes an angular-based quantitative classification system for gait abnormalities and provides a unified, repeatable and comparable set of parametric references for gait assessment in both clinical practice and research. It holds significant value for the diagnosis and therapeutic evaluation of gait using 2D and 3D analysis frameworks.

## Experimental results

The lightweight, small-sample MTSA-former model was trained for 2D-to-3D human pose estimation. In fig. 5 the variations and convergence trends of key loss metrics during the training of the improved MTSA-former model over 23 epochs are illustrated. Both training and validation of mean per joint position error (MPJPE) exhibit consistent convergence, with validation error decreasing from 90.42 mm to a minimum of 41.12 mm at 18$^{th}$ epoch, representing a 54.5 % improvement (see fig. 5, *a*). The total loss curve demonstrates stable optimisation with occasional fluctuations at 10$^{th}$ and 19$^{th}$ epochs, likely due to learning rate scheduling (see fig. 5, *b*). The epoch-to-epoch validation changes reveal that major improvements occur in early training phases (see fig. 5, *c*), while fig. 5, *d*, shows that the best validation performance plateaus after 18$^{th}$ epoch, suggesting model convergence at this point. Overall, plots demonstrate a stable and well-converged training process for the 3D pose estimation model.

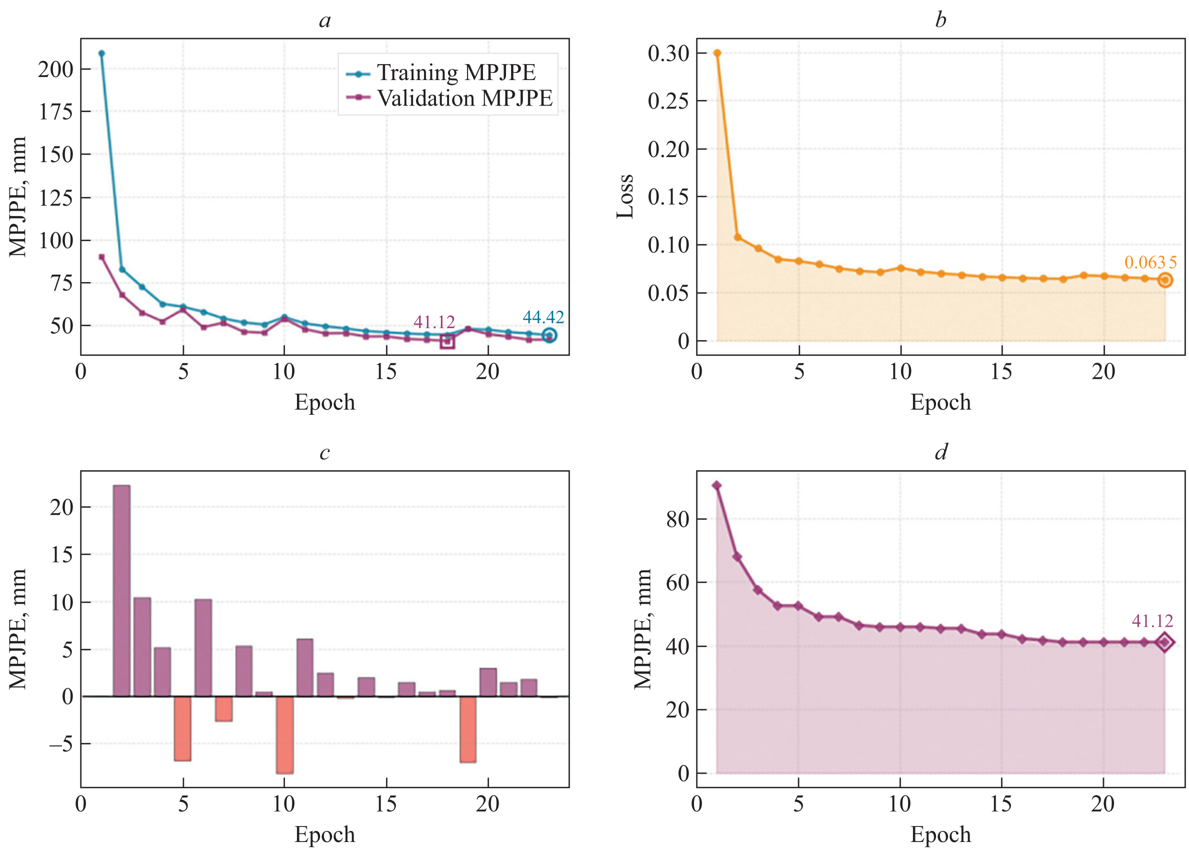


*Fig. 5.* Variations and convergence trends of key loss metrics during the training of the MTSA-former model:
*a* – training and validation MPJPE; *b* – total loss;
*c* – epoch-to-epoch validation change; *d* – best validation MPJPE progress

In fig. 6 the small-sample scenario on the Human3.6 dataset, reflecting the comparative relationship between model performance (measured by MPJPE) and model parameter size among various pose estimation models, based on a training set of 40 thsd samples and a test set of 8 thsd samples using the MTSA-former model framework is showed. The MTSA-former model achieves the lowest MPJPE of 41.12 mm with a relatively small number of parameters, demonstrating performance significantly superior to existing LSTM-based models such as PoseFormer and MHFormer [11; 12]. This indicates that the proposed method effectively reduces model complexity while maintaining high-precision pose estimation, showcasing a more optimised structural design and parameter utilisation efficiency.

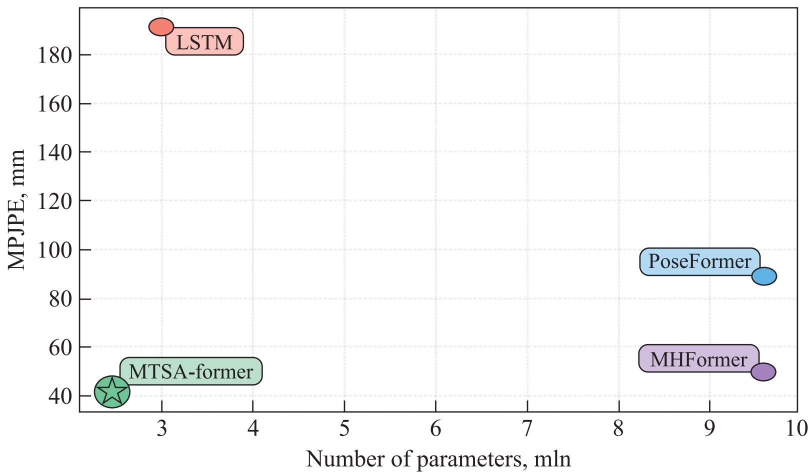


*Fig. 6.* Comparison of model parameters and accuracy.
The size of the circle corresponds to the performance score

The results of the comparative analysis of models LSTM, PoseFormer and MyModel, illustrating their MPJPE performance across 17 different joints, are presented in fig. 7. The comparison is used to evaluate variations and overall trends in spatial localisation accuracy for different parts of the human body. The proposed MyModel achieves the lowest MPJPE for all joints, significantly outperforming the baseline models LSTM and PoseFormer. This performance demonstrates that the model possesses clear advantages in local feature aggregation and global spatial structure preservation, enabling more accurate capture of the hierarchical relationships within human poses.

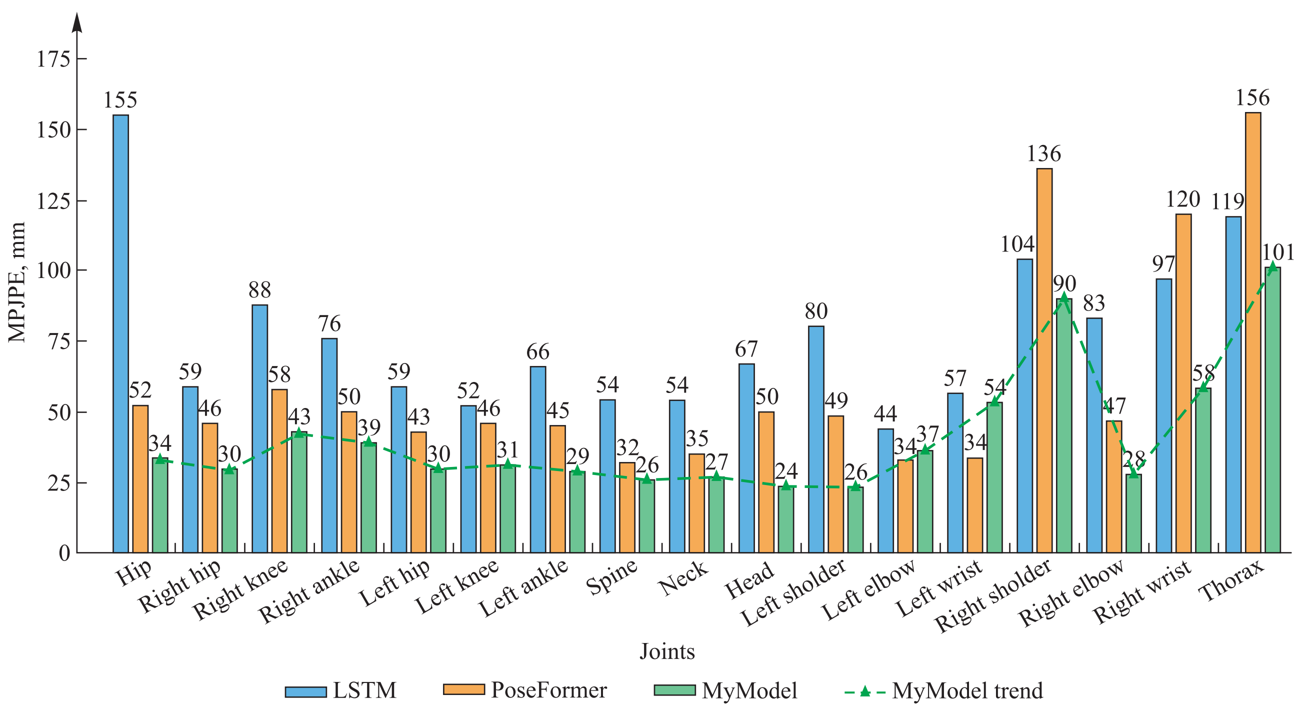


*Fig. 7.* Comparison of skeletal joints across different models

In gait analysis differences in key biomechanical parameters between healthy individuals and patients with Parkinson's disease are further examined (fig. 8). The core indicators, including foot-off events, ankle joint angles, trunk lateral shift and trunk forward inclination, are considered. The time-series analysis of foot-off height shows that the normal group maintains a stable average of 68.1 mm, whereas the Parkinson's disease group exhibits a markedly lower mean of 41.7 mm ($\Delta = 26.34$ mm), indicating reduced foot clearance and increased tripping risk (see fig. 8, *a*). The ankle joint angle distribution reveals a narrow, concentrated pattern in the normal group, while the Parkinson's disease group displays a broader, downward-shifted distribution (mean 151.4°), reflecting diminished mobility and increased rigidity (see fig. 8, *b*). Trunk lateral shift is minimal and smooth in the normal group but increases significantly in the Parkinson's disease group (amplitude 31.8 mm), accompanied by high-frequency oscillations (evidence of impaired trunk stability and hip abductor weakness) (see fig. 8, *c*). The probability density of trunk sagittal inclination indicates a forward shift in the Parkinson's disease group (from −89.2° to −81.1°), with minimal overlap between groups, illustrating the characteristic stooped posture in Parkinson's gait (see fig. 8, *d*). The integrated four-dimensional radar plot, encompassing ankle angle, foot-off height, trunk lateral shift and trunk forward tilt, shows that the normal group forms a balanced, near-boundary diamond, whereas the Parkinson's disease group exhibits a contracted, asymmetric shape, visually highlighting multi-system gait impairment (see fig. 8, *e*).

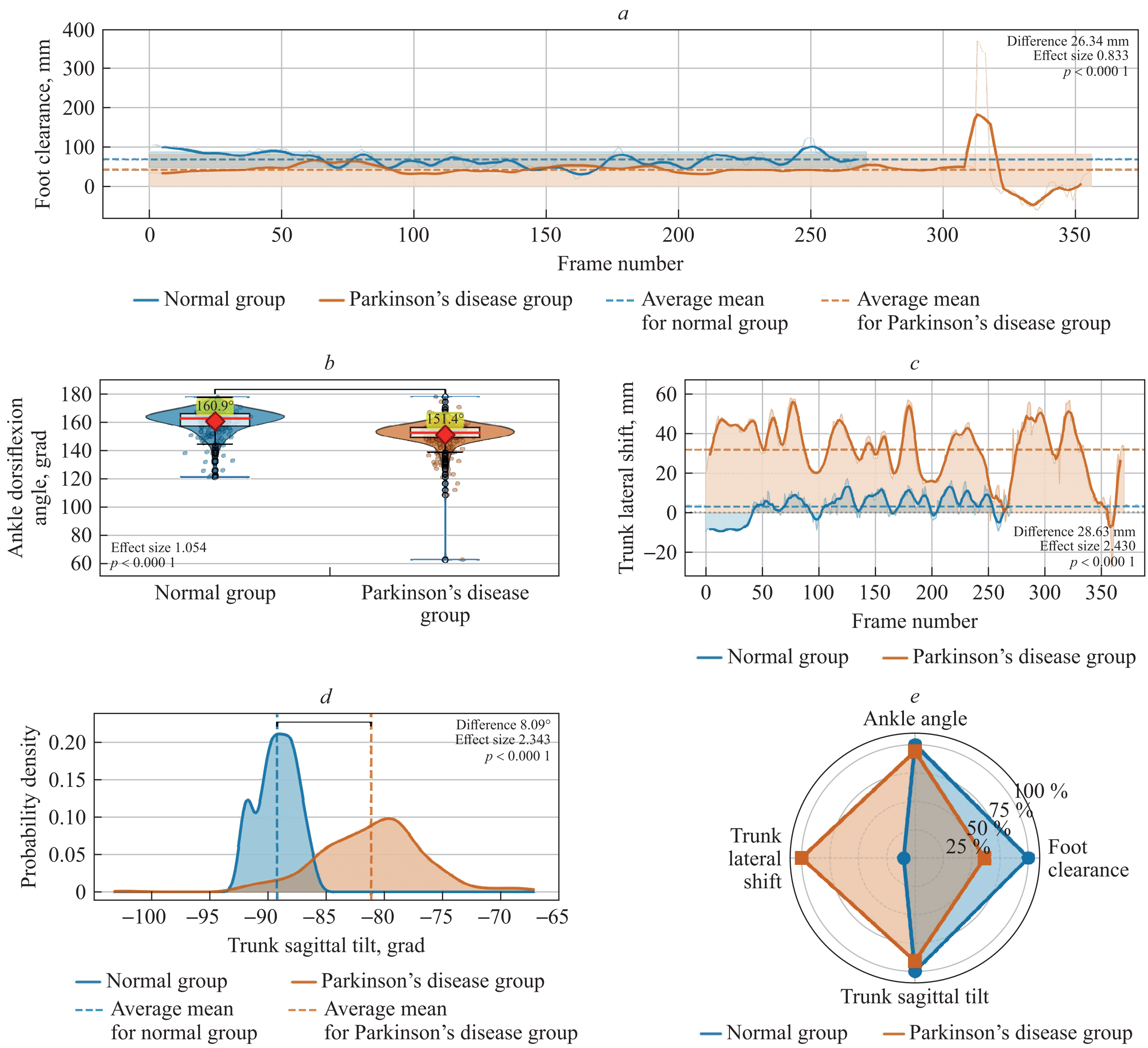


*Fig. 8.* Evaluation of gait in normal group and Parkinson's disease group:
*a* – foot clearance during swing phase; *b* – ankle angle;
*c* – trunk lateral displacement; *d* – trunk forward lean;
*e* – integrated four-dimensional radar plot

The visualisation results of a gait 2D motion analysis system implemented using the lightweight MediaPipe model, showcasing the capability of deep learning-based computer vision methods to perform real-time analysis of human dynamic behaviour under low computational cost conditions, is illustrated in fig. 9. The chart (see fig. 9) presents the dynamic waveform curves generated by the gait event analysis module, corresponding to the temporal motion variations of different body parts (such as the left heel and right toe). This system not only visualises the gait cycle but also enables the extraction of key gait features (including cadence, stride length and stance-to-swing ratio), providing valuable data support for subsequent applications in motor rehabilitation, behavioural monitoring and human biometric identification.

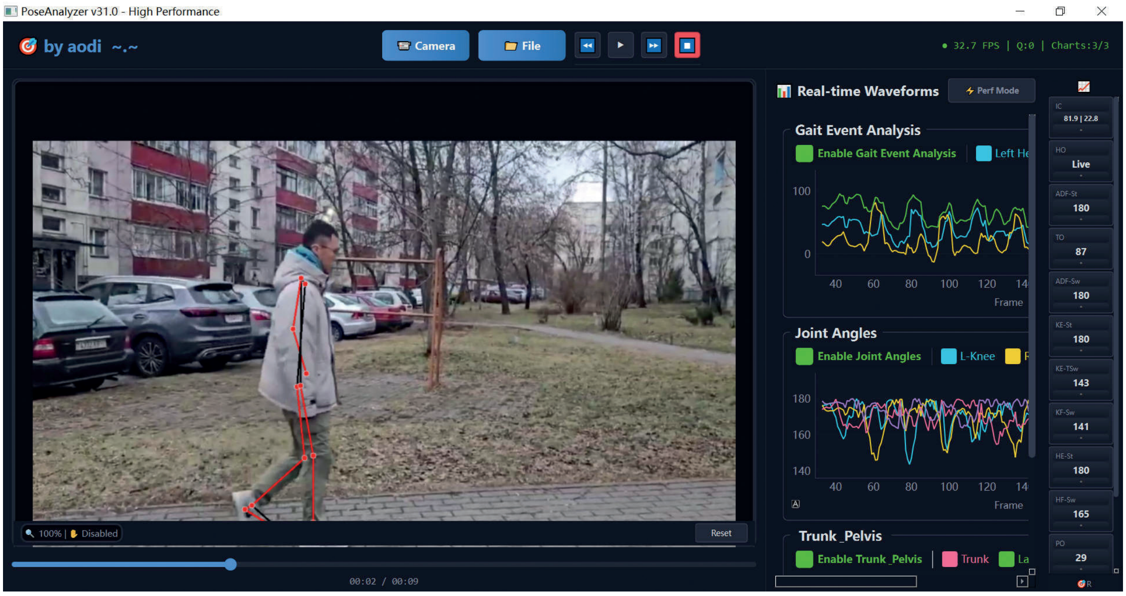


*Fig. 9.* Gait analysis system

A systematic academic analysis comparing healthy gait and Parkinsonian gait, incorporating the medical gait feature assessment table for quantitative inference, is provided in fig. 10. The analysis covers angular variations, motion plane characteristics and clinical significance, aiming to reveal differences in functional control and neuromuscular coordination between the two gait patterns. The normal's group left and right knee angle waveforms are showed mirror symmetry, stable periodicity (approximately one complete cycle per second) and an amplitude of 55–60°. The 3D kinematic characteristics fall within the normal physiological range, demonstrating symmetry, strong periodicity and good dynamic balance. In contrast, the Parkinson's disease group ankle, knee, hip and trunk exhibit are restricted ranges and markedly reduced amplitudes (around 30–35°). The gait cycle is irregular, and waveform peaks display noticeable disturbances. According to clinical assessment scales, this corresponds to a moderate-to-severe gait abnormality.

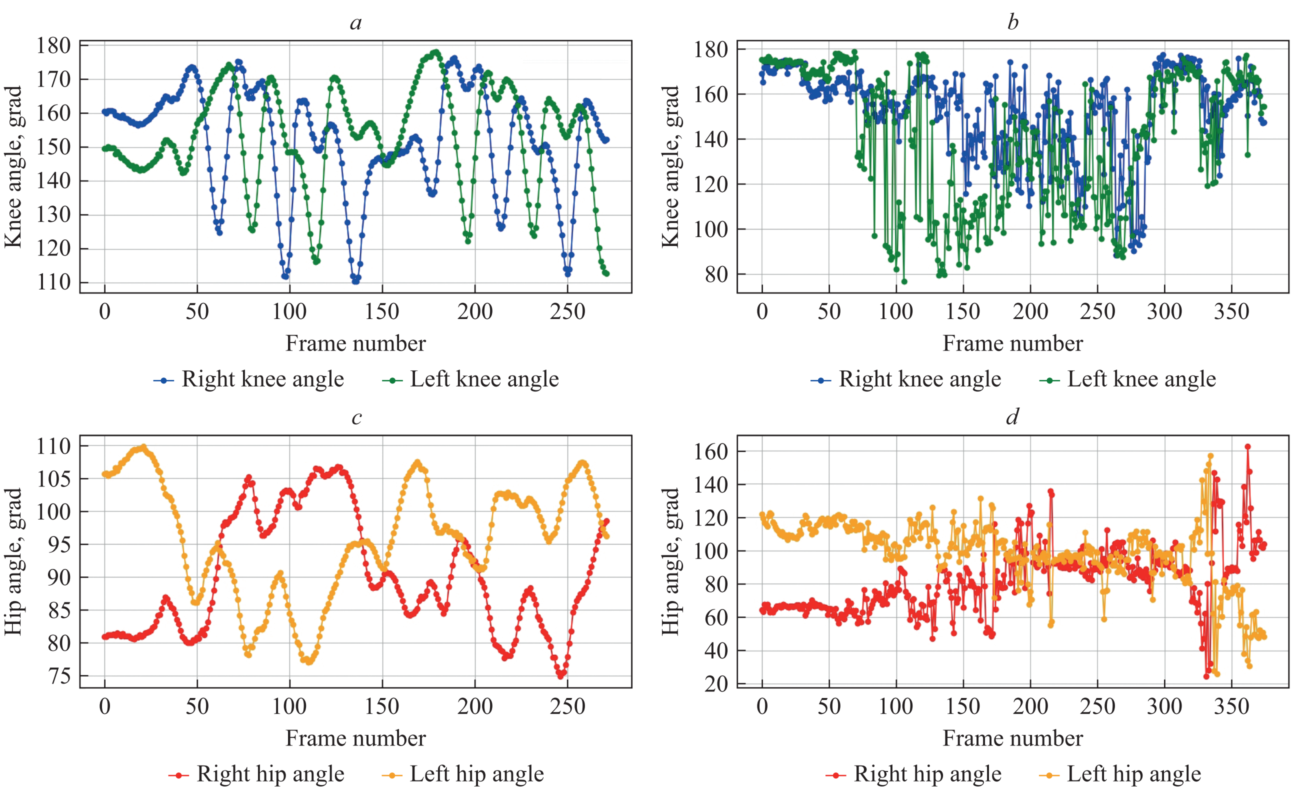


*Fig. 10.* 3D gait analysis, specifically knee joint angle changes (*a*, *b*), hip joint angle changes (*c*, *d*) of normal group (*a*, *c*) and Parkinson's disease group (*b*, *d*) (beginning)

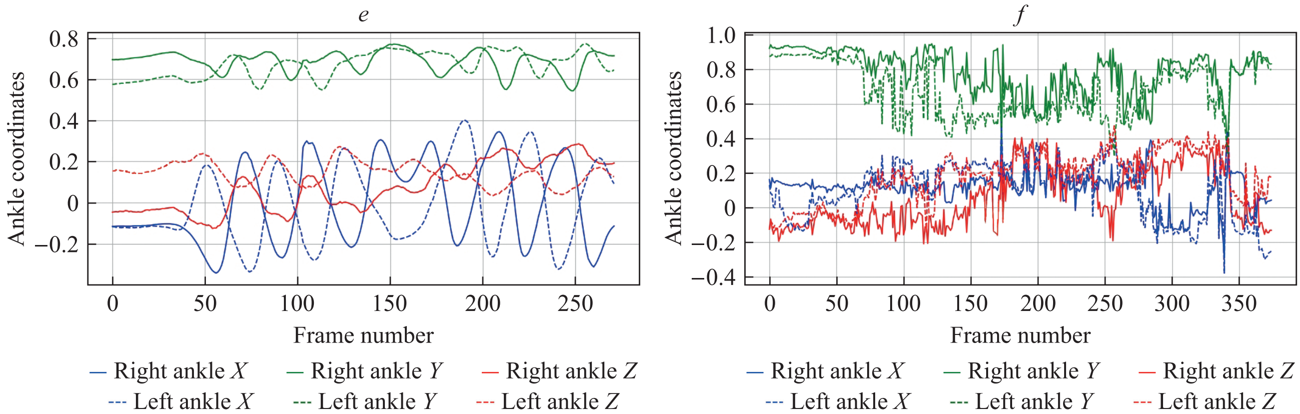


*Fig. 10.* 3D gait analysis, specifically ankle 3D coordinate changes,
of normal group (*e*) and Parkinson's disease group (*f*) (ending)

## Conclusions

This study presents a lightweight 2D and 3D collaborative gait analysis framework enabling high-precision, real-time modelling and quantitative clinical evaluation of human gait. The method integrates MediaPipe-based 2D keypoint detection with an enhanced MTSA-former 3D pose estimation network. Despite containing only 2.5 mln parameters, the model achieves superior 3D reconstruction accuracy (MPJPE = 41.12 mm) with real-time efficiency.

Comparative analyses of gait in normal group and Parkinson's disease group demonstrate the system's sensitivity in detecting abnormal variations in key biomechanical parameters, including ankle dorsiflexion, peak knee flexion, pelvic obliquity and trunk lateral shift. Clinical metrics further confirm its efficacy in distinguishing different motor impairment levels based on 2D kinematic and 3D spatiotemporal features. In gait of Parkinson's disease group the model identifies typical deficits such as limited ankle motion, inadequate knee flexion, reduced pelvic stability and impaired balance control.

## References

1. Rastegari E, Azizian S, Ali H. Machine learning and similarity network approaches to support automatic classification of Parkinson's diseases using accelerometer-based gait analysis. In: Bui TX, editor. *Proceedings of the 52nd Hawaii international conference on system sciences; 2019 January 8–11; Grand Wailea, USA.* Honolulu: ScholarSpace; 2019. p. 4231–4242. DOI: 10.24251/HICSS.2019.511.

2. Ťupa O, Procházka A, Vyšata O, Schätz M, Mařík J, Vališ M, et al. Motion tracking and gait feature estimation for recognizing Parkinson's disease using deep learning. *BioMedical Engineering OnLine*. 2015;14(1):97. DOI: 10.1186/s12938-015-0092-7.

3. Lugaresi C, Tang J, Nash H, McClanahan C, Uboweja E, Hays M, et al. MediaPipe: a framework for building perception pipelines. arXiv:1906.08172 [Preprint]. 2019 [cited 2025 August 24]: [9 p.]. Available from: https://arxiv.org/pdf/1906.08172.

4. Zeng W, Ouyang W, Luo P, Liu W, Wang X. 3D human mesh regression with dense correspondence. In: Institute of Electrical and Electronics Engineers. *Proceedings of the 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR); 2020 June 13–19; Seattle, USA.* [S. l.]: Institute of Electrical and Electronics Engineers; 2020. p. 7052–7061. DOI: 10.1109/CVPR42600.2020.00708.

5. Bazarevsky V, Grishchenko I, Raveendran K, Zhu T, Zhang F, Grundmann M. BlazePose: on-device real-time body pose tracking. arXiv:2006.10204 [Preprint]. 2020 [cited 2025 May 22]: [4 p.]. Available from: https://arxiv.org/pdf/2006.10204.

6. Grishchenko I, Ablavatski A, Kartynnik Y, Raveendran K, Grundmann M. Attention mesh: high-fidelity face mesh prediction in real-time. arXiv:2006.10962 [Preprint]. 2021 [cited 2025 May 1]: [4 p.]. Available from: https://arxiv.org/pdf/2006.10962.

7. Ionescu C, Papava D, Olaru V, Sminchisescu C. Human3.6M: large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2014;36(7):1325–1339. DOI: 10.1109/TPAMI.2013.248.

8. Wandt B, Rosenhahn B. RepNet: weakly supervised training of an adversarial reprojection network for 3D human pose estimation. In: Institute of Electrical and Electronics Engineers. *Proceedings of the 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR); 2019 June 15–20; Long Beach, USA.* [S. l.]: Institute of Electrical and Electronics Engineers; 2019. p. 7774–7783. DOI: 10.1109/CVPR.2019.00797.

9. Baker R, McGinley JL, Schwartz MH, Beynon S, Rozumalski A, Graham HK, et al. The gait profile score and movement analysis profile. *Gait & Posture.* 2009;30(3):265–269. DOI: 10.1016/j.gaitpost.2009.05.020.

10. Kainz H, Carty ChP, Modenese L, Boyd RN, Lloyd DG. Estimation of the hip joint centre in human motion analysis: a systematic review. *Clinical Biomechanics.* 2020;30(4):319–329. DOI: 10.1016/j.clinbiomech.2015.02.005.

11. Zheng C, Zhu S, Mendieta M, Yang T, Chen C, Ding Z. 3D human pose estimation with spatial and temporal transformers. In: Institute of Electrical and Electronics Engineers. *Proceedings of the 2021 IEEE/CVF International conference on computer vision (ICCV); 2021 October 10–17; Montreal, Canada.* [S. l.]: Institute of Electrical and Electronics Engineers; 2021. p. 11636–11645. DOI: 10.1109/ICCV48922.2021.01145.

12. Li W, Liu H, Tang H, Wang P, Van Gool L. MHFormer: multi-hypothesis transformer for 3D human pose estimation. In: Institute of Electrical and Electronics Engineers. *Proceedings of the 2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR); 2022 June 18–24; New Orleans, USA.* [S. l.]: Institute of Electrical and Electronics Engineers; 2022. p. 13147–13156.

*Received 24.12.2025 / revised 19.01.2026 / accepted 29.01.2026.*