Hermes: AI-Powered Mobile App for Automated Gait Pose Estimation and Analysis Using LLMs

Abdullah Magdy, Ahmed Mahmoud, Fady Mohsen, Mohamed Alaa, Mahmoud Mohamed, Ola Sarhan, Aliaa Rehan

***Abstract*--Gait analysis is a fundamental tool for detecting biomechanical abnormalities, creating rehabilitation plans, and enhancing athletic performance. Traditional gait analysis methods, such as visual observation or instrumented techniques, are often subjective, costly, or require the use of specialized equipment. This research presents a novel mobile application for automated temporal gait analysis using the low-computational MediaPipe Pose estimation library combined with Gemini 2.0 Flash, a large language model (LLM), to enable comprehensive gait analysis. The app processes the video data to extract temporal gait parameters and integrates patient-specific details (age, weight, prosthetics, medical condition, and injury) to provide structured, medically informed outputs. The system has a Celery-based backend to handle task queueing effectively such that scalability and responsiveness are guaranteed. This paper provides a summary of the methodology from current research [1] and an overview of the system's potential for clinical and community application. Future work entails the integration of Retrieval-Augmented Generation (RAG) into gait science literature for greater precision.**

***Index Terms*—Artificial Intelligence, Biomechanics, Computer Vision, Gait Analysis, Machine Learning, Medical Diagnostic Imaging, Mobile Applications, Motion Analysis, Pose Estimation, Rehabilitation**

I. Introduction

Gait analysis is at the forefront of clinical biomechanics, rehabilitation, sports science, and forensic investigation because it enables the detection of abnormalities in gait and assessment of intervention effectiveness [1]. Traditional gait analysis techniques, including visual observation, are subject to poor inter-rater reliability; meanwhile, instrumented systems, such as Vicon motion capture, are costly and necessitate specialized staff training [1].

Recent developments in markerless pose estimation, specifically using 2D video-based approaches, provide an affordable and readily available alternative [1, 2]. The current research follows on from the investigation by Hii et al. [1], who assessed *MediaPipe* Pose for temporal gait analysis against the Vicon system and reported a strong to excellent correlation for the majority of parameters. We recommend a mobile application that automatically conducts gait analysis using *MediaPipe* Pose for automatic keypoints extraction and Gemini 2.0 Flash for contextual analysis, incorporating patient-specific variables. Backend processing occurs on a scalable environment using Celery for task queuing, enabling efficient management of computationally intensive AI tasks.

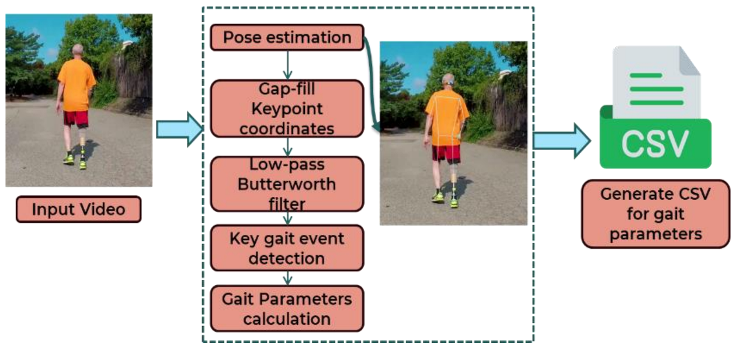

Figure.1. *Pose estimation and gait calculations workflow based on Mediapipe.*

This paper describes the application architecture, methodology, and how it relates to current literature as a practical clinical and community setting tool.

II. Literature Review

Gait analysis forms a fundamental part of clinical biomechanics, rehabilitation, sport science, and forensic investigation, enabling the quantification of human gait patterns to identify abnormalities, determine the effectiveness of interventions, and optimize performance [1], [3]. Conventionally, classical gait analysis techniques have utilized marker-based systems, in which reflective markers positioned over anatomical landmarks are tracked by an array of cameras to recreate three-dimensional joint kinematics [1], [22]. Whittle [3] utilized these systems to calculate accurate joint angles and trajectories, whereas Robertson et al. [22] employed marker-based setups to achieve detailed spatiotemporal gait parameters in controlled laboratory settings. Although these techniques provide significant accuracy, they are time-consuming, expensive, and environment-limited, and therefore their availability on a large scale is restricted [1], [3].

Recent advances in markerless pose estimation have transcended these limitations with computer vision and deep learning to directly infer body keypoints from RGB video without markers [1], [18], [19]. Kidziński et al. [18] demonstrated the use of deep convolutional neural networks for recovering 2D joint positions from single-camera video for gait analysis in everyday environments. Furthermore, Stenum et al. [19] verified OpenPose for the detection of body keypoints and estimation of spatiotemporal gait parameters with performances comparable to marker-based systems without requiring special equipment. Hii et al. [1] extended this idea using MediaPipe Pose, a low-complexity 3D pose estimation model, for temporal gait analysis with automation. They validated MediaPipe against the Vicon motion capture system and reported excellent intraclass correlation coefficients (ICC > 0.90) for stance time (left) and step time (left and right), good ICC (0.75–0.90) for stance time (right), swing time (left), and double support time (left to right), but moderate ICC (0.50–0.75) for swing time (right) and double support time (right to left) [1]. MediaPipe's low computational expense makes it suitable for real-world deployment, addressing the resource-hungry nature of models like OpenPose [1], [4].

Other markerless methods have investigated 3D camera systems, like Microsoft Kinect, for gait analysis [5–7]. Latorre et al. [5] validated Kinect v2 for gait analysis in stroke patients with good sensitivity and reliability for spatiotemporal parameters. However, Kinect-based systems have limitations because they are expensive, have a low frame rate (30 fps), and are sensitive to clothing and are not well-suited to dynamic gaits [8]. Viswakumar et al. [8] demonstrated OpenPose to accurately estimate knee flexion angles with negligible discrepancy, again demonstrating the advantage of 2D video-based methods. Studies with the Azure Kinect, as shown by Guess et al. [12], reported low relative errors in regard to spatiotemporal parameters; however, they recognized limitations in capturing higher walking speeds accurately due to frame rate limitations [12], [13].

Current research has opened up new frontiers in gait analysis by leveraging sophisticated deep learning models and large language models (LLMs) for improving automation and interpretability. Lin et al. [20] presented GaitGL, a two-branch network that combines global contextual information and local details to enhance gait phase classification and recognition performance on benchmarking datasets. In the same way, Li et al. [21] put forward JointsGait, a model-based approach leveraging gait graph convolutional networks and joints relationship pyramid mapping to obtain robust features from 2D joint data, alleviating clothing and viewing angle variations. These combined approaches demonstrate the potential for more general and more robust representations of gait [20], [21].

The use of LLMs represents a novel frontier in gait analysis for transforming complex kinematic data into clinically meaningful, human-readable results [16], [17]. Chivereanu et al. [16] used LLMs to generate comprehensive text reports from quantitative gait data for facilitating communication between clinicians and patients. Ye et al. [17] integrated large vision models with large language models (LLMs) to produce robust gait representations and synchronize them with natural language outputs to enhance diagnostic interpretability. These advancements are consistent with your project's usage of Gemini 2.0 Flash to offer contextual analysis based on inferred gait parameters and patient-specific information (e.g., age, weight, conditions).

Despite these improvements, issues still linger. Marker-based methods, while accurate, are impossible to apply in daily practice because of cost and initialization limitations [1], [22]. Markerless methods, despite being more common, are plagued by accuracy problems in complicated scenes and require validation across a variety of populations [1], [19]. LLM-based techniques, while promising, are at a research level and lack significant clinical verification [16], [17]. Your proposed mobile app leverages the efficacy of MediaPipe Pose [1] and Gemini 2.0 Flash's contextual evaluation [16], [17] in addressing these issues by offering an affordable, scalable solution for temporal gait analysis. By incorporating patient-specific variables, your system can provide personalized information, rendering it a viable solution for both community and clinical settings. Future studies, such as integrating Retrieval-Augmented Generation (RAG) with the gait science literature, will further enhance the accuracy and specificity of the system. This aligns with future trends in AI-based gait analysis [16], [20], [21].

III. Methods

A. System Architecture

The frontend of the mobile app records video, and the backend processes it using a Celery-based task queueing system to run AI computations. The backend, which is constructed using FastAPI, is responsive as it processes a maximum of three concurrent gait analyses, and any other requests are queued based on a first-in, first-out (FIFO) policy. Results are persisted in a PostgreSQL database and exposed within the user's "Gait Session" interface.

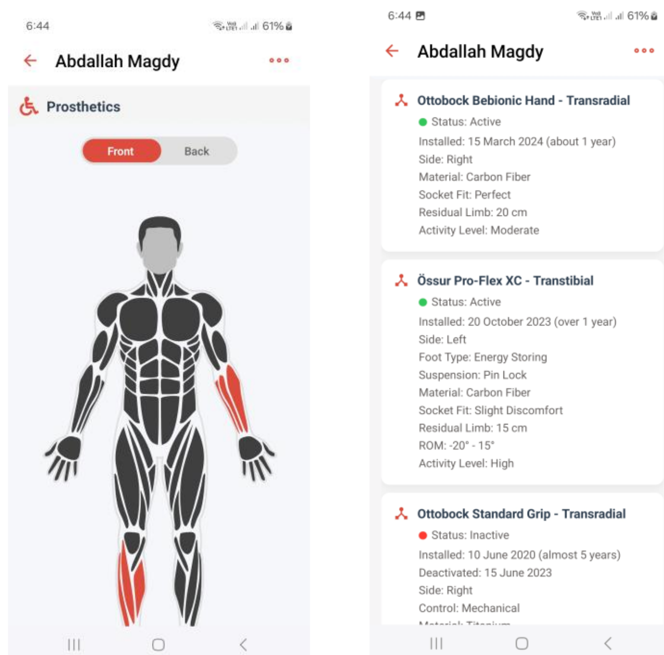

Figure.2. *Mobile applications: Interactive Prosthetics Management and Full Prosthetics History*

B. Analysis Pipeline

The pipeline consists of two stages:

1) MediaPipe-based Gait Data Extraction:

- *Video input (960x540 px, 25 fps) is run through MediaPipe Pose (BlazePose GHUM Heavy model) to identify 33 anatomical keypoints [1, 4]*
- *Gait reference events (toe-off, heel strike) are identified by calculation of relative distance from hip and foot index, and minima and peaks were identified using a 10th-order Butterworth low-pass filter with a cut-off of 0.1752 [1].*
- *the following temporal gait parameters were calculated and saved in a CSV file for each healthy individual:*

(i) Stance time: the duration between heel strike and toe-off of the same leg.
(ii) Swing time: the duration between toe-off and heel-strike of the same leg.
(iii) Step time: the duration between consecutive heel strikes of both feet.
(iv) Double support time: the duration between the heel strike of one leg and the toe-off of the contralateral leg. [1].

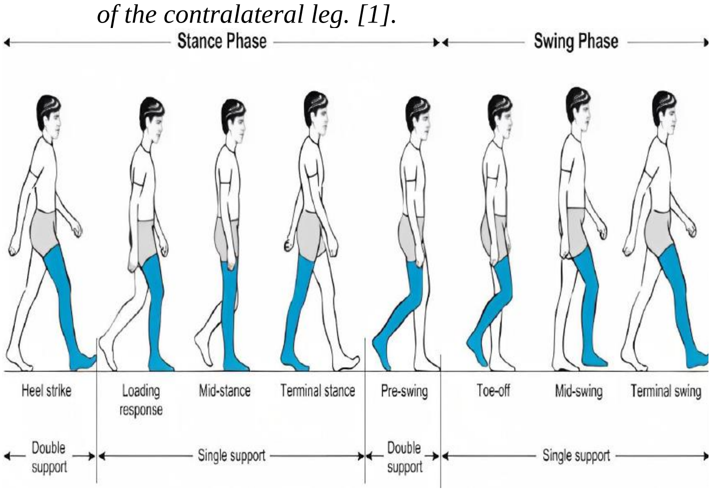

Figure.3. *Phase of the normal gait cycle.*

were recorded in the sagittal plane from a single camera (C1) to give optimum accuracy [1, 9]. Clinical populations, including individuals with cerebellar ataxia, will be added in future releases to assess gait variability [11].

IV. Results

V. Conclusion and Future Work

The proposed mobile application offers an affordable, automated system for temporal gait analysis by leveraging the performance of MediaPipe Pose with the contextual analysis of Gemini 2.0 Flash. By integrating patient-specific data, the system offers personalized results for clinical and rehabilitation settings. Its back-end scalability allows effortless deployment, making gait analysis accessible beyond specialized labs. Future work will include the merging of Retrieval-Augmented Generation (RAG) with gait science literature, including textbooks and journal articles, to enhance the LLM's accuracy and specificity. Incorporating spatial gait parameters and lower limb kinematics will enable pathology-specific evaluation, e.g., for Friedreich ataxia or cervical spondylotic myelopathy [14, 15]. Validation across a broad variety of clinical populations, including individuals with prosthetics or walking aids, will also increase the system's usefulness.

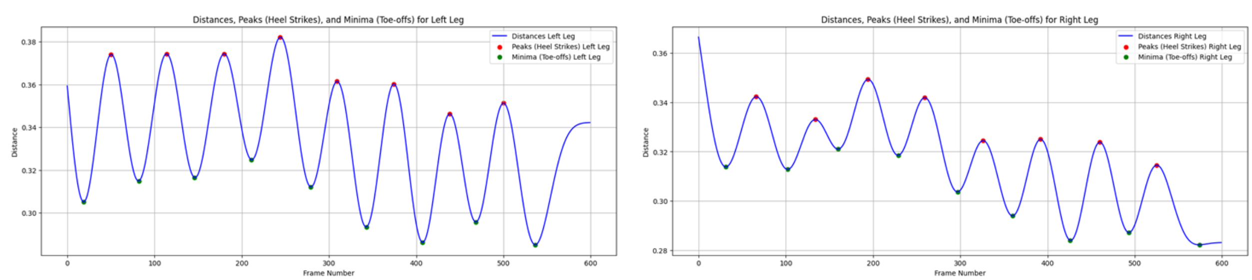

Figure.4. *The relative distance between peaks (heel strikes), and minima (toe-offs) for left and right leg.*

2) *LLM-Based Analysis:*

- *Extracted gait information, in combination with patient-specific inputs (age, weight, prosthetics, medical conditions, injuries), is formatted through Python functions and passed to Gemini 2.0 Flash via a LangChain Structured Output prompt.*
- *The prompt queries:*
  - *Gait pattern characteristics.*
  - *Gait irregularities associated with patient conditions.*
  - *Exercise prescription for gait improvement.*
  - *Weight effect on prosthetics and gait.*
  - *Long-term risks of the current gait pattern.*
- *The LLM answers with a JSON object that includes detailed analysis (Markdown, ≥500 words), summary (100–200 words), lists of abnormalities, recommendations, exercises, and long-term risks.*

C. Data Collection

The system uses an openly available dataset [10] comprising synchronized video and Vicon motion capture data of 31 healthy participants (10 female and 21 male, 20-65 years) [1]. Videos

VI. References

Here are the references with numbers in square brackets:

[1] C. S. T. Hii, K. B. Gan, N. Zainal, N. M. Ibrahim, S. Azmin, S. H. M. Desa, B. van de Warrenburg, and H. W. You, "Automated gait analysis based on a marker-free pose estimation model," Sensors, vol. 23, no. 14, p. 6489, Jul. 2023, doi: 10.3390/s23146489.

[2] D. Seckiner, X. Mallett, P. Maynard, D. Meuwly, and C. Roux, "Forensic gait analysis—Morphometric assessment from surveillance footage," Forensic Sci. Int., vol. 296, pp. 57–66, Mar. 2019, doi: 10.1016/j.forsciint.2019.01.007.

[3] M. W. Whittle, "Clinical gait analysis: A review," Hum. Mov. Sci., vol. 15, no. 3, pp. 369–387, Jun. 1996, doi: 10.1016/0167-9457(96)00006-1.

[4] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 172–186, Jan. 2021, doi: 10.1109/TPAMI.2019.2959571.

[5] J. Latorre, C. Colomer, M. Alcañiz, and R. Llorens, "Gait analysis with the Kinect v2: Normative study with healthy individuals and comprehensive study of its sensitivity, validity, and reliability in individuals with stroke," J. Neuroeng. Rehabil., vol. 16, no. 1, p. 97, Jul. 2019, doi: 10.1186/s12984-019-0566-8.

[6] Y. Ma, K. Mithraratne, N. Wilson, X. Wang, Y. Ma, and Y. Zhang, "The validity and reliability of a Kinect v2-based gait analysis system for children with cerebral palsy," Sensors, vol. 19, no. 7, p. 1660, Apr. 2019, doi: 10.3390/s19071660.

[7] Y. Ma, K. Mithraratne, N. Wilson, Y. Zhang, and X. Wang, "Kinect v2-based gait analysis for children with cerebral palsy: Validity and reliability of spatial margin of stability and spatiotemporal variables," Sensors, vol. 21, no. 6, p. 2104, Mar. 2021, doi: 10.3390/s21062104.

[8] A. Viswakumar, V. Rajagopalan, T. Ray, and C. Farimi, "Human gait analysis using OpenPose," in Proc. 2019 Fifth Int. Conf. Image Inf. Process. (ICIIP), Shimla, India, Nov. 2019, pp. 310–314, doi: 10.1109/ICIIP47207.2019.8985955.

[9] E. D'Antonio, J. Taborri, E. Palermo, S. Rossi, and F. Patane, "A markerless system for gait analysis based on OpenPose library," in Proc. 2020 IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC), Dubrovnik, Croatia, May 2020, pp. 1–6, doi: 10.1109/I2MTC43012.2020.9129357.

[10] C. S. Tony Hii, K. B. Gan, N. Zainal, N. M. Ibrahim, S. A. Md Rani, and N. A. Shattar, "Marker free gait analysis using pose estimation model," in Proc. 2022 IEEE 20th Student Conf. Res. Dev. (SCOReD), Bangi, Malaysia, Nov. 2022, pp. 109–113, doi: 10.1109/SCOReD57089.2022.10087643.

[11] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," J. Chiropr. Med., vol. 15, no. 2, pp. 155–163, Jun. 2016, doi: 10.1016/j.jcm.2016.02.012.

[12] T. M. Guess, R. Bliss, J. B. Hall, and A. M. Kiselica, "Comparison of Azure Kinect overground gait spatiotemporal parameters to marker based optical motion capture," Gait Posture, vol. 96, pp. 130–136, Jul. 2022, doi: 10.1016/j.gaitpost.2022.05.014.

[13] J. A. Albert, V. Owolabi, A. Gebel, C. M. Brahms, U. Granacher, and B. Arnrich, "Evaluation of the pose tracking performance of the Azure Kinect and Kinect v2 for gait analysis in comparison with a gold standard: A pilot study," Sensors, vol. 20, no. 18, p. 5104, Sep. 2020, doi: 10.3390/s20185104.

[14] A. Cimorelli, A. Patel, T. Karakostas, and R. J. Cotton, "Portable in-clinic video-based gait analysis: Validation study on prosthetic users," medRxiv, 2022, doi: 10.1101/2022.06.07.22276110.

[15] A. Gouelle and F. Mégrot, "Interpreting spatiotemporal parameters, symmetry, and variability in clinical gait analysis," in Handbook of Human Motion, B. Müller, S. I. Wolf, G.-P. Brueggemann, Z. Deng, A. McIntosh, F. Miller, and W. S. Selbie, Eds., Cham, Switzerland: Springer, 2016, pp. 1–20, doi: 10.1007/978-3-319-30808-1_104-1.

[16] R. Chivereanu, A. Cosma, A. Catruna, R. Rughinis, and E. Radoi, "Aligning actions and walking to LLM-generated textual descriptions," arXiv preprint arXiv:2404.12192, 2024, doi: 10.48550/arXiv.2404.12192.

[17] D. Ye, C. Fan, J. Ma, X. Liu, and S. Yu, "BigGait: Learning gait representation you want by large vision models," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024, pp. 2002–2010, doi: 10.48550/arXiv.2402.19122.

[18] L. Kidziński, B. Yang, J. L. Hicks, et al., "Deep neural networks enable quantitative movement analysis using single-camera videos," Nat. Commun., vol. 11, p. 4054, Aug. 2020, doi: 10.1038/s41467-020-17807-z.

[19] J. Stenum, C. Rossi, and R. T. Roemmich, "Two-dimensional video-based analysis of human gait using pose estimation," PLoS Comput. Biol., vol. 17, no. 4, p. e1008935, Apr. 2021, doi: 10.1371/journal.pcbi.1008935.

[20] B. Lin, S. Zhang, M. Wang, L. Li, and X. Yu, "GaitGL: Learning discriminative global-local feature representations for gait recognition," arXiv preprint arXiv:2208.01380, 2022, doi: 10.48550/arXiv.2208.01380.

[21] N. Li, X. Zhao, and C. Ma, "JointsGait: A model-based gait recognition method based on gait graph convolutional networks and joints relationship pyramid mapping," arXiv preprint arXiv:2005.08625, 2020, doi: 10.48550/arXiv.2005.08625.

[22] D. G. E. Robertson, G. E. Caldwell, J. Hamill, G. Kamen, and S. N. Whittlesey, Research Methods in Biomechanics. Champaign, IL: Human Kinetics, 2004. [Online].