PostCorrect: Real-Time Posture Correction and Feedback Using Vision Transformers and LLMs

UMKC School of Science and Engineering

Abhinav Kochar, Ahmed Alanazi

Computer Science, School of Science and Engineering

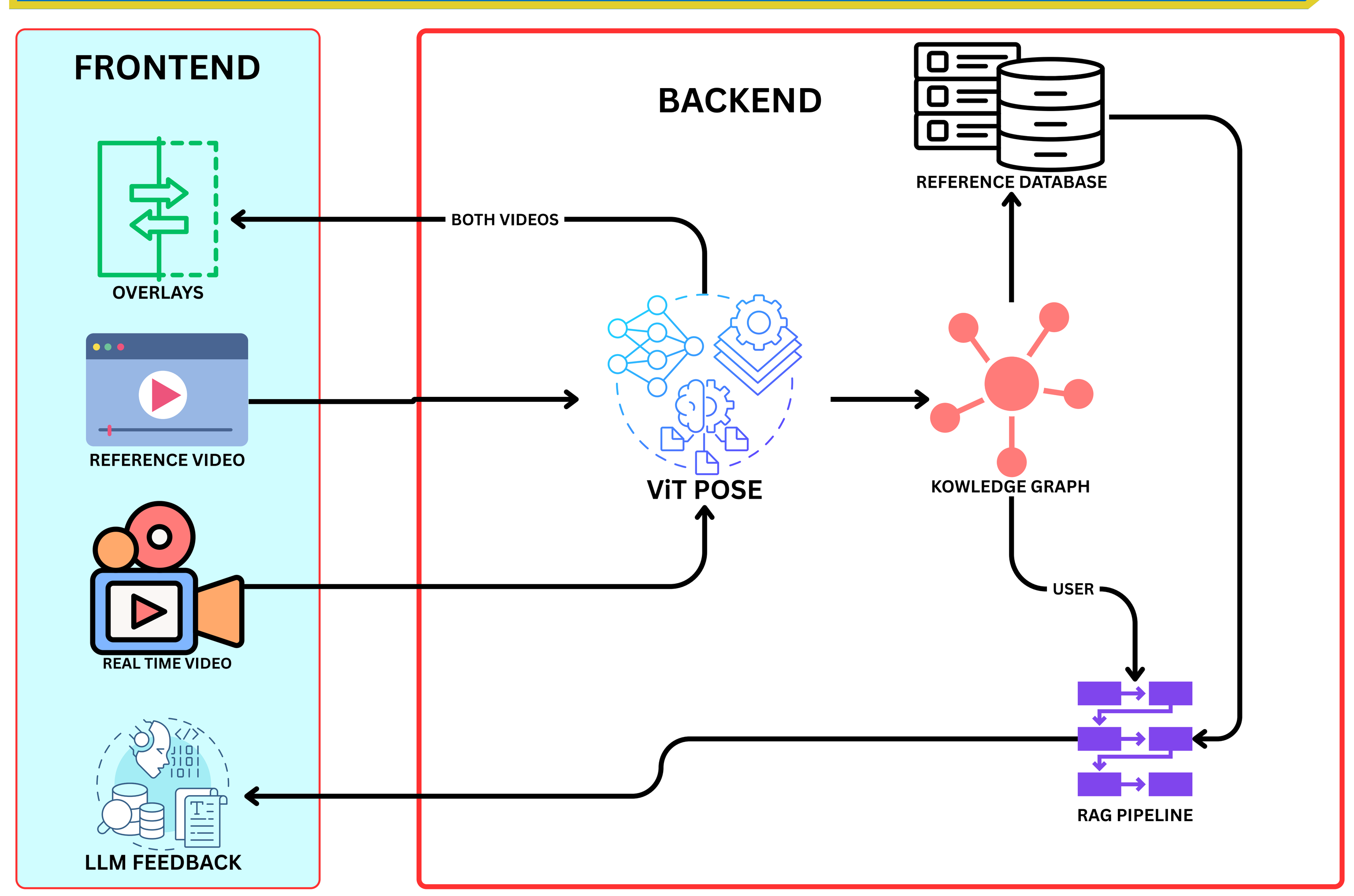
Context

Posture-focused exercises are a cornerstone of both **physical rehabilitation**—where precise form ensures optimal recovery and prevents re-injury—and general fitness, where correct technique boosts performance and efficiency. However, real-time feedback on form and technique is rarely available outside of professional supervision, making it challenging to correct mistakes and adapt movements in the moment. The result can be inefficient workouts, stalled progress, and even increased risk of injury due to repeated errors or overcompensation.

PostCorrect bridges this gap by harnessing a multidisciplinary blend of technologies: it combines computer vision (for extracting body keypoints), temporal analysis (to track movement across time), similarity metrics (to compare user poses with reference models), and large language models (LLMs) to deliver immediate, **personalized posture advice**. This synergy creates an interactive system that adapts to each user's body and movement style, offering not only a color-coded overlay that pinpoints misalignments, but also context-aware text prompts that explain how to fix those issues. By applying a data-driven approach to exercise form, PostCorrect empowers users—whether at home or in a clinic—to refine their technique, reduce the likelihood of injury, and gain more from every exercise session.

This poster provides an overview of the project's architecture, methodology, and key outcomes, highlighting how advanced AI and biomechanical insights can be merged to produce **real-time, user-centric feedback**. From the intricacies of keypoint detection and knowledge graph storage to the user-facing, large language model–based guidance, PostCorrect demonstrates a scalable and accessible way to enhance exercise safety and efficiency for a wide range of users.

System Architecture



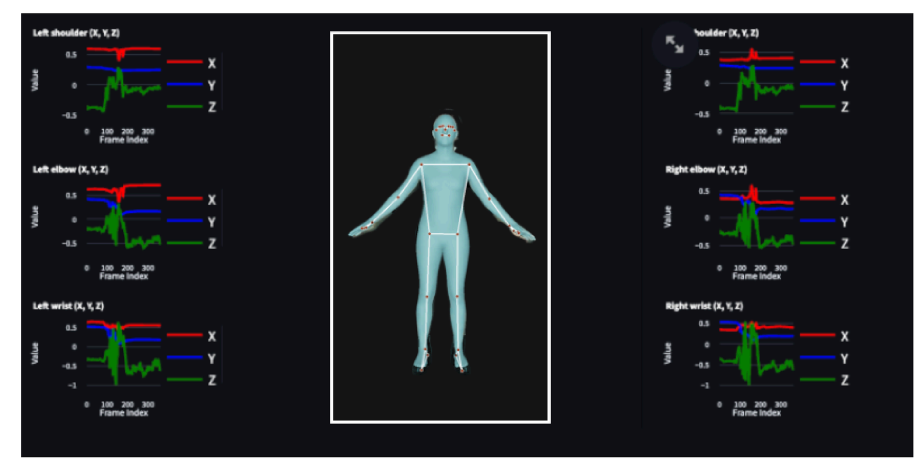
The PostCorrect pipeline begins with the intake of a live video stream or uploaded exercise footage. Each frame is processed through a **Vision Transformer**-based model to extract precise body keypoints, identifying major joints and their spatial coordinates. These keypoints are then filtered using statistical methods such as **Principal Component Analysis (PCA)** to isolate the most relevant joints for the specific exercise being performed.

The refined keypoint data is compared against reference movements stored in a structured knowledge graph, which encodes biomechanical constraints and ideal joint trajectories. A similarity engine calculates the degree of alignment between the user's posture and the reference, identifying any significant deviations. This information is then passed to a **Retrieval-Augmented Generation (RAG)** module, where a large language model interprets the detected errors and generates personalized feedback. The user interface overlays the skeletal structure on the video feed in real time—using green to indicate correct posture and red to highlight misalignment—while simultaneously displaying the LLM-generated correction prompts to guide the user through proper execution.

References

1. Xu, Y., Zhang, Z., Wang, S., et al. (2022). ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation. arXiv:2204.12484.
2. Jolliffe, I. T. (2002). Principal Component Analysis (2nd ed.). Springer.
3. Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS.
4. OpenAI. (2023). GPT-4 Technical Report. https://openai.com/research/gpt-4
5. Cao, Z., Hidalgo, G., Simon, T., et al. (2017). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. CVPR.
6. Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A Review of Knowledge Graphs in AI. Communications of the ACM.
7. Chen, C., et al. (2021). TransPose: Keypoint Localization via Transformer. arXiv:2103.15397.

Keypoint Extraction & Selection



Vision Transformers (**ViTPose**) are used to detect and track up to 33 body keypoints in real time, capturing the user's posture frame by frame. The raw data is normalized to account for differences in body proportions and camera angles, ensuring consistency across users. This normalized set is then passed through an importance-ranking algorithm, such as Principal Component Analysis (PCA), to identify the most influential joints for a given exercise.

Temporal graphs are generated to map the motion of these key joints over time, offering insights into movement flow, symmetry, and deviations. **Keypoint selection graphs** further highlight which joints have the greatest impact on posture evaluation, allowing the system to focus on regions like shoulders, hips, or knees depending on the activity.

This targeted filtering not only improves processing speed but also ensures that feedback is meaningful and specific. It allows the system to distinguish between minor, irrelevant variations and true biomechanical errors. Over time, this approach helps create a **personalized understanding** of user movement, paving the way for adaptive feedback and long-term progress tracking.

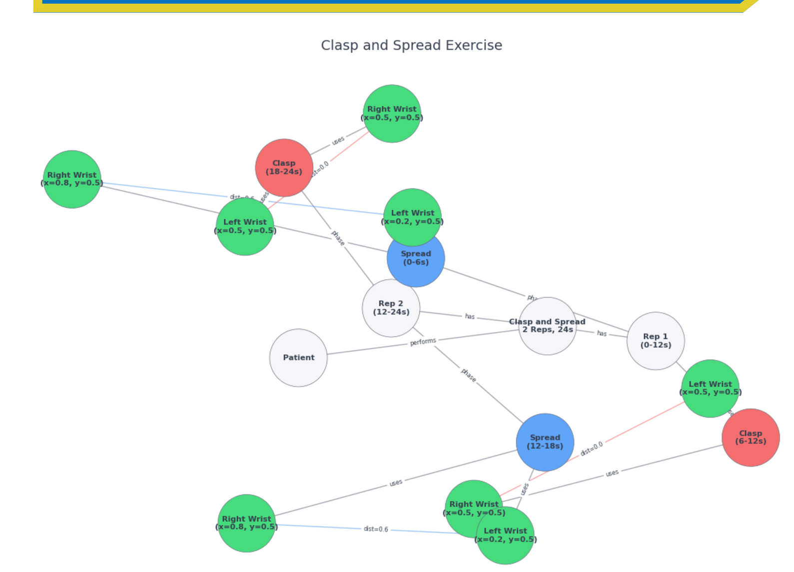
Similarity Metrics & Graphs

To evaluate how accurately a user's posture aligns with the ideal reference, similarity scores are calculated by comparing normalized joint positions and angles against reference patterns stored in the **knowledge graph**. These scores capture how well form is maintained across each phase of movement.

Similarity graphs visualize these scores over time, making it easy to **identify moments where the user deviates from correct posture**—such as at the bottom of a squat or during arm extension. This helps pinpoint specific phases that need improvement and ensures feedback is focused and actionable.

Over time, these comparisons provide a clear picture of technical progress and highlight recurring alignment issues that require attention.

Knowledge Graph



The knowledge graph encodes biomechanical best practices for each exercise, organizing posture checkpoints as nodes—such as joint angles, positions, or movement phases—with edges defining their temporal and spatial relationships.

When a misalignment is detected, the system queries the graph to retrieve the most relevant corrective information based on the specific deviation. This allows feedback to remain **exercise-specific, context-aware**, and biomechanically valid.

The structured nature of the graph also supports scalability, enabling new exercises and variations to be added efficiently while maintaining accurate and personalized feedback.

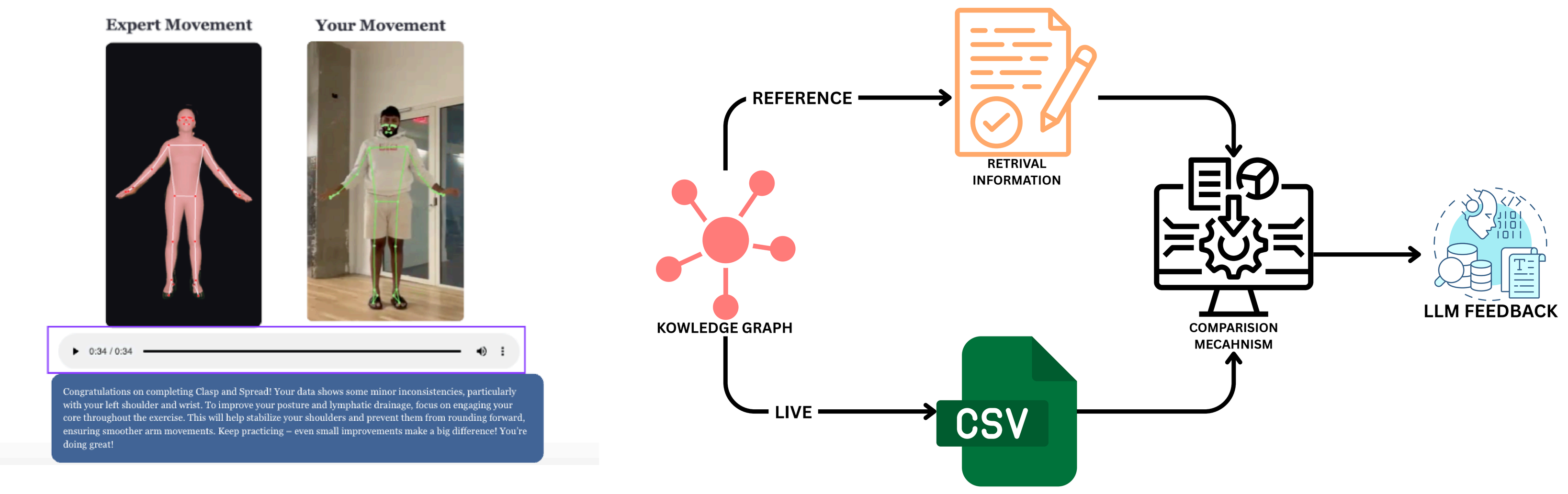
Conclusion & Future Directions

This work demonstrates the potential of combining advanced computer vision, structured knowledge representation, and natural language feedback to support and enhance exercise performance. The system's modular architecture enables adaptability across different exercise types, user profiles, and rehabilitation or fitness environments. By offering real-time corrective feedback through both visual and linguistic channels, the approach encourages body awareness, consistency, and safe movement execution.

Future developments may include the integration of wearable sensors for more detailed biomechanical data, such as muscle activation via **EMG;** support for multi-user sessions in group fitness or **clinical therapy**; adaptive learning that adjusts feedback based on **individual history** and progress; and optimized deployment for mobile and cloud platforms to extend accessibility.

Expanding the knowledge graph to cover more complex or condition-specific exercises can further increase the system's clinical value. Ultimately, this approach lays the foundation for intelligent, **data-driven movement coaching** that is accessible, personalized, and scalable across health, sports, and wellness domains.

LLM-Based Feedback

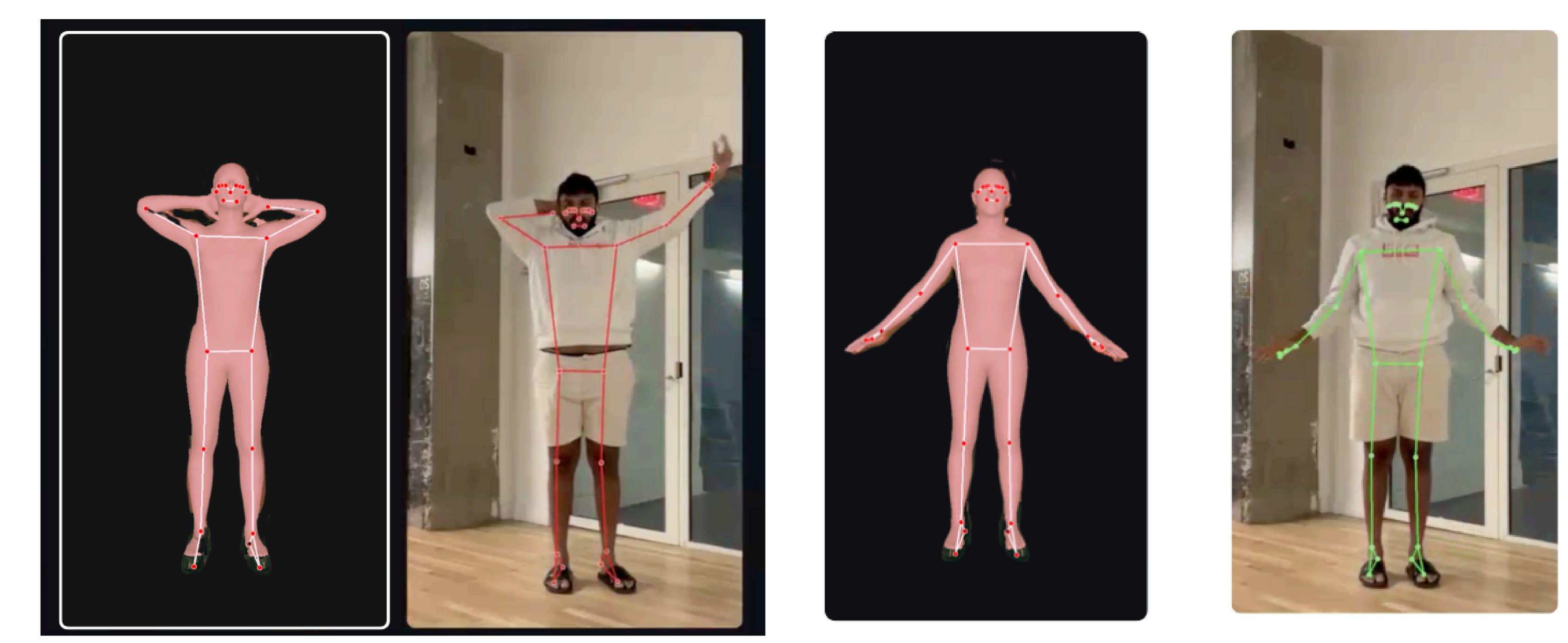


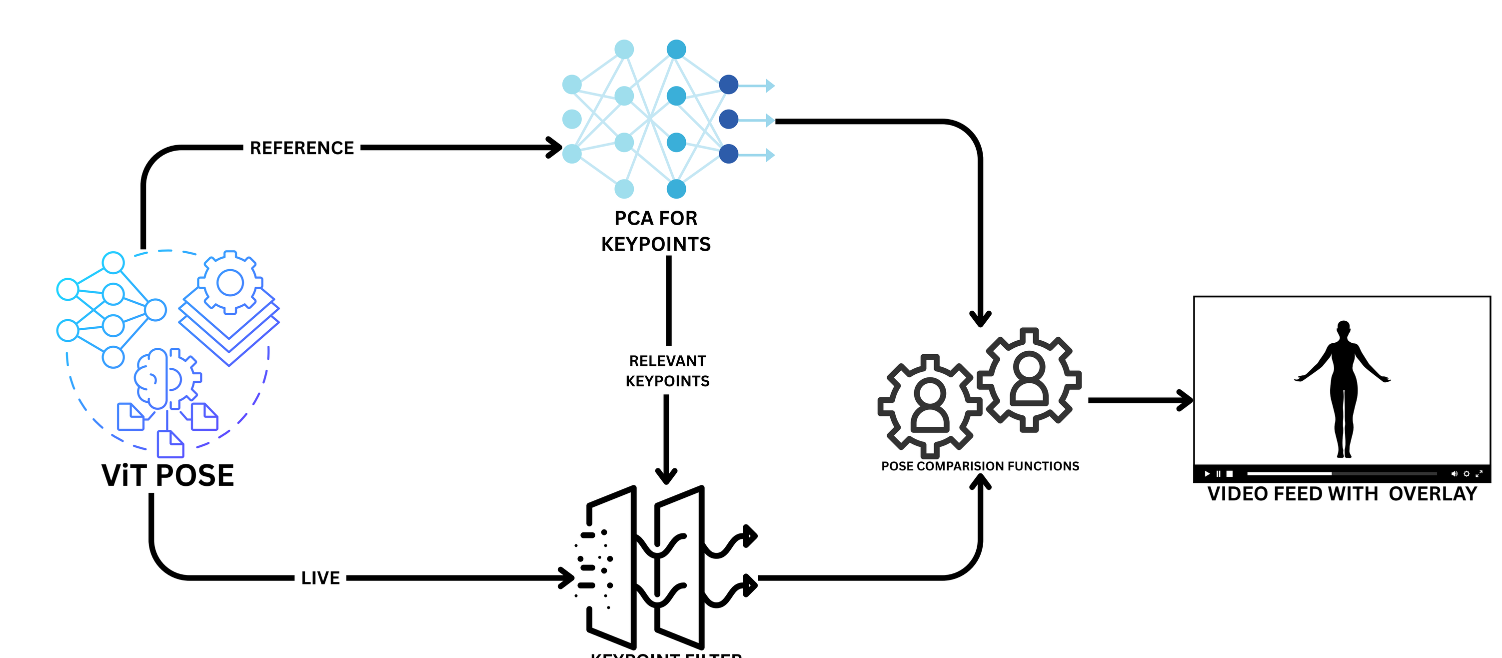
The natural language feedback component distinguishes the system from traditional numeric or visual-only solutions by offering personalized, context-aware guidance. When a posture deviation is detected, the system queries the **knowledge graph** to extract relevant biomechanical insights, which are then processed through a Retrieval-Augmented Generation (RAG) **pipeline**. This pipeline enables a large language model to translate technical posture data into clear, actionable suggestions.

Instead of presenting raw values or vague instructions, the user receives intuitive feedback such as, **"Keep your spine neutral to avoid leaning forward,"** directly tied to their current movement. This enhances understanding and helps the user make immediate corrections during exercise.

The combination of structured knowledge and natural language output improves user engagement and learning, making feedback more accessible to individuals at all fitness or recovery levels.

Real-Time Overlay & Video Feedback





The visual feedback system provides users with an immediate, intuitive understanding of their posture through a real-time skeletal overlay. As the user performs an exercise, joint positions are tracked and color-coded based on alignment accuracy—**green** indicates joints within **acceptable** biomechanical thresholds, while red highlights areas where the form deviates.

This continuous visual feedback enables users to recognize and correct their posture **without delay**, reinforcing proper movement patterns as they exercise. By reducing reliance on external supervision and promoting self-awareness, the system helps **minimize repetitive errors** and supports safer, more effective training or rehabilitation sessions.

Acknowledgements

The authors gratefully acknowledge the mentorship and guidance of **Dr. Yugyung Lee** throughout the project. This work was conducted under the leadership of **Dr. Mei R. Fu** as Principal Investigator. Special thanks are also extended to Dr. Gregory King and Dr. Liu Bowen for their valuable insights and support.