# Literature Review: Real-Time Hand Gesture Recognition Using MediaPipe and Computer Vision Techniques

## 1. Introduction

Hand gesture recognition has become an essential component of modern human–computer interaction (HCI), enabling intuitive interfaces for robotics, augmented reality (AR), virtual reality (VR), and sign-language translation. With advancements in computer vision and deep learning, recognizing hand poses from real-time video streams has become increasingly feasible on consumer-grade devices. The proposed project, Real-Time Hand Gesture Recognition Using MediaPipe and Computer Vision Techniques, focuses on enhancing existing gesture-tracking systems by introducing new gesture classes and improving recognition accuracy through dataset expansion and preprocessing optimization. To situate this work within the broader research landscape, three key papers are reviewed: (1) Zhang et al. (2020), introducing MediaPipe Hands, (2) Li et al. (2023), providing a comprehensive review of deep learning-based gesture recognition, and (3) Ahmed et al. (2021), surveying vision-based dynamic gesture recognition techniques.

## 2. Summary of Related Works

### 2.1 Zhang et al. (2020): MediaPipe Hands – Real-Time Landmark Detection

Zhang et al. (2020) presented MediaPipe Hands, an on-device, real-time hand tracking framework capable of detecting 21 3D landmarks per hand. The system employs a lightweight palm detector followed by a landmark regression model, optimized for performance on both CPU and mobile devices. The framework achieves high frame-rate processing, enabling gesture-based control systems without reliance on heavy GPU computation. This foundational work directly informs the proposed project's base architecture. By leveraging MediaPipe's efficient hand landmark detection, the current project focuses on extending gesture classification capabilities beyond static gestures to a broader set of interactive motions. However, Zhang et al. (2020) primarily focus on hand detection rather than gesture recognition. Their system stops at locating landmarks, leaving gesture interpretation to downstream models—one of the limitations the current project aims to address.

### 2.2 Li et al. (2023): Deep Learning Approaches in Gesture Recognition

Li et al. (2023) conducted an extensive review of deep learning-based hand gesture recognition, categorizing methods into static and dynamic gesture frameworks. The paper compares architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based models, highlighting their effectiveness across datasets like Hagrid, EgoHands, and SHREC. The authors emphasize challenges including lighting variations, occlusions, and user dependency, which often reduce model generalization in real-world conditions. For the proposed project, this review provides the theoretical foundation for designing a CNN-based classifier trained on extended datasets. The paper's comparative analysis also supports incorporating data augmentation and transfer learning strategies to enhance robustness. The main limitation of Li et al. (2023) is that it remains conceptual—while it summarizes advances, it lacks implementation details or real-time constraints that our system explicitly addresses.

### 2.3 Ahmed et al. (2021): Vision-Based Dynamic Gesture Recognition

Ahmed et al. (2021) focused on dynamic gesture recognition using computer vision and deep learning techniques. Their survey analyzed 3D CNNs, optical flow methods, and skeleton-based learning, with an emphasis on modeling temporal dependencies in video sequences. The authors discussed benchmark datasets and compared models' trade-offs between accuracy and computational efficiency. This research is particularly relevant to the proposed project's future extension into continuous gesture recognition from video streams. It reinforces the importance of maintaining computational efficiency while achieving real-time inference, aligning closely with the goals of MediaPipe-based systems. Nevertheless, the study's focus on dynamic gestures limits its immediate applicability to static gesture classification. Still, it provides valuable insight into the evolution of gesture recognition from static images toward motion-based understanding.

## 3. Comparative Analysis and Relation to Proposed Work

All three reviewed studies contribute complementary insights to the proposed hand gesture recognition system. Zhang et al. (2020) provide the technical foundation through MediaPipe's fast and accurate landmark extraction. Li et al. (2023) outline the broader deep learning context and the need for dataset diversity and augmentation. Ahmed et al. (2021) highlight efficiency trade-offs in dynamic gesture modeling that inform future system scalability. Together, these works illustrate a continuum from low-level detection (Zhang et al.) to high-level classification strategies (Li et al.) and sequence modeling (Ahmed et al.). The proposed system bridges these approaches by combining MediaPipe's efficiency with CNN-based classification to recognize additional gesture types. Furthermore, by retraining on the Hagrid dataset with augmented samples, the project addresses limitations in dataset diversity and generalization emphasized across prior literature.

## 4. Conclusion

The reviewed papers collectively demonstrate significant progress in gesture recognition, yet they also expose critical research gaps. Current models either prioritize accuracy over real-time performance or focus solely on hand localization without gesture interpretation. The proposed project seeks to overcome these limitations by developing a lightweight, extendable gesture recognition system capable of operating efficiently in real time. By integrating MediaPipe's detection pipeline with CNN-based classification and dataset expansion, this work aims to improve the accuracy, flexibility, and real-world usability of gesture-based interaction systems.

## References

1. Ahmed, I., Bukhari, S. S., Ali, M., & Malik, M. I. (2021). Vision-based dynamic hand gesture recognition: A survey. Pattern Recognition Letters, 146, 1–13. https://doi.org/10.1016/j.patrec.2021.03.017
2. Li, Y., Yang, X., Wang, P., & Zhang, W. (2023). A review on hand gesture recognition using deep learning. IEEE Access, 11, 28951–28974. https://doi.org/10.1109/ACCESS.2023.3246679
3. Zhang, F., Bazarevsky, V., Vakunov, A., Sung, G., Chang, C. L., & Grundmann, M. (2020). MediaPipe hands: On-device real-time hand tracking. arXiv preprint arXiv:2006.10214. https://arxiv.org/abs/2006.10214