Hand gesture recognition

Arpan Tyagi[1], Dev Joshi[2], Manish Kushwaha[3], Nitu Pandey[4]
IIMT Group of College

***Abstract*—Hand gesture recognition (HGR) has emerged as a significant area of research in the fields of human–computer interaction (HCI), computer vision, and artificial intelligence. It enables natural and intuitive communication between humans and machines by interpreting hand movements and postures as input commands. This paper presents a comprehensive study on hand gesture recognition techniques, focusing on image-based and sensor-based approaches. Traditional methods employing feature extraction and machine learning algorithms are compared with recent advancements using deep learning architectures such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The study also explores challenges including background noise, illumination variations, occlusion, and inter-user variability, which affect system accuracy and robustness. Furthermore, applications of HGR in virtual reality, sign language interpretation, robotics, and assistive technologies are discussed. The results highlight that integrating deep learning with efficient preprocessing and real-time optimization significantly enhances recognition accuracy and computational efficiency, paving the way for more immersive and accessible interactive systems.**

***Index Terms*—Hand Gesture Recognition, Human–Computer Interaction, Deep Learning, Computer Vision, Sign Language Recognition, CNN, Vision Transformer**

I. INTRODUCTION

Hand gesture recognition (HGR) plays a vital role in enabling intuitive and natural human-computer interaction by interpreting human hand movements as input commands. With the rapid advancement of computing power and sensing technologies, HGR has evolved into a crucial component for applications such as sign language translation, virtual and augmented reality, robotics, healthcare, and smart device control[1].

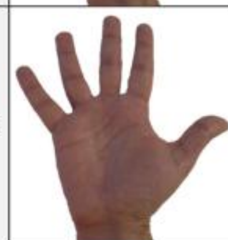
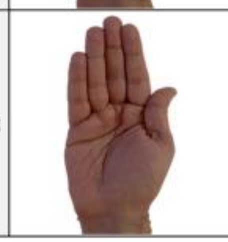
Traditional interaction methods relying on keyboards, mice, or touchscreens are being complemented and, in some cases, replaced by gesture-based interfaces that offer greater accessibility and a more immersive experience. The primary goal of HGR systems is to detect and accurately classify static postures or dynamic sequences of hand gestures captured through various input modes[2].

Initially, sensor-based approaches like data gloves and accelerometers were popular for capturing hand articulations but faced issues like user discomfort and high cost. Consequently, vision-based approaches using cameras—such as RGB, depth sensors, or time-of-flight devices—have become the dominant paradigm. Vision-based HGR presents unique challenges, including segmenting the hand from complex backgrounds, extracting discriminative features that represent the hand's shape and motion, and classifying these gestures under varying lighting conditions and across different users[3].

Over the years, the introduction of machine learning techniques improved classification accuracy by learning from handcrafted features, while recent breakthroughs in deep learning—especially convolutional neural networks and recurrent architectures—have dramatically enhanced the robustness and scalability of HGR systems[4].

Despite these advances, challenges like occlusion, real-time processing constraints, and user variability persist, motivating ongoing research to devise adaptive, efficient, and high-precision algorithms for real-world deployment.

This research paper reviews current methodologies, datasets, and emerging trends in hand gesture recognition, with an emphasis on computer vision and deep learning techniques. Understanding these approaches is essential to advance natural user interfaces that foster seamless communication between humans and machines[5].

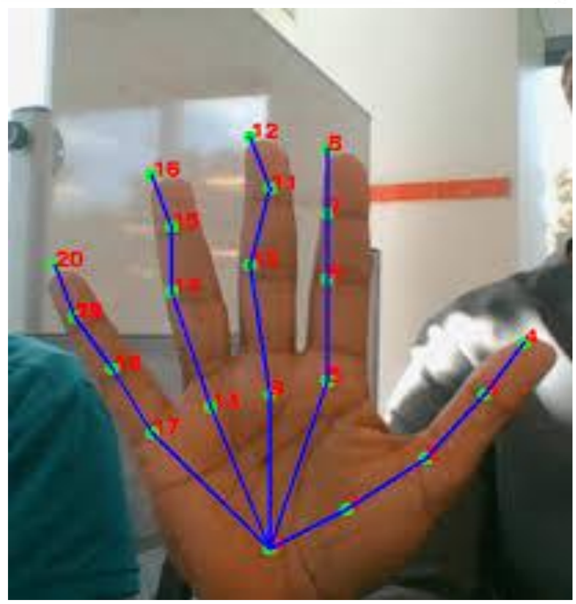
1		Thumbs up	2		Extension of index and middle fingers, flexion of others	3		Flexion of ring and little fingers, extension of others
4		Thumb opposing base of little finger	5		Abduction of all fingers	6		Fingers flexed together in fist
7		Pointing index	8		Adduction of extended finger	9		Wrist extension with closed hand

II. METHODS FOR HAND GESTURE RECOGNITION

Hand gesture recognition systems employ a combination of computer vision, machine learning, and deep learning techniques to interpret human hand movements. Recent advances have favored vision-based approaches, eliminating the need for cumbersome wearable sensors and enabling more natural and flexible interactions[6].

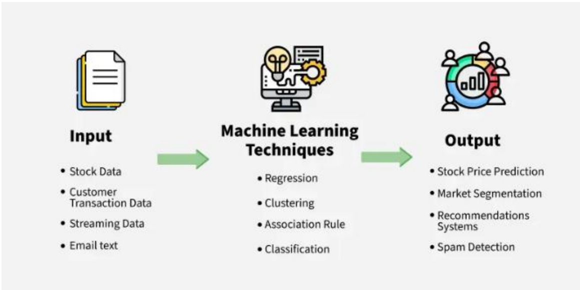
2.1. Vision-Based Gesture Recognition

- This method utilizes cameras (RGB, depth, or 3D sensors) to capture image sequences or video of hand movements. Key steps include:
- Hand Segmentation: Isolating the hand region from the background using skin color models, background subtraction, or depth cues.
- Feature Extraction: Capturing shape, motion, orientation, and texture features of the hand. Features can include contours, histograms of oriented gradients (HOG), or keypoint-based descriptors.
- Gesture Classification: Applying machine learning algorithms or deep neural networks to classify extracted features into predefined gesture categories[7].



2.2. Machine Learning Approaches

- Traditional ML: Support Vector Machines (SVM), k-Nearest Neighbors (kNN), and Random Forest classifiers are employed using handcrafted features from images or sensor data. These require careful feature engineering and perform well on small datasets[8].
- Hidden Markov Models (HMM): Particularly useful for dynamic gesture recognition by modeling temporal sequences of gesture features.

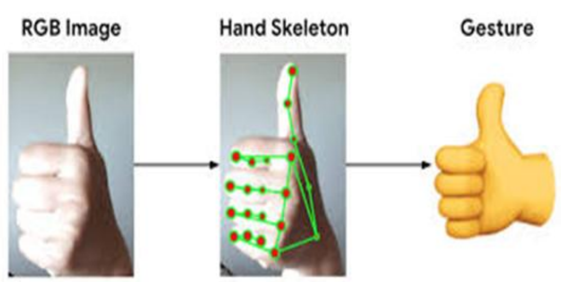


2.3. Deep Learning Techniques[9]

- Convolutional Neural Networks (CNNs): Automatically extract spatial features from images for static or dynamic gesture recognition. CNNs learn hierarchical feature representations, improving accuracy and robustness.
- Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM): Capture temporal dependencies in gesture sequences, enabling recognition of continuous or complex dynamic gestures.
- Hybrid Models: Combining CNNs with RNNs/LSTMs to jointly learn spatial and temporal information.
- Transformers: Emerging architectures for sequence modeling in gesture recognition, capturing long-range dependencies more effectively.

2.4. Sensor-Based Methods (Historical)

- Data gloves, accelerometers, and electromyographic sensors were used to capture hand joint angles or muscle signals. While accurate, these methods require hardware and are less user-friendly compared to vision-based systems.



III. APPLICATIONS

Hand gesture recognition has diverse and impactful applications that span multiple fields, enabling natural and intuitive human-machine interactions. This technology allows users to control devices and systems through simple hand movements and gestures, eliminating the need for traditional input devices such as keyboards and mice[10].

One of the primary applications of hand gesture recognition is sign language interpretation, which serves as a vital communication aid for the deaf and hard-of-hearing community. Automated recognition systems translate signed gestures into text or speech in real time, bridging communication gaps and enhancing accessibility. These systems combine computer vision and natural language processing to understand complex gestures and contextual nuances.

In virtual reality (VR) and augmented reality (AR) domains, gesture recognition allows users to interact naturally within immersive environments. Users can manipulate virtual objects, navigate menus, or control avatars using hand movements, creating an engaging and seamless experience without physical controllers. This enhances gaming, training simulations, and remote collaboration[11].

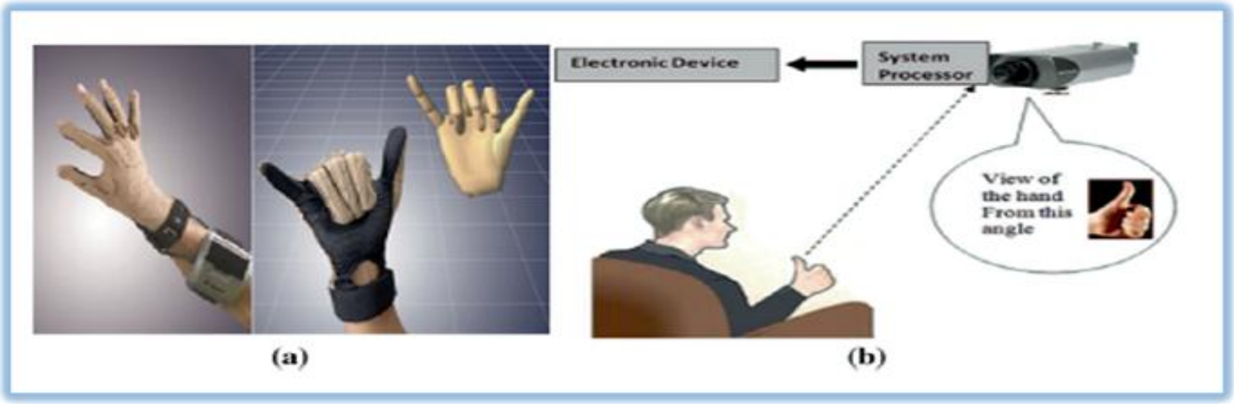
Robotics and automation also benefit greatly from hand gesture recognition. Robots equipped with gesture interpretation can respond dynamically to human commands, enabling safer and more efficient cooperation in manufacturing, healthcare, and service industries. Gesture-based controls offer a non-contact method to manage robotic systems, which is especially useful in hazardous or sterile environments.

In smart home and IoT (Internet of Things) environments, gesture recognition enables users to control appliances, lighting, and entertainment systems effortlessly. Touchless commands improve hygiene and convenience, particularly important in healthcare settings or public spaces.

Healthcare applications leverage hand gesture recognition for assistive technologies, empowering individuals with mobility impairments to control wheelchairs, prosthetics, or communication aids. Gesture-controlled devices enhance independence and quality of life for disabled users.

The technology also finds application in interactive displays and kiosks, where users can navigate information screens or perform transactions with simple gestures, reducing the need for physical contact.

Education and training sectors adopt hand gesture recognition for interactive learning modules and real-time feedback in skill acquisition, such as music or physical therapy exercises.



IV. DATASETS AND BENCHMARKS

Datasets and benchmarks play a crucial role in the development, evaluation, and comparison of hand gesture recognition (HGR) systems. These curated datasets provide standardized, annotated collections of gesture samples captured under various conditions, enabling researchers to train machine learning models and objectively assess their performance.

Several widely adopted benchmark datasets have shaped progress in this field. For instance, the SHREC dataset is popular for 3D hand pose and gesture recognition, featuring a variety of static and dynamic hand poses recorded from multiple viewpoints. The DHG-14/28 dataset offers a challenging set of dynamic hand gestures with varying complexity, designed to test algorithms' ability to handle temporal sequences and intra-class variation. The NVGesture dataset captures natural human gestures with complex backgrounds and includes RGB, depth, and infrared modalities to facilitate multi-modal learning.

Other notable datasets include the IPN Hand dataset, which contains over 4,000 gesture instances and 800,000 frames from 50 subjects, emphasizing complex, continuous gestures in real-world conditions with cluttered backgrounds and varying illumination. This dataset highlights real-life challenges by incorporating natural motions without transitional states and variable gesture speeds, providing a realistic benchmark for deep learning models.

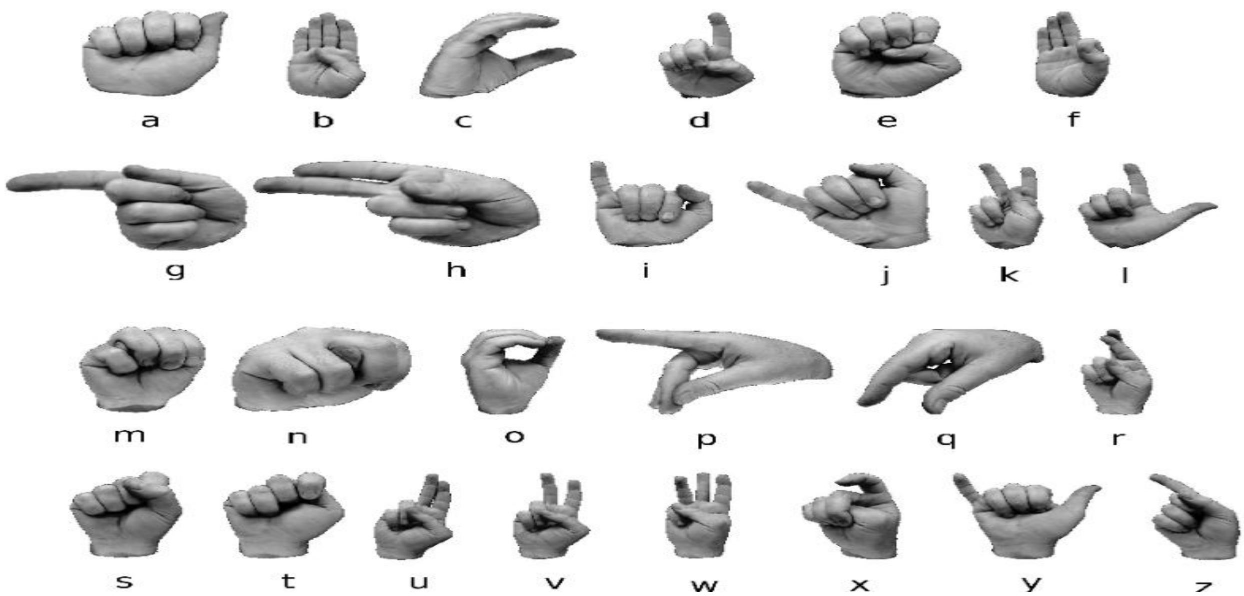
Moreover, datasets like putEMG integrate surface electromyography (sEMG) signals along with video and depth data, enabling multimodal approaches that combine muscle activation patterns with visual inputs. Such datasets are valuable for wearable HGR systems and serve as references for evaluating classification algorithms utilizing physiological signals.

The HaGRID dataset is distinguished by its large scale, containing over half a million labeled images across 18 gesture classes, annotated with bounding boxes for hand detection. It stands out for its diversity with thousands of subjects, wide scene variation, dynamic and static gestures, and extensive image heterogeneity, facilitating robust model training and evaluation.

Benchmarks derived from these datasets typically involve metrics such as accuracy, precision, recall, F1-score, and confusion matrices, with evaluation tasks spanning static classification, temporal sequence recognition, continuous gesture detection, and real-time performance. These benchmarks enable comparative analysis of different algorithms, including classical machine learning models, convolutional neural networks (CNNs), recurrent models, and transformer architectures.

The availability of these comprehensive and challenging datasets has accelerated advancements in HGR by fostering innovation and benchmarking progress in generalization, robustness, and computational efficiency. However, limitations persist, including dataset biases, constrained gesture vocabularies, and limited environmental variability.

Future work continues to expand dataset diversity, incorporate multi-modal sensory data, and establish benchmarks for emerging applications such as sign language translation, human-robot collaboration, and augmented reality control. Such efforts aim to bridge the gap between laboratory experimentation and deployment of reliable hand gesture recognition in real-world scenarios.

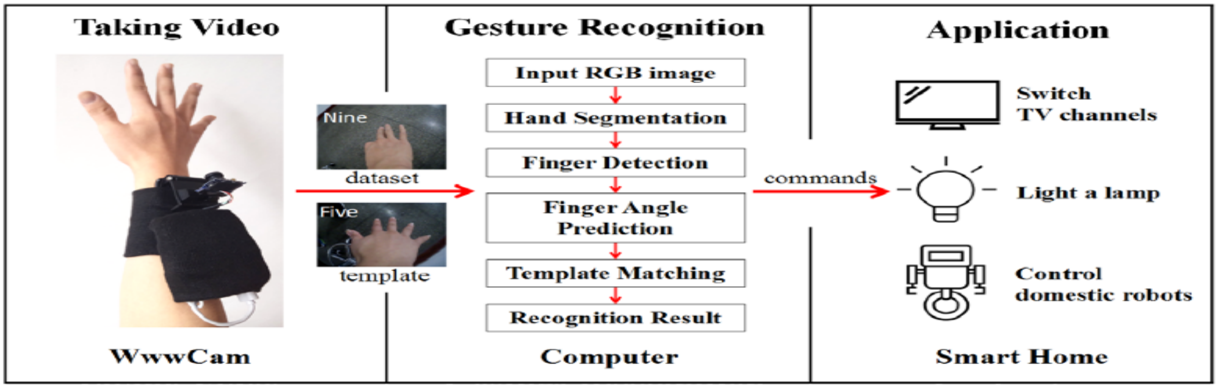


V. CONCLUSION

Hand gesture recognition has established itself as a pivotal technology transforming the way humans interact with machines. By enabling intuitive, natural, and contactless communication, it opens avenues for diverse applications spanning sign language interpretation, virtual reality, robotics, healthcare, and smart environments. This paper has reviewed the evolution of gesture recognition systems from sensor-driven setups to sophisticated, vision-based frameworks empowered by advances in deep learning. The shift towards computer vision-based approaches has significantly enhanced system flexibility, user convenience, and deployment feasibility. Techniques such as convolutional neural networks, recurrent neural networks, and transformer architectures have demonstrated remarkable improvements in extracting spatial and temporal features critical for robust gesture classification. Nevertheless, challenges including occlusions, lighting variations, user diversity, real-time constraints, and generalization to unseen environments remain active areas of research.

Publicly available datasets and standardized benchmarks have been instrumental in pushing the boundaries of accuracy and adaptability. Multi-modal approaches and continual learning paradigms are promising strategies to address existing limitations. Future research is expected to focus on developing lightweight, scalable, and adaptive systems capable of handling complex, dynamic gestures in real-world scenarios.

Overall, hand gesture recognition stands poised to redefine human-computer interaction with more natural, accessible, and immersive interfaces. Continued interdisciplinary innovation across computer vision, machine learning, sensor technology, and human factors will foster the emergence of reliable, widespread applications improving everyday life, communication, and accessibility.



REFERENCES

[1] L. Chen, F. Wang, H. Deng, K. J.-2013 I. conference, and undefined 2013, "A survey on hand gesture recognition," ieeexplore.ieee.org, Accessed: Dec. 18, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6835606/

[2] J. Suarez, R. M.-2012 I. R.-M. the 21st IEEE, and undefined 2012, "Hand gesture recognition with depth images: A review," ieeexplore.ieee.org, Accessed: Dec. 18, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6343787/

[3] H. Cheng, L. Yang, … Z. L. circuits and systems for video, and undefined 2015, "Survey on 3D hand gesture recognition," ieeexplore.ieee.org, Accessed: Dec. 18, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7208833/

[4] A. Sarkar, G. Sanyal, S. M.-I. J. of, and undefined 2013, "Hand gesture recognition systems: a survey," academia.edu, Accessed: Dec. 18, 2025. [Online]. Available: https://www.academia.edu/download/79551762/942ca83601ad2950270b212ed2ca14c8b184.pdf

[5] R. Khan, N. I.-I. journal of artificial, and undefined 2012, "Hand gesture recognition: a literature review," search.proquest.com, Accessed: Dec. 18, 2025. [Online]. Available: https://search.proquest.com/openview/17f752aaef608b544d1fcc2e80da3b53/1.pdf?pq-origsite=gscholar&cbl=646378

[6] M. Oudah, A. Al-Naji, J. C. of Imaging, and undefined 2020, "Hand gesture recognition based on computer vision: a review of techniques," mdpi.com, Accessed: Dec. 18, 2025. [Online]. Available: https://www.mdpi.com/2313-433X/6/8/73

[7] N. Mohamed, M. Mustafa, N. J.-I. access, and undefined 2021, "A review of the hand gesture recognition system: Current progress and future directions," ieeexplore.ieee.org, Accessed: Dec. 18, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9622242/

[8] G. Li et al., "Hand gesture recognition based on convolution neural network," Springer, vol. 22, pp. 2719–2729, Mar. 2019, doi: 10.1007/S10586-017-1435-X.

[9] P. Garg, N. Aggarwal, S. S.-W. academy of science, and undefined 2009, "Vision based hand gesture recognition," academia.edu, Accessed: Dec. 18, 2025. [Online]. Available: https://www.academia.edu/download/105121542/vision_based_gesture_recognition_a_review.pdf

[10] Y. Fang, K. Wang, J. Cheng, H. L.-2007 I. International, and undefined 2007, "A real-time hand gesture recognition method," ieeexplore.ieee.org, Accessed: Dec. 18, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4284820/

[11] "Hand gesture recognition - Google Scholar." Accessed: Dec. 18, 2025. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Hand+gesture+recognition&btnG=