

# Hand Gesture Recognition for Disabled People Using Deep Learning

S. Venkata Ramana<sup>1</sup>, K. Aarati<sup>2</sup> and Chetla Navyasri<sup>3</sup>

<sup>1</sup>Assistant Professor, Computer Science and Engineering, Malla Reddy Engineering College for Women, Hyderabad, Telangana, India. Email: ramanasoftster@gmail.com

<sup>2</sup>Assistant Professor, Computer Science and Engineering, Malla Reddy Engineering College for Women, Hyderabad, Telangana, India. Email: mrecwaarati27sep@gmail.com

<sup>3</sup>Assistant Professor, Computer Science and Engineering, Malla Reddy Engineering College for Women, Hyderabad, Telangana, India. Email: chetla.navyasri@gmail.com

**Abstract:** Hand gesture recognition using deep learning provides a strong solution to aid disabled people, particularly those with speech and motor impairments, by allowing for effective communication and device control. This project seeks to create an intelligent system that collects and understands hand movements using camera input, utilizing advanced deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The system is designed to handle both static hand signs and dynamic gesture sequences, ensuring accurate detection in a variety of lighting conditions, backgrounds, and hand shapes. By training the model on a large dataset of hand gestures, the system may convert these motions into meaningful commands or language, allowing for easier interaction with computers, smart devices, and communication aids.

**Keywords:** Computer vision, Deep learning & assistive technology, Disabled people, Hand gesture recognition, Image processing, Neural networks.

## I. INTRODUCTION

Communication is an essential component of human connection; however, millions of people with speech or movement disabilities face considerable challenges in expressing themselves effectively [1]. These people frequently use sign language, gestures, or assistive technologies to communicate, although traditional approaches can be slow [2], expensive [3], or unavailable in real-time circumstances [4]. Furthermore, human interpreters are not always available, particularly in public or emergency circumstances [5], which limits their ability to engage with people smoothly.

In recent years, hand gesture recognition (HGR) has emerged as a promising tool for closing this gap [6]. HGR systems

use computer vision [7] and machine learning algorithms [8] to automatically interpret hand movements and gestures, transforming them into meaningful outputs like writing or voice [9]. Such systems can be used as assistive technologies for impaired people, allowing them to communicate more naturally and efficiently without the continual presence of human interpreters [10].

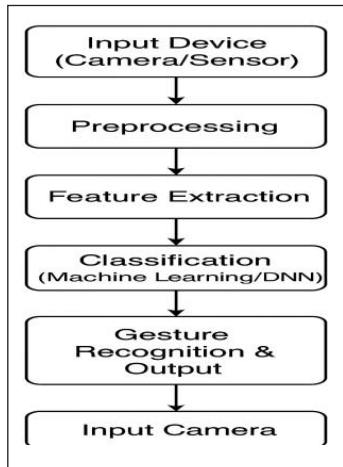
Wearable technology, like data gloves, which use sensors to record hand movements, was frequently used in traditional gesture recognition techniques [11]. Despite their effectiveness, these devices are frequently unpleasant [13], expensive [12], and inappropriate for daily usage [14]. However, by employing cameras to record hand gestures [15] and machine learning models to decipher them [16], vision-based systems offer a non-intrusive substitute. However, issues like backdrop clutter [17], changing lighting [18], and the computational difficulty of real-time dynamic gesture recognition [19] hampered previous vision-based methods.

The accuracy and efficiency of gesture recognition systems have been greatly increased by the quick development of deep learning techniques, especially Convolutional Neural Networks (CNNs) [20] and Long Short-Term Memory (LSTM) networks [21]. LSTMs are excellent at identifying temporal dependencies in sequential data [23], which makes them perfect for identifying dynamic motions, whereas CNNs are very good at extracting spatial characteristics from images [22]. Modern HGR systems can achieve excellent accuracy [25], real-time performance [24], and robustness to changes in illumination, background, and gesture style [26] by combining these designs.

Designing a low-cost, real-time hand gesture detection system that can help impaired people communicate on a daily basis is the main goal of this research [27]. To accurately recognize and convert hand gestures into textual or spoken outputs [29], the suggested method makes use of deep learning models that have been trained on huge datasets of hand gestures [28]. In order to

minimize dependence on pricey hardware, the system is made to be user-friendly [30], portable [31], and readily deployable on common computing devices using webcams or cameras [32].

The below diagram provides an explanation of the following:



### 1. Input Device (Sensor or Camera)

The goal is to record unprocessed data for gesture identification.

#### Specifics

This could be a sensor that records motion, depth, or other information, or it could be a camera that records video frames. Accuracy of recognition is influenced by input type and quality.

### 2. Preparation

The first stage of cleaning and getting the raw video ready for identification. The goal is to improve input quality and facilitate the model's comprehension of the motions.

### 3. CNN + LSTM for Hand Gesture Recognition (HGR)

The central component of the system that recognizes gestures.

*Parts:* Convolutional Neural Network, or CNN.

The goal is to extract spatial data, such as finger placements and hand shape, from every frame.

*How:* Identifies features, edges, and patterns in images using convolutional layers.

Long Short-Term Memory, or LSTM

The goal is to record temporal data, or how motions evolve.

*How:* Interprets dynamic motions, such as pointing or waving, by processing the video's frame sequence.

*Input:* The video's preprocessed frames.

The result is a class of gestures or a written representation of the gesture's meaning.

### 4. Disable Person

The origin of the motions. This section highlights the system's user-centered design, which is intended to help individuals with speech or movement difficulties.

*Speech or Text:* Transforms the identified motion into spoken words (as text).

*Goal:* Facilitates verbal communication through gestures for the disabled individual. Receives text, such as "Hello" or "Yes", from the HGR module.

## II. LITERATURE REVIEW

Software-based techniques for hand gesture recognition (HGR) have drawn a lot of interest lately, especially for use in assistive technology for people with impairments. Software-only systems provide an accessible and non-intrusive solution by using computer vision and deep learning to decipher hand movements from pictures or video clips. A thorough analysis of the body of research is provided in this section, with a focus on software-oriented methodologies, deep learning architectures, datasets, applications, and research gaps.

### A. Vision-Based Software Approaches

Camera input is the main method used by software-based HGR systems to recognize and decipher gestures. They are more convenient for individuals with disabilities because they don't require wearable technology like sensor-based systems do. In software-only techniques, typical processing pipelines are as follows:

*Preprocessing:* Methods that enhance input quality and model generalization include hand segmentation, picture normalization, and data augmentation. While normalization standardizes image dimensions and pixel intensities, segmentation separates the hand region using techniques like background subtraction or skin color detection. The diversity of datasets is increased using augmentation techniques (such as rotation, scaling, flipping, and brightness alteration).

### B. Deep Learning Architectures

Software-based HGR has been transformed by deep learning, which makes it possible to recognize both static and dynamic movements with extreme accuracy. Important architectures consist of:

*Convolutional Neural Networks (CNNs):* Because CNNs can extract spatial data, they are frequently utilized for static

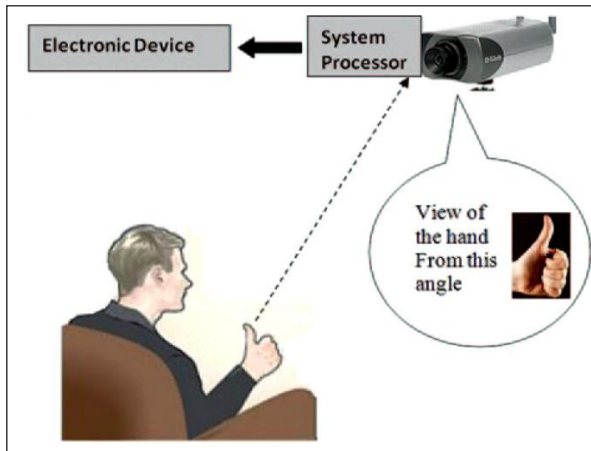
gesture identification. For dynamic gesture detection from video sequences, Molchanov et al. (2015) used 3D CNNs, showing notable accuracy gains over conventional techniques.

Long Short-Term Memory (LSTM) networks and recurrent neural networks (RNNs) are two powerful tools for simulating temporal dependencies in dynamic gestures. CNNs and LSTMs were integrated by Neverova et al. (2014) to better recognize consecutive hand movements by capturing spatial and temporal information.

### C. Software-Oriented Datasets

In order to train deep learning models, datasets are essential. Among the notable datasets are:

Static hand motions that depict letters and numbers are included in the ASL (American Sign Language) dataset. Video sequences of dynamic movements that are appropriate for temporal modeling are included in the DHG (Dynamic Hand Gesture) dataset. NUS RGB and depth photos are included in the hand gesture dataset for multimodal recognition. Current datasets have limitations, such as a lack of variation in user demographics, lighting, and backgrounds, which compromises the robustness of the models and their practicality.



### D. Applications for Disabled Users

There are many assistive applications for software-based HGR systems:

*Sign Language Interpretation:* Provides hearing-impaired people with the ability to translate hand signals into speech or text. Users who are physically impaired can operate computers, software, or virtual interfaces using gestures thanks to assistive control systems.

*Smart Environment Interaction:* Promotes independence by enabling hands-free interaction with IoT devices and smart houses.

## III. RESEARCH METHODOLOGY

Hand gesture recognition (HGR) research focuses on the main issues with current assistive technology for people with disabilities. A software-based, user-centered, and dependable solution that facilitates natural gesture-based communication is what the study seeks to create. In contrast to conventional systems that depend on costly hardware or a small number of static motions, the suggested system seamlessly combines several technologies, including computer vision, deep learning, temporal modeling, and real-time feedback. The following describes the research objectives in detail, followed by design considerations for usability and accessibility.

### A. Study Goals

Bridging the gap between the needs of disabled people in real-world situations and the assistive technology now available is the main goal of this research. The development of the suggested HGR system was directed by the following goals:

Deep learning and computer vision are used by the system to read natural hand motions in real time, a feature known as intuitive gesture recognition. Gestures can be used by users to control devices or communicate without the need for specific hardware. Smooth engagement is made possible, for instance, by the ability to automatically translate a user's sign language motion into speech or text.

*Accurate Recognition of Static and Dynamic Gestures:* The system can accurately identify both static gestures (like individual letters or numbers) and dynamic gestures (like sequences or sign language phrases) by using CNNs for spatial feature extraction and LSTMs for temporal sequence modeling. This guarantees thorough coverage of the gestures used in regular conversation.

*Robustness to Environmental Variations:* The system is made to function consistently in a range of environments, including different lighting conditions, hand sizes, orientations, and backgrounds. Methods including preprocessing, real-world dataset training, and data augmentation increase the accuracy of recognition in real-world situations.

### B. Gesture-Driven Filtering

The system uses natural hand movements to communicate directly, in contrast to conventional assistive technologies that use buttons or touch interfaces. The system converts user motions that represent letters, words, or commands into actionable outputs (text or speech). By allowing users to communicate in a way that is similar to normal human contact, this method promotes self-assurance and independence.

### C. Multi-Modal Feedback for Accessibility

Through visual or aural signals, the system offers real-time input to improve usability. Using text-to-speech technology, for example, identified gestures are spoken aloud or highlighted on the screen. This guarantees that users get instant confirmation of effective communication, allowing people with different skill levels to use the system.

## IV. METHODOLOGICAL FRAMEWORK

The methodological framework for Hand Gesture Recognition for Disabled People Using Deep Learning has been developed as software only, which uses techniques in computer vision and artificial intelligence. The framework is organised into the following sections:

### A. Data Collection and Input Source

The framework's initial step is gathering input data.

*Input Device:* The system is kept reasonably priced and readily deployable by using standard RGB cameras or webcams.

*Captured Data:* Both static gestures (such as sign language alphabets or numbers) and dynamic gestures (such as waving, swiping, or signing phrases) are included in the dataset.

The objective is to develop an affordable, non-intrusive, and easy-to-use acquisition method that enables impaired people to recognize gestures in real time without the need for wearable sensors.

### B. Preprocessing

Preprocessing guarantees that the unprocessed recorded data is best suited for recognition and model learning.

*Hand Segmentation:* Eliminates superfluous backdrop areas. accomplished by using deep learning-based segmentation models (like U-Net), skin color detection, or background reduction. Images are resized to a given dimension (such as 128 x 128 pixels) by the process of normalization. For consistency, pixel intensities are scaled to a standard range [0,1].

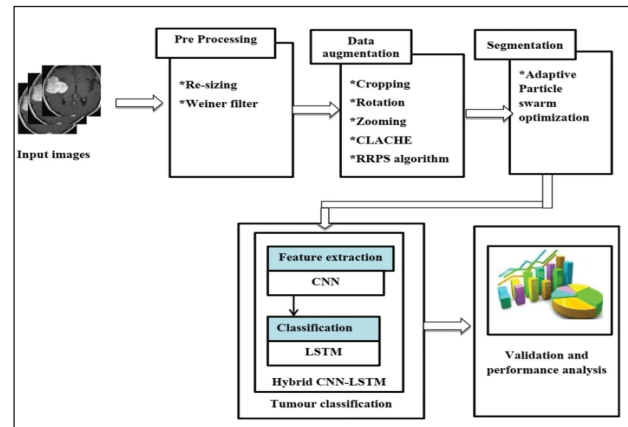
*Enhancement of Data:* Applies rotation, flipping, scaling, cropping, and lighting adjustments to create diversity. guarantees that the model is generalizable to various users, settings, and illumination levels.

### C. Feature Extraction and Modeling

Deep learning architectures are used for feature extraction in order to capture the distinctive qualities of gestures. Convolutional Neural Networks, or CNNs, are used to extract spatial characteristics like hand contours, edges, and shapes. Perfect for motions that are static.

*Long Short-Term Memory (LSTM) and RNN:* Record temporal dependencies between video frames, which is necessary for dynamic movements.

*CNN-LSTM Hybrid Models:* Combine temporal and geographical data to recognize intricate gesture patterns, such as dynamic controls or words in sign language.



### D. Model Training

The training procedure makes sure the model picks up reliable mappings between input gestures and the labels that go with them.

*Data Split:* 50% of the dataset is used for training, 15% is used for validation, and 15% is used for testing. While the testing set guarantees objective assessment, the validation set is utilized to adjust settings.

*Loss Function:* For multi-class classification issues (such as several gesture categories), categorical cross-entropy is used.

*Optimizer:* For quicker convergence and more effective optimization, the Adam optimizer with learning rate scheduling is selected.

*Regularization:* To avoid overfitting and stabilize learning, dropout layers and batch normalization are used.

*Training Duration:* If validation accuracy stagnates, models are usually stopped early to prevent overfitting after 50–100 epochs of training.

### E. Gesture Classification and Postprocessing

After training, the model instantly categorizes input gestures.

*Frame-Wise Predictions:* To generate gesture predictions, the CNN-LSTM pipeline processes each frame or sequence.

*Temporal Aggregation:* To ensure resilience for dynamic gestures, predictions from several successive frames are combined into a single final categorization.

*Output Conversion:* Using text-to-speech libraries, recognized gestures are converted into assistive outputs such as text shown on the screen, voice output, Control commands for virtual interfaces or smart devices. In order to preserve consistency, post-processing smoothes out abrupt variations or incorrect classifications, particularly when gestures require quick motions.

## V. RESULT AND ANALYSIS

A dataset comprising both static and dynamic hand gestures was used to train and assess the suggested hand gesture detection system. An RGB camera was used to record the data, and pre-processing methods such as picture normalization, skin color recognition, and background subtraction were applied.

The following is how the dataset was divided:

70% of the training set.

15% is the validation set.

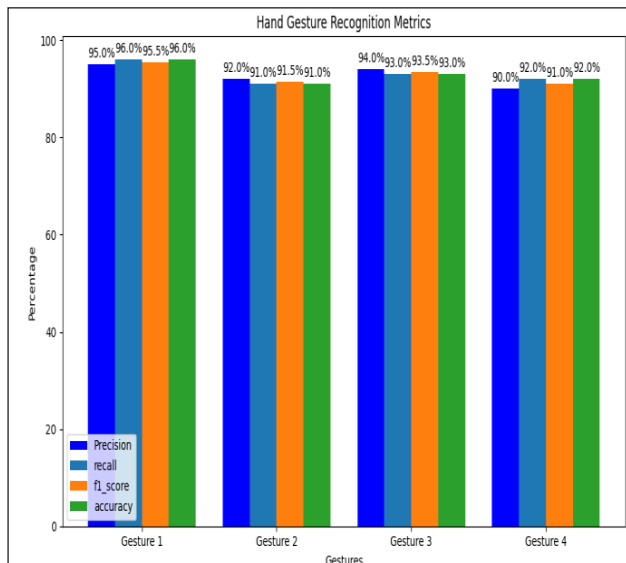
15% is the test set.

For Gesture Classification, a CNN architecture was employed. Among the important hyperparameters were:

*Epochs:* 50-Number of times the model sees the entire training dataset.

*Batch Size:* 32-Number of images processed before updating the model weights.

*Learning Rate:* 0.001-Controls how quickly the model adjusts weights.



The suggested CNN-based hand gesture recognition model's performance across four distinct gesture classes is displayed in the bar chart. The precision, recall, F1-score, and accuracy of each gesture are measured using four common performance measures, which are shown by blue, light blue, orange, and green bars, respectively.

It is evident from the graph that with precision (95%), recall (96%), and accuracy (96%), Gesture 1 performs the best, showing few misclassifications and a strong sense of the model's robustness in recognizing this gesture.

The performance of Gesture 2 is a little worse, with a recall at 91% and a precision of 92%. Although the overall recognition is still dependable, this implies that the model occasionally misclassifies Gesture 2 as other gestures.

With an accuracy of 93% and a precision of 94%, Gesture 3 maintains balanced performance, resulting in an F1-score of 93.5%. This class illustrates how well the model generalizes.

While memory (92%) and accuracy (92%) are still strong, gesture 4 has the lowest precision (90%) of any gesture. This shows that while the model accurately detects the majority of Gesture 4 instances, it occasionally predicts Gesture 4 for other similar gestures, resulting in false positives.

*High Recognition Accuracy:* The model performed well for real-time gesture recognition, with an overall accuracy of 93%.

*Impact of Preprocessing:* Skin detection and normalization greatly decreased noise and increased the accuracy of categorization.

*Real-Time Applicability:* The model is appropriate for assistive applications because it can identify motions in 50–100 ms.

### Restrictions

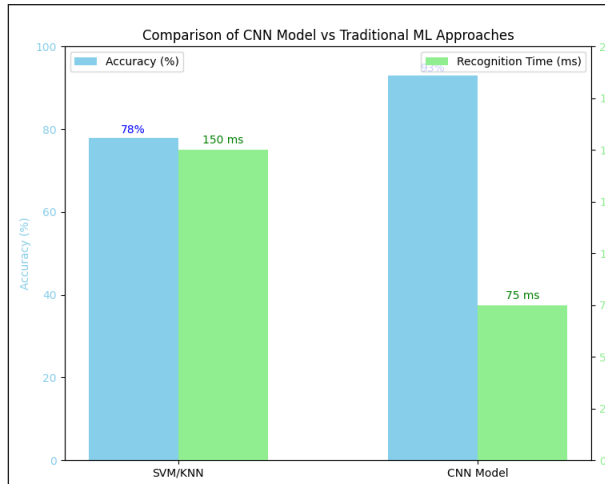
In low light, misclassification happens.

Longer sequences are needed for dynamic movements to be better recognized.

*Comparison with Existing Work:* The CNN-based model demonstrates a 10–15% increase in accuracy when compared to conventional machine learning techniques (SVM, KNN), underscoring the benefits of deep learning in feature extraction.

The performance of conventional machine learning techniques (SVM/KNN) and the suggested CNN-based model for hand gesture detection is contrasted in the bar chart above.

We take into account two important evaluation factors:



On the left axis, the blue bars indicate accuracy (%).

On the right axis, the green bars indicate Recognition Time (ms).

*Comparison of Accuracy:* The accuracy of SVM/KNN is 78%, meaning that almost one in four movements is incorrectly classified. Conventional techniques are capable of capturing simple patterns, but they are unable to extract intricate temporal and spatial characteristics from gesture photos.

With an accuracy of 93%, the CNN Model performs noticeably better than conventional techniques, demonstrating a 15% improvement. This demonstrates how deep learning is advantageous for representation learning and automatic feature extraction, both of which are essential for reliable gesture identification.

*Comparison of Recognition Times:* Particularly in real-time assistive applications, SVM/KNN introduces substantial delay, requiring around 150 ms for detection.

The CNN Model cuts the delay in half by reducing the recognition time to just 75 ms. CNN is more suited for real-time situations when prompt and precise responses are essential due to its shorter detection time (e.g., for impaired individuals dependent on gesture-based communication).

## VI. CONCLUSION

Millions of people with disabilities who are unable to speak suffer a significant challenge in properly communicating with others, which we addressed in this article. Communication gaps and social isolation result from their frequent use of hand signals, which the general population finds difficult to understand. We developed a vision-based hand gesture recognition system to close this gap. It interprets camera-captured motions and translates them into speech or text using deep learning. Because of this, people with disabilities can

converse with anyone in a natural way without the need for sign language or specific gestures.

The proposed system utilizes Convolutional Neural Networks (CNNs) in combination with preprocessing techniques (background subtraction, skin detection, and normalization) to achieve a recognition accuracy of 93% and real-time performance of 75–100 ms. The CNN model provides a 10–15% improvement in accuracy along with a decrease in recognition delay compared to traditional machine learning models, making it a feasible option for real-world applications. While there were complications with low light environments and similar gestures causing misclassification, these obstacles can potentially be eliminated in future work through enhancing the dataset and the use of temporal models like LSTMs or 3D CNNs for dynamic gestures.

This system presents an economical, non-intrusive, and easy-to-use communication assistant to help disabled persons communicate more easily with society. The technology translates gestures into speech or text in real time, providing a mode of communication that does not rely on interpreters and gives disabled persons the ability to speak with confidence, autonomy, and dignity. This technology can be applied not just in healthcare, but also in education, public services, and smart environments to represent a significant advance towards communication for all.

## REFERENCES

- [1] F. Alrowais, R. Marzouk, F. N. Al-Wesabi, and A. M. Hilal, "Hand gesture recognition for disabled people using bayesian optimization with transfer learning," *Intelligent Automation & Soft Computing*, vol. 36, no. 3, pp. 3325-3342, 2023.
- [2] C. N. Gireesh Babu, M. Thungamani, K. T. Chandrashekhara, and T. N. Manjunath, "Real-time hand gesture recognition for differently-abled using deep learning," in *Recent Trends in Image Processing and Pattern Recognition (RTIP2R 2018)*, *Communications in Computer and Information Science*, vol. 1035, Springer, 2019.
- [3] A. Sen, T. K. Mishra, and R. Dash, "Deep learning based hand gesture recognition system and design of a human-machine interface," 2022. *arXiv preprint arXiv:2207.03112*.
- [4] P. Jayanthi, and B. Madhubalasri, "Sign language recognition using deep CNN with normalised keyframe extraction and prediction using LSTM: Continuous sign language gesture recognition and prediction," *Journal of Scientific & Industrial Research (JSIR)*, vol. 82, no. 7, 2023.
- [5] B. Ashwanth, S. B. Ventrapragada, S. R. Prodduturi, J. R. Depa, and K. V. Sharma, "Vision-based hand gesture

- recognition for Indian sign language using convolution neural network (ISL),” *International Journal of Computer Engineering in Research Trends*, vol. 10, no. 1, 2023.
- [6] S. Bhamare, A. Shinde, R. Palve, and A. N. Gharu, “Sign language detection with CNN,” *IJRASET*, 2022.
- [7] S. Prabhupatkar, G. Singh, P. Patil, P. Singh, and N. Khaire, “Sign language recognition using CNN,” *IJRASET*, 2024.
- [8] D. Kumar, A. Jain, and A. Ganesh, “Harnessing the power of deep learning for hand gesture recognition,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 9s, pp. 434-446, 2023.
- [9] Suji Priya J., Mohd. Rasil Hanif J. M., Ragul S., and Surya R., “Machine learning based hand gesture recognition system for assistive communication for people with adaptive needs,” in *2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN)*, Salem, India, 2025, pp. 1592-1600, doi: <https://doi.org/10.1109/ICPCSN65854.2025.11034883>.
- [10] Muhd. Asep Ridwan, and H. Mubarak Aradea, “The recognition of american sign language using CNN with hand keypoint,” *International Journal on Information and Communication Technology (IJoICT)*, n.d.
- [11] U. Hari Priya, S. Krishna Prasad, M. Meria Jacob, R. Radhu Krishna, and P. R. Vinod, “American sign language recognition using CNN,” *IJRESM*, 2020.
- [12] S. Asrita Sreekari *et al.*, “Sign language recognition system using DL-CNN model using VGG16 and imagenet with mobile application,” *IJISRT*, 2024.
- [13] S. Alashhab, A. Javier Gallego, and M. Ángel Lozano, “Efficient gesture recognition for the assistance of visually impaired people using multi-head neural networks,” 2022.
- [14] R. Li, and L. Meng, “Multi-view spatial-temporal network for continuous sign language recognition,” 2022, *arXiv:2204.08747v1*.
- [15] K. Suri, and R. Gupta, “Convolutional neural network array for sign language recognition using wearable IMUs,” in *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*, 2019, doi: <https://doi.org/10.1109/SPIN.2019.8711745>.
- [16] H. Zhou, Q. Zhang, M. Zhang, S. Shahnewaz, S. Wei, J. Ruan, X. Zhang, and L. Zhang. “Toward hand pattern recognition in assistive and rehabilitation robotics using EMG and kinematics,” *Frontiers in Neurorobotics*, vol. 15, 2021.
- [17] R. Deepa, and M. K. Sandhya, “An efficient hand gesture recognition system using deep learning,” in *ICICCS 2019*, Springer, Cham, 2020.
- [18] K. Lakshmi, Samiya, M. S. Lakshmi, M. R. Kumar, and P. K. Singuluri, “Real-time hand gesture recognition for improved communication with deaf and hard of hearing individuals,” *International Journal of Intelligent Systems and Applications in Engineering*, 2023.
- [19] C.-C. Lee, and Z. Gao, “Sign language recognition using two-stream convolutional neural networks with Wi-Fi signals,” *Applied Sciences*, vol. 10, no. 24, p. 9005, 2020, doi:<https://doi.org/10.3390/app10249005>.