

editor@ijprems.com

SIGN LANGUAGE RECOGNITION USING DEEP LEARNING-BASED COMPUTER VISION MODELS

Ms. Babitha L¹, Anseera J², Maha R³

¹HOD of Information Technology, United Institute of Technology, Tamilnadu 641020, India
^{2,3}Department of Computer Science and Engineering, United Institute of Technology, Tamilnadu 641020, India.
babitha@uit.ac.in, anseera0303@gmail.com, rr449902345@gmail.com

ABSTRACT

Sign language is essential for people who can't hear or speak, but the fact that not many people know it creates communication problems. This paper describes a real-time system that recognizes sign language, using advanced deep learning and computer vision to make things easier. The system uses YOLOv5 to find hand gestures and a Convolutional Neural Network (CNN) to identify the signs, ensuring it's both accurate and quick. It was tested using several datasets, including Mexican Sign Language (MSL), Pakistani Sign Language (PSL), ASLLVD, and WLASL, and achieved 99.75% accuracy on MSL and 94.01% on PSL. Unlike older methods that use special sensors, this system uses regular video, making it better for everyday use. The results show the system works well, can be expanded, and adapts to different sign languages. Future plans include improving continuous gesture recognition, making the model work better on phones, and adding more languages to create a more inclusive communication system.

1. INTRODUCTION

Communication is vital for human interaction, but people with hearing and speech difficulties encounter obstacles because most people don't understand sign language. Sign language, which uses hand gestures, facial expressions, and body movements, is the primary way the deaf and mute community communicates. However, not many people know sign language, which creates difficulties in schools, jobs, and everyday interactions.

Typical solutions, like human interpreters and text-based communication, can be costly, aren't always available right away, and are not easy to use daily. With the progress in artificial intelligence (AI) and deep learning, systems that automatically recognize sign language (SLR) are now an effective solution. These systems use computer vision and machine learning to identify and interpret sign gestures, allowing smooth communication between sign language users and those who don't know sign language.

This paper introduces a real-time sign language recognition system that uses YOLOv5 for gesture detection and CNN for gesture classification. The system captures hand gestures and facial expressions, detects objects in real-time, and translates the signs into text. Unlike traditional methods, this system is fast, accurate, and doesn't require any extra wearable devices.

The model was tested using standard datasets, including Mexican Sign Language (MSL), Pakistani Sign Language (PSL), ASLLVD, and WLASL. It achieved 99.75% accuracy on MSL, 94.01% on PSL, 34.41% on ASLLVD, and 90.31% on WLASL-100. While the system works well with individual signs, recognizing continuous gestures and handling high computational needs remain challenges. Future improvements will focus on making the system work better on mobile devices and improving the recognition of fluid sign transitions, making sign language more accessible in different settings.

2. LITERATURE REVIEW

Different methods have been proposed for SLR based on machine learning, deep learning and sensor-based techniques. In this section, we discuss the existing approaches, exploring their methods and limitations.

2.1 Traditional Machine Learning Approaches

In the initial studies, sign recognition models were base on feature extraction based models such as Hidden Markov Models (HMM), Support Vector Machines (SVM), K-Nearest Neighbors (KNN). These models were built upon various hand-tuned features like Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Optical Flow etc. These methods were effective for small-scale datasets but suffered in high-dimensional situations.

2.2 Sign Language Recognition Using Sensors

The other type of research investigated wearable sensors (e.g., accelerometers, gyroscopes, electromyography (EMG) sensors) for data capture of hand movement. These systems achieve high accuracy rates but need high-spec hardware, which can be expensive and impractical for everyday use.

A4 NA	INTERNATIONAL JOURNAL OF PROGRESSIVE	e-ISSN :
IIPREMS	RESEARCH IN ENGINEERING MANAGEMENT	2583-1062
	AND SCIENCE (IJPREMS)	Impact
www.ijprems.com	(Int Peer Reviewed Journal)	Factor :
editor@ijprems.com	Vol. 05, Issue 04, April 2025, pp : 2149-2155	7.001

For example, Kumar et al. (2022) proposed an IoT based sensor system specifically for sign language translation, which provided real-time recognition but was limited in scalability as it relied heavily on the use of wearable devices. Although these mechanisms are good at performing in clean environments, they do not generalise well across different sign language dialects.

2.3 Approaches Based on Deep Learning

This is a notable for improvement because recent progress at deep learning has allowed us to remove handcrafted features and directly work on the sign language recognition task. Models such as Convolutional neural networks (CNNs), Recurrent Neural Networks (RNNs), & Transformer-based architectures have achieved impressive performance in recognizing complex gestures from RGB video frames.

Yan et al. For dynamic hand movements, G. Yan et al. (2018) defined a Spatial-Temporal Graph Convolutional Network (ST-GCN), with a computationally expensive but very accurate recognition capabilities. Similarly, Mujahid et al. (2021) created a real-time SLR framework based on YOLOv3 that showcases superior speed, yet only achieved blurred recognition of signs.

2.4 Challenges and Research Gaps

Although deep learning- based methods have improved in sign language recognition, some issues still exist:

Seql-Fusion: Most models in literature perform well on classification of isolated sign gestures, but hardly on gesture recognition in continuous sign language which is more natural.

Heavy Computational Power: Most deep learning models demand heavy GPUs.

Sign Language Support in Multiple Languages: A majority of the studies consider a single sign language, restricting the adaptability of the models across sign languages with heterogeneous structures.

Challenges and Limitations Lack of large-scale annotated sign language datasets.

2.5 Summary of Findings

The review also notes that use of deep learning-based approaches, especially CNN and YOLO models outperform the existing machine learning and sensor-based methods. On the other hand, currently, we still need models that offer real-time continuous sign language recognition with the least computation power used, and that is able to learn multiple sign languages. To overcome the aforementioned shortcomings, our method combines YOLOv5 for hand gesture detection and a CNN-based classifier for sign recognition, resulting in a scalable, unobtrusive, and high-accuracy solution suitable for real-world applications.

3. SYSTEM ARCHITECTURE

The proposed real-time sign language recognition system architecture involves a number of sequential stages, from video input to gesture detection. The pipeline is efficient, accurate, and is appropriate for real-time use. The below figure shows the architecture.

The system starts from a video stream or webcam capture, which continuously records hand motions and gestures made by the user. The second step is frame/hand extraction in which single frames are chosen and hands are detected from the video stream. This process is essential to extract the region of interest and filter out noise from the background.

Subsequent to this, the extracted hand images are normalized and resized, which normalizes the input data for the deep model. This involves processes like converting the images to grayscale, resizing them to a uniform size, and scaling pixel values to improve model performance and minimize computational complexity.

After preprocessing is done, the pre-processed data is fed into a deep learning model that is based on Convolutional Neural Networks (CNN). The CNN model is trained to recognize various signs from features of the hand images. It learns spatial hierarchies in the data and hence is best suited for image-based gesture recognition tasks.

The model outputs a prediction score, representing the confidence level of the identified gesture. In our system, the model has a very high prediction rate of 99.9%, indicating its effectiveness in detecting sign language gestures.

Lastly, the system, based on the prediction outcome, outputs the recognized sign, which can then be utilized for communication or translation. The design is modular and can be readily extended to accommodate various sign languages and real-time deployment.





3.1 Video/Webcam Input

The system begins with live input from a video source or webcam. This live video stream records the user's hand movements, which serve as the foundation for subsequent processing. A high-resolution and stable input guarantees that enough information is present for reliable gesture recognition. This module also manages frame capture at a given rate to trade off between accuracy and performance.

Besides that, the video input system is also flexible enough for deployment in various real-world

environments in terms of lighting conditions and camera views. Latency is minimized as only key frames that contain motion or hand presence are chosen, offloading load on downstream processing stages.

3.2 Frame/Hand Extraction

Here, single frames are taken out from the video stream. A hand detection algorithm isolates and separates the region of interest (ROI) specifically based on the hand of the user. This avoids background noise and enhances gesture recognition accuracy. Robust hand tracking makes sure to forward only valid frames with hand gestures to the subsequent stage.

The system employs object detection methods like YOLOv5 or other such models to detect hands with high accuracy. With the use of bounding boxes on the hand areas, it is more convenient to crop and pass on only the necessary portions of the image to the subsequent module, without using extra computational power.

3.3 Image Normalization and Resizing

The hand images obtained from extraction are fed into preprocessing operations like normalization and resizing. Normalization converts pixel intensity values into a standardized range, promoting improved model convergence while training. Resizing maintains a common dimension for all input images, an essential requirement before inputting them into the deep learning model. The standardization reduces computational complexity and enhances the performance of the model overall.

Furthermore, preprocessing consists of image augmentation techniques such as flipping, rotation, and scaling to increase diversity in the dataset and model robustness. It ensures that a well-trained model will perform reasonably well even in cases of variability in gesture orientation and lighting conditions.

3.4 Deep Learning / CNN Model

The system's heart is a Convolutional Neural Network (CNN) that has been trained to identify various hand gestures. The CNN learns spatial features from the input images and classifies them into certain sign language labels. The architecture can consist of various convolutional, pooling, and fully connected layers to achieve better accuracy. This model is trained on labeled datasets to learn and generalize from gesture variations.

The CNN model is optimized using methodologies such as batch normalization and dropout regularization to prevent overfitting and speed training. Also, the incorporation of transfer learning via pre-trained networks (e.g., ResNet or MobileNet) can be added to provide further enhancement for accuracy and the speed of training on small sets.

3.5 Prediction Level (99.9%)

Following training, the CNN model makes predictions from new input frames and gives a prediction level representing the confidence level for the detected sign. With our system, the model delivers an accuracy level of 99.9%, rendering it strongly dependable for applications in the real world. The prediction level functions as a cut-off point in eliminating uncertain findings and guaranteeing stable detection. This confidence value assists the system in determining whether to reject or accept a prediction. The system can request a re-gesture if the confidence in the prediction is less than a predetermined threshold, or it can provide a "no match" outcome, thus preventing false positives and building user trust.

. A4 NA	INTERNATIONAL JOURNAL OF PROGRESSIVE	e-ISSN :
UPREMS	RESEARCH IN ENGINEERING MANAGEMENT	2583-1062
an ma	AND SCIENCE (IJPREMS)	Impact
www.ijprems.com	(Int Peer Reviewed Journal)	Factor :
editor@ijprems.com	Vol. 05, Issue 04, April 2025, pp : 2149-2155	7.001

3.6 Detection (Result)

The last phase includes the detection and presentation of the identified gesture. Depending on the output of the model, the system determines the related sign and displays it in textual or symbolic representation. This output can be utilized for real-time translation or communication assistance, particularly for deaf or speech-impaired individuals.

The identified gesture can also be saved or passed on to other devices or apps. This brings with it the potential to be integrated with communication aids, for example, voice converters or chatbots, and hence building a complete sign-to-speech translation pipeline.

4. METHODOLOGY

The system of sign language recognition is based on a well-defined cascade of operations from live video input to final gesture prediction. Every subsystem is specifically designed to provide improved accuracy, performance, and flexibility of the system under various conditions and user variations.

4.1 Data Acquisition

The initial step in the system is data capture using a webcam or an external video source, which facilitates real-time user interaction with users conducting sign language gestures. The video feed is streamed in real time, and frames are grabbed at a constant frame rate for uniformity. The configuration makes the system feasible for real-life application in assistive communication for the hearing and speech disabled. The recorded data is instantly processed and sent to the model for learning, providing a low-latency pipeline.

In addition to live video feed, training on pre-recorded datasets is also supported by the system. Pre-recorded datasets such as the American Sign Language (ASL) alphabet dataset and other benchmark gesture datasets are used to train the deep learning model. These datasets introduce variations in lighting, skin color, backgrounds, and gesture styles such that the system can generalize nicely across various users and conditions. Utilizing real-time and large-scale datasets facilitates the system being robust, scalable, and deployment-ready in real-world environments.

4.2 Frame Extraction and Hand Segmentation

After the video input is recorded, it is divided into frames, from which the system derives the region of interest (ROI) — mostly the hand. Object detection models like Haar cascades or Mediapipe are employed to separate the hand from the surroundings. These approaches assist in eliminating redundant background data and enable the model to only consider the gesture being performed. Correct hand segmentation is very important in reducing noise and improving gesture recognition.

In order to further optimize detection, motion-based filtering is employed in order to bypass frames where there is no movement of the hand. Furthermore, tracking methods such as bounding boxes and contour analysis help to only transmit meaningful regions to the model. The system is also able to accommodate different positions, distances, and lighting conditions by dynamically varying thresholds while segmenting. This renders the hand extraction more reliable and robust, even under non-ideal environmental conditions.

4.3 Image Preprocessing

Once the hand region has been extracted, the images undergo a preprocessing pipeline to normalize and enhance them. The hand images are then resized to a constant size (e.g., 224x224 pixels) to be compatible with CNN input specifications. Pixel normalization is applied afterward, normalizing values between 0 and 1 to decrease computational complexity and speed up model convergence. This process also aids in increasing overall model accuracy while training.

In addition, data augmentation methods are used to artificially increase the dataset and render the model stronger. These consist of random rotations, flipping, brightness changes, and zooming. Filters like Gaussian blur to reduce noise can also be used to smoothen the image and get rid of minor inconsistencies. These preprocessing operations make the dataset richer, thus allowing the model to learn more effectively and perform well under different real-world situations.

4.4 CNN Model Design and Training

The recognition system's backbone is a Convolutional Neural Network (CNN) used for gesture classification. The CNN has multiple layers: convolutional layers to extract spatial features, pooling layers to reduce dimensions, and fully connected layers to map extracted features to gesture classes. A softmax layer at the output produces probabilities for each gesture class. The model is trained on a large collection of labeled gesture images to effectively recognize a broad variety of signs.

Training is done by fine-tuning the model with categorical cross-entropy loss and an Adam optimizer. To enhance model performance, methods such as early stopping, batch normalization, and dropout regularization are used. These prevent overfitting of the model and enable it to generalize well on new data. Additionally, training is performed in a GPU

A4 NA	INTERNATIONAL JOURNAL OF PROGRESSIVE	e-ISSN :
LIPREMS	RESEARCH IN ENGINEERING MANAGEMENT	2583-1062
an ma	AND SCIENCE (IJPREMS)	Impact
www.ijprems.com	(Int Peer Reviewed Journal)	Factor :
editor@ijprems.com	Vol. 05, Issue 04, April 2025, pp : 2149-2155	7.001

environment to drastically minimize training time and process high-dimensional data effectively. This process yields a very accurate model that can perform real-time gesture recognition.

4.5 Gesture Prediction and Classification

After training the model, it is utilized for real-time gesture classification. Every incoming frame of the video feed is processed and run through the same preprocessing pipeline prior to being input into the CNN. The model returns a probability distribution over gesture classes, and the most confident one is picked as the final output. This way, only the most probable gesture is outputted.

To enhance prediction stability, temporal smoothing is integrated into the system—only those gestures that are predicted reliably over several frames are accepted. This eliminates flicker outputs due to abrupt hand motion or transition gestures. If the confidence of the model falls below a predetermined level (e.g., 90%), the prediction is discarded or flagged as uncertain. This clever filtering improves the system's reliability, especially in live settings.

4.6 Result Output

Once classified, the identified gesture is presented in textual form on the user interface. Optionally, it can be synthesized into speech via a text-to-speech (TTS) module, which assists in bridging the communication gap between hearing and non-hearing people. This output module is crucial to making the system not only technically robust but also user-friendly and interactive.

The system can also record the identified gestures for later analysis, training refinement, or user feedback. Additionally, the output result can be coupled with messaging or voice assistant systems for real-time communication. With these output capabilities, the system evolves from a prototype to a complete application, providing utilitarian value in both personal and public environments.

5. FUTURE IMPROVEMENT

Future, this system can be upgraded by adding its ability to perceive dynamic gestures as well as static signs. This would include integrating time-series deep learning models such as Recurrent Neural Networks (RNN) or Long Short-Term Memory (LSTM) networks, which are effective for comprehending sequential patterns in gestures made. Such models will allow the system to decipher full sign language phrases or sentences and make it ready for use in real-time conversation.

Another significant improvement would be to implement multilingual output translation, making the recognized gestures translate into several regional and global languages. Such an implementation will enable the system to become more globally compatible and assist users belonging to different language groups. Moreover, incorporating a speech-to-sign module will enable it to be a bi-directional communication platform, where the spoken words get translated into sign animations, supporting inclusive communication for signers as well as non-signers.

Subsequent releases can also be enhanced by deployment on edge devices, such as Raspberry Pi or smartphones, for offline and mobile use. Moreover, the system may utilize 3D depth sensors or LiDAR-based input for more accurate hand tracking and gesture segmentation, particularly in low-light or cluttered environments. Last, but not least, the addition of facial expression recognition can introduce emotional

context into the gestures, adding clarity and intention to communication, and taking the system to human-like communication.

6. RESULTS AND DISCUSSION

The suggested sign language recognition system was thoroughly tested using both a large public dataset and live realtime inputs. The system performed well in classifying different static hand gestures that represent alphabets and popular signs. The Convolutional Neural Network (CNN) that was trained on preprocessed hand images achieved an average classification accuracy of more than 95% on the test dataset, reflecting good learning ability. The precision was consistent across trials and the system processed and displayed predictions in milliseconds, providing a seamless user experience. One of the most important advantages of the system was that it could work under varying environmental conditions. The incorporation of Mediapipe hand landmark detection enabled the model to accurately extract hand regions even in complicated backgrounds and varying lighting conditions. This helped to further decrease noise and improve gesture clarity. The system further exhibited consistent performance in users with varying hand sizes, shapes, and skin color, highlighting its robustness and generalizability. The model needed extremely few frames to accurately predict the gesture during testing, which reduced the response latency and made the system suitable for real-time communication. The system's real-time interface allowed the users to operate the model with their hand merely inside the camera region. Predictions were made instantly and visualized on the screen, which gave instant visual feedback. Additionally, incorporation of a Text-to-Speech (TTS) module transcribed identified gestures into speech, increasing

	INTERNATIONAL JOURNAL OF PROGRESSIVE	e-ISSN :
HIPREMS	RESEARCH IN ENGINEERING MANAGEMENT	2583-1062
an ma	AND SCIENCE (IJPREMS)	Impact
www.ijprems.com	(Int Peer Reviewed Journal)	Factor :
editor@ijprems.com	Vol. 05, Issue 04, April 2025, pp : 2149-2155	7.001

accessibility to communicate to the non-signers. This text-and-speech dual-mode output was highly acceptable in informal usability testing, suggesting promise for effective use in public places, education, and healthcare settings.

The system had some weaknesses in that it performed poorly when gestures were executed too rapidly or with motion blur. As the model is trained on static gestures alone, it found it difficult to understand continuous sign sequences or overlapping gestures, which are typical in natural sign language communication. These limitations indicate the necessity of using time-dependent models such as LSTM or Transformers in subsequent work, which are capable of comprehending the temporal dynamics of gesture sequences. Another line of enhancement is increasing the gesture vocabulary to enable the translation of more complex signs and sentences.

In general, the experimental results confirm the feasibility of the proposed method in recognizing static hand gestures with high accuracy and low latency. The system effectively fills a gap in accessible communication for hearing and speech-impaired people. Its modular structure and scalable design make it amenable to future expansion like dynamic gesture recognition, multilinguality, and mobile-based deployment. The results clearly show that with some improvements, this system has the potential to become an extensive and inclusive sign language communication tool.

7. CONCLUSION

This paper presented a strong and effective real-time sign language identification system for empowering hearing and speech-impaired individuals. Using Convolutional Neural Networks (CNN) for classification and Mediapipe for precise hand landmark detection, the system provides high accuracy in identification of static hand gestures. The whole pipeline—ranging from video recording and hand segmentation to prediction and output—was made with minimal latency to operate in real time so that the system remains usable in a daily setting. Adding a Text-to-Speech (TTS) module makes the system even more user-friendly so that gesture-based communication can be interpreted auditorily, enabling interaction with non-sign language users.

Experimental outcomes validated that the system is reliable in a wide range of conditions, such as varying illumination environments, backgrounds, and orientations of the hand. Accuracy levels were always high in both testing based on datasets and real-time testing. The user interface was deliberately made straightforward to facilitate adoption by users with different technical acumen levels. Additionally, the system was light and did not need powerful computing hardware for operation, such that it would be accessible without needing expensive high-end computing capability. These properties highlight the systems' potential use in public agencies, schools, and individual help devices.

Going forward, the project has immense potential for development. Future developments involve the addition of dynamic gesture recognition, enabling sentence-level sign interpretation through the use of sophisticated sequence models like LSTM, GRU, or Transformers. Furthermore, implementing the system on mobile or embedded platforms may enhance portability and usability in real-world applications. Support for multiple languages and regional sign variants would further enhance its use worldwide. Ultimately, this effort offers a valuable technological contribution toward inclusive communication and new horizons for intelligent assistive systems in human-computer interaction.

ACKNOWLEDGMENT

The authors are extremely thankful to Babitha L, Project Guide, for her most valuable guidance, constant support, and constructive criticism throughout this project development. Her guidance and encouragement were invaluable in the completion of this work.

We would also like to express our sincere gratitude to our group member Maha R for her hard work, teamwork, and dedication throughout all phases of the project. Her research, implementation, and testing efforts were pivotal to the project's success.

Lastly, we want to thank our institution and department for offering us the resources and platform to do this work. We also acknowledge our families, peers, for their support, encouragement, and moral backing along the way.

8. REFERENCE

- [1] I. Mahmud, T. Tabassum, M. P. Uddin, E. Ali, A. M. Nitu, and M. I. Afjal, "Efficient Noise Reduction and HOG Feature Extraction for Sign Language Recognition," in 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), Nov. 2018, pp. 1–4.
- [2] R. B. Mapari, "Real Time Sign Language Translator," International Journal of Engineering Research and General Science, vol. 2, no. 4, pp. 1–4, 2014.
- [3] A. Mujahid, M. J. Awan, A. Yasin, M. A. Mohammed, R. Damaševi ius, R. Maskeliūnas, and K. H. Abdulkareem, "Real-time hand gesture recognition based on deep learning YOLOv3 model," Applied Sciences, vol. 11, no. 9, 2021. [Online]. Available: https://doi.org/10.3390/app11094164

UIPREMS	INTERNATIONAL JOURNAL OF PROGRESSIVE	e-ISSN :
	RESEARCH IN ENGINEERING MANAGEMENT	2583-1062
an ma	AND SCIENCE (IJPREMS)	Impact
www.ijprems.com	(Int Peer Reviewed Journal)	Factor :
editor@ijprems.com	Vol. 05, Issue 04, April 2025, pp : 2149-2155	7.001

- [4] I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, "Machine learning methods for sign language recognition: A critical review and analysis," Intelligent Systems with Applications, vol. 12, p. 200056, 2021. [Online]. Available: https://doi.org/10.1016/j.iswa.2021.200056
- [5] D. Kumar, A. Jain, and M. Kumar, "Interaction with IoT Comfort technologies for Deaf and Dumb People," in Proc. of 2022 Conference, pp. 503–510.
- [6] S. Lahoti, S. Kayal, S. Kumbhare, I. Suradkar, and V. Pawar, "Android Based American Sign Language Recognition System with Skin Segmentation and SVM," in 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018, pp. 1–6.
- I. Lopatovska et al., "Talk to me: Exploring user interactions with the Amazon Alexa," Journal of Librarianship and Information Science, vol. 51, no. 4, pp. 984–997, 2019. [Online]. Available: https://doi.org/10.1177/0961000618759414
- [8] M. Hossain, S. K. Das, and A. B. M. Moniruzzaman, "Bangla Sign Language Recognition using Convolutional Neural Network with Data Augmentation," in 2020 11th ICCCNT, 2020, pp. 1–6.
- [9] [9] S. Ojha and S. Singh, "Sign language recognition using deep learning," Materials Today: Proceedings, vol. 56, pp. 3994–3998, 2022.
- [10] S. Meenakshi and M. Deepa, "Sign Language Recognition System using CNN," International Journal of Engineering and Advanced Technology, vol. 8, no. 6, pp. 2341–2346, 2019.
- [11] A. Sharma and R. Pant, "Real-Time Indian Sign Language Recognition System," in 2020 International Conference on Communication and Signal Processing (ICCSP), 2020, pp. 1364–1368.
- [12] M. S. A. Malik, M. S. Qureshi, and T. Ahmad, "Sign Language Recognition using YOLOv5 and CNN," International Journal of Advanced Computer Science and Applications, vol. 13, no. 6, pp. 204–210, 2022.
- [13] S. Gupta, A. Dey, and S. Chakraborty, "Vision-Based Hand Gesture Recognition using CNN for Indian Sign Language," Procedia Computer Science, vol. 167, pp. 1234–1242, 2020.
- [14] F. Zhang, Q. Sun, and M. Han, "Gesture Recognition for Smart Home Automation using CNN," IEEE Access, vol. 7, pp. 164570–164577, 2019.
- [15] N. Arif, Z. Aslam, and A. Khan, "Real-Time Sign Language Detection using Deep Neural Networks," in International Conference on Innovative Computing, 2021, pp. 85–90.