IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

SmartSign: Real-Time Hand Gesture Recognition and Speech Generation Platform

[1] Mr. Animesh Chandrashekhar Agarkar, [2] Mr. Rahul Mahadev Kharap, [3] Mr. Sahil Sharma,
[4] Mr. Vaibhav Sahadeo Korde, [5] Dr. Abrar S. Alvi
Students, Department of Information Technology[1-4]
Guide, Department of Information Technology[5]
Prof. Ram Meghe Institute of Technology & Research (Autonomous), Badnera, Amravati, Maharashtra, India.

Abstract: *Sign language is an essential means of communication for individuals with hearing and speech impairments. However, communication barriers often arise between sign language users and people who are unfamiliar with sign language. To address this challenge, this paper presents a SmartSign: Real-Time Hand Gesture Recognition and Speech Generation Platform that uses computer vision and deep learning techniques to recognize hand gestures and convert them into understandable text. The proposed system is developed using Python, Flask, OpenCV, MediaPipe, and a trained machine learning model. It provides multiple detection methods, including image-based detection, video-based detection, multi-image processing, and real-time webcam recognition. The system also incorporates interactive learning features such as sign language practice, tutorials, a chatbot, progress tracking, history, and analytics to support continuous learning. Additionally, an administrative panel enables administrators to manage users, tutorials, detection history, and application settings. The system processes hand landmarks and extracts relevant features to identify corresponding signs and display the predicted results through a user-friendly web interface. By integrating sign recognition with learning and progress-monitoring facilities, the proposed application aims to improve accessibility, support sign language education, and reduce communication barriers. The system demonstrates the potential of combining computer vision, deep learning, and web technologies to develop an interactive and accessible sign language recognition platform*

Keywords: Sign Language Detection, Sign Language Recognition, Computer Vision, Deep Learning, Machine Learning, MediaPipe, OpenCV, Hand Landmark Detection, Real-Time Gesture Recognition, Flask Web Application, Accessibility, Assistive Technology

I. INTRODUCTION

Sign language is one of the primary communication methods used by individuals with hearing and speech impairments. It involves the use of hand gestures, finger movements, facial expressions, and body movements to convey information and emotions. However, communication between sign language users and individuals who do not understand sign language can be challenging. This communication gap creates difficulties in education, healthcare, workplaces, and everyday social interactions. Therefore, developing an automated sign language detection and recognition system is an important step toward improving accessibility and inclusive communication[1].

Recent advancements in Artificial Intelligence (AI), Machine Learning (ML), Computer Vision, and Deep Learning have created new opportunities for recognizing and interpreting human gestures. Computer vision techniques can identify hand movements and extract meaningful features from images and video frames. MediaPipe can be used to detect hand landmarks, while a trained machine learning or deep learning model can classify the extracted features into corresponding sign classes. OpenCV supports image and video processing, and Python provides the necessary libraries for implementing the recognition process. Flask enables the integration of these technologies into a web-based application that users can access through a browser[2].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

The proposed Sign Language Detection and Recognition System aims to provide an interactive platform for detecting hand gestures through different input methods, including single images, videos, multiple images, and real-time webcam input. The system processes the input, identifies relevant hand features, predicts the corresponding sign, and displays the recognition result. By supporting multiple input methods, the application provides flexibility for users to test and understand sign language gestures in different situations[3].

In addition to gesture recognition, the proposed system includes learning and user-management features to improve the overall user experience. The learning section provides practice activities, tutorials, a chatbot, and progress monitoring. The data section maintains detection history and presents analytics to help users review their activities and learning progress. Furthermore, the administrative panel supports user management, tutorial management, history monitoring, and application settings[4].

The main objective of this research is to develop a user-friendly and accessible sign language recognition platform that combines computer vision, trained classification models, and web technologies. The system aims to support sign language learning, simplify gesture recognition, and reduce communication barriers between sign language users and the wider community. Although the proposed application can assist communication and learning, its recognition performance depends on the training dataset, supported gestures, lighting conditions, hand positioning, and model accuracy. The system therefore represents a practical step toward developing more inclusive and intelligent assistive communication technologies[5].

II. LITERATURE ANALYSIS

Several research studies have focused on improving communication and learning for hearing and speech-impaired individuals using sign language technologies. Chaithraj and Kavana H. R (2019) proposed a vision-based system that converts sign language gestures into text and speech using image processing techniques for two-way communication. Sanyukta Bhavsar et al. (2023) developed a mobile-based learning application with interactive modules and real-time feedback to make sign language education more accessible and user-friendly. Vishnu Priya A. K et al. (2023) introduced a deep learning-based approach using Convolutional Neural Networks (CNN) and OpenCV for real-time gesture recognition and translation between sign language and speech. Additionally, Rekha V et al. (2023) presented an AI-based multilingual system integrating advanced models such as CNN, LSTM, and NLP to support inclusive learning and communication across different languages. Together, these studies highlight the growing role of computer vision, artificial intelligence, and user-centric design in bridging communication gaps and promoting social inclusion for the deaf and mute community

TABLE I.: LITERATURE WORK

Author and Year	Methods	Future Scope
Chaithraj, Kavana H R (2019)	Vision-based system using webcam; Image processing for hand coordinate extraction; Matching with pre-stored dataset; Gesture-to-text and speech conversion; Two-way communication (sign-to-text/speech and audio-to-sign/text); Sentence formation using continuous gestures	Improve performance in different lighting/backgrounds; Support regional languages and complex gestures; Use advanced ML/DL models; Develop mobile applications; Enhance real-time communication; Implement continuous sentence recognition
Sanyukta Bhavsar, Dhanashree Dhondge, Siddhi Khande, Sejal Zend, Prof. M. D. Sanap (2023)	Mobile-based learning application; Interactive modules for alphabets and phrases; User-friendly interface; Real-time feedback system; User-centric design with continuous updates; Focus on Indian Sign Language learning	Improve gesture accuracy and representation; Add advanced learning levels; Enhance cultural sensitivity; Integrate AI-based personalized learning; Expand to multiple languages; Add real-time communication features

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

Vishnu Priya A. K, Jayashri S., Sivanjali V., Sneha V., Thamaraiselvi K (2023)	Two-way sign language translator using Deep CNN; Real-time processing with OpenCV; Gaussian Mixture Model for background removal; RGB camera-based gesture recognition; Sign-to-text and speech-to-sign conversion	Develop Android/mobile application; Improve speed and accuracy; Use advanced deep learning models; Support complex gestures and sentence recognition; Enhance real-world usability
Mrs. Rekha V, Shreya M C, Thanushree H C, Sonu Chouhan, Sumathi V (2023)	AI-based multilingual system; Use of CNN, LSTM, Transformers, hybrid models; Integration of Computer Vision and NLP; ISL gesture to multilingual text/speech conversion; Supports learning and communication	Develop universal communication system; Improve multilingual and real-time performance; Expand large-scale deployment; Enhance collaboration with experts; Advance AI accuracy; Promote inclusive digital environments

III. WORKING METHODOLOGY

The proposed Sign Language Detection and Recognition System uses computer vision, hand landmark detection, and a trained machine learning or deep learning model to recognize sign language gestures. The system is implemented using Python, Flask, OpenCV, and MediaPipe. It provides multiple input methods, including image, video, multi-image, and real-time webcam detection. The complete methodology consists of the following stages.

A. Input Acquisition

The system accepts input through different methods to provide flexibility for users. Users can upload a single image, submit a video, process multiple images, or use a webcam for real-time gesture recognition. The Flask web application receives the selected input and forwards it to the appropriate processing module. This approach enables the system to recognize gestures from both static images and video-based inputs.

B. Image and Video Preprocessing

After receiving the input, the system performs preprocessing to prepare the data for recognition. OpenCV is used to read images and video frames, convert color formats when required, and resize or process the input according to the model's requirements. For video and webcam detection, frames are extracted continuously or at selected intervals. This step helps maintain a consistent input format for subsequent hand detection and feature extraction.

C. Hand Landmark Detection and Feature Extraction

MediaPipe is used to identify hand landmarks from the input images or video frames. It can detect 21 key landmarks on each hand, representing important points such as the wrist, finger joints, and fingertips. These landmarks provide spatial information about hand shape and finger positions. The coordinates are extracted and organized into a numerical feature vector. When using 21 three-dimensional landmarks, the feature vector can contain 63 coordinate values representing the (x), (y), and (z) coordinates of each landmark. These features are then supplied to the trained recognition model.

D. Sign Language Classification

The extracted hand landmark features are passed to a previously trained machine learning or deep learning model. The model analyzes the feature patterns and predicts the corresponding sign class from the gestures it has learned during training. The predicted class is converted into a readable label that represents the recognized gesture. The recognition process depends on the classes included in the training dataset and the accuracy of the trained model.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

E. Multiple Detection Methods

The system provides four primary detection methods. Image detection identifies a gesture from a single uploaded image. Video detection processes frames from an uploaded video to recognize gestures over time. Multi-image detection allows users to process several images in one session. Webcam detection captures live video frames and performs gesture recognition continuously, enabling users to observe predictions in real time. These methods improve the flexibility and usability of the application.

F. Web Application Integration

Flask acts as the backend framework and connects the detection modules with the web interface. It manages user requests, processes uploaded files, invokes the recognition model, and returns the results for display. The dashboard provides access to the image, video, multi-image, and webcam modules. This integration allows users to perform sign language detection through a browser without directly interacting with the underlying machine learning code.

G. Learning and Progress Monitoring

The application also incorporates learning features, including practice exercises, tutorials, a chatbot, and progress monitoring. Users can explore sign language concepts, practice supported gestures, and review their learning activities. The history module stores previous detection records where supported by the application, while analytics presents available activity information in a summarized form. These features aim to improve user engagement and support continuous learning.

H. Administrative Management

An administrative panel is provided to manage the application. It includes a dashboard, user management, tutorial management, detection history, and settings. These facilities help administrators maintain user information, organize learning resources, and monitor application activities. Administrative access is intended to be restricted to authorized users.

I. Output Generation and Evaluation

After classification, the system displays the predicted sign label on the web interface. For video and webcam inputs, predictions can be displayed for individual frames or across the processed sequence, depending on the implementation. The system can be evaluated using test images and videos by measuring classification accuracy, precision, recall, and F1-score, along with inference time for real-time detection. Testing under different lighting conditions, hand positions, and backgrounds helps identify limitations and opportunities for improvement.

Overall, the working methodology integrates input acquisition, preprocessing, landmark extraction, gesture classification, web-based result presentation, and learning support into a single application. This modular approach makes the system suitable for sign language recognition experiments and provides a foundation for developing more advanced assistive communication applications.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

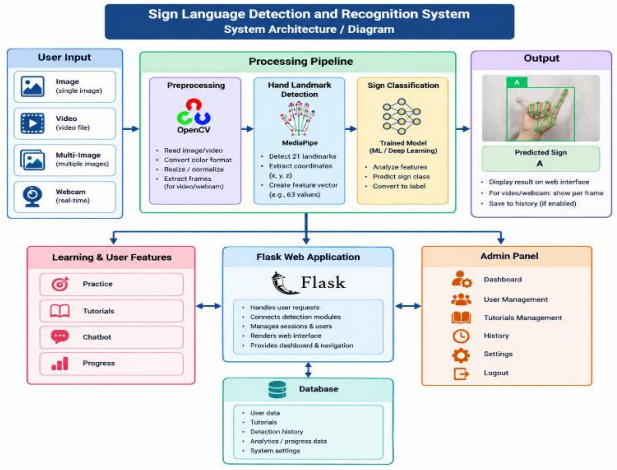


Figure 1. System Diagram

IV. RESULTS AND DISCUSSION

The proposed Sign Language Detection and Recognition System provides a web-based platform for recognizing hand gestures and supporting sign language learning. The application integrates computer vision, hand landmark detection, a trained recognition model, and Flask to provide multiple detection methods through a user-friendly interface. The system includes image detection, video detection, multi-image processing, and real-time webcam recognition, along with learning and administrative functionalities.

A. Sign Language Detection Results

The image detection module allows users to upload an image containing a hand gesture and obtain the predicted sign label. The video detection module processes video frames to identify supported gestures, while the multi-image module enables the processing of multiple images in a single session. The webcam module captures live frames and performs gesture recognition, allowing users to observe predictions during real-time interaction. These modules provide different ways to interact with the trained recognition model.

B. Hand Landmark Detection and Classification

MediaPipe is used to identify hand landmarks and extract relevant spatial features from input images and video frames. The extracted features are passed to the trained model, which predicts the corresponding sign class. This approach reduces the need to process every image pixel directly during classification and provides a structured representation of hand posture. However, prediction quality depends on the training data, the supported sign classes, landmark detection quality, and environmental conditions.

C. Learning and User Interaction

The system includes practice exercises, tutorials, a chatbot, and progress monitoring to support sign language learning. The history module allows users to review previous detection activities, while analytics provides a summary of available usage and progress information. These features extend the application beyond basic gesture classification by combining detection capabilities with interactive learning resources.

D. Administrative Features

The administrative panel provides facilities for managing registered users, tutorials, detection history, and application settings. These functions help maintain learning resources and monitor application activity. Separating administrative operations from regular user functionality also supports more organized application management.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

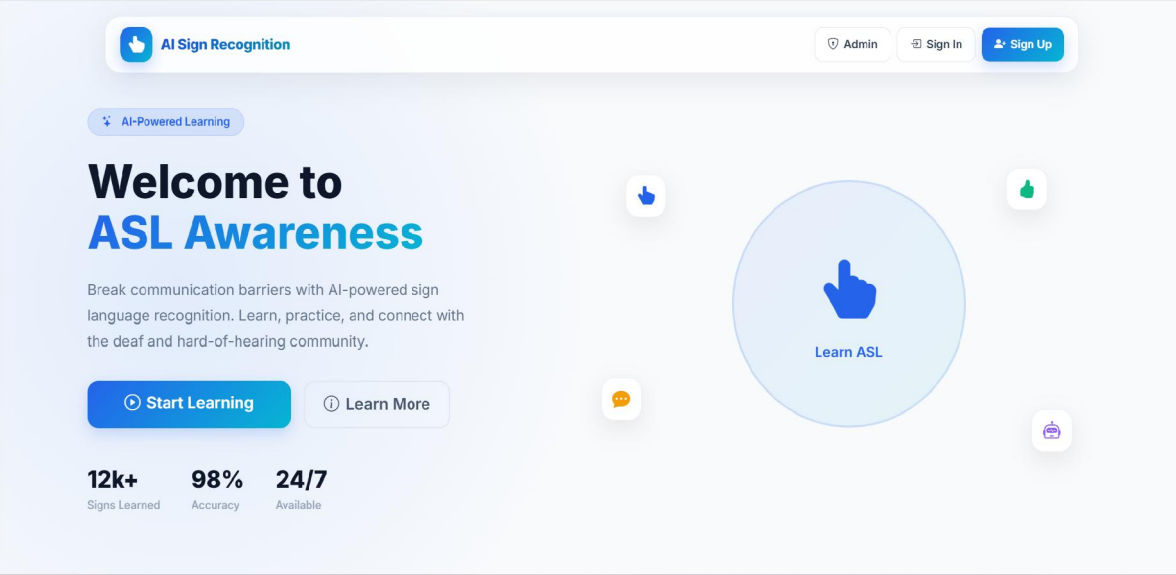
E. Discussion

The integration of computer vision and a trained recognition model with a Flask web application demonstrates a practical approach to developing an accessible sign language recognition platform. Supporting multiple input methods improves flexibility, while the learning and progress-monitoring modules provide additional educational value. Nevertheless, the system may encounter difficulties when hand landmarks are partially obscured, lighting is insufficient, backgrounds are complex, or gestures are visually similar. Recognition of continuous sign language also requires consideration of movement over time, facial expressions, and body posture where relevant.

The performance of the proposed system should be assessed using a separate test dataset and quantitative evaluation metrics, including accuracy, precision, recall, F1-score, and inference time. Since measured experimental results have not been provided here, no specific accuracy percentage or performance improvement is claimed. Future improvements may include expanding the training dataset, supporting additional signs, improving temporal gesture recognition, and evaluating the model across diverse users and environmental conditions.

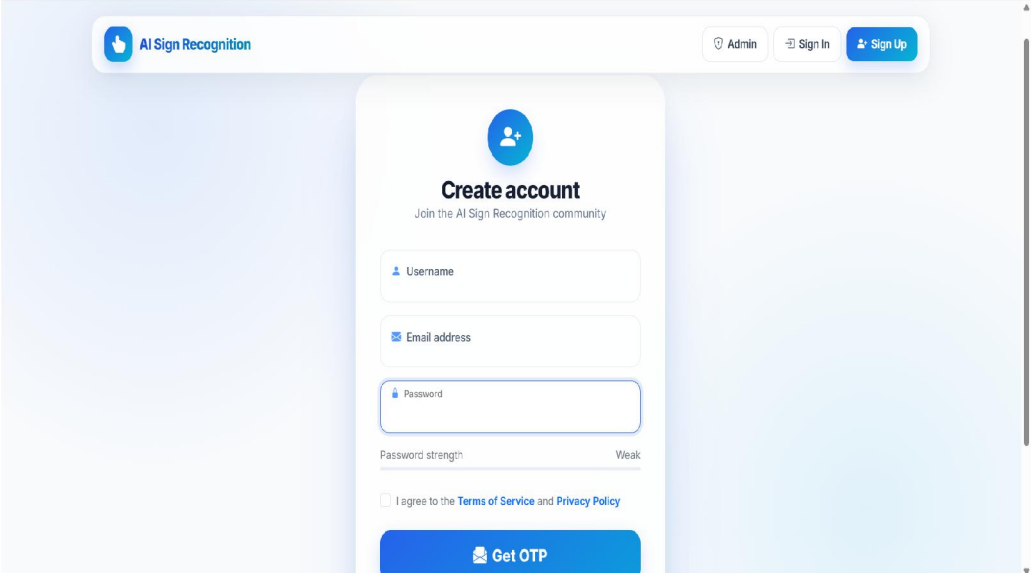
Overall, the system combines sign language detection, interactive learning, activity tracking, and administrative management in a unified web application. The proposed approach has the potential to support sign language education and assistive communication, subject to further validation of its recognition accuracy and real-world performance.

Screenshots:

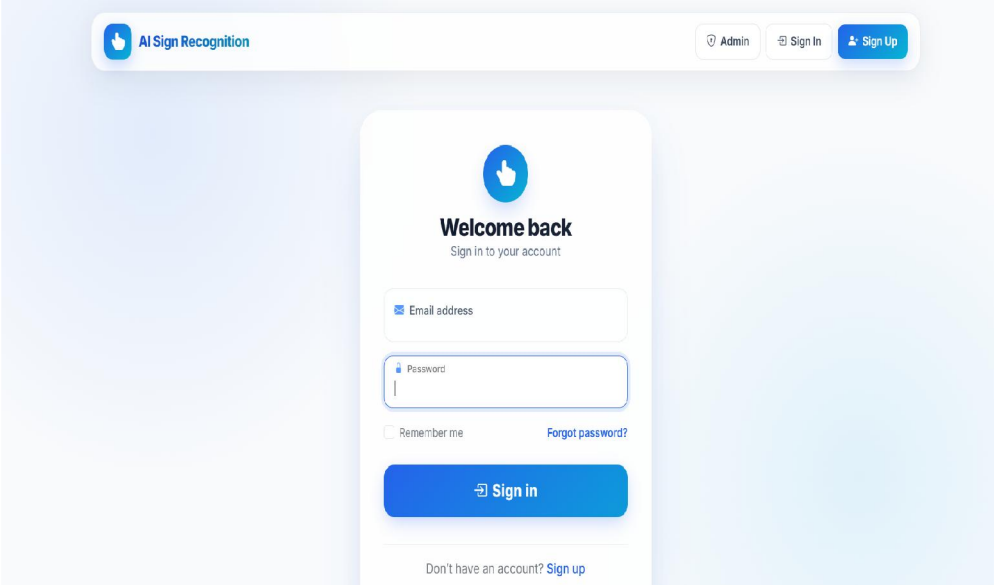


Screenshot 5.1 Home Page

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

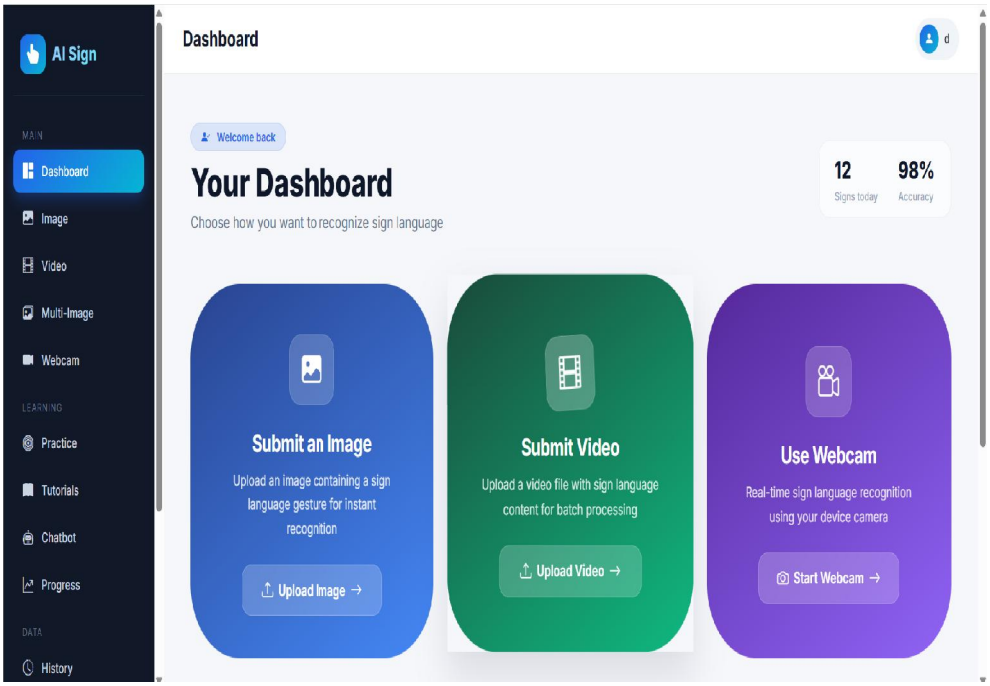


Screenshot 5.2 Registration Page

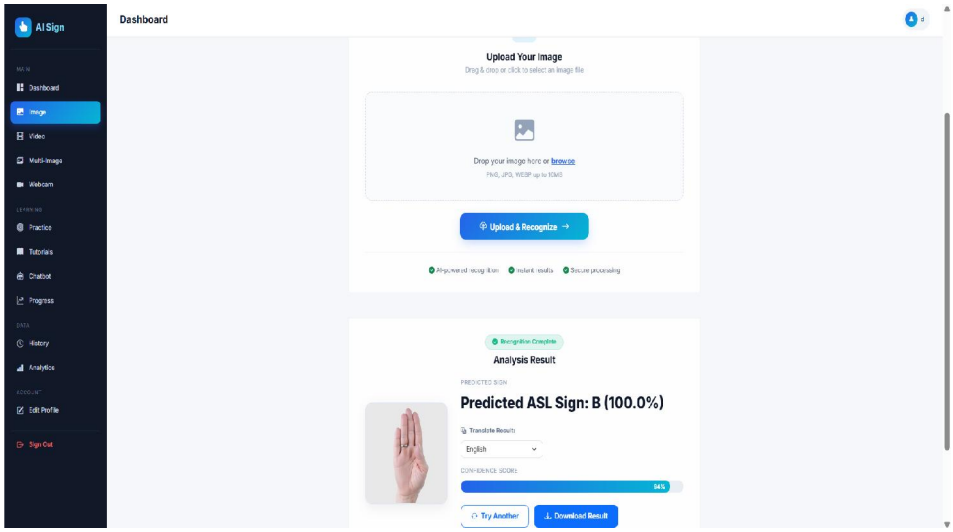


Screenshot 5.3 Login Page

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

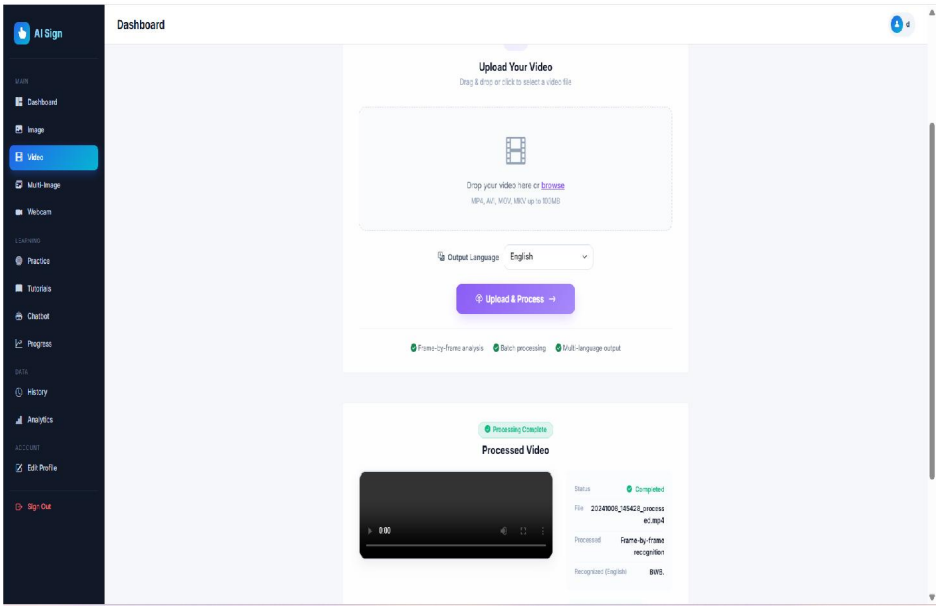


Screenshot 5.4 User Dashboard

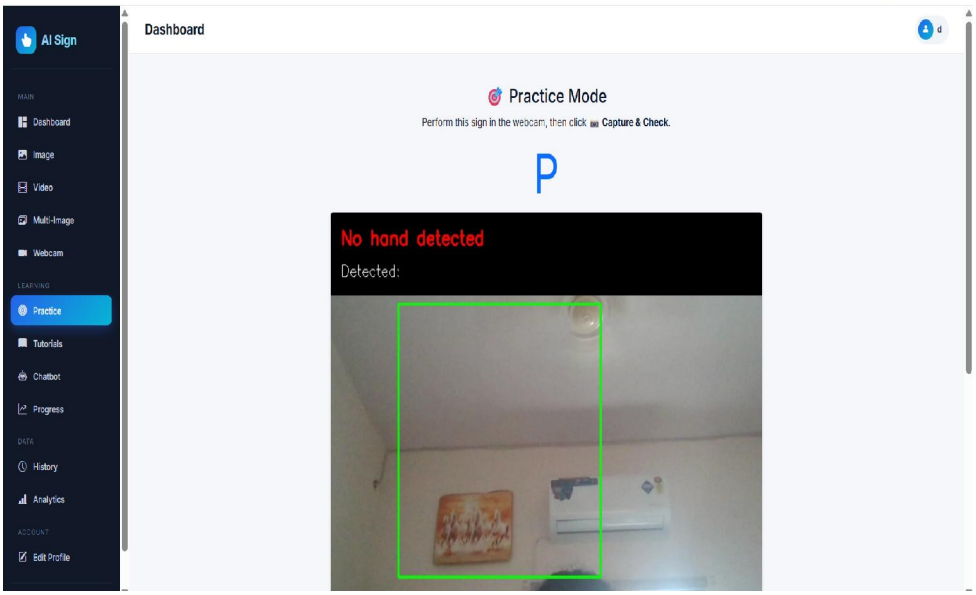


Screenshot 5.5 Image Detection

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

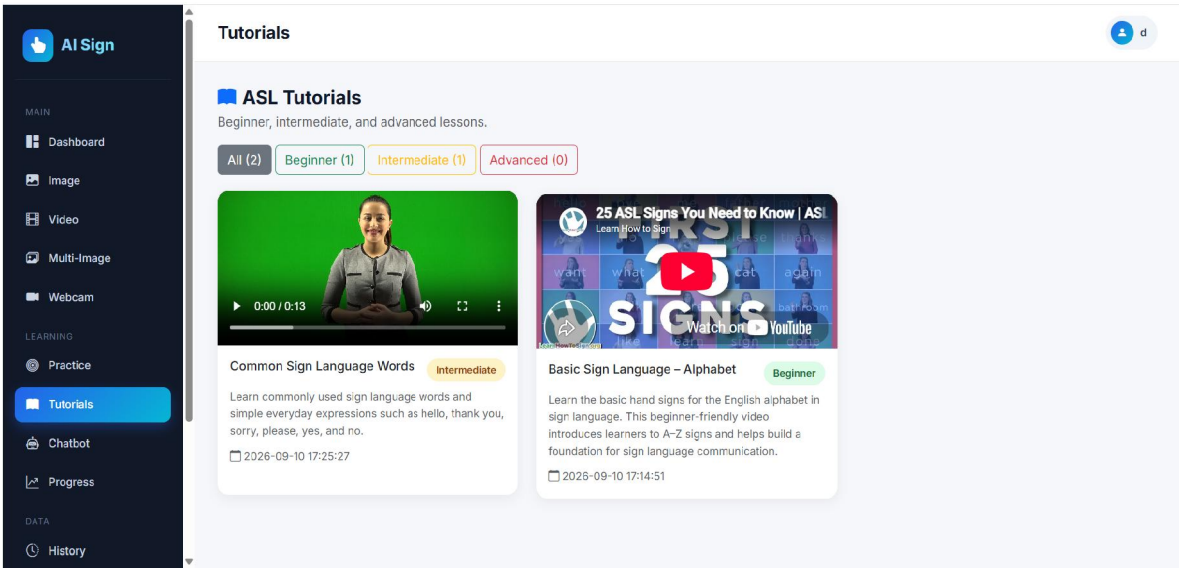


Screenshot 5.6 Video Detection

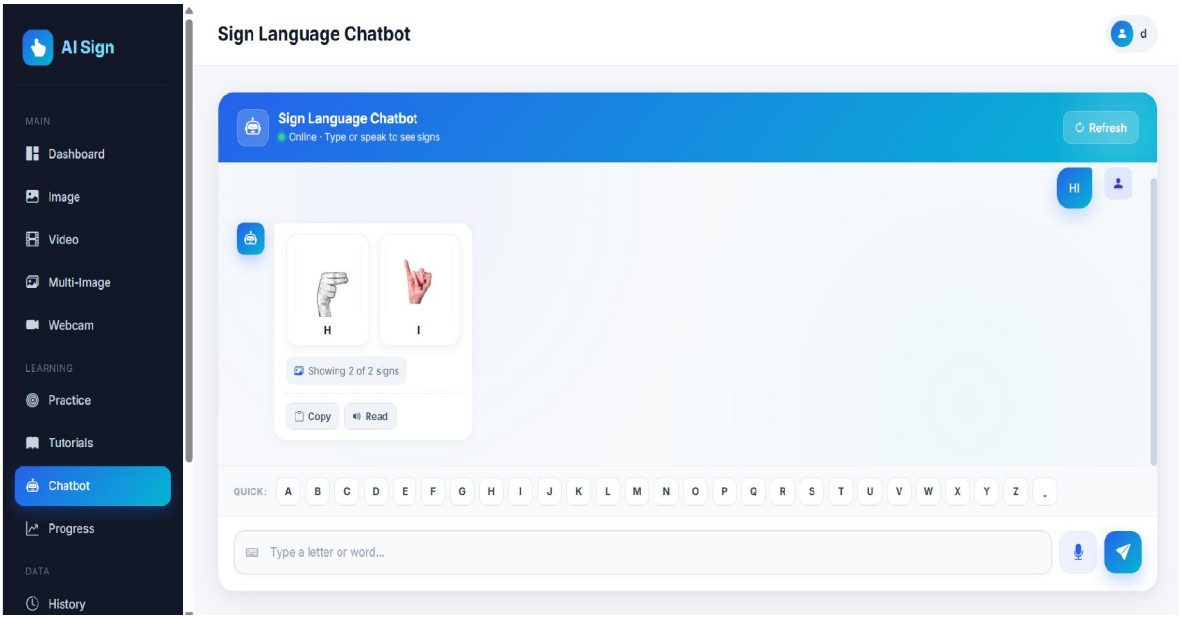


Screenshot 5.7 Practise

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

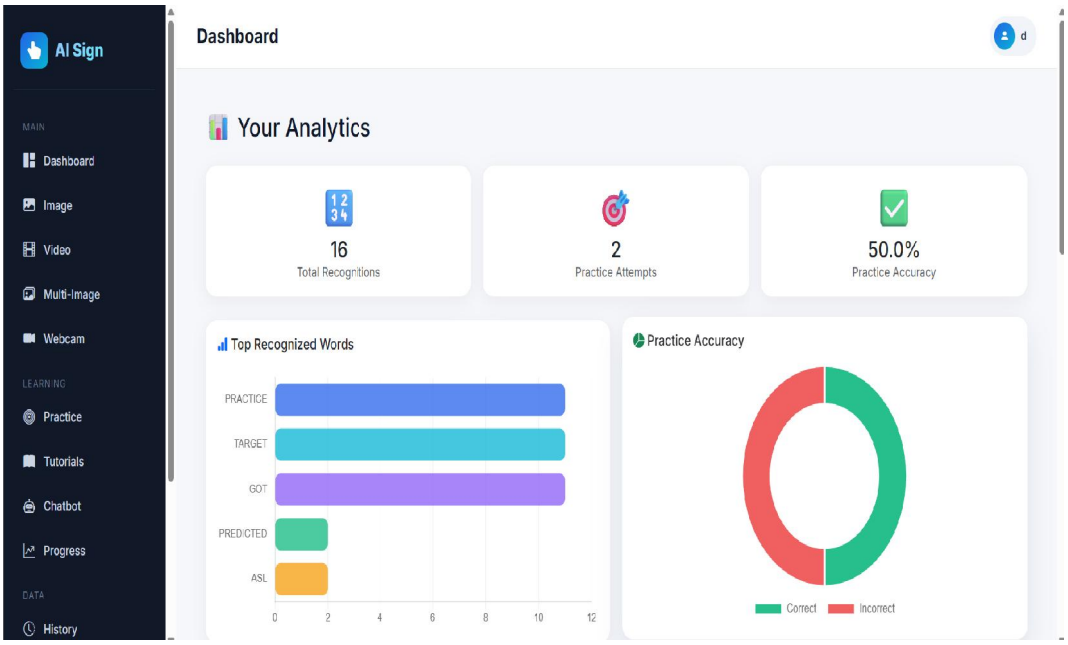


Screenshot 5.8 Tutorial Videos

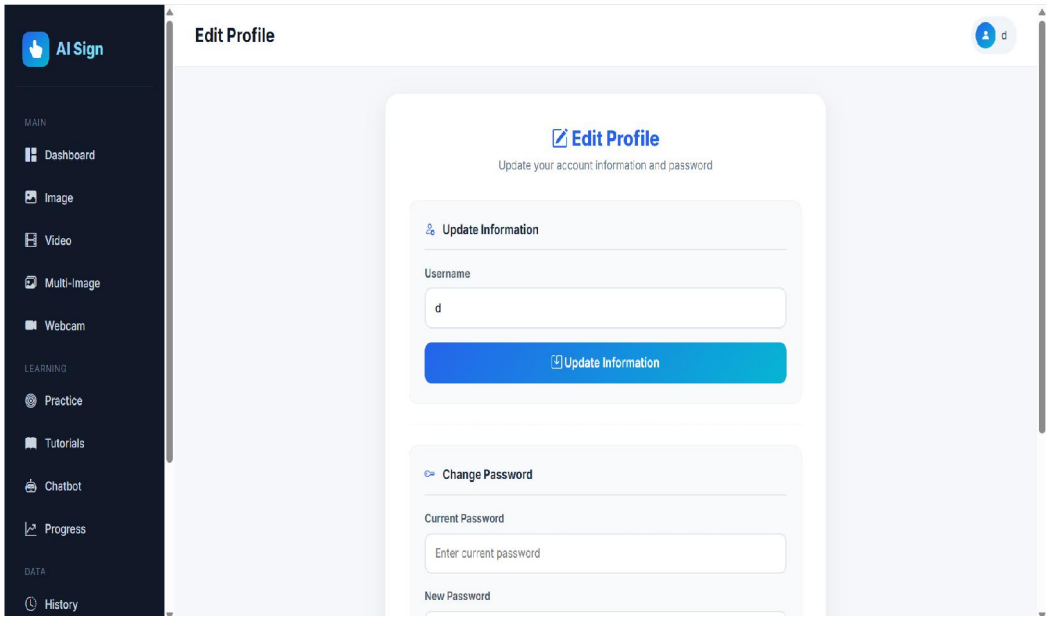


Screenshot 5.9 ChatBot

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2



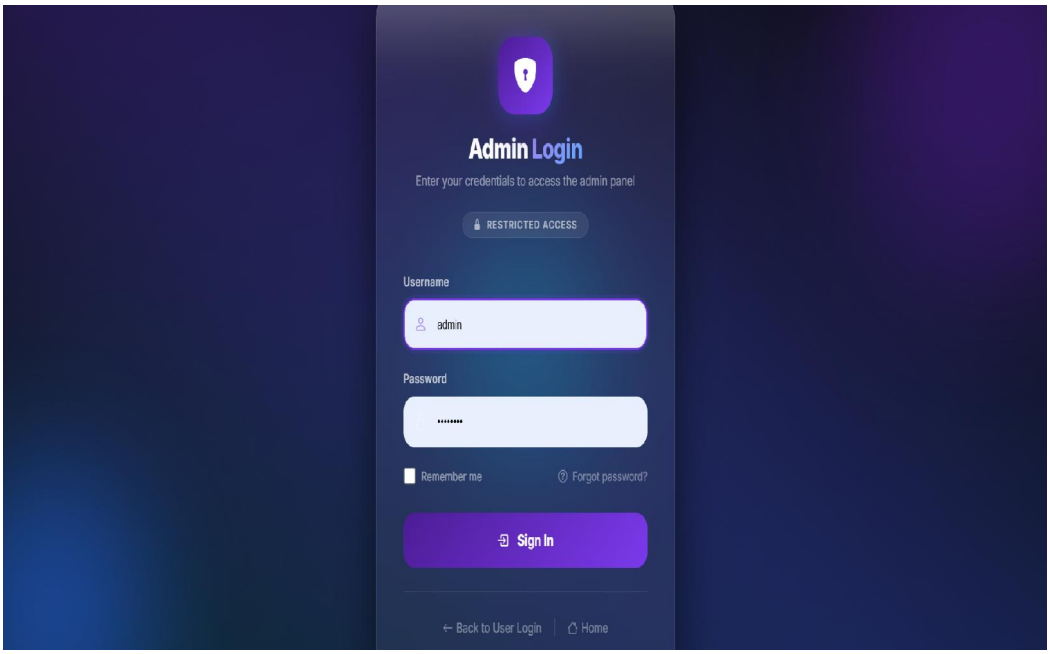
Screenshot 5.10 Analytics



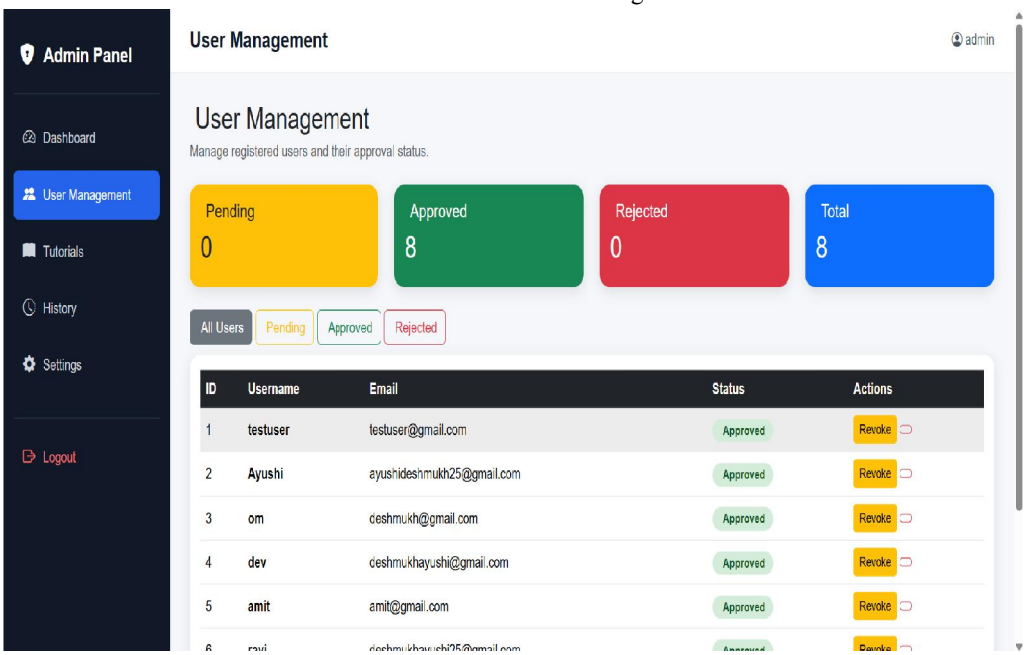
Screenshot 5.11 Edit Profile

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-38861

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2



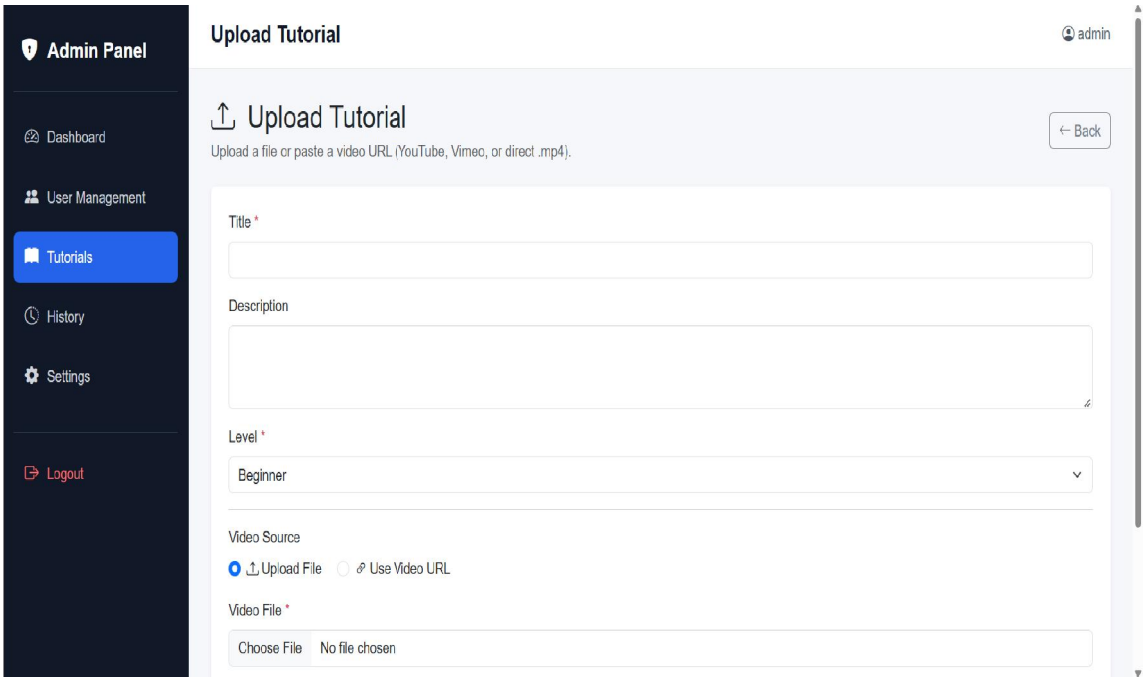
Screenshot 5.12 Admin Login



Screenshot 5.13 Manage User

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-38861

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2



Screenshot 5.14 Upload Tutorial Video

V. CONCLUSION

The proposed SmartSign: Real-Time Hand Gesture Recognition and Speech Generation Platform provides a web-based solution for recognizing hand gestures and supporting sign language learning through computer vision and machine learning techniques. The system integrates Python, Flask, OpenCV, MediaPipe, and a trained recognition model to process images, videos, multiple images, and real-time webcam inputs. By extracting hand landmarks and classifying the resulting features, the application identifies supported sign gestures and displays the predicted labels through an interactive web interface.

In addition to gesture recognition, the system incorporates practice exercises, tutorials, a chatbot, progress monitoring, detection history, and analytics to enhance the learning experience. The administrative panel provides facilities for managing users, tutorials, history, and application settings. These integrated features make the application more comprehensive than a basic sign detection tool by combining recognition, learning, and user management within a single platform.

Although the proposed system offers a practical approach to assistive communication and sign language education, its effectiveness depends on the quality of the training dataset, the accuracy of the recognition model, and environmental conditions such as lighting, background complexity, and hand positioning. Further improvements can include expanding the range of recognizable signs, incorporating facial expressions and body movements, enhancing continuous gesture recognition, and evaluating performance across diverse users and real-world environments.

In conclusion, the proposed system demonstrates the potential of combining artificial intelligence, computer vision, and web technologies to develop accessible sign language recognition applications. With further testing and improvement, it can contribute to reducing communication barriers, promoting inclusive learning, and supporting the development of intelligent assistive technologies for individuals with hearing and speech impairments.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-38861

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

REFERENCES

[1] Chaithraj and Kavana H. R, "Sign Language Converter for Deaf and Dumb People in Two Way Communication for Regional Languages," Dept. of ISE, NIEIT, Mysuru, India, 2019.
A vision-based system using image processing to convert sign language into text and speech for effective two-way communication.
[2] S. Bhavsar, D. Dhondge, S. Khande, S. Zend, and M. D. Sanap, "Sign Language Learning Application for Deaf and Dumb People," International Journal of Creative Research Thoughts (IJCRT), vol. 11, no. 5, May 2023.
A mobile-based application providing interactive modules and real-time feedback to enhance sign language learning and accessibility.
[3] V. Priya A. K, J. S., S. V., S. V., and T. K., "Two Way Sign Language for Deaf and Dumb using Deep Convolution Neural Network," International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 11, issue IV, Apr. 2023.
A deep learning-based system using CNN and OpenCV for real-time two-way sign language translation.
[4] R. V., S. M. C., T. H. C., S. C., and S. V., "Empowering Deaf and Mute Students through Multilingual Sign Language Learning," International Journal of Research Publication and Reviews (IJRPR), 2023.
An AI-based multilingual system integrating CNN, LSTM, and NLP for inclusive sign language learning and communication.
[5] Harshitha C, Sendil Vadivu D, and Narendran Rajagopalan, "A Survey on Machine and Deep Learning Approaches in Sign Language Recognition: Techniques and Future Trends," Procedia Computer Science, vol. 218, pp. 143–152, 2023.
[6] Arun Singh, Ankita Wadhawan, Usha Mittal, Ahmed Al Ahdal, Manik Rakhra, and Shambhu Kumar Jha, "Indian Sign Language Recognition using Dynamic Signs," Journal of Ambient Intelligence and Humanized Computing, vol. 14, pp. 3351–3364, 2022.
[7] Shashidhar R, Ankit Priyesh, Surendra R. Hegde, A. S. Manjunath, Chinmaya K, and Arunakumari B. N, "Indian Sign Language to Speech Conversion using Convolutional Neural Network," International Journal of Engineering Research & Technology (IJERT), vol. 10, no. 6, pp. 1–6, 2021.
[8] Nidhi Chandarana, Nidhish Chhajed, Shreya Manjucha, Monica G. Tolani, Priyansi Chogale, and Mani Roja M. Edinburgh, "Indian Sign Language Recognition with Conversion to Bilingual Text and Audio," International Research Journal of Engineering and Technology (IRJET), vol. 9, no. 4, pp. 1452–1458, 2022.
[9] Marinee Devi, Anusha, Ankitha, Ishwarya, and Kaviya, "Learning App for Deaf-Mute using Sign Language English/Gujarati Converter," International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), vol. 10, no. 5, pp. 2156–2163, 2022. [10] Soumen Das, Saroj Kr. Biswas, and Biswajit Purkayasth, "Occlusion Robust Sign Language Recognition System for ISL using CNN," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 7, no. 2, pp. 435–446,2023.

Author Information

Mr. Animesh Chandrashekhar Agarkar, Department of Information Technology, Prof. Ram Meghe Institute of Technology & Research (Autonomous), Badnera, Amravati, Maharashtra, India
Mr. Rahul Mahadev Kharap, Department of Information Technology, Prof. Ram Meghe Institute of Technology & Research (Autonomous), Badnera, Amravati, Maharashtra, India
Mr. Sahil Sharma, Department of Information Technology, Prof. Ram Meghe Institute of Technology & Research (Autonomous), Badnera, Amravati, Maharashtra, India
Mr. Vaibhav Sahadeo Korde, Department of Information Technology, Prof. Ram Meghe Institute of Technology & Research (Autonomous), Badnera, Amravati, Maharashtra, India
Guide Dr. Abrar S. Alvi, Department of Information Technology, Prof. Ram Meghe Institute of Technology & Research (Autonomous), Badnera, Amravati, Maharashtra, India