

Literature Review

1

Conversion of Sign Language into Text), Mahesh Kumar, India (2018):

This paper shows the sign language recognizing of 26 hand gestures in Indian sign language using MATLAB. The proposed system contains four modules such as: pre-processing and hand segmentation, feature extraction, sign recognition and sign to text. By using image processing the segmentation can be done. Otsu algorithm is used for segmentation purposes. Some of the features are extracted such as Eigen values and Eigen vectors which are used in recognition. The Linear Discriminant Analysis (LDA) algorithm was used for gesture recognition and recognized gestures are converted into text and voice format. The proposed system helps to dimensionality.

Link : https://www.ripublication.com/ijaer18/ijaerv13n9_90.pdf

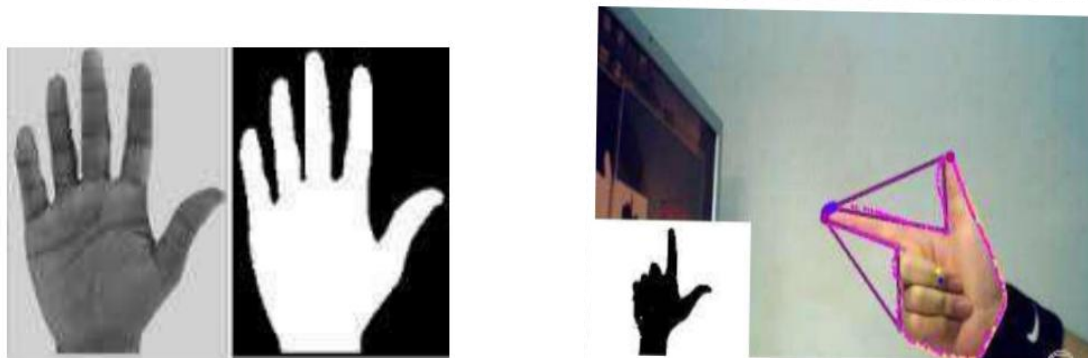


Figure 1 explains about the sign language of 26 gestures.

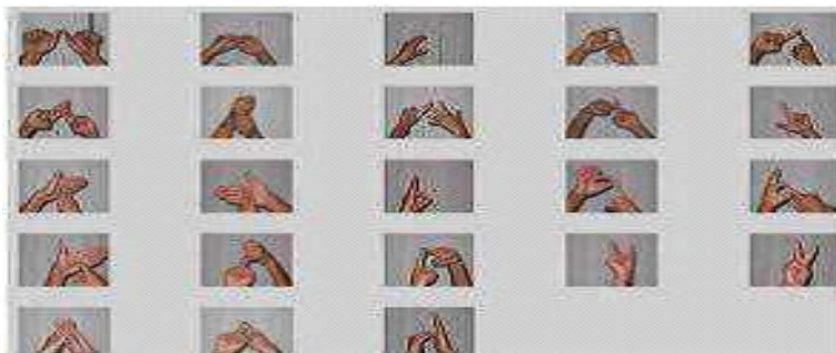


Figure 1. Gestures of sign recognition

2

Sign Language Recognition using Deep Learning, Mohamed Mahyoub (2023)

This paper investigates the effectiveness of deep learning techniques in recognizing sign language words from videos. It compares five different architectures ranging from simple CNN models to advanced hybrid architectures combining 3D CNNs, Transformers, and GRUs.

The study demonstrates that:

- Simple CNN models struggle with complex video-based sign recognition.
- Transformer-based sequence modeling significantly improves results.
- 3D CNNs (I3D) capture temporal information better than 2D CNNs & Deep hybrid architectures achieve near state-of-the-art performance.

Link :

<https://researchonline.ljmu.ac.uk/id/eprint/19754/1/1.%20Sign%20Language%20Recognition%20Using%20Deep%20Learning.pdf>

TABLE I. SIGN LANGUAGE DATASETS USED

Sign Language	INCLUDE Indian Sign Language (ISL)	WLASL American Sign Language (ASL)	AUTSL Turkish Sign Language (TSL)
# Words	263	2,000	226
# Word Categories	15	-	-
# Videos	4,292	21,083	38,336
# Frames	270,000	-	-
Video Resolution	1920x1080	-	512x512
# Signers	7	119	43



figure 2.

3

Recognizing American Sign Language gestures efficiently and accurately using a hybrid transformer model , Mohammed Aly& Islam S. Fathi (2024)

This paper presents a machine learning–based sign language recognition system designed to facilitate communication between deaf or mute individuals and people who do not understand sign language. The proposed system integrates computer vision and machine learning techniques to recognize hand gestures and convert them into text and speech in real time. The system uses **MediaPipe** to extract hand landmarks and detect gesture features, while a **Convolutional Neural Network (CNN)** is employed for feature extraction. A **Random Forest classifier** is then used for gesture classification. The model is integrated with a **Leap Motion Controller** to capture hand movements accurately and enable real-time recognition. The methodology includes data collection, preprocessing (resizing, grayscale conversion, normalization), feature extraction using hand keypoints, model training with Random Forest, and system deployment with a graphical user interface built using Tkinter. Additionally, the system incorporates text-to-speech functionality using pyttsx3 to provide auditory feedback. Although the paper reports successful real-time gesture recognition and system integration, the exact numerical accuracy is not clearly specified. The evaluation metrics mentioned include accuracy, precision, recall, and F1-score. Overall, the study demonstrates a practical implementation of a hybrid machine learning system that combines hardware (Leap Motion Controller) and software (CNN + Random Forest) to develop an assistive communication tool.

Link : <https://www.nature.com/articles/s41598-025-06344-8.pdf>



figure 3.

Analysis And Design

3.1 Comparison Table

Author name	Mahesh Kumar	Mohamed Mahyoub	Mohammed Aly& Islam S. Fathi
Dataset	26 hand gestures in Indian sign language	American Sign Language Lexicon Video Dataset (ASLLVD)	American Sign Language
Algorithm & Model	FAST + SURF feature extraction + KNN classifier	4-Layer CNN, VGG16, VGG16+Attention, VGG16+Transformer+ GRU, I3D+Transformer+GRU	CNN + Random Forest + MediaPipe + Leap Motion
Accuracy	80%	Up to 98% (INCLUDE), 96% (AUTSL)	Not clearly specified (X%)
Year	2018	2023	2024
Framework	MATLAB + Kinect	Python, OpenCV, Numpy	Python, OpenCV, Scikit-learn, Tkinter
Methodology	Image segmentation → ROI extraction → Feature detection → KNN classification → Text-to-Speech	Video dataset split (70/10/20) → CNN feature extraction → Transformer encoding → GRU decoding → Accuracy evaluation	Preprocessing → Landmark extraction → CNN feature extraction → Random Forest classification → Real-time gesture recognition

Literature Review

4

Sign Language Recognition for Arabic Alphabets Using Transfer Learning Technique, Mohammed Zakariah, Yousef Ajmi Alotaibi (2022)

This study proposes an automated system to interpret **Arabic Sign Language (ArSL)** alphabets and convert them into textual information. To address the complexity and dialects within Arabic sign language, the researchers employed a **Transfer Learning** approach using several pre-trained models. The final system aims to provide a low-cost communication solution for the deaf and mute communities.

Link : <https://pdfs.semanticscholar.org/5739/0c1b6dfd7a89b91d102c4019e39ff20cc488.pdf>

Figure 1 explains about the sign language of 26 gestures.



Figure 1. Gestures of sign recognition

5

A machine learning-driven web application for sign language learning, Hope Orovwode ,Oduntan Ibukun (2024)

This research presents an interactive **web application** designed specifically for learning the **American Sign Language (ASL)** alphabet. Unlike systems focused solely on translation, this application uses a "gamified" methodology where users are assigned words to spell out via their webcam. The system provides live feedback, awarding points when letters are signed correctly, thereby fostering inclusivity by making sign language education more accessible through common web browsers.

Link : <https://public-pages-files-2025.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1297347/pdf>

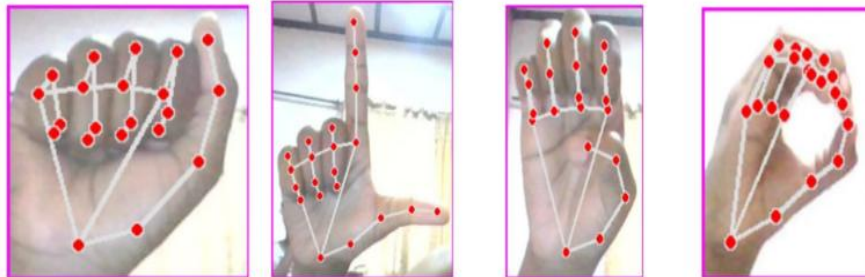


FIGURE 1
Sample of the dataset (Orovwode et al., 2023).

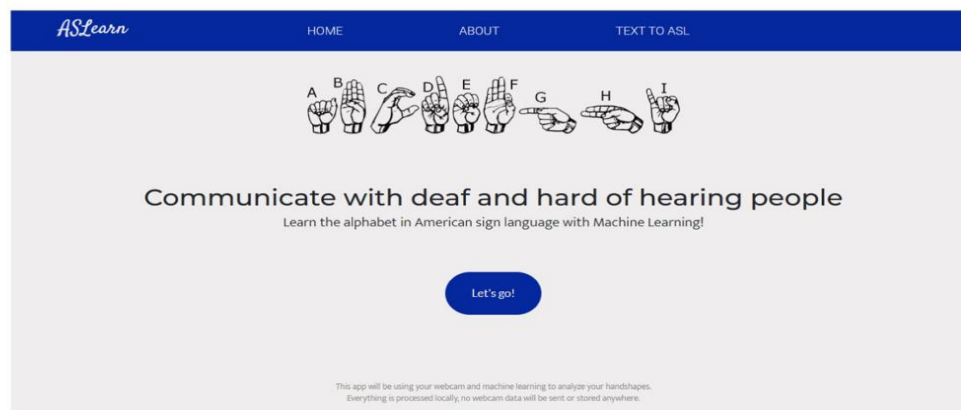


FIGURE 4
Landing page of the application.

figure 2.

6

Translation of Sign Language Finger-Spelling to Text using Image Processing, Krishna Modi, Amrita More (2013)

his is a foundational study using **classical Image Processing** techniques for ASL finger-spelling. Unlike the newer papers that use Deep Learning, this system relies on **BLOB analysis** and skin-color detection to isolate the hand. It works by capturing video frames, converting them to binary images, and comparing the "blob" features against a statistical database. It represents the "traditional" approach to computer vision, where manual feature extraction (like hand area and perimeter) is the primary method of recognition.

Link : <https://research.ijcaonline.org/volume77/number11/pxc3891313.pdf>

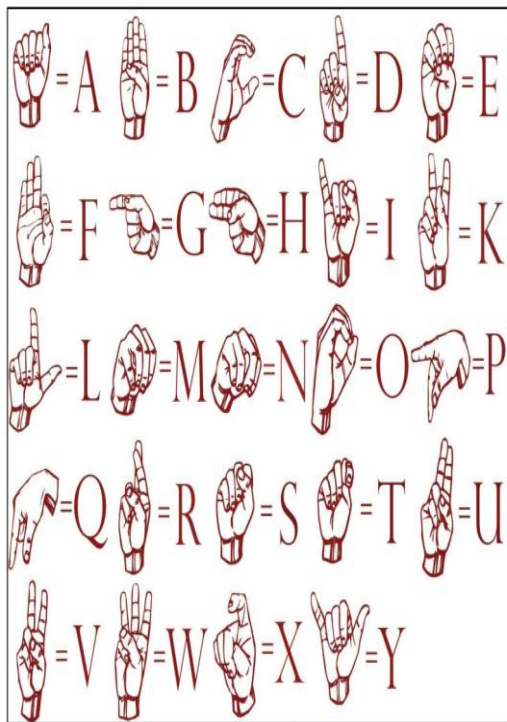


Fig 5: Collection of the American Sign Language Finger-spellings with their corresponding alphabets

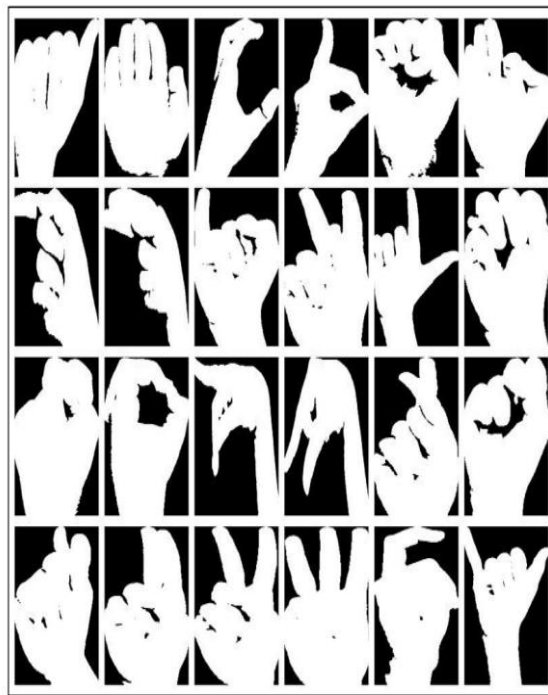


Fig 6: Collection of the images stored in the database

figure 3.

Analysis and Design

3.1 Comparison Table

Author name	Mohammed Zakariah& Yousef Ajmi	Hope Orovwode, Oduntan Ibukun	Krishna Modi & Amrita More
Dataset	Arabic Sign Language consists of 54049 images	American SL Alphabet Dataset: A collection of static images and real-time webcam captures.	American Sign Language (Custom Dataset: 26 letters (A-Z) captured via webcam.)
Algorithm & Model	FAST + SURF feature extraction + KNN classifier	Convolutional Neural Network (CNN): Integrated into a web-based delivery system.	Image Processing & BLOB Analysis (Statistical approach).
Accuracy	Training accuracy: 98%; Testing accuracy: 95%	CNN models achieved 99%	On an average, we obtained that there is 96%
Year	2022	2024	2013
Framework	Keras (running on top of TensorFlow).	Flask (Python) for the backend, with HTML, CSS, and JavaScript for the frontend.	AForge.NET framework (C#/.NET).
Methodology	1. Preprocessing: rescaling images to (32x32). 2. Data Augmentation 3. Transfer Learning: Custom-trained pre-trained ImageNet models.	Interactive Gamification: Users spell words via webcam; the system predicts signs in real-time and grants points for correct sequences.	Classical Vision: Frame capture, thresholding, and geometric feature extraction.