Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 15 Issue 01, 2026

Tribal Sign Translation System for Common People

[1]T.Binkey Mol, [2]Anishma R. G., [3]Anuja R., [4]Ashmi S. K.
[1]*Assistant Professor,Department of Artificial Intelligence and Data Science*.
[2,3,4] *Department of Artificial Intelligence and Data Science*.

Peer Review Information

Submission: 08 March 2026

Revision: 26 March 2026

Acceptance: 05 April 2026

Keywords

Communication, Social Interaction, Tribal Communities, Hand Signs / Gestures, Communication Gap, Cultural Differences, Healthcare, Education, Public Services, Miscommunication, Tribal Sign Translation System, Computer Vision, Deep Learning, Gesture Recognition, Real-Time System, Text and Voice Output, Image Preprocessing, Recognition Accuracy, Environmental Conditions, Inclusive Communication, Accessibility, Low Literacy Support, Cultural Preservation, Digital Interpretation, Human-Computer Interaction, Assistive Technology, Service Delivery, Social Inclusion, Technology Integration.

Abstract

Communication between tribal communities and the general population is often difficult due to the use of unique hand signs and gestures that are not commonly understood outside their communities. To bridge this communication gap, this project introduces a Tribal Sign Translation System that converts tribal hand gestures into easily understandable text or voice output. The system captures images using a camera and processes them with OpenCV, applying techniques such as Gaussian filtering and image normalization to improve image clarity and consistency. For gesture recognition, it utilizes MobileNetV2, a lightweight and efficient deep learning model based on Convolutional Neural Networks (CNN), which ensures accurate and fast detection of hand signs. Additionally, the system can be enhanced with real-time translation capabilities, user-friendly interfaces, and support for multiple tribal gesture datasets to improve usability and inclusivity. It also has the potential to integrate text-to-speech functionality for voice output, making communication even more accessible. Overall, this solution aims to promote better interaction, preserve tribal communication methods, and enable seamless understanding between tribal and non-tribal communities.

Objectives

The main objective of this project is to design a system that translates tribal hand signs into understandable text and voice output, enabling effective communication between tribal communities and common people. The system aims to capture and process hand gesture images using advanced computer vision techniques and implement a deep learning model, specifically MobileNetV2, for accurate and efficient hand gesture recognition. It also focuses on reducing communication barriers and supporting inclusive communication in critical areas such as healthcare, education, and public service environments. In addition, the project seeks to enable real-time gesture recognition and translation for faster interaction, while providing a user-friendly interface that can be easily used by both tribal and non-tribal users. Efforts are made to improve the accuracy and reliability of gesture detection under varying lighting and background conditions, and to develop a scalable system capable of supporting multiple tribal languages and gesture sets. The integration of text-to-speech functionality further enhances accessibility, especially for illiterate users, while minimizing the dependency on human interpreters. Moreover, the system contributes to preserving tribal communication methods through digital technology and ensures efficient performance using a lightweight model suitable for real-time applications. Overall, the project aims to promote social inclusion and enable better interaction in everyday life and essential service sectors.

Introduction

Communication plays a vital role in social interaction and access to essential services, especially in diverse communities where language and cultural differences exist. However, many tribal communities rely on unique hand signs and gestures as their primary mode of communication, which creates a significant communication gap with common people who are unfamiliar with these expressions. This gap often leads to serious challenges in essential sectors such as healthcare, education, and public services, where accurate understanding is crucial for effective service delivery. Miscommunication in such scenarios can result in delays, misunderstandings, and reduced quality of support for tribal individuals. To address this issue, the Tribal Sign Translation System is proposed, which leverages computer vision and deep learning techniques to recognize tribal hand gestures and convert them into understandable text and voice output. The system is designed to work in real time, enabling faster and more efficient interaction between tribal and non-tribal communities. It also incorporates preprocessing methods to enhance image quality and improve recognition accuracy under varying environmental conditions. Furthermore, this project aims to promote inclusive communication by making interactions more accessible and user-friendly, even for individuals with limited literacy. By digitally interpreting and preserving tribal gesture-based communication, the system contributes to safeguarding cultural identity while integrating it with modern technological advancements. Overall, the proposed solution helps bridge social gaps, enhances mutual understanding, and supports equal access to essential services for tribal populations.

Problem Statement

Tribal communities often communicate using unique hand signs and gestures that are not easily understood by common people, creating a significant gap in interaction. This lack of understanding results in communication barriers, especially in critical areas such as healthcare, education, and public services, where clear and accurate communication is essential. As a result, tribal individuals may face difficulties in expressing their needs, receiving proper services, and participating fully in society. Furthermore, existing communication tools and technologies are primarily designed for widely recognized languages and do not support tribal sign languages, which further increases the chances of misunderstandings and social exclusion. In many situations, the absence of proper communication can lead to delays in treatment, misinterpretation of information, and reduced quality of services provided to tribal communities. Additionally, reliance on human interpreters is not always feasible due to their limited availability and the time required, which further complicates the communication process. These challenges highlight the need for an efficient and automated solution that can accurately recognize tribal gestures and convert them into understandable formats. Such a system can significantly reduce communication gaps, improve accessibility to essential services, and ensure that tribal communities are better integrated into mainstream society while preserving their unique communication methods.

There is a need for an automated system that can accurately recognize tribal hand gestures

and translate them into understandable text and voice output for effective communication

Existing System

Communication between tribal communities and common people mainly depends on human interpreters or manual explanation methods, which are often time-not support the unique hand signs used by tribal communities, making them ineffective in this context.General sign language translation systems are available, but they are primarily designed for standard sign languages such as ASL and not always reliable. Traditional methods lack automation, accuracy, and real-time translation capabilities, which limits their usefulness in fast-paced environments. Additionally, many existing systems do not provide voice output support, making communication particularly difficult for illiterate users who rely more on audio-based understanding. Existing approaches are also time-consuming and prone to errors due to human involvement and lack of technological support, and they are not suitable for use in critical public service environments such as healthcare, education, and government offices. Furthermore, these systems often lack scalability and adaptability to include new gestures or support diverse tribal languages. They may also struggle to perform effectively under varying environmental conditions such as poor lighting or complex backgrounds. All these limitations emphasize the need for a more advanced, automated, and reliable system that can provide accurate, real-time translation with both text and voice output, ensuring efficient and inclusive communication.

Disadvantages of Existing system

- Requires trained interpreters.
- No support for tribal-specific gestures.
- Limited accuracy and scalability.
- Not user-friendly for healthcare and education sectors.

Proposed System

The proposed system captures tribal hand gestures in real time using a camera, ensuring continuous and immediate input for processing. The captured images are then preprocessed using OpenCV techniques such as Gaussian filtering and normalization to enhance image quality, reduce noise, and maintain consistency in lighting and contrast. This step helps in improving the clarity of important hand features required for accurate recognition. A MobileNetV2-based Convolutional Neural Network (CNN) model is employed to recognize tribal hand signs efficiently, as it is lightweight and well-suited for real-time applications while maintaining high accuracy. Once the gesture is identified, it is converted into readable text that represents the meaning of the tribal sign in a commonly understood language. The generated text is further transformed into audio output using Text-to-Speech (TTS) technology, allowing users to understand the message through voice as well. The system operates in real time, enabling quick and seamless translation of gestures, which significantly improves communication between tribal communities and common people. Additionally, the combination of text and voice outputs enhances accessibility, making the system effective for users with different literacy levels and practical usage in environments such as healthcare, education, and public services.

Advantages:

- Eliminates the need for human interpreters.
- Supports tribal-specific hand gestures, unlike existing systems.
- Provides both text and voice output for better understanding.
- Offers real-time and accurate translation.
- Easy to use in healthcare, education, and public service environments.
- Promotes inclusive and accessible communication.

System Architecture

The system captures tribal hand gestures using a camera and preprocesses the images using OpenCV for noise removal and feature enhancement.A MobileNetV2-based deep learning model performs feature extraction and hand gesture recognition with high accuracy.The recognized gesture is converted into readable text and speech output, enabling effective communication for common people.

Modules

- Image Capture
- Image Preprocessing
- Feature Extraction
- Hand Gesture Recognition
- Text Conversion

Image capture

Captures real-time images of tribal hand gestures using a digital camera or webcam.Ensures proper hand positioning and visibility for accurate gesture detection.Provides input images to the preprocessing module for further analysis.

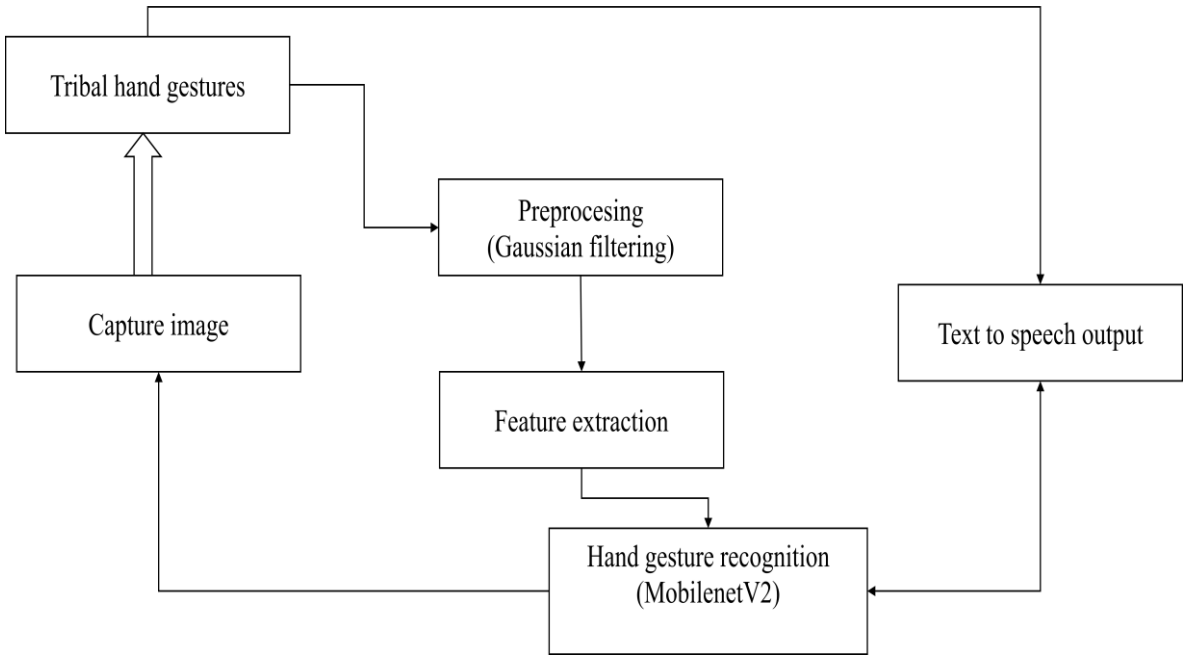


Image Preprocessing

The captured hand gesture images are processed to improve overall image quality.Gaussian filtering is applied to remove noise and smooth the image.Image normalization is performed to maintain consistent brightness and contrast.Images are resized to match the input requirements of the MobileNetV2 model.Preprocessing helps enhance important hand features and improves recognition accuracy

Gaussian Filtering

Gaussian filtering is an image smoothing technique used to reduce noise in images.It works by applying a Gaussian function–based kernel that gives more weight to nearby pixels.This process removes high-frequency noise while preserving important hand edges.

Feature Extraction

Feature extraction is used to identify important patterns from the preprocessed hand gesture images.The MobileNetV2 model extracts meaningful features such as hand shape, edges, and finger orientation.It reduces the dimensionality of image data while retaining essential gesture information.Extracted features help distinguish one tribal hand sign from another accurately.This module improves recognition performance and reduces computational complexity.

MobileNetV2

MobileNetV2 uses depthwise separable convolutions, which split standard convolution into depthwise and pointwise operations, significantly reducing computation and model size.It introduces inverted residual blocks, where the input is first expanded, then filtered, and finally compressed back to a lower dimension Linear bottlenecks are used instead of nonlinear activation at the final layer of each block to preserve important feature information.The network efficiently extracts spatial and semantic features such as edges, hand contours, and finger positions.

Convolutional Layer

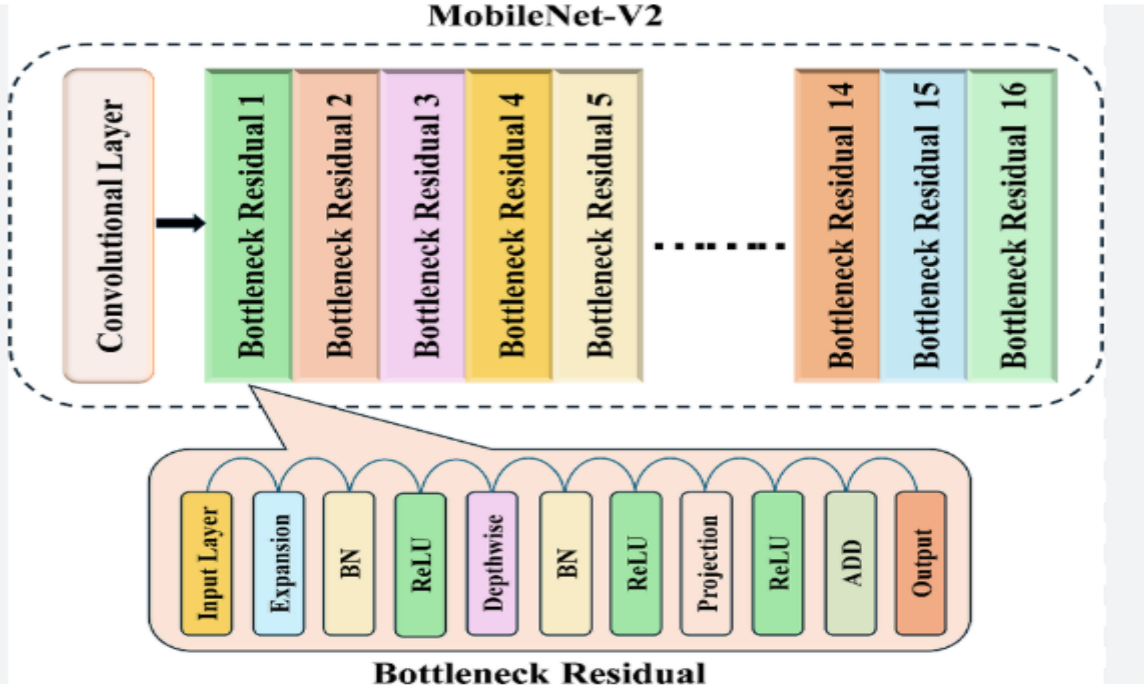
The input hand gesture image obtained after preprocessing is first passed through a standard convolutional layer.This layer extracts basic visual features such as edges, curves, and hand boundaries.It prepares the image for deeper feature extraction in bottleneck residual blocks.

Bottleneck Residual Blocks:

MobileNetV2 mainly consists of multiple Bottleneck Residual blocks.Each bottleneck block efficiently extracts hand gesture features such as:

- Finger orientation
- Hand shape
- Gesture patterns

These blocks allow the model to be deep yet computationally efficient, which is ideal for real-time gesture translation.Each Bottleneck Residual block in the architecture consists of the following stages:



a) Expansion Layer:
Expands the input feature channels to a higher dimension.Helps capture **detailed hand features** like finger joints and palm structure.

b) Batch Normalization (BN) + ReLU :
Batch Normalization stabilizes training.ReLU activation introduces non-linearity to learn complex gesture patterns.

c) Depthwise Convolution:
Performs convolution separately on each channel. Efficiently extracts spatial features of hand gestures such as contours and edges.Reduces computational cost compared to standard convolution.

d) Projection Layer:
Compresses the expanded features back to a lower dimension.Uses a linear activation (Linear Bottleneck) to preserve important gesture information.Prevents loss of subtle hand gesture details.

e) Residual Connection (ADD):
The input of the block is added to the output when dimensions match.Helps prevent vanishing gradients and improves training efficiency.Enhances gesture recognition accuracy by retaining original feature information.Hand gesture recognition: Hand gesture recognition is the core module of the Tribal Sign Translation System.The preprocessed hand gesture images are given as input to the trained MobileNetV2 model.MobileNetV2 extracts high-level features such as hand shape, finger orientation, and gesture patterns.The extracted features are passed through fully connected layers for classification.Each hand gesture is matched with a predefined tribal sign label.The recognized gesture is converted into corresponding text output.The text is further processed by the Text-to-Speech (TTS) module to generate voice output.

Text Conversion
The Text Conversion module converts the recognized hand gesture into meaningful textual information.After hand gesture recognition, the classified output label is mapped to a predefined tribal sign dictionary.Each recognized gesture corresponds to a specific word or sentence in a common language.The generated text represents the meaning of the tribal hand sign in a readable format.This module ensures accurate and understandable communication for common people.The converted text is displayed on the screen and forwarded to the Text-to-Speech module.

Text-to-Speech
The Text-to-Speech module converts the generated text into audible voice output.The text obtained from the Text Conversion module is given as input to the TTS engine.The TTS system processes the text and synthesizes natural-sounding speech.The voice output helps common people easily understand the meaning of tribal hand signs.This module enables real-

time and hands-free communication.It improves accessibility in healthcare, education, and public service environments.

Conclusion

The Tribal Sign Translation System effectively bridges the communication gap between tribal communities and common people.By using computer vision and deep learning techniques, the system accurately recognizes tribal hand gestures.MobileNetV2 provides an efficient and lightweight solution for real-time hand gesture recognition.The conversion of gestures into both text and voice output ensures better understanding and accessibility.This system can significantly improve communication in healthcare, education, and public service sectors.Overall, the project promotes inclusive communication while preserving tribal cultural expressions through technology.

Future work

The system can be extended to support a larger variety of tribal hand signs and gestures.Advanced deep learning models can be explored to improve recognition accuracy.The system can be enhanced to work under different lighting and background conditions.Multilingual text and voice output can be added for wider usability.Integration with mobile and wearable devices can enable real-time field usage.The system can be expanded to support two-way communication between tribal and common people.

References

Priya Sharma & Vikram Rao, "Voice-Based Assistive Technology for Healthcare Using Speech Recognition and NLP," Springer, 2024.

Hasan M. & Li Zhang, "A Conversational Medical Assistant Using Natural Language Understanding," IEEE Access, 2024.

Abini M.A., Divya Lakshmi P., Sharan K.S., Sulphiya V.N., "Real-Time Sign Language Recognition & Translation Using Raspberry Pi," International Journal of Computer Trends and Technology (IJCTT), 2024.

Aditya Raj Verma et al., "Enhancing Sign Language Detection with MediaPipe & CNN," Preprint arXiv, 2024.

"Hybrid CNN-LSTM with Attention for Isolated Sign Recognition," Electronics (MDPI), 2024.

Abini M. A., Divya Lakshmi P., Sharan K. S., and Sulphiya V. N.,"Deep Learning Based Real-Time Sign Language Recognition Using Raspberry Pi," International Journal of Computer Trends and Technology (IJCTT), 2024.

A. R. Verma et al., "Enhancing Sign Language Detection with MediaPipe and Convolutional Neural Networks," arXiv preprint, 2024.

Diksha Kumari; Radhey Shyam Anand", Isolated Sign Language Recognition Using CNN-LSTM with Attention Mechanism,"Electronics (MDPI), 2024.

S. Raskar, N. Rane, J. Kulkarni, and N. D. Kuchekar,"Sign Language Recognition with Transfer-Learning MobileNetV2," International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2025.

Srushti Raskar, Neha Rane, Janhavi Kulkarni & N.D. Kuchekar, "Sign Language Recognition with Transfer-Learning MobileNetV2," International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2025.