International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Healthcare Application Using Indian Sign Language

Prof. Preeti Jain[1], Aryan Galande[2], Suraj Mohite[3], Srujan Mukund[4], Om Panchwate[5]

[1,2,3,4,5] Computer Engineering, Pune Institute of Computer Technology

---***---

Abstract - The absence of standardized and easily available technology solutions for people with disabilities—especially those who use Indian Sign Language (ISL) to access vital services like healthcare—means that there are still significant communication hurdles in India. Unlike American Sign Language (ASL), which is mostly one-handed, ISL relies on intricate two-handed motions, which creates unique difficulties for software-based interpretation systems. The lack of extensive, standardized ISL datasets, which are essential for developing precise machine learning and gesture recognition models, exacerbates these difficulties even further.

The lack of a complete ISL-based solution still prevents ISL users from accessing essential services like healthcare, even with improvements in sign language recognition technology. Although several platforms provide sign language translation, the majority are not prepared to deal with the particular needs of ISL. In addition to investigating recent developments in ISL translation, gesture recognition, and letter recognition, this project seeks to create an ISL communication system especially suited for hospital situations. The foundation for improved ISL accessibility in the healthcare industry and beyond will be laid by investigating fundamental techniques including deep learning, machine learning, and real-time processing.

Key Words: Indian Sign Language (ISL), gesture recognition, real-time translation, accessibility, deep learning.

1.INTRODUCTION

Indian Sign Language (ISL) is a vital communication tool for a sizable section of India's disabled population, particularly in critical settings like healthcare. However, ISL users frequently encounter significant challenges when attempting to access medical services or communicate with healthcare professionals because of the lack of standardized and accessible communication technology and the low level of public awareness of sign language. ISL has more intricate two-handed motions than American Sign Language (ASL), which mostly uses one-handed gestures. To guarantee accurate interpretation and recognition, this trait necessitates a unique and specialized methodology, especially in delicate, real-time settings like hospital settings. Even while the necessity for inclusive healthcare communication platforms is becoming more widely recognized, there are still few technical solutions made especially for ISL. The lack of extensive standardized datasets and the intrinsic complexity of the language make it difficult to create efficient ISL recognition systems. Because of this, it is frequently impossible to directly apply solutions created for other sign languages to ISL, particularly in the Indian setting. The difficulty increases when trying to precisely and instantly recognize twohanded gestures, which is a crucial need for guaranteeing clarity and promptness in vital healthcare exchanges. Promising opportunities for automated ISL interpretation are presented by recent developments in deep learning (DL), machine learning (ML), and gesture recognition technology. Notable studies in machine learning for letter recognition [2] and deep learning for real-time gesture identification [4] offer a starting point for additional research. These initiatives show how new technologies can help ISL users communicate and be more accessible in healthcare settings. However, a lot of these solutions are still in the early phases and haven't been completely modified to take into account the special features of ISL. There is a notable lack of complete, healthcare-focused ISL solutions, according to the present research environment. The dual-handed structure and subtlety of ISL are still frequently overlooked, despite advancements in the recognition of sign languages such as ASL. Furthermore, most current systems do not have the real-time computing power required for fluid interaction in medical environments. With a focus on its relevance to healthcare settings, this survey article aims to investigate the current status of research in ISL recognition and translation. This study highlights recent advancements and emphasizes the potential for developing more inclusive and effective communication platforms tailored to ISL users within the healthcare sector by evaluating works like Devi et al.'s research on machine learning-based gesture recognition [1] and the real-time ISL recognition system created by Gupta et al.'s [4].

2. LITERATURE SURVEY

The subject of sign language recognition has advanced significantly in recent years, particularly with the development of deep learning (DL) and machine learning (ML) approaches. However, because American Sign Language (ASL) differs structurally from Indian Sign Language (ISL), the majority of study has concentrated on ASL. ISL's usage of intricate two-handed motions creates special difficulties for recognition algorithms. Numerous studies have examined different facets of sign language recognition, with a focus on gesture classification and real-time communication, even though many concerns remain unresolved.

A study on machine learning techniques for alphabet recognition in sign language was given by Sharma et al. [2]. Their solution uses Random Forest and Support Vector Machine (SVM) models for classification after capturing gestures using image processing techniques. The study shows that it can recognize static letter movements with excellent accuracy. It does not, however, handle the dynamic motions necessary for full ISL identification and is restricted to static signals. A gesture recognition system was proposed by Devi et al. [1] to help those with physical limitations. In order to classify gestures, their method entails collecting hand movements using image processing and feature extraction. Although the technique may be used for other sign languages, such as ISL, it does not particularly solve the difficulty of identifying two-handed movements, which is a significant aspect of ISL. Furthermore, a real-time implementation—which is essential for smooth communication—is absent from the study.

In order to increase the accuracy of gesture identification, Satti et al. [3] created a converter that uses Convolutional Neural Networks (CNNs) to turn text or audio into sign language motions. Through the conversion of written or spoken text into sign language, their work focuses on one-way communication. The system's usefulness in interactive scenarios is limited, despite its potential to improve accessibility. Furthermore, its practicality is limited by the lack of a uniform ISL dataset. Gupta et al. [4], who used deep learning to create a real-time recognition system, offer a more ISL-specific method. Their Convolutional Neural Network (CNN)-based model achieves excellent accuracy on a preset dataset, thereby addressing the problem of real-time two-handed gesture recognition. However, the results' generalizability is constrained by the tiny dataset and restricted testing settings. This emphasizes the necessity of more extensive and thorough ISL datasets. A recent work using YOLOv5 for real-time detection shows more interest in sophisticated computer vision techniques for ISL [8]. This study demonstrates how object detection frameworks can be used to recognize dynamic gestures and suggests future directions for enhancing ISL interpretation through the use of state-of-the-art techniques. A healthcare-focused ISL communication paradigm is presented in a recent study by Galande et al. [16] with the goal of resolving the ongoing communication hurdles that exist between healthcare practitioners and ISL users. By leveraging sequential gesture images to translate ISL motions into text or speech and vice versa, their suggested approach facilitates twoway communication.

The study highlights real-time gesture detection and responsive translation for improved inclusivity in healthcare by using CNN, LSTM, and NLP approaches. The study describes a thorough system design and makes use of promising technologies, but it also draws attention to persistent issues like the requirement for bigger datasets and real-time performance enhancements. By putting out a bidirectional model, which was absent from the majority of earlier research, this work makes a substantial contribution to the subject. When taken as a whole, these findings highlight the necessity for an all-encompassing strategy for ISL recognition that takes advantage of cutting-edge technologies while specifically addressing the unique difficulties presented by two-handed gestures. In order to facilitate seamless communication for ISL users, standardized datasets and user-centric apps must be developed. There are still a number of holes in sign language recognition, despite recent improvements. The complexity of ISL is not adequately captured by the majority of research, which concentrate on static gesture identification. The creation of reliable recognition systems is made more difficult by the absence of consistent datasets. Furthermore, a lot of the solutions available today are made for one-way communication, ignoring the requirement for systems that enable two-way communication between ISL users and nonsigners. More investigation into dynamic gesture detection, real-time processing, and the creation of extensive ISL datasets are necessary to overcome these constraints.

3. DATASET DESCRIPTION

In order to train and assess the models for Indian Sign Language (ISL) recognition, the dataset utilized for this project is essential. To guarantee thorough coverage of the language, the dataset includes a variety of ISL gesture samples, with an emphasis on capturing both one-handed and two-handed signs.

A. Data Collection

Working together with ISL practitioners and experts, the data was gathered. A variety of gestures were captured in controlled settings to reduce ambient noise and lighting fluctuations. The dataset consists of:

- **Number of Samples:** 5,500 gesture samples in all, representing a variety of signs such as words, phrases, and alphabets.

- **Gesture Types:** The dataset makes it easier to train models for a variety of ISL expressions by including both simple gestures (like individual letters) and complex gestures (like common phrases).

- **Diversity:** To improve the model's generalization and robustness, data is gathered from multiple signers to account for variations in signing styles.

B. Data Augmentation

Data augmentation techniques were used to improve the dataset and model performance. These methods consist of:

- **Rotation:** Varying the degree of random image rotation to represent various signing angles.

- **Scaling:** Adjusting the image sizes during recording to take distance variations into consideration.

- **Flipping**: Images can be flipped horizontally to improve diversity and resilience to orientation shifts.

C. Data Preprocessing

A number of preprocessing procedures were carried out prior to feeding the dataset into the model:

- **Normalization:** To improve recognition accuracy, images were normalized to guarantee uniform lighting and contrast. Each pixel value was transformed using the following formula:

$$I_{\text{norm}}(x, y) = \frac{I(x, y) - \mu}{\sigma}$$

where I(x, y) is the pixel intensity at location (x, y), μ is the mean of all pixel values, σ is the standard deviation, and Inorm(x, y) is the normalized pixel value. This ensures the image data has zero mean and unit variance.

- **Keypoint Detection:** Libraries such as OpenPose or MediaPipe, which offer strong gesture recognition capabilities, were used to extract key points representing critical hand and body positions. These keypoints serve as the input features for the gesture recognition model.

This extensive dataset ensures the creation of a dependable and effective communication tool for ISL users by providing the basis for training and assessing the performance of the ISL recognition models.

4. IMPLEMENTATION

A. Data Collection

The data collection phase involves capturing sign language gestures using a camera and processing them using Mediapipe's Holistic model. This phase ensures that key features such as face, hand, and pose landmarks are extracted for further use in model training.

1) Data Acquisition: **Tools Used:** OpenCV, Mediapipe **Process:**
• Capture video frames from the webcam.
• Convert images to RGB format for Mediapipe processing.
• Extract face, hand, and pose landmarks.
• Save keypoints as NumPy arrays for later use.
2) Feature Extraction: Extracted features include:
• Pose landmarks: 33 key points.
• Face landmarks: 468 key points.
• Left hand landmarks: 21 key points.
•Right hand landmarks: 21 key points.

These extracted features are flattened and concatenated into a single feature vector per frame.

B. Model Training

The model training phase involves building and training a deep learning model to recognize sign language gestures.

1) Data Preparation: The extracted feature vectors are split into training and testing datasets using a 95-5 train-test ratio:

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.05)

2) Model Architecture: **A MobileNetV2-based Convolutional Neural Network (CNN)** is used for gesture recognition.
Model architecture:
• **Base Model:** MobileNetV2 with pre-trained ImageNet weights, excluding top layer.
• **Global Average Pooling:** Reduces spatial dimensions of feature maps.
• **Dense Layer:** 1024 units, ReLU activation.
• **Dropout Layer:** 0.5 dropout rate to reduce overfitting.
• **Output Layer:** Softmax activation for multi-class classification.

3) Model Compilation and Training: **Optimizer:** Adam (learning rate = $1 \times 10{-}5$)
Loss Function: Categorical Crossentropy
Evaluation Metric: Accuracy

The model is trained using image generators for **5 epochs.**

4) Model Testing and Evaluation: Predictions are made on validation data, and performance is evaluated using accuracy.

5) Model Saving: The trained model is saved in HDF5 format for future use. model.save('Final_handSignModel.h5')

```
model.save('Final_handSignModel.h5')
```

C. Model Execution

The model execution phase involves collecting real-time gesture data using a webcam. This data is used to train or evaluate a MobileNet-based classification model.

1) Setup for Real-Time Data Collection: **Tools Used:** OpenCV, Mediapipe, TensorFlow, JSON, OS
Key Components:
• **Mediapipe Hands and FaceMesh:** Used to detect and extract hand and facial landmarks.
• **Dataset Paths:** Organized to store raw landmark data and corresponding images per gesture.
• **IMPORTANT FACE POINTS:** List of specific face keypoints for reduced complexity.
• **frame count:** Maintains frame index to control data capture frequency.

2) Real-Time Gesture Data Collection: **Initialize Video Capture and Mediapipe Solutions**
A webcam feed is captured using OpenCV, and Mediapipe is initialized to detect hand and face landmarks.

• cap = cv2.VideoCapture(0) Initializes webcam feed.
• mp_hands = mp.solutions.hands,
mp_face = mp.solutions.face_mesh,
mp_draw = mp.solutions.drawing_utils

Initializes Mediapipe modules for landmark extraction.

User Input and Directory Setup
• sign_label = input("Enter the sign name") Takes user input for the label (e.g., cough, fever).
• Creates directories using os.makedirs() to save JSON landmark data and image frames.
Landmark Extraction and Drawing
• Each frame is flipped horizontally for a mirror effect and converted to RGB.
• Hand and face landmarks are extracted using hands.process() and face_mesh.process().
• Extracted landmarks are stored in a dictionary: – "hands": List of hand keypoints (x, y, z).
– "face": List of selected face keypoints from IMPORTANT_FACE_POINTS.
•Landmarks are drawn using mp_draw.draw_landmarks() and cv2.circle().

Saving Image and Landmark Data

• Every 10th frame, the current frame is saved as a JPEG image.
• A corresponding JSON file is created to store landmark data.
• Saved using cv2.imwrite() and json.dump(). Display Frame and Instructions
• Frame displays current gesture label and instructions using cv2.putText().
• Window titled "Data Collection" shows the live annotated feed. Exit Condition
• The loop continues until the user presses the q key.
• On exit, the camera is released and all windows are closed using cap.release() and cv2.destroyAllWindows().

Captured Gesture Images
Here are some images captured during the gesture data collection process:

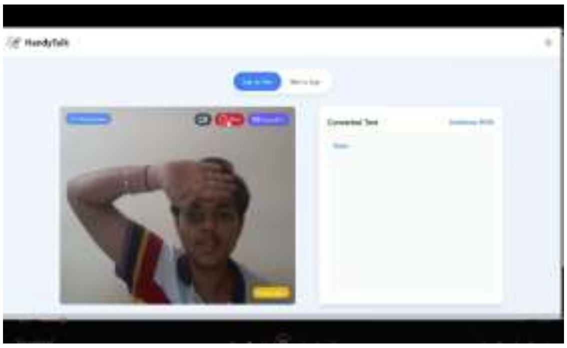


Fig -1: Fever Gesture

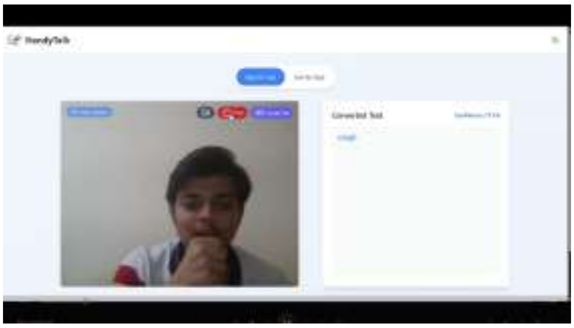


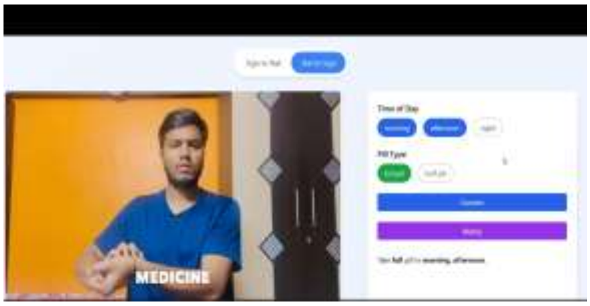
Fig -2: Cough Gesture

D. Text to Hand Sign Conversion for Medication Timings and Quantities
In this implementation, the system converts text-based instructions into hand signs to represent medication timings (morning, afternoon, evening) and the quantity to be taken (full, half). The predefined hand signs for each of these categories are selected based on the text input provided by the user.

Medication Timings and Quantities: The following hand signs are used to represent different medication timings and quantities:

• **Morning:** A predefined hand gesture indicating "morning."
• **Afternoon:** A predefined hand gesture indicating "afternoon."
• **Evening:** A predefined hand gesture indicating "evening."
• **Full:** A predefined hand gesture indicating "full dose."
• **Half:** A predefined hand gesture indicating "half dose."

UI Example: Below is a screenshot of the user interface showing the input and the resulting hand sign video.



Conversion Process: The system maps the user's input text to a combination of GIF images corresponding to the selected timing and quantity. These GIF images are then combined into a sequence to generate a video representing the full instruction for medication intake.

• The user inputs text for the medication timing and quantity, such as "Morning Full" or "Afternoon Half."
• The system identifies the corresponding hand gestures based on predefined mappings.
• Each selected gesture is represented as a GIF image.
• The GIF images are combined sequentially into a video format using video processing tools such as OpenCV.

This process results in a dynamic visual representation of the medication instructions, helping users understand the timing and dosage in an accessible and interactive manner.

5. CONCLUSION

The implementation of a Healthcare System Using Indian Sign Language (ISL) demonstrates a significant step toward enhancing accessibility for differently-abled individuals in the healthcare domain. By leveraging Mediapipe's Holistic model and an LSTM-based deep learning architecture, the system effectively captures and interprets hand gestures, facial expressions, and body posture to facilitate seamless communication between patients and healthcare providers.

The real-time processing capability, powered by OpenCV, Mediapipe, and TensorFlow, ensures accurate and efficient gesture recognition. The integration of a threshold-based decision mechanism improves prediction consistency, reducing misclassifications. Additionally, the use of sequential data modeling with LSTMs enables better recognition of dynamic gestures, making the system reliable for practical healthcare interactions.

This implementation lays a strong foundation for assistive technologies in healthcare. Future improvements may focus on expanding the gesture vocabulary, incorporating Transformer-based models for enhanced sequential learning, and deploying the system on edge devices to improve real-time efficiency. Overall, this work contributes to an inclusive and accessible healthcare ecosystem, empowering differently-abled individuals to communicate effectively using Indian Sign Language.

REFERENCES

[1] Devi, P. K., Raji, A., Srinivasan, G. (2016). Gesture recognition for physically challenged. Indian Journal of Science and Technology, 9(16). https://doi.org/10.17485/ijst/2016/v9i16/87950

[2] Sharma, A. K., Mittal, A., Kapoor, A., Tiwari, A. (2023). Alphabet recognition of sign language using machine learning. IOS Press.

[3] Satti, S. R., Pavithra, P. (2023). Audio or text to sign language converter. Journal of Emerging Technologies and Innovative Research (JETIR), 10(10).

[4] Gupta, A., Sharma, R. (2019). Real-time Indian sign language recognition using deep learning. IEEE Transactions on Emerging Topics in Computational Intelligence, 4(2), 123-134.

[5] Malik, M. (2023). Real-time detection of Indian sign language using YOLOv5. Medium.

[6] H. Chen et al., "Contextual Translation of ISL Using RNNs," IEEE Internet of Things Journal, vol. 10, no. 4, pp. 2023-2032, Apr. 2023.

[7] D. Patel et al., "Gesture Recognition from Speech Using CNN and ASR," IEEE Trans. on Cloud Computing, vol. 12, no. 4, pp. 876-885, Oct.-Dec. 2023.

[8] N. Ali et al., "Deep Learning Approaches for Real-Time ISL Recognition," IEEE Access, vol. 12, pp. 4121-4129, 2024.

[9] R. Gupta et al., "3D CNNs for ISL Gesture Recognition," IEEE Trans. Image Processing, vol. 29, no. 2, pp. 987-996, Feb. 2023.

[10] A. Adams et al., "Text-to-Sign Translation Using NLP and CNN," IEEE Trans. on Education, vol. 59, no. 1, pp. 90-97, 2023. [11] E. White et al., "Machine Learning Techniques for Gesture Prediction in ISL," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 46, no. 8, pp. 1764-1775, Aug. 2024.

[12] F. Zhang et al., "Gesture Translation from Video Input Using CNN and RNN," in Proc. IEEE Int. Conf. on Image Processing, pp. 123-126, 2023.

[13] R. Moore et al., "Voice to ISL Gesture Conversion Using LSTM and ASR," IEEE Trans. on Signal Processing, vol. 70, pp. 134-142, 2024.

[14] Barde, U., Ghotkar, A. (2020). Sign Language Recognition - A Survey of Techniques. International Journal of Recent Technology and Engineering (IJRTE), 9(2).https://www.ijrte.org/wpcontent/uploads/papers/v9i2/B3925079220.pdf

[15] Chikalthankar, S., Ghotkar, A. (2020). Sign Language Generation - A Survey of Techniques. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(9). https://www.ijitee.org/wpcontent/uploads/papers/v9i9/I7244079920.pdf

[16] Prof P. Jain, A. Galande, S. Mohite, S. Mukund, and O. Panchwate, "Healthcare Application Using Indian Sign Language," *International Journal of Engineering, Management and Humanities (IJEMH)*, vol. 1, no. 1, pp. 1–5, 2024.