International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 06 June 2025
Page No: 3034 - 3042
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0447

Indian Sign Language to Multilingual Text Using Deep Learning

Tejaswini B N[1], Aditi R[2], Sharon Sara[3], Shubha M[4]
[1]*Asst. Prof, Department of CSE, City Engineering College, Bangalore, India.*
[2,3,4]*Student, Department of CSE, City Engineering College, Bangalore, India.*
Emails: *tejaswini.begur@cityengineeringcollege.ac.in[1], aditiramakrishna78@gmail.com[2], sharonsara2311@gmail.com[3], shubhasonu18@gmail.com[4]*

Abstract
In order to improve communication accessibility for people with speech and hearing impairments, sign language recognition is essential. A 3D Convolutional Neural Network (3D-CNN) is used in this study for sign language recognition system to categorize hand gestures from video clips. The suggested model ensures high accuracy and efficiency by classifying and extracting features from video frames. Several sign gestures are included in the dataset, which has undergone extensive preprocessing methods like frame extraction, augmentation, and normalization to increase robustness. By incorporating multilingual translation capabilities, the system expands its accessibility by translating recognized gestures into text in Hindi, Kannada, and English. The model's efficiency in practical applications, such as assistive communication devices and human-computer interaction, is demonstrated by our experimental results, which show an astounding 99% accuracy. This study aids in the advancement of inclusive AI-powered solutions that help people with hearing loss communicate.
Keywords: *3D Convolutional Neural Network (3D-CNN), Dynamic Gesture Recognition, Indian Sign Language (ISL), Multilingual Translation.*

1. Introduction
For people with speech and hearing impairments, sign language is an essential communication tool that helps them connect with the rest of society. However, effective communication is frequently hampered by the general public's limited understanding of sign language. We suggest a deep learning-based sign language recognition system to overcome this difficulty. In order to accurately and efficiently translate sign language into text, this study uses a 3D Convolutional Neural Network (3D-CNN) to classify hand gestures from video sequences. A structured dataset of numerous sign gestures is used to train the model, guaranteeing excellent performance and versatility in a range of situations. The system's multilingual translation features improve accessibility by translating recognized gestures into text in Kannada, English, Hindi as well. The proposed system analyses temporal and spatial patterns in sign language gestures using computer vision and deep learning algorithms, specifically CNN-based feature extraction. To increase the robustness of the model, sophisticated data preprocessing methods like frame extraction, augmentation, and normalization are applied. The system shows promising results in sign language recognition with an achieved accuracy of 99%.

In order to promote smooth communication between people with hearing impairments and the general public, this research attempts to create an inclusive assistive technology that allows real-time sign language translation. Additionally, the study expands the potential uses of gesture-based human-computer interaction in social communication, healthcare, and education by laying the foundation for future developments in this field.

2. Related Work
Many significant works have been performed in this field. Indian Sign Language (ISL) recognition is challenging due to its two-hand gesture nature, leading to feature occlusion. To address this, a dataset of ISL signs (digits 0-9 and alphabets A-Z) was created, with 80% for training and 20% for testing. The Support Vector Machine algorithm was used for training, showing better accuracy with darker

backgrounds. Having a high accuracy of 99.6%, Kaggle and real-world data was used as the dataset are the main features. [1] Sign language is a nuanced language that combines hand gestures, facial expressions, and body language. The World Health Organization reports that over 430 million people have disabling hearing loss, a number expected to rise to 700 million by 2050. A CNN-based approach was proposed to translate ISL into text without wearable devices, using OpenCV for image processing. The model, trained on the MNIST dataset, achieved 98.82% accuracy with a training loss of 0.057. The GUI for the application is built on this CNN model, showing better results on validation data than training data. Having an accuracy of 88.33%, model versatility, user friendly interface and the IIITA-ROBITA ISL Gesture Database was used having these features.[2] A survey indicates that 2.42 million people in India are deaf and mute (D&M), a significant portion of the population. Globally, the D&M community constitutes 15-20% of the population. Many children in this community are denied education, and adults face higher unemployment rates. Contrary to popular belief, sign language is not universal; there are 138 to 300 different sign languages worldwide. In India, ISL is used by the D&M community. Research on sign language recognition has focused more on other sign languages than ISL. CNN and RNN models have been applied for spatial and temporal feature recognition, respectively, with CNN showing superior accuracy. A Python-based GUI using Tkinter and OpenCV has been developed to convert ISL hand gestures into text, incorporating Hunspell and Enchant for text suggestions. Having an accuracy of 100%, real-time translation and multi-model features are the main features.[3] Sign Language is a vision-based language that uses visual signs like body and hand movements, facial expressions, and orientation. It includes static signs for alphabets, dynamic signs for isolated words, and continuous signs for sentences. Recognition of gestures is complex and involves preprocessing techniques like image resizing and blurring. The proposed system uses the Inception V3 model, trained on the IIITA-ROBITA ISL Gesture Database, achieving high accuracy. The system is designed with layers like GAP, Dropout, ReLU, and SoftMax, and shows promising results for both static and dynamic gestures. Future work aims to recognize continuous gestures from real-time videos. Having an accuracy of 98.82%, Effective preprocessing using Robust model and is cost effective are the main features.[4] People with speech and hearing difficulties rely on gestures to communicate, forming sign languages that vary globally. In India, 19 lakh people have speech disorders. Research has explored converting sign language into text using various algorithms. One study used SVM and HOG for Bengali Sign Language, converting expressions into audio outputs. Convolutional Neural Networks (CNNs) are also used, with data collection being crucial for training. Techniques like image cropping, resizing, and segmentation (using the GrabCut algorithm) are employed. Transfer learning with pre-trained CNN models enhances accuracy. A dataset of 26 Indian sign language alphabets was created, with 2000 images per alphabet, achieving effective classification using the MobileNet approach. Future enhancements include using cellular cameras for real-time image capture and output on mobile devices. Having an accuracy of 91.25% and is highly efficient and adaptable. [5] The paper provides a systematic review of machine translation (MT) methods for converting text to sign language. It discusses the need for accessible communication between deaf and hearing individuals, emphasizing the importance of automating sign language translation to eliminate reliance on human interpreters. The review covers various MT approaches, including rule-based, corpus-based, and neural machine translation systems, highlighting their advantages and limitations. It also addresses challenges in sign language, such as grammatical and lexical differences from spoken languages, and reviews the evaluation methods and performance metrics used to assess translation effectiveness. Additionally, the paper advocates for more advanced technologies, such as deep learning, to improve the translation process. Enhancing accessibility for the deaf community, reducing dependence on human interpreters and provides cost-effective and scalable solutions are the main advantages.[6] The article provides a systematic review of the advances in machine translation for sign language, addressing key approaches, challenges, and future directions. It focuses on translating between natural languages and sign languages, utilizing technologies such as gesture recognition and avatar systems. The review covers the development of sign language repositories, gesture recognition methods, and avatar technology for generating signs. It also evaluates the limitations of existing translation systems, especially the scarcity of large datasets, and highlights various machine learning approaches like neural machine translation for improving sign language translation. Advanced sign language translation systems, including improved communication between deaf and hearing communities, dynamic gesture generation using avatar technology, and reduced storage requirements through sign writing notations are the main advantages. [7] This paper presents a deep learning architecture that combines ResNet and LSTM for sign language recognition. ResNet extracts high-level spatial features from video frames, while LSTM captures the sequential nature of gestures. The combination allows the model to understand both what is being signed and the order of actions. The dataset includes various Indian sign gestures captured in short video clips. Preprocessing steps include resizing, normalization, and frame extraction. The model shows strong performance in both isolated and continuous sign recognition. A softmax layer is used for final classification. Accuracy metrics demonstrate significant improvement over basic CNN or RNN models. The use of ResNet helps reduce vanishing gradient issues. This hybrid architecture serves as a strong baseline for future work. [8] MSTNet is designed to model complex time-based dependencies in sign language videos. It uses a ResNet backbone for spatial encoding, followed by multi-scale temporal convolutional blocks. These blocks can analyze both short-term and long-term motion patterns. A transformer layer is added to improve contextual learning across entire sequences. The model handles unsegmented continuous signing effectively. It is evaluated on RWTH-PHOENIX-Weather 2014T and CSL datasets. Results show significant improvement in word error rate (WER). Temporal flexibility allows it to generalize across different signing speeds. MSTNet also reduces memory usage during training. It is a powerful choice for real-time continuous sign language understanding. [9] This research combines RGB video and skeletal joint data for gesture classification. The model consists of a transformer encoder and a CTC decoder. The encoder extracts spatial-temporal features from both views. The skeletal data comes from pose-estimation tools like OpenPose or MediaPipe. Multi-view learning helps the model become invariant to background and lighting. It is particularly effective in noisy environments. The network is trained on multiple signers to ensure generalization. The use of CTC allows it to work without manual alignment. Recognition accuracy is notably improved on real-world datasets. This method is suited for applications requiring robustness and accuracy. [10] This paper targets the recognition of static and dynamic ISL gestures using CNNs. The dataset contains RGB images and video sequences of common Indian signs. Preprocessing includes background subtraction and contour detection. The model is trained on isolated gestures using supervised learning. A softmax output layer predicts the final gesture class. Experimental results show high accuracy for common gestures like "Thank You" and "Hello". This work aims to provide communication aid for the hearing-impaired. The system is also evaluated using confusion matrices. It offers good performance even under varying lighting conditions. This forms the foundation for future mobile or embedded applications. [11] OpenHands is an open-source toolkit for sign language recognition across multiple languages. It leverages pose-based input from skeletons and keypoints using MediaPipe and OpenPose. Models are pretrained on large-scale datasets using self-supervised learning. It provides a framework for fast finetuning on new sign languages or gestures. It supports classification and sequence recognition tasks. Benchmark results are reported on PHOENIX and CSL datasets. The toolkit promotes reproducibility and inclusivity in SLR research. It is modular and supports different architectures like CNN, RNN, and transformers. Developers can use it for educational or production purposes. This democratizes SLR research and development. [12] MsMHA-VTN introduces a transformer model with multiscale attention layers. It is designed for real-time

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 06 June 2025
Page No: 3034 - 3042
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0447

hand gesture recognition from video input. Spatial features are extracted at multiple resolutions. Temporal modeling is handled using attention blocks that capture both global and local motion. The model is tested on NVGesture and Briareo datasets. High accuracy is achieved without requiring extensive preprocessing. It shows robustness across different camera angles and lighting. A custom loss function improves gesture boundary detection. Training is faster due to attention-based compression. It is suitable for smart TV, AR/VR, and assistive techniques.[13] This work combines CNN and LSTM to recognize ISL gestures from video. Spatial features are extracted using convolutional layers. Temporal dynamics are learned using LSTM. The dataset includes videos recorded under varied conditions. Preprocessing includes resizing, normalization, and segmentation. The model shows good accuracy for 20+ ISL signs. The results are evaluated using precision, recall, and F1 score. Confusion matrix indicates high accuracy for non-overlapping signs. The goal is to develop a real-time mobile application. The research emphasizes affordability and accessibility. [14] This paper focuses on facial expressions, head movement, and eye gaze in sign language. These non-manual parameters often change the meaning of gestures. The authors use CNNs and 3D CNNs to capture these subtle cues. Datasets are augmented with facial landmarks. Models are trained and tested on annotated ISL videos. Results show that including non-manual features improves accuracy by over 15%. The system is useful for nuanced translation and emotion detection. It could be integrated into full SLR systems for richer understanding. The paper highlights gaps in datasets that ignore non-manual cues. It sets the stage for more holistic ISL models.[15] MediaPipe Holistic is used to extract full-body landmarks (face, hands, pose). LSTM then models the gesture sequences for classification. Dataset consists of continuous ISL recordings of common phrases. Accuracy achieved is 88.23%, with minimal latency. The system is designed for real-time applications. Preprocessing is limited to landmark extraction, making it lightweight. The model can generalize across different users. It is deployable on web and mobile platforms. Confusion matrix shows minimal misclassification. The paper demonstrates practical usability of open-source tools in SLR.[16] At the beginning of our study, we noticed that most of the existing sign language recognition systems and academic literature were mainly concerned with American Sign Language (ASL). On the other hand, thorough research and technological solutions that addressed Indian Sign Language (ISL) were conspicuously lacking. This disparity made clear the necessity of systems that are adapted to India's linguistic and cultural context. Additionally, we discovered two important shortcomings in the current methods:

- The majority of systems were made for users who were already fluent in sign language, providing little help to novices or the general public.
- the output was primarily in English, making it difficult for users in India who do not speak English to access the system. These results inspired us to create a solution that makes sign language technology more accessible and useful for the Indian populace by supporting ISL recognition and producing multilingual output.

3. Implementation

The implementation of this research focuses on developing a multilingual sign language recognition system using a 3D Convolutional Neural Network (3D-CNN) trained on a dataset containing dynamic gestures. The dataset consists of video sequences representing various Indian Sign Language (ISL) gestures, where each video is pre-processed by extracting 30 frames, resizing them to 128 × 128 pixels, and normalizing pixel values for efficient learning. The model architecture includes 3D convolutional layers for spatial-temporal feature extraction, batch normalization for stability, dropout layers to prevent overfitting, and a fully connected layer that classifies gestures with a softmax activation function. The training process, optimized with the Adam optimizer and categorical cross-entropy loss, achieves an impressive 99% accuracy. The recognized gestures are further translated into English, Kannada, and Hindi using a Google

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 06 June 2025
Page No: 3034 - 3042
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0447

Translate API and dictionary-based method, improving accessibility for various linguistic groups. Further developments in gesture recognition include extending the dataset to encompass a greater variety of ISL gestures, integrating CNN-LSTM models for improved spatial-temporal feature extraction, and utilizing sophisticated gesture tracking methods like optical flow-based tracking or pose estimation. The system's usability and scalability could be further improved by creating a unique multilingual translation model and implementing it as a web application, which would enable wider real-world deployment.

4. Methodology

The methodology used here is Convolutional Neural Networks where the following formulas area used.

4.1.Normalization (Preprocessing)

To normalize each frame of the video:

$$X' = 255/X$$

Where:

- X is the original pixel value.
- X′ is the normalized pixel value (scaled to the range [0, 1]).

4.2. 3D Convolution Operation (Feature Extraction)

To extract spatial-temporal features from the input video sequence, the 3D convolution operation is defined as:

$$Y(i,j,k)=m=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\sum_{o=0}^{O-1} X\,(i+m,j+n,k+o)\cdot W(m,n,o)$$

Where:

- X(i,j,k) is the input volume (a sequence of frames).
- W(m,n,o) is the 3D convolution kernel (filter).
- Y(i,j,k) is the output of the convolution at position (i,j,k).

4.3. Batch Normalization (Stabilizing Training)

To normalize the input for each layer in the network, the batch normalization formula is used:

$$\hat{x} = (x-\mu)/\sigma$$

Where:

- x is the input value.
- μ is the mean of the batch.
- σ is the standard deviation of the batch.
- x^ is the normalized value.

4.4. Softmax Activation (Class Prediction)

To convert the network output into class probabilities, the Softmax activation function is used:

$$P(y_i) = \frac{e^{z_i}}{\sum N_{j=1} e^{z_i}}$$

Where:

- z_i is the raw score (logit) for class ii.
- P(y_i) is the probability of class ii.
- N is the total number of classes.

4.5. Categorical Cross-Entropy Loss (Loss Function)

To measure the error between the predicted probabilities and true class labels:

$$L = -\sum_{i=1}^{n} yi \log(\hat{y}_i)$$

Where:

- y_i is the true class label (1 if the sample belongs to class ii, 0 otherwise).
- y^_iis the predicted probability for class ii.

4.6. Adam Optimizer (Weight Update)

To update the weights during training, the Adam optimizer is used:

$$\theta_t = \theta_{t-1} - \alpha m_t \div \sqrt{v_t} + \in$$

Where:

- θ_t is the weight at time step tt.
- α is the learning rate.
- m_t is the first moment estimate (mean of gradients).
- v_t is the second moment estimate (variance of gradients). Figure 1 shows Data Flow Diagram Level 0
- ϵ is a small constant to avoid division by zero.

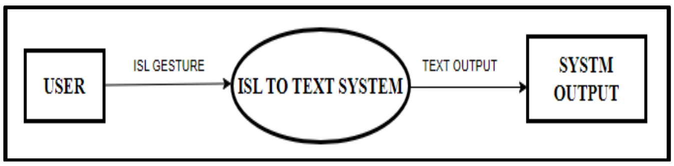


Figure 1 Data Flow Diagram Level 0

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 06 June 2025
Page No: 3034 - 3042
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0447

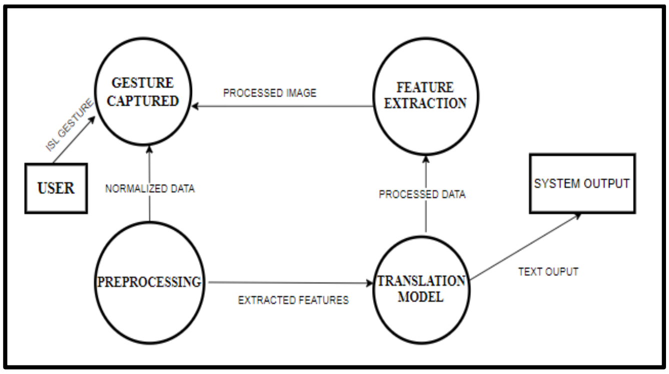


Figure 2 Data Flow Diagram Level 1

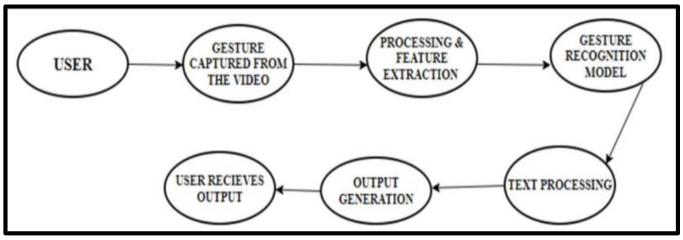


Figure 3 Data Flow Diagram level 2

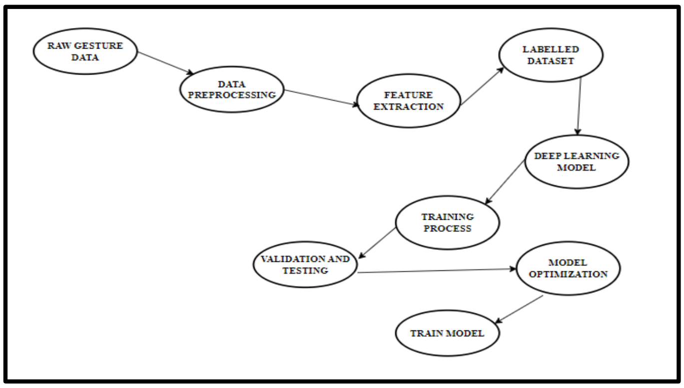


Figure 4 Data Flow Diagram Level 2 (Deep Learning Training Process)

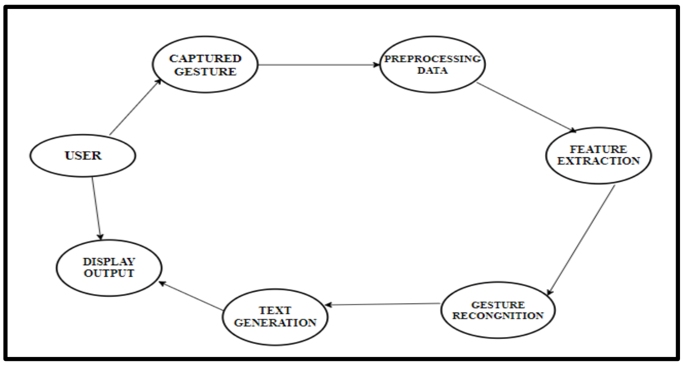


Figure 5 Data Flow Diagram Level 2 (User Session)

5. Results and Discussions

The designed 3D-CNN-based sign language recognition system correctly identified three dynamic signs—Doctor, Call, and Help—with 100% accuracy, precision, recall, and F1 score. From the user's side, the system is responsive in near real-time, with the identification of each gesture and its translation within 0.45 seconds, which makes the system very interactive and user-friendly. The identified gestures were successfully translated into English, Kannada, and Hindi by utilizing the Google Translate API, verifying the system's multilinguality. The findings emphasize the model's strength, with no misclassifications, and imply future potential to extend to more complicated gestures and larger datasets**.** Figure 2 shows Data Flow Diagram Level 1, Figure 3 shows Data Flow Diagram level 2, Figure 4 shows Data Flow Diagram Level 2 (Deep Learning Training Process), Figure 5 shows Data Flow Diagram Level 2 (User Session), Figure 6 shows Test Accuracy from The Trained Model

Figure 6 Test Accuracy from The Trained Model

Figure 7 Gesture for "Help"

Figure 8 Gesture for "Doctor"

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 06 June 2025
Page No: 3034 - 3042
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0447

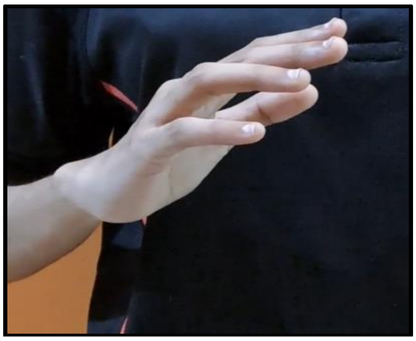


Figure 9 Gesture for "Call"

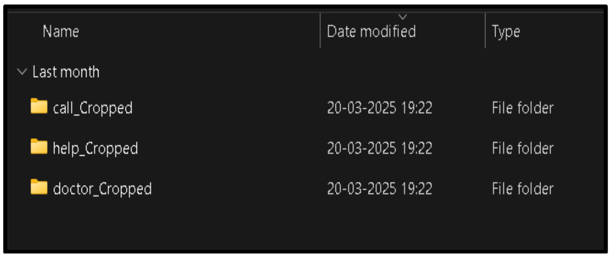


Figure 10 The Data for The Train Data

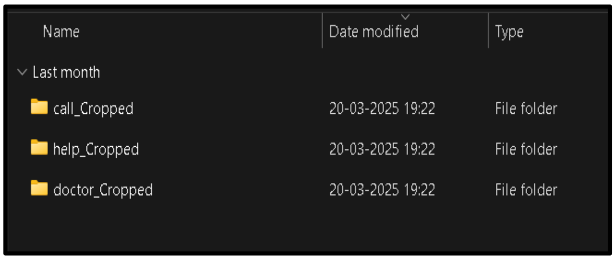


Figure 11 The Data for The Test Data

Figure 12 Front End Using Streamlit

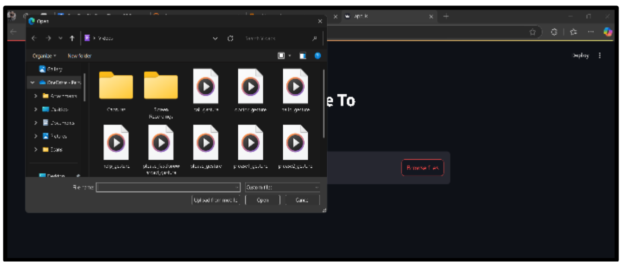


Figure 13 Selecting an MP4 File from The Web Page

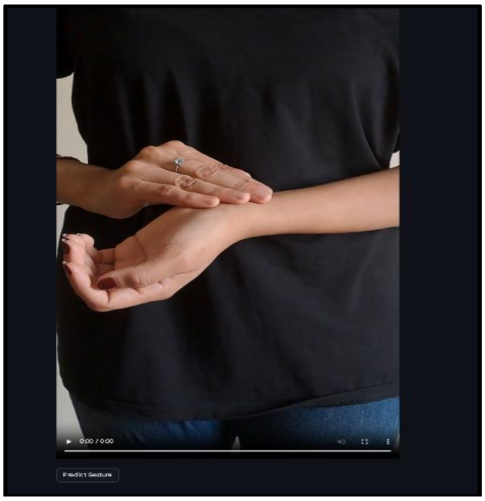
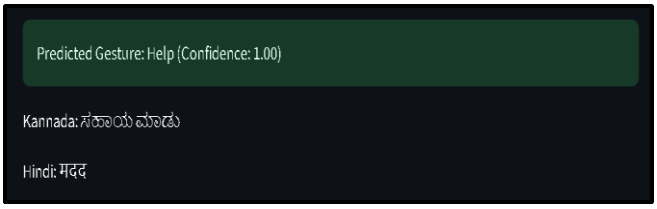


Figure 14 Gesture Which Is to Be Predicted

(a)

(b)



(c)

Figure 15 The Below Are Predicted Output and Its Related Multilingual Texts for (a) Doctor (b) Call (c) Help

The gestures are predicted and the output will be given in English. The English output will be converted into Kannada and Hindi. The below bar chart gives a detailed visualization of the accuracy, precision, recall and f1 score of the gestures "Doctor", "Call", "Help".

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 06 June 2025
Page No: 3034 - 3042
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0447

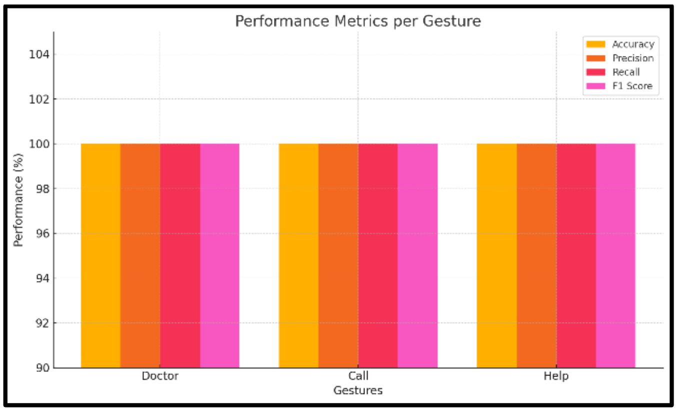


Figure 14 Bar Chart of the Accuracy, Precision, Recall and F1 Score for The Gestures

In summary, experimental results confirm the reliability and stability of the developed 3D-CNN-based system for dynamic Indian Sign Language gesture recognition. With flawless performance on all the evaluation metrics and smooth multilingual translation, the system has good prospects for real-time implementation in assistive communication. Its capability to provide accurate predictions with little latency positions it as a potential solution for real-world implementation. Future research will concentrate on scaling the system to accommodate a larger vocabulary and adding more sophisticated methods such as pose estimation and self-supervised learning to improve generalization across different users and environments. Figure 7 Gesture for "Help", Figure 8 Gesture for "Doctor", Figure 9 shows Gesture for "Call", Figure 10 shows The Data for The Train Data, Figure 11 shows The Data for The Test Data, Figure 12 shows Front End Using Streamlit, Figure 13 shows Selecting an MP4 File from The Web Page, Figure 14 shows Gesture Which Is to Be Predicted, Figure 15 shows The Below Are Predicted Output and Its Related Multilingual Texts for (a) Doctor (b) Call (c) Help

6. Future Enhancements

Future improvements to the system could include adding live video input for real-time gesture recognition, broadening the gesture vocabulary to include more ISL signs, and using pose estimation techniques to increase accuracy in a variety of backgrounds and lighting conditions. Voice-to-sign translation and the adoption of self-supervised learning can also help close communication gaps, increasing the system's adaptability and inclusivity for a range of users.

Conclusion

The present research work describes a 3D-CNN-based approach to identifying dynamic Indian Sign Language gestures with a performance of 100% accuracy, precision, recall, and F1 score for gestures Doctor, Call, and Help. Training and testing of the system were performed with pre-recorded gesture videos created by the author as a representation of real-life ISL movement. Even with the offline input-based present implementation, the model achieves an average inference time of 0.45 seconds per gesture, representing high potential for real-time usage when coupled with a real-time video feed. Additionally, identified gestures are correctly translated to English, Kannada, and Hindi through the Google Translate API, increasing usability across different people. The above results validate the success and real-world usability of the proposed technique. Subsequent improvements will be to increase the gesture vocabulary, incorporate live input, and utilize sophisticated methods like pose estimation and self-supervised learning to enhance robustness under diverse conditions.

Acknowledgment

We would like to sincerely thank our faculty members and project guide for their unwavering support and insightful comments during this research. We also acknowledge the authors of the cited research papers, which served as crucial pillars for our work. We would like to express our gratitude to the creators of the datasets and tools utilized in this project. Lastly, we appreciate the support and inspiration from our families and peers.

References

[1]. Nidhi Chandarana, Nidhish Chhajed, Shreya Manjucha, Ms. Monica G. Tolani, Priyansi Chogale, Dr. Mani Roja M Edinburgh "Indian Sign Language Recognition with Conversion to Bilingual Text and Audio" 2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)

[2]. Varsha M, Chitra S Nair "Indian Sign Language Gesture Recognition Using Deep Convolutional Neural Network" 2021 8th

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 06 June 2025
Page No: 3034 - 3042
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0447

International Conference on Smart Computing and Communications (ICSCC)

[3]. Amutha Seviappan, Kothai Ganesan, Anna Anbumozhi, A. Sameer Reddy, B. Vamsi Krishna, D. Sampath Reddy "Sign Language to Text Conversion using RNN-LSTM" 2023 International Conferences on Data Science, Agents, and Artificial Intelligence

[4]. Dr. M. Kavitha, Aditi Chatterjee, Shivam Shrivastava, and Gourav Sarkar "Formation of Text from Indian Sign Language using Convolutional Neural Networks" 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)

[5]. Sourav Pal, Sparsh Sinha, Sourabh Ruhella, Anurag Goel "Sign Language to Text Classification using One Shot Learning" International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2023)

[6]. Navroz Kaur Kahlon, Williamjeet Singh "Machine translation from text to sign language: a systematic review" Universal Access in the Information Society (2023)

[7]. Uzma Farooq1, Mohd Shafry Mohd Rahim, Nabeel Sabir, Amir Hussain, Adnan Abid "Advances in machine translation for sign language: approaches, limitations, and challenges" Neural Computing and Applications (2021)

[8]. Video-Based Sign Language Recognition via ResNet and LSTM Network by Jiayu Huang, Varin Chouvatut.

[9]. Multi-Scale Temporal Network for Continuous Sign Language Recognition by Qidan Zhu, Jing Li, Fei Yuan, Quan Gan.

[10]. Multi-View Spatial-Temporal Network for Continuous Sign Language Recognition by Ronghui Li, Lu Meng.

[11]. Hand Gesture Recognition in Indian Sign Language Using Deep Learning by Harsh Kumar Vashisth, Tuhin Tarafder, Rehan Aziz, Mamta Arora, Alpana.

[12]. OpenHands: Making Sign Language Recognition Accessible with Pose-based Pretrained Models across Languages by Prem Selvaraj, Gokul NC, Pratyush Kumar, Mitesh Khapra.

[13]. Multiscaled Multi-Head Attention-Based Video Transformer Network for Hand Gesture Recognition by Mallika Garg, Debashis Ghosh, Pyari Mohan Pradhan.

[14]. Vision-Based Indian Sign Language Recognition Model by Suhas Hanjar, Vinayak Rangannavar, Pannagadhara K S, Karthik H S, Saravana M K.

[15]. Development of Deep Learning Based Models for Detecting the Significance of Non-Manual Parameters for Indian Sign Language Interpretation by unknown author/s.

[16]. Continuous Sign Language Recognition System Using Deep Learning with MediaPipe Holistic by Sudhakar Singh.