IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153143

MULTI-LEVEL SIGN LANGUAGE RECOGNITION SYSTEM

NALINA P[1], Dr REVATHI A[2]

M.Sc. Data Science and Business Analysis, Rathinnam College of Arts and Science Coimbatore – 641021[1]

Assistant Professor, Department of Computer Science,

Rathinam College of Arts and Science Coimbatore – 641021[2]

Abstract: Sign language is an important medium of communication for hearing and speech-impaired people. Nevertheless, there still exist communication gaps between sign language speakers and the general public owing to a lack of understanding of sign language gestures. This paper presents a Multi-Level Indian Sign Language (ISL) Recognition System that can identify word-level as well as sentence-level gestures from video inputs. The proposed system employs a hybrid deep learning model that combines Convolutional Neural Networks (CNN) for spatial feature extraction, Bi-Directional Long Short-Term Memory (BiLSTM) for modeling temporal sequences, and an Attention Mechanism to selectively, concentrate on important frames in the gesture sequence. The system analyzes video frames, extracts spatial-temporal features, and classifies them to predict the corresponding word and sentence. Besides text output, the system also offers multimodal feedback in terms of synthesized speech and visualization of output, thus improving accessibility and usability. The proposed method is expected to fill the communication system.

Keywords: Indian sign language (ISL), CNN, BiLSTM, Attention mechanism, video gesture recognition, multimodal assistive system.

INTRODUCTION

Communication plays an important role in human life, enabling people to express their thoughts, feelings, and share information. For deaf and speech-impaired people, however, the only means of communication is through sign language. Unfortunately, the majority of the people are not aware of this kind of communication. This is one of the biggest gaps in the communication of differently abled people. But with the rapid growth of artificial intelligence, computer vision, and the deep learning algorithms, it is possible to solve this problem by developing an automated sign language recognition system.

The initial research done on the sign language recognition was based on the static image classification using Convolutional Neural Network (CNN), which can easily extract spatial features such as hand shape, orientation, and motion. However, the static sign language cannot recognize dynamic gestures and signs at the sentence level, which is based on the temporal motion of hands in different frames. To overcome this problem, sequence modeling algorithms such as Long Short-Term Memory (LSTM) and Bi-Directional LSTM networks were introduced to handle the temporal dependency in the video frames. Recently, attention models have also been introduced in the deep learning algorithms to focus on specific frames in the sign language gestures.

However, despite these developments, the existing systems are mostly based on word-level and sentence-level recognition systems that have multimodal output. These assistive systems are based on the text, speech, and the visual output. This is especially true in the case of Indian Sign Language (ISL), which has less research work done in comparison to other sign languages. This creates a need to develop a reliable system based on a multi-level recognition system that can recognize both the word-level and sentence-level gestures in sign language and also has multimodal output. The proposed system aims to overcome these difficulties by using a combination of CNN, BiLSTM, and attention to develop an efficient and assistive Indian Sign Language recognition system.

PROBLEM STATEMENT

The communication between hearing and the speech-impaired people and the general public still poses a challenge because of the lack of understanding of sign language. Although various sign language recognition systems have been proposed, most of the existing techniques are mainly focused on the static gestures or isolated work recognition, which makes them less capable of recognizing continuous sentence-level gestures. The existing models are mainly dependent on spatial feature extraction and fail to incorporate the temporal dependencies that exist in the dynamic sign language

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026
DOI: 10.17148/IJARCCE.2026.153143

video sequences.

Moreover, most of the existing systems are only capable of producing text output without incorporating multimodal output, such as visual display, text, and the speech synthesis for the Indian Sign Language (ISL), the research contributions and the availability of large-scale annotated datasets are relatively less, which makes it difficult to design robust recognition systems, Thus, there is a need for an advanced multi-level that is capable of effectively incorporating spatial and temporal modeling techniques and producing multimodal output.

PROPOSED WORK

The proposed system provides a multi-level Indian Sign Language recognition system, which can recognize word-level and sentence-level signs. The system can also process video inputs and retrieve the frames. Then, using a Convolution Neural Network (CNN), the system can learn spatial features. Further, using a Bi-Directional LSTM network, the system can learn temporal dependencies. Additionally, the system can make use of the attention mechanism to focus on the relevant frames. This helps the system to increase the classification accuracy. The results are obtained using a fully connected layer and provided as text, speech, and visual.

3.1 Indian Sign Language

Indian Sign Language (ISL) is a visual language that is used by hearing and speech-impaired people in India. It consists of hand gestures, facial expressions, and body language. Compared to other sign languages, Indian Sign Language has few large-scale datasets and contributions to research.

3.2 Convolutional Neural Network

Convolutional Neural Network (CNN) is a deep learning algorithm that is used for extracting spatial features from images or video frames. In this system, CNN is used to detect prominent patterns in the images.

3.3 Bi-Directional Long-Term memory

BiLSTM is a type of neural network that can process data both forward and backward. This helps the neural network recognize dynamic gestures performed by a user on a video.

3.4 Attention Mechanism

The attention mechanism is used by the neural network. This mechanism helps the neural network focus on the most important frames of the gesture. This helps the neural network be more accurate by giving more importance to the relevant information.

3.5 Video Gesture Recognition

Video gesture recognition is the process of analyzing video to recognize hand gestures. Gesture recognition involves both spatial and temporal modeling.

3.6 Multimodal Assistive System

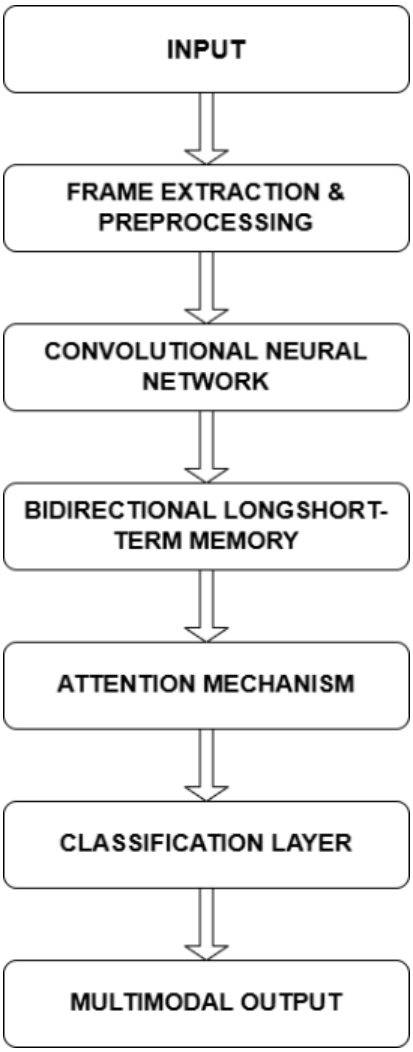
A multimodal assistive system is a system that offers output in the different forms and can be used to improve the accessibility of communication.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153143



SYSTEM ARCHITECTURE

The proposed system will work by analyzing the sign language video. It will be based on the spatial and temporal features of the deep learning approach. In this approach, the video will be preprocessed by extracting the frames. It will be analyzed using a Convolutional Neural Network (CNN), which will be used to extract the spatial features from the video. The features will be further analyzed using a Bi-Directional Long-term Memory (BiLSTM) to understand the temporal features of the gestures. It will be further analyzed using a fully connected layer to classify the word-level and sentence-level. The system will be able to display the result in multimodal form.

4.1 Dataset Collection

Both the word-level and sentence-level datasets were collected using the publicly available online sources like Kaggle. The Indian Sign Language video dataset was downloaded, and separate word-level and sentence-level datasets were created for training and testing. The dataset was preprocessed according to the requirements.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153143

4.1.1 Word Dataset

The word-level module identifies the sign gestures of individual words and produces output in the form of three ways, first showing the image of the sign language word, second predict the text, and finally synthesized speech recognition.

4.1.2 Video Dataset

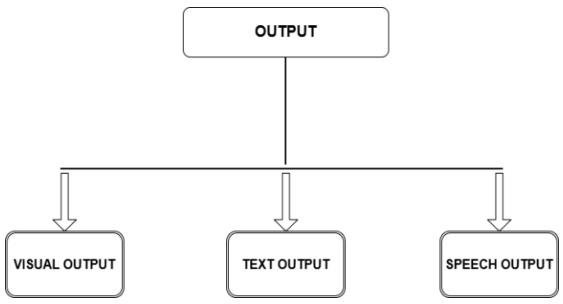
The sentence-level module identifies complete sign language videos to predict complete sentences, and it produces output in the form of three ways,first showing the image of the sign language word, second predict the text, and finally synthesized speech recognition.

EXPERIMENTAL SETUP

The proposed model has been implemented using the Python language Programming and the PyTorch library. The implementation of the Indian Sign Language system uses a dataset consisting of word-level and sentence-level video data. Then the dataset has been divided into two types, train set and test set. The resized and normalized video frames have been given as input to the proposed **Convolutional Neural Network (CNN) - Bi-Directional Long-Term memory (BILSTM) - Attention Mechanism.** The model was trained using the Adam optimizer and the cross-entropy loss function, and this model has been tested using the accuracy of the classification of the word-level and sentence-level in Indian Sign Language.

RESULTS AND DISCUSSION

The proposed CNN, BiLSTM, Attention mechanism model was successfully implemented to recognize both the word-level and the sentence-level gestures of Indian Sign Language. The spatial feature exaction and temporal feature extraction of the basic CNN models. The accuracy was further enhanced by using the attention mechanism to focus model was also capable of generating multimodal output in the form of visual display, text prediction, and speech recognition.



6.1 Visual Display Output

The system also provides the user with feedback in the form of the corresponding word-level image and sentence-level input video. This helps the user understand in a better way and increases the level of their confidence.

6.2 Text Output

The text of the word or sentence that has been predicted by the classifier is displayed on the screen. This helps the user

DOI: 10.17148/IJARCCE.2026.153143

to understand the sign language gesture that has been identified.

6.3 Audio Speech Output

The text of the word-level and sentence-level has been predicted by the classifier and is converted to audio speech using the text-to-speech module.

FUTURE WORK

The system that has been proposed can be further enhanced by extending it to real-time sign language recognition by using live camera feed. This can be done in the future by adding the transformer model to increase the accuracy of the system for real-time sentence recognition. The dataset can be further extended by adding more Indian Sign Language images. The proposed system can be further extended to support the development of a mobile or web application. Emotion recognition and facial expression can be further incorporated into the system to increase the understanding of the gestures in a specific context. The system can be deployed on the cloud and further optimized to make it more efficient.

CONCLUSION

This paper has demonstrated a Multi-Level Indian Sign Language Recognition System based on a hybrid CNN-BiLSTM-Attention Mechanism model. The proposed model was able to extract spatial and temporal features from the input for word-level and sentence-level gesture recognition. The combination of the attention mechanism helped to improve the classification accuracy by focusing on the relevant frames of the gesture sequence. The system provided multimodal output in terms of visual display, text prediction, and synthesized speech, making it more accessible and assistive. The proposed system is a scalable solution for future real-time and large-scale sign language recognition systems.

REFERENCES

[1]. N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, "Sign Language Transformers: Joint End-to-nd Sign Language Recognition and Translation," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[2]. N. C. Camgoz, S. Hadfield, O. Koller, and R. Bowden, "Neural Sign Language Translation," *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[3]. O. Koller, N. C. Camgoz, H. Ney, and R. Bowden, "Deep Sign: Enabling Robust Continuous Sign Language Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[4]. J. Huang, W. Zhou, Q. Zhang, H. Li, and W. Li, "Video-Based Sign Language Recognition Without Temporal Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[5]. P. Pigou, S. Dieleman, P. Kindermans, and B. Schrauwen, "Sign Language Recognition Using Convolutional Neural Networks," *European Conference on Computer Vision Workshops*, 2015.

[6]. R. Rastgoo, K. Kiani, and S. Escalera, "Sign Language Recognition Using Multi-View Convolutional Neural Networks," *IEEE Access*, vol. 8, pp. 151260–151269, 2020.

[7]. A. Kumar and S. Singh, "Sign Language Recognition Using CNN-BiLSTM Architecture," *International Journal of Computer Vision and Image Processing*, 2020.

[8]. G. Verma and P. Gupta, "Indian Sign Language Recognition Using Deep Learning," *Procedia Computer Science*, vol. 167, pp. 1075–1084, 2020.

[9]. J. Pu, W. Zhou, and H. Li, "Iterative Alignment Network for Continuous Sign Language Recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[10]. H. Cooper, B. Holt, and R. Bowden, "Sign Language Recognition," *Visual Analysis of Humans*, Springer, 2011.

[11]. A.Graves and J. Schmidhuber, "Framewise Phoneme Classification with Bidirectional LSTM Networks," *IEEE International Joint Conference on Neural Networks*, 2005.

[12]. A. Vaswani et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems*, 2017.

[13]. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *International Conference on Learning Representations*, 2015.

[14]. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, pp. 1735–1780, 1997.

[15]. A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Neural Information Processing Systems*, 2012.

[16]. T. Starner, J. Weaver, and A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer-Based Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153143

[17]. R. Rastgoo, S. Escalera, and K. Kiani, "Hand Gesture Recognition Using Deep Learning for Sign Language," *IEEE Access*, 2021.

[18]. J. Huang, W. Zhou, and H. Li, "Attention-Based Sign Language Recognition," *IEEE Transactions on Multimedia*, 2019.

[19]. C. Pu, W. Zhou, and H. Li, "Sign Language Recognition Using Temporal Convolution Networks," *IEEE International Conference on Multimedia and Expo*, 2019.

[20]. S. Kishore, P. Kumar, and A. Kumar, "Indian Sign Language Recognition Using CNN," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 3, pp. 2249–8958, 2020.

[21]. H. Cooper and R. Bowden, "Learning Signs from Subtitles: A Weakly Supervised Approach to Sign Language Recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[22]. Y. Yin et al., "Continuous Sign Language Recognition Using Hierarchical Recurrent Neural Networks," *IEEE International Conference on Image Processing*, 2016.

[23]. S. Cui, W. Zhou, and H. Li, "Attention-Based Sign Language Recognition Using Deep Neural Networks," *IEEE Transactions on Multimedia*, 2019.

[24]. M. Joze and O. Koller, "MS-ASL: A Large-Scale Dataset and Benchmark for Understanding American Sign Language," *British Machine Vision Conference*, 2019.

[25]. A. Moryossef et al., "Real-Time Sign Language Translation Using Neural Networks," *ACL Conference on Computational Linguistics*, 2020.