Real-Time Sign Language Translation for Accessibility: A Review of Vision-Based Recognition and Text-To-Speech Technologies

Chakane Trupti[1], Ghodake Sakshi[2], Bangar Payal[3], Kale Sayali[4], Prof. K. K. Raikar[5]
[1,2,3,4]*Undergraduate Students, Department of Artificial Intelligence and Data Science, Shri Chhatrapati Shivaji Maharaj College of Engineering, Nepti, Maharashtra, India*
[5]*Asst. Professor, Department of Artificial Intelligence and Data Science, Shri Chhatrapati Shivaji Maharaj College of Engineering, Nepti, Maharashtra, India*

***Abstract*—Indian Sign Language (ISL) is very important for people who can't hear or speak. There is still a big problem when people who use signs talk to people who don't. New technology in learning and computer vision has made it possible to create systems that can recognize and translate sign language automatically. This paper looks at the ways to recognize Indian Sign Language in real time and turn signs into text and then into speech. The focus is on hand movements taking out parts recognizing gestures making text and turning that text into speech. Different deep learning methods like Convolutional Neural Networks (CNNs) Recurrent Neural Networks (RNNs) Long Short-Term Memory (LSTM) networks and ways to track hands using vision are studied to see if they work for sign language recognition. The paper also talks about real-time video processing, how good the data is, changes in the background lighting, similar gestures and how fast the system can work. Also, how to use the signs that are recognized to make text. Then speech is looked at to help people communicate better. Some problems that are still there include not ISL data, different ways of signing recognizing signs that keep going and understanding the meaning of signs. The paper also suggests work to make the systems better stronger bigger and easier to use. This review is meant to help people who work on making easier and real-time Indian Sign Language translation systems.**

***Index Terms*—Indian Sign Language (ISL), Deep Learning, Hand Gesture Recognition, Computer Vision, Convolutional Neural Networks (CNN), Real-Time Sign Language Recognition, Text-to-Speech (TTS) Conversion, Speech Synthesis, Assistive Technology, Sign Language Translation.**

I. INTRODUCTION

Communication is a part of human life. It allows people to share their thoughts, feelings and information. For individuals who are deaf or hard of hearing communication can be difficult when others do not know sign language. In India many deaf and hard-of-hearing people use Sign Language or ISL. This language uses hand movements, facial expressions, body posture and space to convey meaning. Even though ISL is important many people in the population do not know it. This lack of understanding creates difficulties in schools, hospitals, workplaces and daily life. Because of this researcher are working on automated systems that can recognize sign language and turn it into text or speech. Recent progress in Artificial Intelligence Machine Learning and Computer Vision has made it possible to build systems that understand hand gestures from video. Traditional methods for recognizing sign language often depend on created features, special sensors or fixed gesture patterns. These methods can work well in controlled settings. They often fail in real-life conditions. They are not flexible, hard to scale. May not perform well outside laboratory environments. Deep learning has become popular because it can automatically learn features from images and videos. This reduces the need for work in designing features. Many deep learning models are used for sign language recognition. Convolutional Neural Networks or CNNs are effective at picking out details from images of hand gestures. Recurrent Neural Networks or RNNs and Long Short-Term Memory networks or LSTMs are

good at handling time-based data, such as video sequences. Hand tracking techniques can also be used to find hand positions in videos. This helps improve gesture recognition. These methods allow systems to work using a camera without needing wearable devices. Still their performance depends on things—like how large the training data is, how clear the images are, how bright the lighting is, hand angles and differences in how people sign. Real-time recognition of Sign Language starts with capturing hand movements using a camera. Video frames are processed to find features and then the gestures are classified into signs. The result can be shown as text. Turned into spoken words using a Text-to-Speech system. This helps connect sign language users with people who do not know sign language. Recognizing single gestures is not the same as understanding continuous signing. In conversation signs flow together and movements, facial expressions and context all affect meaning. A reliable translation system must go beyond just recognizing hand shapes. This paper looks at research on learning-based sign language recognition especially for Indian Sign Language and real-time text-to-speech translation. It reviews techniques, computer vision methods, dataset types and ways to combine gesture recognition with speech output. It also highlights challenges, such as the lack of large and varied ISL datasets, high computing demands, difficulty with continuous signing and how well models work across different users and settings. The main goal of this review is to look at what existing systems can do and what they cannot. It also aims to find paths for future research. The hope is to support the building of systems that're easy to use, accurate and work well in real life. Such systems can help make communication more inclusive, for hard-of-hearing people.

II. METHODOLOGY AND LITERATURE REVIEW

The methodology of this review paper focuses on analyzing existing deep learning techniques for real-time Indian Sign Language (ISL) recognition and their integration with text-to-speech systems. Particular attention is given to video-based gesture recognition, multimodal feature extraction, temporal sequence modeling, and the conversion of recognized signs into understandable text and speech.

A. Literature Review and Classification

Existing research on sign language recognition is examined to identify the methods used for gesture detection, feature extraction, sequence recognition, and output generation. The reviewed approaches are classified into image-based recognition, video-based recognition, pose-based recognition, and multimodal recognition. Deep learning architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Transformer-based models, are considered to understand their suitability for real-time ISL recognition.

B. Video Preprocessing and Feature Extraction

Video preprocessing is an important stage in continuous sign language recognition. The reference study uses video frame sampling and alignment correction to prepare input sequences for processing. RGB video frames provide visual information about hand gestures and body movements, while pose and hand landmarks provide complementary spatial information. The reference approach uses a 3D convolutional neural network to extract spatiotemporal features from RGB video and MediaPipe-based pose and hand features to represent body and hand movements. These feature representations are combined to provide information about both visual appearance and skeletal movement.

C. Deep Learning-Based Gesture Recognition

The extracted features are processed using a sequence-learning architecture to recognize signs from continuous video. In the reference study, RGB features and pose-based features are integrated using a Transformer encoder. Connectionist Temporal Classification (CTC) is used to train the model to align input video sequences with sign-label sequences without requiring a separate frame-level label for every input frame.This approach is relevant to real-time ISL recognition because signs occur over time and may differ in duration, speed, and movement pattern. The review compares this multimodal approach with other deep learning techniques to understand their advantages and limitations.

D. Sign-to-Text Conversion

After gesture recognition, the predicted sign labels are converted into readable text. In isolated sign

recognition, the system identifies individual signs, whereas continuous sign recognition must process a sequence of signs and generate an appropriate sequence of words. The review therefore considers recognition accuracy, temporal alignment, word error rate, and the handling of signs that are not represented in the training vocabulary.

E. Text-to-Speech Translation

Text-to-speech translation is considered as a subsequent stage in the proposed application workflow. Once the recognized signs have been converted into text, a speech synthesis module can generate spoken output from the resulting sentence. This stage can improve communication accessibility for users who prefer or require audio output.

It is important to distinguish this application-level pipeline from the reference study: the cited multimodal recognition paper primarily addresses continuous ISL recognition and does not itself establish a complete text-to-speech translation system. Therefore, speech generation is discussed separately in this review.

F. Comparative Analysis

The reviewed methods are compared according to their input modality, feature extraction technique, deep learning architecture, recognition task, evaluation measures, and suitability for real-time applications. Metrics such as Word Error Rate (WER), recognition accuracy where reported, processing speed, and robustness to variations in signing conditions are considered when available.

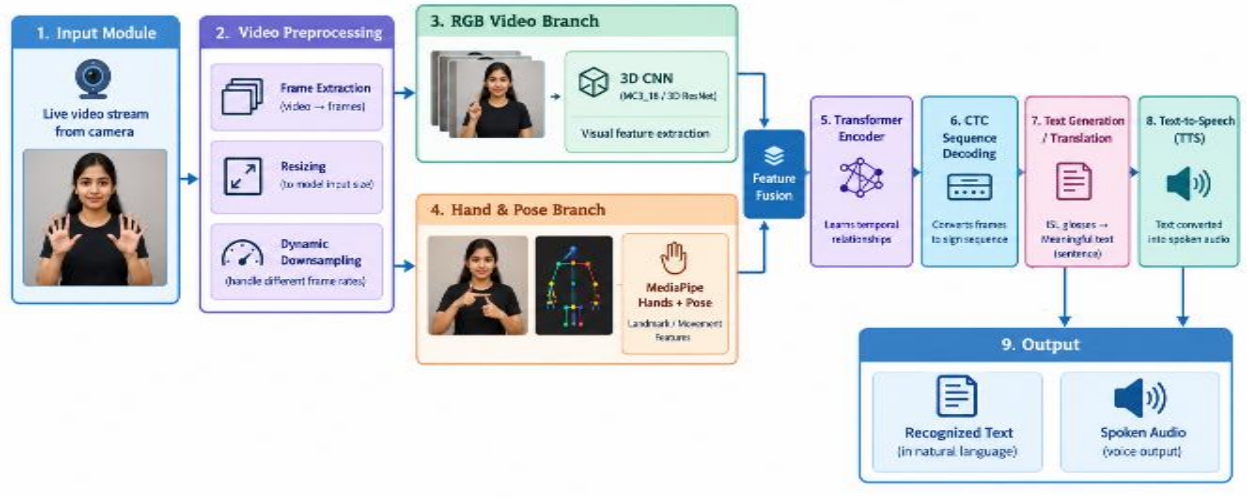
The reference study highlights challenges associated with signing speed, background variation, hand dominance, capture conditions, and out-of-vocabulary signs. These issues are used to identify research gaps and possible improvements for future ISL recognition and speech translation systems.

G. Overall Workflow

The methodology followed in this review can be summarized as:

Literature Review → Technique Classification → Video and Pose Feature Analysis → Deep Learning Model Comparison → Sign-to-Text Analysis → Text-to-Speech Integration Review → Research Gap Identification.

This methodology provides a structured basis for evaluating current ISL recognition approaches and identifying opportunities to develop more accurate, robust, and accessible real-time sign-to-text and speech translation systems.



III. MATHEMATICAL FORMULATION

A mathematical description makes it easier to see how a learning-based Indian Sign Language (ISL) recognition system works. The pipeline has five stages: capturing the image, extracting features, classifying the gesture, generating text, and synthesizing speech. Each stage can be written as a function, which shows how raw visual input is gradually turned into understandable communication.

A. Input Image Representation

Let X be a single image or video frame captured by the camera. It is represented as a three-dimensional tensor:

$X \in \mathbb{R}^{H \times W \times C}$

where H is the height, W is the width, and C is the number of color channels (C = 3 for an RGB image). Before the frame is passed to the model, its pixel values are usually normalized. This keeps the input consistent and helps the model train faster:

$$X' = \frac{X}{255}$$

B. Feature Extraction Using Deep Learning

A deep learning model learns useful features from the image on its own, without hand-crafted rules. In a Convolutional Neural Network (CNN), each convolution operation is expressed as:

$$F(i,j) = \sum_m \sum_n X\,(i+m,\, j+n)\, K(m,n) + b$$

Here X is the input feature map, K is the convolution kernel, b is the bias term, and F(i,j) is the output feature at position (i,j). As more convolution layers are stacked, the network moves from simple edges to richer patterns such as hand shape, finger placement, and overall gesture structure.

C. Gesture Classification

The extracted features are passed to a classification layer that estimates which gesture is being shown. The probability of each gesture class is computed with the Softmax function:

$$P(y = c \mid X) = \frac{e^{z_c}}{\sum_{j=1}^{N} e^{z_j}}$$

where z_c is the raw score the model assigns to class c, and N is the total number of gesture classes. The class with the highest probability is taken as the final prediction:

$$\hat{y} = \arg\max_{c \in \{1,\ldots,N\}} P(y = c \mid X)$$

where $\hat{y}$ denotes the predicted gesture label.

D. Model Training and Loss Function

The model is trained on labeled gesture images or video clips. The gap between its predictions and the true labels is measured with the categorical cross-entropy loss:

$$L = -\sum_{c=1}^{N} y_c \log(p_c)$$

where y_c is the true (one-hot) indicator for class c, and p_c is the probability the model predicts for that class. During training, the weights are updated to reduce L, which improves the model's ability to recognize gestures it has not seen before.

E. Gesture-to-Text Conversion

Once a gesture is classified, its label is mapped to the corresponding word or phrase:

$$T = g(\hat{y})$$

where T is the resulting text and g(.) is the mapping function from gesture labels to words. For continuous signing, additional temporal modeling and language-processing steps may be needed to join individual predictions into complete, meaningful sentences.

F. Text-to-Speech Conversion

The generated text is then sent to a Text-to-Speech (TTS) engine, which produces audible output:

$$S = h(T)$$

where S is the synthesized speech signal and h(.) is the TTS function. This lets listeners hear the meaning of the signed gestures.

Together, these equations outline the stages of a deep learning-based ISL recognition and speech translation system. The specific architecture, equations, and processing steps will differ depending on the approach each study adopts.

IV. DEEP LEARNING TECHNIQUES FOR ISL RECOGNITION

Deep learning has become an approach for sign language recognition because it enables automatic feature extraction and classification from images and video sequences. In Sign Language (ISL) recognition deep learning techniques are used to identify hand shapes, finger positions, movement patterns and other visual characteristics associated with signs. The selection of a model depends on the type of gestures, dataset characteristics, computational resources and whether recognition is performed on static images or continuous video. This section reviews commonly used deep learning architectures and vision-based methods for ISL recognition.

A. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) are widely used for image-based hand gesture recognition. They learn features through convolutional and pooling layers followed by classification layers that predict the corresponding gesture category. In ISL recognition

CNNs can identify patterns such as finger arrangements, hand orientation and palm shape from camera images. CNN-based models are particularly suitable for recognizing static gestures. Their performance depends on the quality and diversity of training data, image preprocessing and the ability to handle variations in lighting, background and hand position.

However, a conventional image-based CNN does not independently model the relationships between successive video frames. Consequently, additional sequence-processing methods may be needed for dynamic sign recognition.

B. Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNNs) are designed to process data by maintaining information from previous time steps. For video-based sign language recognition extracted features from frames can be supplied to an RNN to learn temporal patterns associated with gesture movements. RNN-based approaches can support the recognition of dynamic gestures in which the order and duration of movements contribute to the signs meaning. However conventional RNNs may have difficulty learning long-term dependencies in sequences. Their effectiveness also depends on the quality of the input features and the availability of sequential training data.

C. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network designed to learn longer-term dependencies in sequential data. They use memory cells and gating mechanisms to control the retention and updating of information over time. In ISL recognition LSTM networks can process sequences of image features or hand landmarks extracted from video frames. When combined with a CNN or a hand-tracking method an LSTM can learn both information and temporal changes in hand movement. Such combinations are useful for recognizing gestures although their performance depends on sequence length, model complexity and training data diversity.

D. CNN–LSTM Hybrid Models

CNN–LSTM models combine the feature-extraction capability of CNNs with the temporal modeling capability of LSTMs. The CNN processes frames to extract visual representations while the LSTM analyzes the sequence of representations to identify gesture patterns over time. This hybrid architecture is an approach for real-time ISL recognition because it considers both hand appearance and movement. Nevertheless, processing frames can increase computational requirements and latency. Efficient frame selection, model optimization and appropriate hardware are therefore important for real-time deployment.

E. Vision-Based Hand Tracking

Vision-based hand-tracking methods detect hand landmarks, such as fingertips, finger joints and wrist positions from camera images. These landmarks can be represented as coordinate vectors. Used as input features for a classification model. Compared with processing images landmark-based recognition may reduce the dimensionality of the input and focus on hand geometry. The extracted coordinates can be classified using machine learning or deep learning models, including neural networks and LSTM networks. However landmark detection may become less reliable when hands are occluded gestures overlap, lighting is poor or the hand moves rapidly.

F. Research Considerations

Although deep learning techniques have demonstrated potential for sign language recognition several challenges remain. The availability of diverse and well-annotated ISL datasets is limited compared with some widely studied sign languages. Differences, in signing styles, hand orientation, background conditions and camera quality can also affect model generalization. Furthermore, recognizing gestures does not necessarily provide complete sign language translation because continuous signing may involve transitions, facial expressions and contextual information. Future research should investigate architectures improved ISL datasets, robust hand tracking and integrated spatial-temporal models. These developments may help improve recognition reliability and enable real-time systems that convert recognized signs into text and speech.

V. TEXT-TO-SPEECH TRANSLATION

Text-to-Speech translation is a part of an Indian Sign Language recognition system because Text-to-Speech

translation changes recognized gestures into spoken sound. Adding gesture recognition to speech synthesis can lower the communication wall between sign language users and people who do not read ISL. In a system the hand gestures that are seen are first put into clear labels then turned into written words and finally sent to a Text-to-Speech engine that makes spoken words. The success of this chain depends on how well gesture recognition works, how correct the written words how good the speech sound is.

A. Gesture-to-Text Conversion

The first step of the translation chain is to turn the guessed gesture class into the matching written word. When a camera records hand motion the recognition model looks at the image. Guesses the related gesture label. Then a mapping step turns that label into a word, a letter or another written piece depending on the data set and the chosen recognition way. The mapping process can be represented as: $T = g(\hat{y})$ where T denotes the generated text, $\hat{y}$ represents the predicted gesture label and $g(\cdot)$ denotes the mapping function. When the gesture is stand-alone the output can be a letter or word. When the signing is flowing several predictions must be. Processed to make clear sentences. Thus, looking at time order and using language context can help make the translation more accurate.

B. Text-to-Speech Conversion

Once the text is made the system sends it to a Text-to-Speech engine. That engine turns the written text into a sound that can be heard. The process can be shown as: $S = h(T)$ where S represents the synthesized speech signal, T is the input text and $h(\cdot)$ denotes the TTS function. Text-to-Speech systems can use synthesis, statistical parametric synthesis or neural speech synthesis. Old style synthesis builds speech from recorded sound pieces or acoustic models. New style neural TTS learns links between written words, how they sound and speech traits. Choosing a Text-to-Speech method depends on which languages it can do how natural the voice sounds, how fast it works and the computer power that is available.

C. Integration with Real-Time ISL Recognition

Adding Text-to-Speech to an ISL recognition system needs the parts that grab video classify gestures write text and play sound to work together. The system keeps looking at camera frames finds gestures that are recognized and turns predictions into written words. The written words are then sent to the Text-to-Speech engine for sound. The normal flow looks like this: Camera Input → Hand Gesture Detection → Gesture Classification → Text Generation → Text-to-Speech Engine → Speech Output For use the system must keep the pause between seeing a gesture and hearing the sound very short. When the same gesture is guessed again the system should avoid saying the word or sound over and over. When the signing keeps going the system might need to split gestures handle the sequence and think about sentences before it makes sound.

D. Language Support and Speech

Quality Language support is a point for making ISL translation systems easy to use. The written words that come out may have to be produced or changed into a language chosen by the user before they turn into sound. Indian Sign Language is not just a straight copy of spoken words. Sign languages have their grammar and language traits. The sound that the Text-to-Speech engine makes depends on how the words are pronounced, the rise and fall of tone how clear it is and how good the chosen Text-to-Speech engine is. Systems that will be used in life should think about Indian languages, such as English and other local tongues if good speech engines exist for them. Doing the right text cleaning and picking the language can make the sound even better.

E. Challenges in Text-to-Speech Translation

problems can hurt how well the ISL recognition and Text-to-Speech system works. Mistakes, in recognizing gestures can make written words, which can then give wrong spoken sound. Unclear gestures flowing signing, missing context and different signing styles can lower how exact the translation is. Also needing a network not having languages and the high computer cost of neural speech synthesis can make the system slower. To address these challenges future systems should focus on gesture classification, context-aware text generation, efficient speech synthesis and reduced processing latency. Evaluation should consider end-to-end translation quality, speech intelligibility, response time and the usability of the system for real-world communication. In summary Text-to-Speech technology provides a link between

recognized ISL gestures and audible communication. Its integration, with learning-based recognition methods can support accessible communication applications provided that gesture interpretation, text generation and speech synthesis are designed and evaluated as interconnected stages.

VI. CHALLENGES AND RESEARCH GAPS

Indian Sign Language ISL recognition has problems. One problem is that there are not big and varied data sets. Another problem is that people sign at speeds. Sometimes there is background noise. Also, hand movements can be different from person to person. It is hard to recognize signs that happen one after another because the signs might overlap or last different amounts of time. Signs that are not in the vocabulary can be a problem. Lighting, camera angles and the way people sign can also make it harder to recognize signs correctly. A big problem is making the system work quickly while still being accurate. Combining gestures with text and turning that text into speech needs reliable processing. Current methods also need to work for different people and different places. Future work should focus on making data sets better models that use more than one type of data better ways to recognize signs that happen in a row and faster ways to turn text into speech. These improvements will help make communication easier and more natural, for everyone.

VII. COMPARATIVE ANALYSIS OF EXISTING RESEARCH

The development of Sign Language (ISL) recognition systems has used various methods, including traditional machine learning, Convolutional Neural Networks (CNNs) recurrent neural networks, hybrid deep learning architectures and vision-based hand tracking. These methods are different in their ways of taking out features ability to recognize moving signs the amount of computing power they need and if they work well for real-time use. Looking at these methods helps to see what they can do what they can't do and how they can work with text-to-speech translation systems.

A. Comparison of Recognition Approaches machine learning methods usually take out features by hand like how the hand looks the shape of the edges and the way the hand is shaped. These methods can work well when there are not signs to recognize but may need a lot of work to get the right features and may not work well when there are more complicated changes. CNN methods take out features from images of hand signs. They work well for signs that do not move much and can reduce the need for designing features by hand. However, image-based CNNs alone may not capture the time part of the signs, which is needed for moving signs. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks look at information over time. Can show how movements happen over video frames. When used with CNNs they can learn both how things look and how they move. However, mixing these methods can make the computer work harder and need a lot of data to train properly. Vision-based hand landmark methods show the hand using points like the tips of fingers the joints and the wrist. These features can be used in a model that classifies signs so the model does not need to look at the image. However, if the hand is not seen clearly like when something blocks the view or if the lighting is bad or if the hand moves fast this can make the model less.

B. Comparison Based on Real-Time Applicability Real-time ISL recognition needs the system to take pictures from a camera recognize signs make text and speak it all in a time. So, when choosing a model, it is important to look at not how well it works but also how fast it can run how power it needs and how well the whole translation system works. Lightweight CNNs and landmark-based models may work well on devices. CNN–LSTM models can look at how signs move but may need some changes to work fast. Also making text and then speaking it adds time that should be measured when testing. A fair comparison should use the data set to test and report the right numbers, like accuracy, precision, recall, F1-score how fast it runs and how long it takes to translate from sign to speech. For signs that happen in a row it's important to look at the sequence not just each sign by itself.

C. Research Findings and Future Directions The comparison shows that there are chances for more research. First more ISL data sets are needed to test how well the system works with people and in different places. Second methods that look at both

space and time should be checked for signs. Third models that are simpler and hand tracking that works better could help with computing power. Finally, signs should be connected to text and speech that works with the situation to help make communication. Overall, the current methods give a start for ISL recognition but to make real-time translation work the system needs to work well with signs classifying them making text and speaking it. Future research should test systems in real situations and share results that can be checked to show how good they are, for helping people talk.

VIII. FUTURE SCOPE AND CONCLUSION

In the future, this project will aim to improve the accuracy and speed of the gestures using advanced Deep Learning techniques. This can be extended to detect full sentences in sign language, facial expressions and body motions. The system can be improved by adding multi-lingual text to speech, creating an app for mobile devices and having an animated 3D avatar to display the sign gestures. This project will help break down barriers between people using Indian Sign Language and people who don't understand by detecting hand gestures, converting them to meaningful text and creating speech. There are still issues to overcome such as varying gestures, lighting and accuracy but using better data sets and deep learning models will improve the system. Overall, this project has great potential to help create accessible communication within education, healthcare, work places and daily life. It will help increase inclusion within the deaf and hard of hearing community.

ACKNOWLEDGMENT

We would like to thank our mentor Prof. K. K. Raikar and project coordinator Prof. G. P. Joshi for their guidance and motivation throughout the review paper. We would also like to thank the faculty members of the AI & DS department for their continuous support. We would like to thank our college Shri. Chhatrapati Shivaji Maharaj College of Engineering, Nepti for providing us with the necessary academic facilities and learning environment. We would like to thank all the researchers who have written papers on Indian Sign Language, Deep Learning, Computer Vision and Text to Speech technologies which helped us gain an understanding of the current scenarios, issues and future scopes.

REFERENCES

[1] R. Damdoo and P. Kumar, "Survey on Indian sign language recognition and translation," *IET Image Processing*, vol. 19, no. 1, Art. no. E70000, 2025, doi: 10.1049/ipr2.70000.

[2] A. V. Nair and B. V., "Sign language recognition," International Journal of Computer Applications, vol. 73, no. 22, pp. 33–38, 2013, doi: 10.5120/13037-0260.

[3] S. Dour and M. M. Sharma, "Literature review for sign language recognition system development," International Journal of Computer Applications, vol. 132, no. 5, pp. 27–34, 2015, doi: 10.5120/ijca2015907360.

[4] P. B. V. and S. Srinath, "Review on Indian sign language recognition system using vision-based approaches," International Journal of Computer Applications, vol. 184, no. 43, pp. 52–58, 2023, doi: 10.5120/ijca2023922548.

[5] N. K. Bahia and R. Rani, "A [level unspecified] taxonomy review for sign language recognition with emphasis on Indian sign language," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 22, no. 1, 2023, doi: 10.1145/3530259.

[6] R. Rastgoo, K. Kiani, and S. Escalera, "Sign language recognition: A survey," Expert Systems with Applications, vol. 164, Art. no. 113794, 2021, doi: 10.1016/j.eswa.2020.113794.

[7] S. Subburaj and S. Murugavalli, "Sign language recognition: A survey of vision-based and learning approaches," Measurement: Sensors, vol. 23, Art. no. 100385, 2022, doi: 10.1016/j.measen.2022.100385.

[8] I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, "Machine learning methods for sign language recognition: An analysis," Intelligent Systems with Applications, vol. 12, Art. no. 200056, 2021, doi: 10.1016/j.iswa.2021.200056.

[9] A. Wali, R. Shariq, S. Shoaib, S. Amir, and A. A. Farhan, "Progress in sign language recognition: A review," Machine Vision and Applications, vol. 34, Art. no. 127, 2023, doi: 10.1007/s00138-023-01479-y.

[10] "Machine and deep learning approaches in sign language recognition: Techniques and future trends," in Proc. 2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT), 2024, doi: 10.1109/IConSCEPT61884.2024.10627910.