Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Sign Language Translator

Naga Durga Saile K, Haripriya S, Pranuthi S, Vaishnavi C, Manoja Nanded

Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, 500090
Email: saileknd3@gmail.com

Sign language translation is necessary to increase inclusivity and communication within the Deaf and hard-of-hearing communities [1]. Communication difficulties with non-sign language speakers affect social, professional, and educational relationships for deaf people [1]. These issues can be solved by utilizing technology, such as sign language recognition, which offers substitute communication channels [2]. This work presents a novel method for recognizing and translating sign language motions using MediaPipe Holistic and Long Short-Term Memory (LSTM) neural networks [3]. After extensive testing, our LSTM-based model shows impressive accuracy, which makes it a useful tool for real-time translation [4]. We discuss problems encountered and their solutions, providing insights into the model architecture, training protocols, and data preprocessing [5]. Our results suggest that systems for translating and recognizing sign language have a lot of room for improvement. This method encourages a thorough assessment of its effectiveness and has potential uses for the Deaf community, sign language interpreters and technology developers. This research lays the foundation for future sign language translation technology developments, closing the communication gap between the Deaf and hearing communities.

Keywords: LSTM, Sign Language Translation, Deaf Community, Media Pipe Holistic, Dataset, Temporal aspects.

1. Introduction

For the Deaf and hard-of-hearing communities, sign language is an essential communication tool that links the spoken and auditory worlds. However, sign language experts face many challenges when interacting with non-native speakers. These issues come up in various contexts throughout daily life, including social, professional, and educational ones. These communication barriers have made innovative technological solutions necessary that could facilitate an easy exchange of ideas and feelings between the Deaf and hearing communities

[1]. Conventional approaches to sign language interpretation mostly use human interpreters and manual translation, which is also labor-intensive and constrained by the availability of qualified interpreters [2]. There is a clear need for a more effective and widely available system that can facilitate seamless communication between sign language users and the larger community and real-time sign language translation.

In recent years, machine learning techniques have shown promise in solving these issues. Machine learning techniques, which leverage neural networks, data analysis, and pattern recognition [3], could create accurate, real-time sign language translation systems. These tools open up new communication channels between non-signers and those learning sign language.

Comprehension and dialogue.

This article investigates how MediaPipe Holistic can accurately translate sign language by highlighting essential details, compiling a large dataset, and using LSTM (Long Short-Term Memory) models to draw important landmarks on sign language gestures. The study's performance evaluation uses pertinent indicators to ascertain the model's applicability and accuracy in real-time to complete an exhaustive assessment.

This research project attempts to close the gap between contemporary data-driven methods and traditional sign language interpretation. The objective is to use machine learning-based models that can accurately recognize and translate signs to enable more effective communication and comprehension among sign language users. This will result in increased diversity in the neighborhood.

2. Literature Review

In the paper Development of Sign Language Motion Recognition System for Hearing Impaired People Using Electromyography Signal, a real-time motion recognition system based on an electromyography signal was proposed to help hearing people learn sign language and to help deaf people communicate with others [2]. This system is designed to recognize authentic ASL hand motions. One example of contemporary sensor-based technology is a hand glove system that can translate hand gestures into characters displayed on an LCD screen. Another illustration is the hand glove-based sign-to-text translation system. Deaf and mute people can use this method to communicate with non-sign language speakers. The system consists of an LCD, a CPU, and flex resistors.

Although much noteworthy research has been done on this subject, hand identification and tracking using different machine-learning approaches is a comparatively fresh development. The k-nearest neighbors approach is used to classify hand poses in real-time Indian Sign Language (ISL) recognition after features are tracked using a grid-based framework technique that converts them into a feature vector [3]. Another study, Sign Language Recognition Using Two-Stream Convolutional Neural Networks with Wi-Fi Signals, offers a Wi-Fi-based method for recognizing sign language that uses singular value decomposition to preprocess Channel State Information. Using a convolutional neural network, the technique mixes motion and spatial data and incorporates a feature selection process. Compared to competing approaches, the strategy demonstrated exceptional recognition accuracy rates on the SignFi dataset, indicating improved performance and generalization capabilities [5].

A technique for identifying signers utilizing a wide vocabulary of sign language is presented in Continuous sign language recognition (Koller et al., 2006) [6]. This method can be applied in real-world scenarios. They perform studies centered around five topics to demonstrate the usefulness of tracking, multimodal sign language features, and temporal derivatives in identifying sign language. They examine visual modeling approaches, use class language models, and adapt CMLLR to address signer reliance. This method produces low word error rates on two publicly available large vocabulary databases that reflect lab data and unrestricted real-life sign language.

Camgoz et al. were the first ones to present the Sign Language translation problem (2018). It attempts to translate spoken English films from sign language while accounting for the intricate grammatical and syntactic patterns of sign language. Together, they trained the language model, the spatial representations, and the mapping between spoken and sign language as they formalized SLT inside Neural Machine Translation (NMT) framework. They collected the first "Continuous SLT dataset, RWTH- PHOENIX-Weather 2014T" to assess the efficacy of their Neural SLT. This dataset includes over 67K signs from a sign vocabulary of over 1K and over 99K words from a German vocabulary of over 2.8K. At the frame and gloss levels, their end-to-end tokenization networks received scores of 9 and 18, respectively.

The suggested methodology uses the most recent advancements in machine learning and dataset curation to offer a fresh approach to sign language translation and recognition. Additionally, by giving dataset curation priority, the model is guaranteed to recognize a broad range of sign language motions commonly used in communication. In the end, this work presents a thorough plan for improving accessibility and communication for hearing loss, offering new insights and state-of-the-art methods to advance the field of sign language translation and recognition.

3. Dataset Curation

A series of 10 videos were recorded for every action, capturing different gesture nuances and angles [7]. Every video comprised 10 frames to guarantee stability and precision and offer a thorough depiction of the sign language movements. To replicate real-world scenarios, the dataset was painstakingly selected to account for variations in lighting conditions, backgrounds, and hand orientations [8].

The MediaPipe Holistic framework was used to extract the essential features required for training [9]. With the aid of this vital instrument, essential points about the hands, body, and face could be extracted, supplying rich spatial-temporal data necessary for precise gesture recognition.

The carefully labeled and arranged dataset included ten actions, each with ten videos and ten frames. This extensive dataset was used to train the LSTM (Long Short-Term Memory) deep learning model, an essential part of the Sign Language Translator system. The dataset ensured the model was exposed to a wide variety of variations in sign language, which prepared it for accurate and dependable gesture recognition.

This carefully chosen dataset is a vital tool for progressing the study of sign language interpretation and serves as a solid basis for upcoming investigations into assistive technology for the hard of hearing.

The compiled dataset consists of 10 actions, each of which is painstakingly captured in 10 videos for 100 videos. Ten frames make up each video, which offers a comprehensive temporal representation of sign language gestures [10]. The cutting-edge MediaPipe Holistic framework was used to process these videos to extract precise vital points corresponding to movements of the hands, body, and face. To guarantee that every action was accurately represented, the 1,000-frame dataset underwent meticulous labeling [11].

The dataset is roughly 2.5 GB and stored in JPEG format [12]. After extensive testing and analysis, the model proved remarkably effective at differentiating between various sign language gestures. This demonstrates how well the model can identify minute differences in hand, body, and facial configurations. The dataset was further divided into separate sets for the hands, body, and face after key point extraction was finished, enabling more specialized training to improve gesture recognition accuracy [13]. This enhanced dataset is offered as a starting point for the later phases of model training. It is an essential tool for figuring out the Sign Language Translator system's quality parameters.

4. Research Methodology

4.1 LSTM

The Long-Short-Term Memory (LSTM) layer is a crucial component of our model for understanding sign language. LSTMs, a recurrent neural network architecture type, are well known for their effective sequential data modeling. LSTMs are crucial to our project because they enable precise identification of the temporal relationships present in sign language gestures.

The networked LSTM units that comprise the LSTM layer process input sequences successively while preserving their internal states. Because these units enable the model to learn and retain patterns progressively, they are perfect for tasks involving sequential data.

The LSTM layer in our model for sign language recognition processes a set of extracted key point features corresponding to sign language gestures. Because of its sequential design, the LSTM can identify motions by identifying the relationships between important point frames over time.

Through gradient descent and backpropagation, the LSTM layer can modify its parameters (weights and biases) during training. To maximize the model's performance, hyperparameters like the batch size and learning rate are adjusted.To generate final predictions for sign language recognition, the output of the LSTM layer is integrated with other architectural elements and processed further in succeeding layers (see Fig. 1).

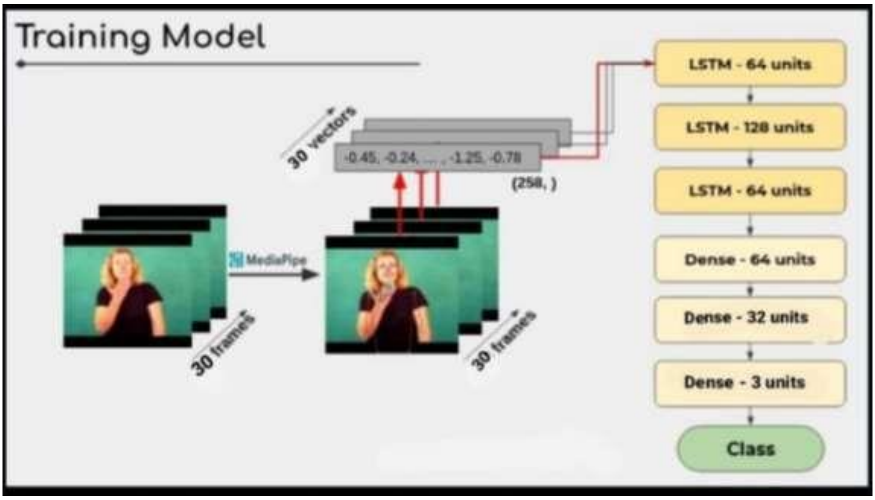


Fig. 1. Training LSTM model

5. Experimental Findings and Analysis

This study implements the Sign Language Translator on VS Code using a 10-action dataset. The computing system's configuration is Windows 11, with a CPU Intel(R) Core (TM) @1.20 GHz, 1190 MHz, 2 Core(s), 4 Logical Processor(s) [7], and 64GB RAM.

5.1 Workflow

The workflow is described in Fig. 2 and is explained as follows:

Data Preparation. Data preparation is essential to creating a reliable sign language translation system. This module describes the procedure for gathering key point values to train and evaluate the machine learning model. The process entails setting up the camera, capturing action-specific video sequences, and obtaining key point values from the identified landmarks.

Webcam Initialization: The cv2.The video Capture function in OpenCV initializes the webcam and captures live video frames. This makes it possible to record sign language movements made in front of the camera.

Holistic Mediapipe Initialization: The Mediapipe holistic model (mp_holistic) is used for each video frame to identify and display the hand, facial, and posture landmarks.

Video Sequence Capturing: The module records a certain amount of video sequences for every action. The Mediapipe holistic model is used to identify landmarks for each frame in the video stream. These key point values are then retrieved and saved for later processing.

Data Storage. The extracted key point values are stored in the appropriate folders to produce an organized model training and testing dataset. Every action has a folder of its own that is further subdivided into sequences and frames.

Data preprocessing and labeling. The gathered key values are converted into structured input data for the LSTM model's training using the preprocessing code. It generates a label map for every action and sets the sequence length to thirty. After that, it loads the matching key point values and arranges them into sequences for every action, sequence, and frame. The final sequences and their labels are transformed into NumPy arrays, and the labels undergo one-hot encoding.

Development and Training of the LSTM Neural Network. The Keras API provided by TensorFlow is used to build the LSTM model. It comprises tightly linked layers for classification and several LSTM layers to capture temporal relationships. The model uses a softmax activation function to generate probabilities for each action in sequences of 30 frames, each containing 1662 characteristics (key point values).

Model Assessment. The testing dataset is used to assess the trained LSTM model. Predictions and ground truth labels are compared to evaluate the model's performance. The evaluation measures show how well the model generalizes and anticipates sign language gestures.

Real-time Translation and Recognition of Gestures. This last module includes real-time gesture translation and recognition. Based on the identified key points, the model makes a real-time prediction about the action. As the user proceeds with the tasks, the anticipated action appears on the screen, and a phrase is generated. The phrase is updated as long as the model's prediction confidence is higher than a predetermined threshold.

5.2 Experimental Strategy

The train_test_split function from scikit-learn divides the dataset into training and testing sets. 95% of the data used for training and 5% for testing are split during the split process. This ensures enough data for training and a small but representative amount left over to assess the model's performance. The training set is employed for model learning, optimization, and parameter changes. The validation set is implicitly implicated during model training, particularly when Tensor-Board callbacks are used. The testing set is essential for assessing the model's generalization performance and gauging its correctness on never-before-seen data. Backpropagation on the training set optimizes the model during the training phase. In order to minimize the categorical cross-entropy loss function, the LSTM layers' weights are adjusted during the training phase. Training metrics are visualized in real-time using TensorBoard callbacks. The testing set is used to assess the model's performance following training. The number of LSTM units, activation functions, and optimizer settings are examples of hyperparameters that can be changed and altered during the experimental approach to enhance the model's performance. In order to determine the ideal collection of hyperparameters, the model is iteratively trained using various configurations, and the results are assessed.

This all-encompassing approach guarantees a full investigation of the model's- capabilities, culminating in an accurate and well-optimized sign language translation system.

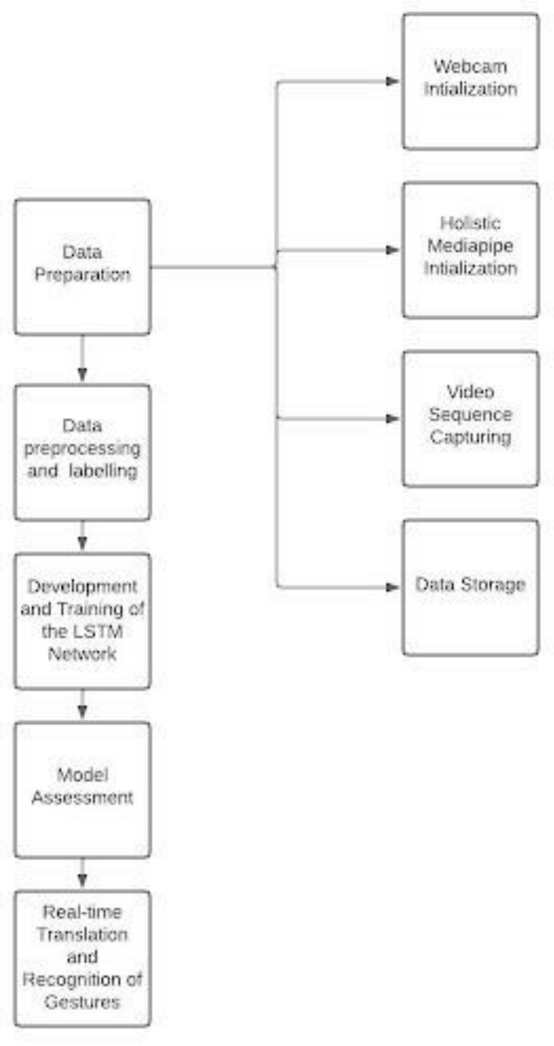
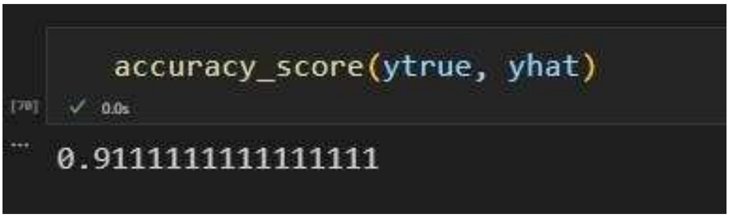

Data Preparation
Data preprocessing and labelling
Development and Training of the LSTM Network
Model Assessment
Real-time Translation and Recognition of Gestures
Webcam Intialization
Holistic Mediapipe Intialization
Video Sequence Capturing
Data Storage

Fig. 2. Workflow

5.3 Results

Fig. 3 shows the accuracy score that the LSTM model produced. This score is the main metric used to assess the model's performance

The ratio of accurately predicted samples to the total number of samples in the testing set is used to compute accuracy. The experiment's accuracy score of 0.91 shows that the model accurately classified all samples in the testing dataset.



```
accuracy_score(ytrue, yhat)
[70]  ✓ 0.0s
...   0.9111111111111111
```

Fig. 3. Accuracy of LSTM Model

The Tensor Board logs provide an epoch-by-epoch display of the categorical accuracy and loss during the training process. The graphs shed light on the model's convergence and optimization throughout the training process.

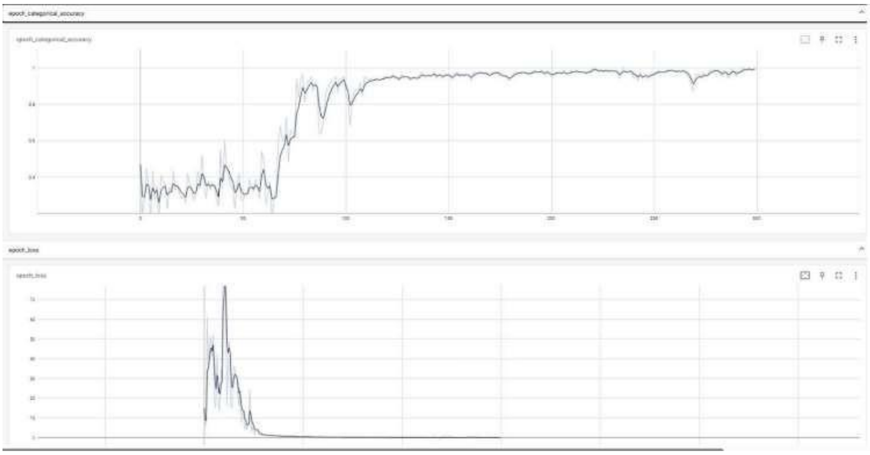
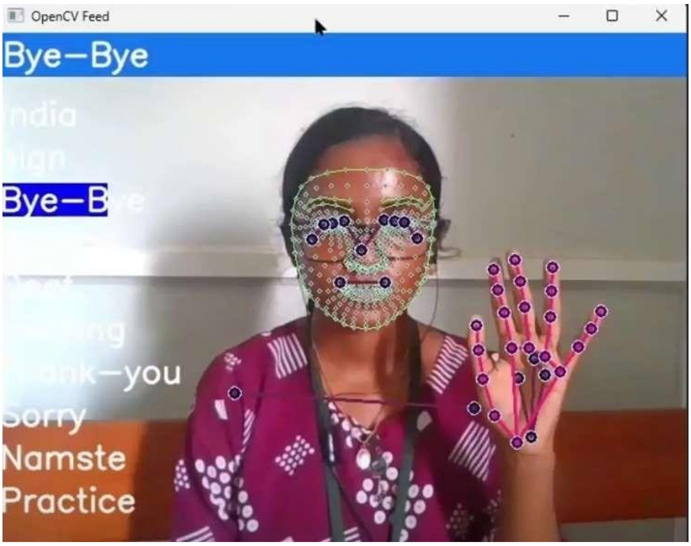


Fig. 4. Epoch Categorical Accuracy and Epoch Categorical Loss Graphs

Fig. 4 displays the epoch categorical accuracy graph, which shows how well the model classified the training data over 300 epochs. At first, the accuracy exhibits a low and erratic trend. However, a noticeable improvement is noted when the precision reaches a high and stable level at specific epochs. This shows that the model picked up on the training set quickly and adjusted to it, producing reliable and accurate predictions.

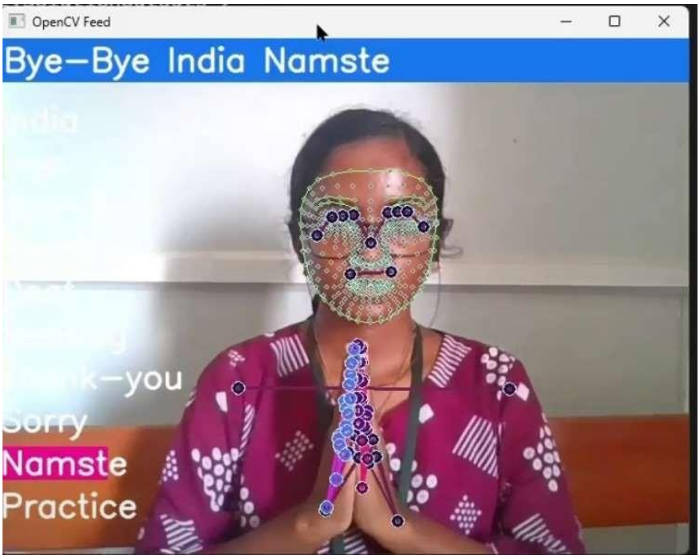
The epoch_categorical_loss graph in Fig. 4 shows the training method's capacity to minimize the loss function. The loss is not zero in the early epochs, but it varies. After some periods, the loss shows a discernible decrease and eventually approaches zero. This decrease implies that the model optimized its performance and convergence by successfully reducing mistakes and inconsistencies in its predictions.

To evaluate the model in real time, webcam frames are taken, landmarks are identified using the holistic model, and predictions are made based on the key points that are identified. Users can make sign language movements in front of the camera and watch the model's predictions; this interactive real-time testing allows users to evaluate the model's capacity to recognize gestures on the spot. Fig. 5.1. and Fig. 5.2. display the system's real-time testing.



Bye—Bye

Fig. 5.1 Real Time Testing (Sign: Bye-Bye)



Bye—Bye India Namste

Fig. 5.2 Real Time Testing (Sign: Namaste)

The system offers an interactive representation of the model's predictions for sign language motions during the real-time testing phase. The screen dynamically displays all recognized sign language activities, enabling users to see and follow the ongoing predictions. A probability bar indicates how confident the model is in its forecast for each recognized action. Users can see changes in real-time as the height of the bar reflects the probability. The detected actions are gradually gathered into a text box at the top of the screen. The system adds the predicted actions with the highest likelihood to this stack, thereby creating a sequential history of the recognized gestures. The intuitive interface allows users to observe the dynamics of the sign language recognition system in real-time, enhancing the model's predictability and interpretability. The stacked text box provides a convenient way to access the order of recognized activities during the testing phase.

6. Discussion

The study has made great strides toward developing assistive devices for the deaf and hard of hearing, with an astounding accuracy of 91.11%. The model's excellent accuracy in identifying and translating sign language motions is a sign of its ability to support inclusive communication.

The study exhibits a remarkable speed of 6.4 seconds to correctly estimate the sign, indicating its potential for real-time applications. This could lead to smooth and instantaneous sign language translation in various contexts. In situations like classrooms, customer service, and everyday interactions, when prompt communication is essential, this efficiency is critical. Our sign language translator's precision and efficiency make it a viable option for improving communication and accessibility for the hard of hearing, fostering inclusivity and understanding in a world where hearing people predominate.

6.1 Comparative Analysis

LSTM is a preferable deep learning technique for translating sign language because it can handle long-term memory, essential for dynamic movements, and efficiently record temporal connections. CNNs are great at recognizing static images, whereas LSTM sequential modeling is more suited to the temporal aspect of sign language. Vanishing gradient problems are handled by RNNs, such as LSTM, which guarantees effective learning over long sequences. Despite their strength, attention processes, and transformer models might not offer significant benefits for translating between sign languages. In conclusion, LSTM plays a critical role in the current work's accurate and context-aware sign language detection because to its unique architecture for sequential data.

7. Conclusion

In summary, the Sign Language Translator System represents a significant advancement in assistive technology for the deaf. The project's use of LSTM-based classification and key point detection allows it to identify motions in sign language with an excellent 0.91 accuracy score. The system's intuitive user interface (UI) and real-time testing capabilities enhance its practicality and usefulness by providing users with prompt and clear feedback while they make gestures. The project's extensive methodology, which entails training models, preparing datasets, and conducting real-time testing, contributes to developing a reliable communication tool to close communication gaps and advance inclusivity.

The intuitive interface makes the dynamics of the sign language recognition system visible to users in real time, enhancing the model's predictability and comprehensibility. The stacked text box provides a convenient way to refer to the order of recognized activities during the testing phase.

8. Current & Future Development

This study uses Indian Sign Language to demonstrate how technology-based solutions can improve communication and accessibility for India's deaf and hard-of-hearing population. To the extent that these technologies can meet the needs of this community, however, there is still more work to be done in their development and implementation. Future research endeavors may concentrate on converting the movement sequence into text, words, and phrases and subsequently converting that text into speech that can be heard. Other machine learning and deep learning techniques can be investigated for sign language identification and translation tasks in addition to the LSTM methodology used in this study. Convolutional neural network (CNN) models, for example, are appropriate for gesture detection tasks because they are good at extracting spatial characteristics from picture input. Furthermore, three-dimensional convolutional neural network (3DCNN) models can capture spatial and temporal dependencies in video data, which could lead to increased accuracy and robustness. Subsequent research endeavors may examine the potential for enhancing Indian Sign Language proficiency through the utilization of augmented reality (AR) and virtual reality (VR) technologies. It may be easier for users of AR and VR to practice and improve their sign language skills because of the more interactive and immersive experiences these technologies provide. Future research could also enhance the reliability and accuracy of sign language recognition systems, particularly for Indian Sign Language, which has its syntax and lexicon. It might be necessary to develop brand-new machine learning algorithms tailored explicitly to Indian Sign Language and improve data collection and annotation methods to ensure that systems are trained on a comprehensive and representative set of signs. The social and cultural factors that influence the adoption and utilization of sign language recognition technologies by deaf and hard-of-hearing individuals in India could also be examined in this study. This could include investigating how users feel about these technologies and how societal and cultural barriers affect their usability and effectiveness.

References

1. Deafness in India. In Wikipedia, The Free Encyclopedia, (2023).
2. Tateno S, Liu H, Ou J. Development of Sign Language Motion Recognition System for Hearing-Impaired People Using Electromyography Signal. Sensors, Kitakyushu 808-0135, Japan (2020).
3. K. Shenoy, T. Dastane, V. Rao and D. Vyavaharkar, "Real-time Indian Sign Language (ISL) Recognition," 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India, pp. 1-9, (2018).
4. H. El Hayek, J. Nacouzi, A. Kassem, M. Hamad and S. El-Murr, "Sign to letter translator system using a hand glove," The Third International Conference on e-Technologies and Networks for Development (ICeND2014), Beirut, Lebanon, pp. 146-150, (2014).
5. Lee, C.-C.; Gao, Z. Sign Language Recognition Using Two-Stream Convolutional Neural Networks with Wi-Fi Signals. Appl. Sci., 10, 9005, (2020).
6. Koller, O., & Ney, H., Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. Computer Vision and Image Understanding, 101(3), 108-125, (2006).
7. Y. Liu, R. Wang, S. Shan and X. Chen, "Structure Inference Net: Object Detection Using Scene-Level Context and Instance-Level Relationships," 2018 IEEE/CVF Conference on

Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 6985-6994, (2018).

8. Samaan, G.H.; Wadie, A.R.; Attia, A.K.; Asaad, A.M.; Kamel, A.E.; Slim, S.O.; Abdallah, M.S.; Cho, Y.-I. MediaPipe's Landmarks with RNN for Dynamic Sign Language Recognition. Electronics, 11, 3228, (2022).
9. Shagun Katoch, Varsha Singh, Uma Shanker Tiwary, Indian Sign Language recognition system using SURF with SVM and CNN, Array, Volume 14, 100141, ISSN 2590-0056, (2022).
10. CARDENAS, Edwin J. Escobedo ; CERNA, Lourdes Ramirez; CAMARA-CHAVEZ, Guillermo. Dynamic Sign Language Recognition Based on Convolutional Neural Networks and Texture Maps. In: CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES (SIBGRAPI), 32., 2019, Rio de Janeiro. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, (2019).
11. Sharma, S., Singh, S. Recognition of Indian Sign Language (ISL) Using Deep Learning Model. Wireless Pers Commun 123, 671–692 (2022).
12. Siddhartha Pratim Das, Anjan Kumar Talukdar, Kandarpa Kumar Sarma, Sign Language Recognition Using Facial Expression, Procedia Computer Science, Volume 58, Pages 210-216, ISSN 1877-0509, (2015).
13. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (June 2017), 84–90, (2017).
14. Subramanian, B., Olimov, B., Naik, S.M. et al. An integrated media pipe-optimized GRU model for Indian sign language recognition. Sci Rep 12, 11964 (2022).
15. Napier, J., & Leeson, L., Sign Language in Action. Springer Science and Business Media LLC, (2016).
16. Ko, S-K., Kim, C.J., Jung, H., & Cho, C., Neural Sign Language Translation Based on Human Keypoint Estimation. Applied Sciences, (2019).
17. von Agris, U., Zieren, J., Canzler, U. et al. Recent developments in visual sign language recognition. Univ Access Inf Soc 6, 323–362 (2008).
18. Shagun Katoch, Varsha Singh, Uma Shanker Tiwary, Indian Sign Language recognition system using SURF with SVM and CNN, Array, Volume 14, 100141, ISSN 2590-0056, (2022).
19. Alzubaidi, Mohammad & Otoom, Mwaffaq & Rwaq, Areen., A Novel Assistive Glove to Convert Arabic Sign Language into Speech. ACM Transactions on Asian and Low-Resource Language Information Processing. 22, (2022).
20. M. Garg, P. M. Pradhan and D. Ghosh, "Multiview Hand Gesture Recognition using Deep Learning," IEEE 18th India Council International Conference (INDICON), Guwahati, India, 2021, pp. 1-6, (2021).
21. P. Sheth and S. Rajora, "Sign Language Recognition Application Using LSTM and GRU (RNN)," in Proceedings of the IEEE 18th India Council International Conference (INDICON), Guwahati, India, 2023, pp. 1-6, doi: 10.13140/RG.2.2.18635.87846 (2023).