Sign Language Translation: A Study

Ajeet Singh Rajput, Sumitra Sureliya, Nishil Vani, Mohit Singh Tiwari, Pankaj Yadav

Computer Science & Engineering, Medicaps Univerty, Indore, India

Corresponding author: Ajeet Singh Rajput, Email: ajeetsinghrajput@gmail.com

Individuals without hearing or speech problems find it challenging to interact and communicate with those who do not have these impairments without the help of a translator. The sole means by which deaf and sign language is one way that non-verbal individuals can converse. Indian Sign Language is a unique language that contains its own syntax, grammar, vocabulary, and phonemes. The suggested study yields encouraging results and is validated on common datasets. A variety of techniques for data augmentation and mediapipe based feature extraction are used to assess the deep neural network.

Keywords: Sign language, Indian Sign Language, Speech and deaf people, Deep learning, LSTM, American Sign Language, BERT Transformer, INCLUDE-50, GASLA, CNN, SVM.

in: *Advancements in Communication and Systems*. Ed. by Ashish Kumar Tripathi and Vivek Shrivastava. Computing and Intelligent Systems, SCRS, India., 2023, pp. 529–540. DOI: https://doi.org/10.56155/978-81-955020-7-3-47

1 Introduction

Information is transferred from one person to another mostly through communication. The general population uses a variety of communication channels to express themselves. Thoughts are primarily communicated through language. Even though our community is multilingual, communication between speakers of different languages has greatly increased because to real-time language translation technologies. There are several tools that can translate local languages into English and vice versa, but very little attention is paid to translating sign language, which is used by the deaf and mute.

The World Health Organization (WHO) reports that more than 5% of people worldwide suffer from a speech or hearing impairment. Given the large number of deaf and silent individuals, an automated system is required to support the deaf community's learning, particularly for younger members of the population. Children with hearing loss or partial deafness have a major impact on their academic performance. Deafness and muteness lead to social and emotional imbalance in addition to intellectual loss. This results in social circle isolation because of a lack of communication channels. Financial losses are another consequence of hearing impairment, since the productivity of an individual is directly influenced by the efficiency with which he/she can communicate. Individuals with hearing impairments find it most challenging to interact with others in general. Being born deaf limits a person's ability to learn to talk. To help people communicate more effectively, several sign languages include hand gestures to represent verbal equivalents. The sign languages used differ between nations. The primary means of communication for the deaf and mute is this sign language. However, most people cannot comprehend this sign language, and conversing with someone who is hard of hearing is nearly difficult. We suggest utilizing the commonly utilized smartphone technology to bridge the communication gap that exists between the hearing impaired and the general population.

2 Literature Review

Numerous strategies have been investigated in latest years with reference to sign-language reputation, which has been the point of interest of tons studies and improvement [1, 2, 3]. Even with those developments, more efficient algorithms and systems designed for sign-language recognition are still required. Capturing and understanding complicated hand gestures and motions correctly is one of the hurdles in sign-language recognition. Various methodologies had been tested, along with the use of deep learning knowledge of models, including CNN, SVM, and LSTM, to enhance the performance and accuracy of recognition. Additionally, latest research targeted on improving communication accessibility for human beings using sign language through integrating real-time audio output and gesture-to-textual content conversion features.

Mhatre et al. [1] provided a method for real-time sign language recognition. The procedure comprises collecting videos of Marathi sign language gestures, dividing in the training and testing dataset, and training the model uses LSTM and dense layers in order to achieve 90–96% classification accuracy. The system produces words and audio in real-time using Google Text-to-Speech. We examine its precision, recall, and accuracy. Facial expressions are likely preserved in future plans for better interaction. One of the main contributions of the work is helping people who have difficulties with hearing talk with one another.

Sabunwala et al. [2] Examines the use of ML for the detection of Indian Sign Language (ISL). The authors discussed several methods such as Convolutional Neural Networks (CNN), K-nearest neighbors, and Surface Vector Machine (SVM) for gesture recognition. In the recommended method 210 gestures are collected as images, and then after the SVM model, SIFT descriptors and gray scale transformation, and the real-time gesture recognition algorithm is put to the test in several situations the system confirmed adaptability to different lighting conditions and angles, and achieved accuracy of 98%. The study shows that gestures can be successfully translated into text Using sign language recognition.

Prathap et al. [3] Propose two methods for sign language translation: bidirectional LSTM and BERT Transformer, both using pre-trained pose detection models. The INCLUDE-50 dataset was used for validation, and the best model achieved an accuracy of 89.5 %. The system uses smart phone cameras to convert hand movements into sign language and translate video frames into individual character. This technology can greatly benefit individuals who are deaf and mute by increasing their ability to communicate and learn.

M. Faisal et al. [4] introduce the Saudi Deaf Companion System (SDCS), a bidirectional communication system between deaf and mute. SDCS has three important modules: the Sign Recognition Module (SRM) for sign language interpretation, the Speech Recognition Synthesis Module (SRSM) for simple speech-to-text conversion, and the Avatar Module (AM) for visual representation for sign language The extensive Saudi Sign Language Database (KSU-SSL), the system contains 293 characters in various fields.

Jiao Li et al. [5] present a novel approach for American Sign Language (ASL) recognition system, addressing the excessive data collection issue by introducing GASLA, a wearable sensor-based solution which generates sentence-level perceptual data from word-level data. Their acceleration approach optimizes data generation speed. While their template strategy guarantees that generated sentences closely reflect the collected sentence data, their acceleration approach maximizes the speed at which data is generated. Studies show that GASLA reduces the setup time and significantly reduces maintenance costs, requiring only two samples per sentence compared to 10 samples in current systems, improving efficiency.

S. B. Abdullahi and K. Chamnongthai [6] present another example of dynamic sign word recognition. The model handles depth feature inconsistencies in spatio-temporal models by introducing a multivariate pairwise consistency feature ranking (PairCFR) approach. This new approach improves feature selection, and thus model performance spike. The effectiveness of the IDF-Sign is verified through real life tests using leap motion sensors. The IDF-Sign model's optimized forest classifier demonstrated the model's robustness and effectiveness with an amazing number of detections across a range of datasets.

O. El Ghoul et al. [7] provide the JUMLA-QSL-22 dataset, a data set that includes 6300 records of 900 sentences, obtained by utilizing true depth cameras to record continuous handwriting from various sources. Their research takes into account the progress made in handwriting methods as well as the need for a thorough sign language identification system to assist communities of people with hearing loss.

Z. Guo et al. [8] proposed a Locality-Aware Transformer (LAT) for video-based sign language translation. Their method uses a multi-stride position encoding (MSPE) system to improve local dependencies, an adaptive temporal interaction (ATI) module to capture local and non-local frame correlations, and a gloss computation function to aid overall understanding. The LAT framework proved to be effective, outperforming existing methods by incorporating flexibility local correlation while maintaining non-local temporal sampling modeling capabilities

R. Zhu et al. [9] present a soft fiber optic glove system with machine learning support for sign language recognition. The proposed system integrates fiber optic sensors and gyroscopes as a dual-glove system and is capable of translating ASL into text. Using deep learning, the system achieves high recognition accuracy for static 98.6% and for dynamic gestures 95%. This innovative approach enhances human-machine communication and provides a cost-effective solution to break down communication barriers between signers and non-signers.

F. Morillas-Espejo and E. Martinez Martin [10] developed a low-cost software, Sign4all, to facilitate communication between deaf and hearing to individuals the usage of the Spanish Sign Language (LSE) alphabet. Their work features a Sign Language recognizer and a digital avatar for sign interpretation. Utilizing Convolutional Neural Networks (CNNs), especially ResNet50, the system carried out an accuracy of 79.96% in sign recognition from RGB images. This innovative tool objectives to reduce social exclusion by supplying a realistic answer for everyday interactions without the need for interpreters.

Hezhen Hu et al. [11] present SignBERT+, a self-supervised pre-training framework designed to enhance sign language understanding (SLU). The technique consists of a model-aware hand prior into the pre-training method, making use of hand pose as a visible token derived from an off-the-shelf detector. The framework uses multi-level masked modeling strategies to model the statistics of sign data and improve interpretability. In terms of results, extensive experiments on isolated and continuous sign language recognition (SLR), and signal language translation (SLT) shows that SignBERT+ achieves state-of-the-art performance, supplying notable profits in SLU duties.

L. Chaudhary et al. [12] presented SignNet II, a two-way sign language interpretation method that uses a transformer-based model to facilitate communication between hearing and deaf or hard of hearing people. To improve performance, the model uses multi-feature transformers and, dual learning and metric embedding for sign similarity. Using the key point-based pose feature to translate dynamically across changes in video quality. Tested on the largest German sign language database, SignNet II shows leading BLEU scores, indicating significant improvements in automatic sign language processing.

Jiwei Hu et al. [13] propose a novel approach to sign language translation that optimally fuses spatial and temporal features. Their technique, called STFE-Net, enhances word identification and position estimation in sign language by merging a temporal feature extraction network (TFE-Net) and a spatial feature extraction network (SFE-Net). On continuous Chinese sign language datasets, the network outperformed various state-of-the-art techniques with promising BLEU scores. This development could greatly enhance hearing-impaired people's ability to communicate.

Table 1. Provides a summary of recent sign language translation systems.

Paper Title	Year	Methodology	Dataset	Output Format	Sign Language	3D Capabilities	Key Results
Sign Language Detection using LSTM[1]	2022	LSTM model	Marathi signs used daily	Sentence, Audio	Marathi	No	High training accuracy of 90-96%, could classify seven gestures
Indian Sign-Language Detection using Tensor-Flow and Deep-Learning [2]	2022	SVM modeling and cluster modeling	Indian Sign Language	Not specified	Indian	No	Converts signs to text using SVM, K-means clustering, and bag-of-words
ISLR: Indian Sign-Language Recognition [3]	2023	Bidirectional LSTM and BERT Transformer	INCLUDE -50 dataset	Sentence	Indian	No	Accuracy of 89.5%, development of a web application for ISLR

Paper Title	Year	Methodology	Dataset	Output Format	Sign Language	3D Capabilities	Key Results
Enabling 2 Way Communication of Deaf Using Saudi Sign-Language [4]	2023	3 modules: Sign Recognition Module, Speech Recognition and Synthesis Module, Avatar Module	King Saud Universit y Saudi (KSU) SSL database	Sentence, Audio	Saudi	Yes	System aids speech and hearing-impaired individuals by translating Saudi sign language into text and audio
Enhancing the applicability of sign-language-translation [5]	2024	GASLA solution, Acceleration Approach, Template Strategy	Not specified	Word, Sentence	America n	Yes	Enhanced ASL system effectiveness with reduced overhead costs
IDF Sign: Addressing Inconsistent Depth Features for Dynamic-Sign Word Recognition[6]	2023	IDF-Sign model, multivariate pairwise consistency feature ranking (PairCFR) approach	3D skeletal hand joint coordinat es	Word	America n, Greek, German	Yes	High recognition performance across various datasets
JUMLA-QSL-22: A Novel Qatari Sign-Language Continuous Data-set [7]	2023		JUMLA-QSL-22 dataset	Sentence	Qatari	No	Dataset is publicly accessible, promoting research and development in sign language recognition systems
Locality Aware Transformer for Video Based Sign-Language Translation [8]	2023	Adaptive temporal interaction, multi-stride position encoding, and gloss counting task	Not specified	Sentence	Chinese	No	Shows effectiveness on benchmark datasets, outperforming state of the art methods in sign-language translation tasks
Machine Learning Assisted Soft Fiber-Optic Glove System for Sign-Language Recognition [9]	2024	Soft fiber optic glove system for each hand, deep learning techniques	Not specified	Not specified	America n	Yes	98.6% accuracy in static motions and 95% in dynamic gestures for recognition

Paper Title	Year	Methodology	Dataset	Output Format	Sign Language	3D Capabilities	Key Results
Sign4all: A Low Cost Application for Deaf-People Communication [10]	2023	Sign Language recognizer, virtual avatar, ResNet50 CNN	Not specified	Sentence	Spanish	No	79.96% accuracy in sign recognition
SignBERT+: Hand Model Aware Self Supervised Pre Training for Sign-Language Understanding [11]	2023	Self-supervised pre-training, masked modeling techniques, downstream task adaptation	Not specified	Word, Sentence	America n	No	State of the art results in SLR and SLT tasks
SignNet II: A Transformer Based Two Way Sign-Language-Translation Model [12]	2023	Dual learning mechanism, sign similarity concept	German Sign-Language (GSL) benchmar k dataset	Word, Sentence	German	No	Significant improvements in BLEU scores over singly-trained networks
STFE Net: A Spatial Temporal Feature Extraction Network for Continuous Sign-Language-Translation [13]	2022	Relative position encoding with Transformer, Spatial-temporal feature extraction	Chinese continuou s sign language dataset	Sentence	Chinese	No	High BLEU scores on both the newly created dataset and two public datasets

3 Methodology

3.1 Long Short-Term Memory (LSTM)

A dataset is trained using an LSTM model for real-time signal language reputation. For the categorization of sign language activities, the LSTM model is employed, which has similarities to recurrent neural networks. The generation facts person-made hand alerts the use of a digital camera, tactics and analyzes them later. A general of 543 landmarks are obtained via extracting crucial hand and facial points the use of the Mediapipe holistic library. 30 sample videos of each sign gesture have been used to train the model, and the system's efficiency is evaluated based on accuracy, precision, and recall. The sign gesture produces textual content and audio output, with the Google Text-to-Speech library used to generate the audio output. The proposed system was able to classify seven gestures with a training accuracy of 90-96%. (see Figure 1.)

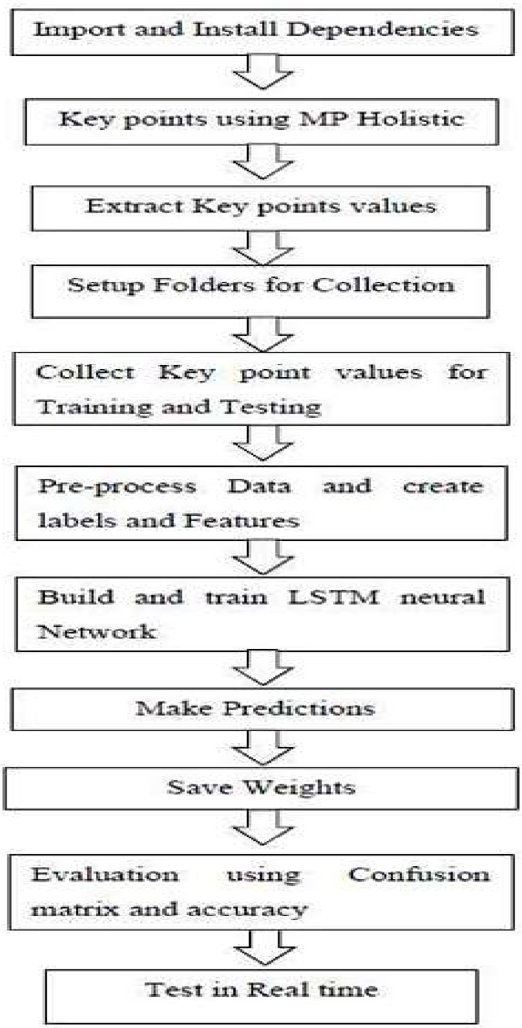

Figure 1. Flowchart of LSTM Methodology [1]

3.2 SVM Modeling and Cluster Modeling

It consists of amassing information through taking photographs of gestures with Python, pre-processing the images to turn them into grayscale and separate training and test units, and extracting feature using the SIFT descriptors. The SVM model classifies gestures, while a Mini batch K-means cluster version builds a visible bag of words for gesture detection. The system's accuracy is tested using metrics like precision scores and recall, with the results displayed on a confusion matrix. The methodology aims to create an efficient sign language detection system that can work under various conditions. (see Figure 2.)

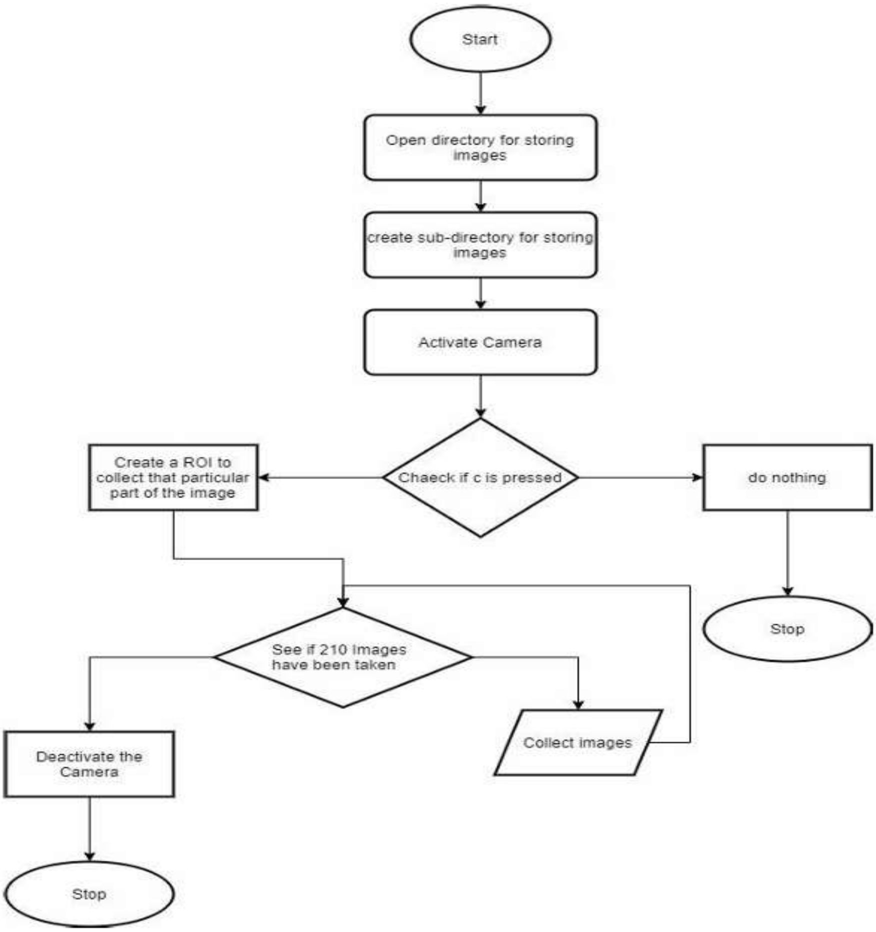

Figure 2. Flowchart of SVM Methodology [2]

3.3 BERT Transformer

The study presents the BERT Transformer technique, a deep learning strategy for the recognition of Indian Sign Language. Key points are extracted from video frames of sign language gestures the usage of pre-trained Media pipe models. The BERT (Bidirectional Encoder Representations from Transformers) architecture, that may deal with data sequences and collect context from both directions (past and future), then processes these key points. In order to translate sign language into text, the BERT layers examine the order of key points. The publication achieved an accuracy of 89.5% with the INCLUDE-50 dataset utilizing this method. (see Figure 3.)

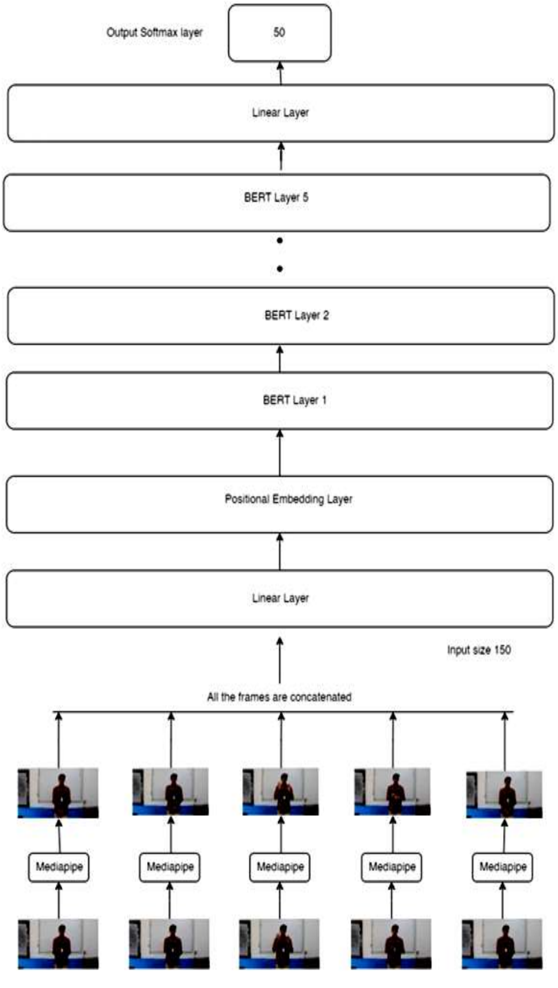

Figure 3. Flowchart of BERT Transformer Methodology [3]

3.4 GASLA Solution

A progressive strategy to improve the utility of American Sign Language (ASL) translation system is the GASLA method. It proposes a wearable sensor-based technique that overcomes the challenge of too much data being gathered for every word and phrase by producing sentence-level sensing data from word-level records. GASLA makes use of a template method to make created sentences extra comparable to terms which might be directly gathered, as well as an acceleration algorithm to maximize the pace at which data is generated. This technique substantially shortens the time wished for preliminary setup and preservation, boosts universal gadget efficiency, and makes ASL translation structures extra effective and user-friendly. (see Figure 4.)

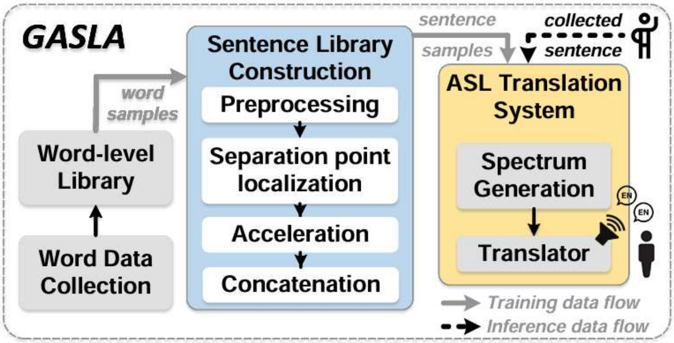

Figure 4. Overview of GASLA Methodology [5]

3.5 IDF-Sign Model

IDF-Sign, a spatial-temporal model that makes use of a multivariate pairwise consistency feature ranking (PairCFR) technique to pick the best model features. Temporal features are derived from 3D skeletal hand joint coordinates, while spatial capabilities are averaged from 3D video frames. The PairCFR approach ranks functions at distinct thresholds, and a threshold selection algorithm computes a midpoint value for every function based totally on its weight. The receiver operating characteristics (ROC) scheme identifies sensitive parameters and their relationship with the features, which are then used as modeling inputs. The methodology is showcased via experiments that utilize a leap motion sensor to record depth videos of sign language. These videos are analyzed in real-time. Machine learning classifiers are used to find the performance of the IDF-Sign model on different datasets, showcasing how accurate it is in detecting sign language. (see Figure 5.)

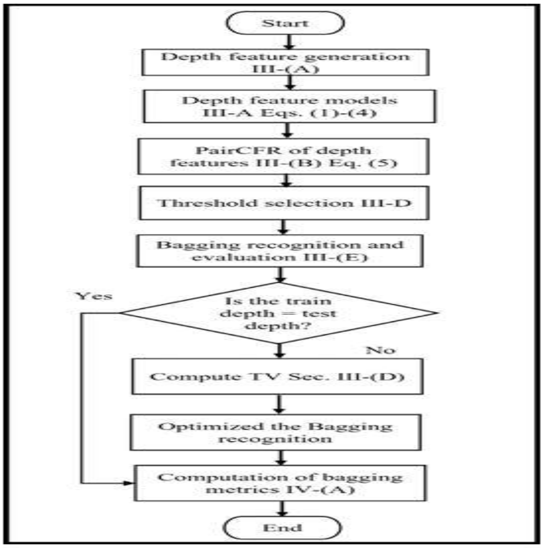

Figure 5. The IDF-Sign model approach procedure [6]

3.6 Fiber Optic Glove System

This sign language recognition system includes a glove system for each hand with fiber optic sensors, gyroscopes, printed circuit boards with wi-fi and batteries. The fiber optic sensors are multimode and have a soft liquid-center. It makes use of deep learning strategies for data analysis to recognize static and dynamic gestures of American Sign Language. The glove device can interpret a variety of sign language, which includes numbers, alphabets, phrases, and sentences, with high recognition accuracy. The sensors discover finger movements and interactions, while gyroscopes measure dynamic hand motions. The data accumulated is processed with the use of ML algorithms to get accurate gesture recognition, this is then implemented to the movement of a virtual avatar in a VR interface. The approach emphasizes the mixing of soft sensors and ML for effective human-device interaction. (see Figure 6.)

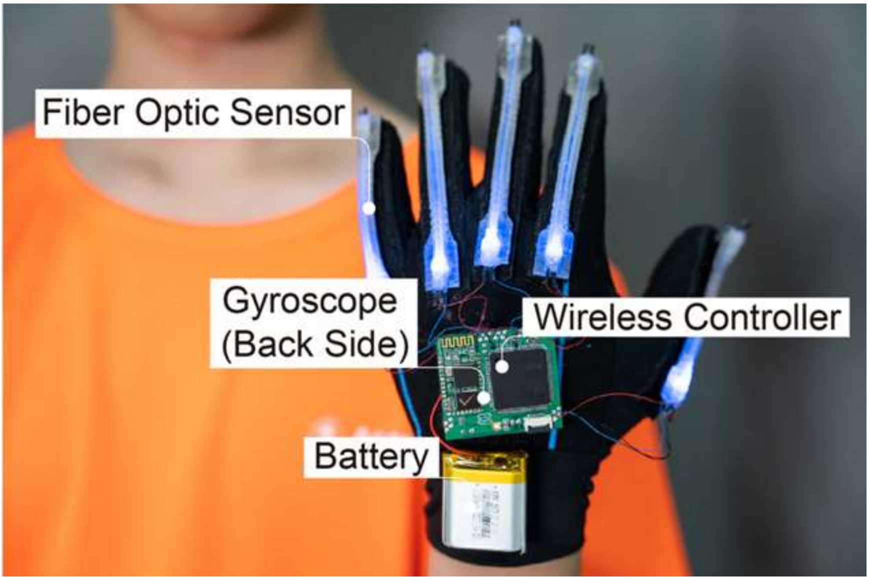

Figure 6. Soft fiber optic glove that can be put on both hands [9]

4 Conclusion

In the end, this study has shed significant insight on the subtleties and complexity of translating from sign language. We have outlined the advancements and obstacles that still need to be overcome in order to achieve precise and efficient translation between spoken and sign languages by looking at a variety of approaches and technologies. Notwithstanding the development made in ML and computer vision, extra enhancements and adjustments are nonetheless required to these technologies with the intention to accommodate the distinct features of sign languages. Additionally, to guarantee that translation structures are inclusive, available, and culturally appropriate, continuous collaboration among researchers, translators, and the deaf community is necessary. This study lays the basis for future improvements in sign language as we maintain to work closer to better conversation and comprehension across linguistic limitation.

References

[1] S. Mhatre, S. Joshi, and H. B. Kulkarni, "Sign language detection using LSTM," 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET), Dec. 2022. doi:10.1109/ccet56606.2022.10080705

[2] H. Sabunwala, R. Kurhe, A. Sarkar, A. Khetan, and M. Bachute, "Indian sign language detection using tensor flow and deep learning," 6th Smart Cities Symposium (SCS 2022), Dec. 2022. doi:10.1049/icp.2023.0659

[3] K. B. Prathap, G. D. Swaroop, B. P. Kumar, V. Kamble, and M. Parate, "ISLR: Indian sign language recognition," 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS), Apr. 2023. doi:10.1109/pcems58491.2023.

[4] M. Faisal et al., "Enabling Two-way communication of deaf using Saudi sign language," IEEE Access, vol. 11, pp. 135423–135434, 2023. doi:10.1109/access.2023.3337514

[5] J. Li, J. Xu, Y. Liu, W. Xu, and Z. Li, "Enhancing the applicability of sign language translation," IEEE Transactions on Mobile Computing, pp. 1–15, 2024. doi:10.1109/tmc.2024.3350111

[6] S. B. Abdullahi and K. Chamnongthai, "IDF-sign: Addressing inconsistent depth features for dynamic sign word recognition," IEEE Access, vol. 11, pp. 88511–88526, 2023. doi:10.1109/access.2023.3305255

[7] O. El Ghoul, M. Aziz, and A. Othman, "Jumla-QSL-22: A novel qatari sign language continuous dataset," IEEE Access, vol. 11, pp. 112639–112649, 2023. doi:10.1109/access.2023.3324040

[8] Z. Guo, Y. Hou, C. Hou, and W. Yin, "Locality-aware transformer for video-based sign language translation," IEEE Signal Processing Letters, vol. 30, pp. 364–368, 2023. doi:10.1109/lsp.2023.3263808

[9] R. Zhu et al., "Machine-learning-assisted soft fiber optic glove system for sign language recognition," IEEE Robotics and Automation Letters, vol. 9, no. 2, pp. 1540–1547, Feb. 2024. doi:10.1109/lra.2023.3347133

[10] F. Morillas-Espejo and E. Martinez-Martin, "Sign4all: A low-cost application for deaf people communication," IEEE Access, vol. 11, pp. 98776–98786, 2023. doi:10.1109/access.2023.3312636

[11] H. Hu, W. Zhao, W. Zhou, and H. Li, "Signbert+: Hand-model-aware self-supervised pre-training for sign language understanding," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 9, pp. 11221–11239, Sep. 2023. doi:10.1109/tpami.2023.3269220

[12] L. Chaudhary, T. Ananthanarayana, E. Hoq, and I. Nwogu, "SignNet II: A Transformer-based two-Way Sign Language Translation model," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–14, 2022. doi:10.1109/tpami.2022.3232389

[13] J. Hu, Y. Liu, K.-M. Lam, and P. Lou, "STFE-net: A spatial-temporal feature extraction network for Continuous Sign Language Translation," IEEE Access, vol. 11, pp. 46204–46217, 2023. doi:10.1109/access.2023.3234743