International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0341

e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2359-2364

SIGNLINK: Linking Audio, Text, and Hand Gestures Using Machine Learning Algorithms

Naheed Fatima[1], *M. Sridevi*[2]

[1,2]*Department of IT, GNITS, Hyderabad, India.*

Email ID: *naheedfatima2934@gmail.com*[1], *m.sridevi@gnits.ac.in*[2]

Abstract

Sign language, an important means of communication for the hearing impaired, often encounters barriers when interacting with non-signing people. Our "Sign-Link" aims to bridge this gap by seamlessly integrating audio and text communication with American Sign Language (ASL) through advanced hand gesture recognition technology. By utilizing state-of-the-art techniques, our project facilitates the conversion of multilingual audio input into text or translates text into ASL sign language animations with English dubbed voices. By using a convolutional neural network (CNN), our system accurately identifies and reads aloud the hand gestures enabling real-time recognition and interpretation whose accuracy ranges between 98% to 100% percent respectively. This innovative approach not only improves accessibility and inclusion, but also promotes meaningful communication between people, regardless of their hearing abilities.

Keywords: *ASL Animations; CNN; Gesture recognition; Multilingual conversion; Speech recognition*

1. Introduction

Signing exemplifies the persistence and adaptability of human communication. It is a lifeline for persons who happen to be deaf or hard of hearing, providing a complex and subtle mode of expression that goes beyond the limitations of spoken language [1]. Despite its importance, the use of signs often encounters communication hurdles, particularly when interacting with those who are new to its intricate lexicon of gestures and movements. The world has a complex tapestry of sign languages, each reflecting its speaker's distinct cultural and linguistic past. Sign languages offer a rich mosaic of expression, encompassing regional variations and dialects that mirror the diversity of spoken languages [2]. Despite their differences, these sign languages share a common purpose: to facilitate communication and foster connection within the deaf community and beyond. In this context, our proposed project, Sign Link, focuses on the integration of ASL, one of the most widely used sign languages in the United States, with audio and text-based communication modalities. ASL, with its distinct grammar and syntax, offers a robust foundation for our endeavor to develop seamless hand gesture recognition technology. Through SignLink, we aim to create a platform that enables the conversion of multilingual audio input into text and, subsequently, translates this text into ASL sign language animations. The aforementioned new solution employs modern neural network learning strategies, notably neural networks made up of convolutions (CNNs), to reliably recognize and comprehend the hand motions in actual time. [3-7]

2. Methodology

Our study intends to facilitate communication among individuals with hearing problems through the integration of speech and text via the use of American Sign Language (ASL) through flawless gesture recognition for hands. This is achieved as follows,

2.1. Text/audio to ASL Animation Conversion

This component allows users to input multilingual text or audio, which is then converted into ASL sign language animations. The process involves several steps:

2.1.1. Dataset Preparation

We utilized a dataset comprising 200 labeled videos with English dubbed voices.

2.1.2. Data Input

- **Audio:** Input through audio can be multilingual, comprising nine different languages with English as the primary language. These languages include

French, German, Spanish, Korean, Japanese, Hindi, Urdu, Telugu, and Arabic. The audio is converted to text and then sent for preprocessing.

- **Text:** Input for text is directly sent for preprocessing, as shown in Figure 1

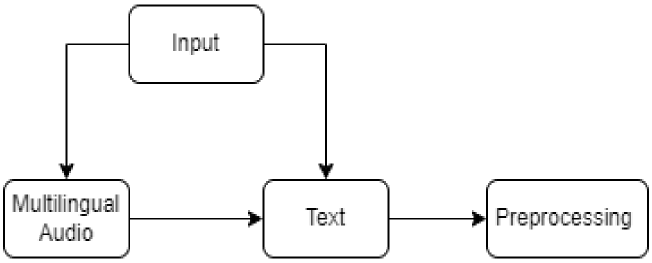


Figure 1 Input Illustration

2.1.3. Text Preprocessing

Before converting text to ASL, we preprocess it using techniques like tokenization, lemmatization, and removal of stop words.

- **Tokenization** is splitting apart the text into separate sentences or characters.
- **Lemmatization** shrinks phrases to their fundamental or base form, preserving uniformity in language representation.
- **Stop words** are words that include "the" or "and," are eliminated to improve the precision and effectiveness of linguistic analysis.

2.1.4. ASL Animation Generation

Once the text is processed, it is translated into ASL animations. This involves mapping each word or phrase to corresponding ASL gestures and movements. The animations are then generated to visually represent the input text in sign language, Refer Figure 2.

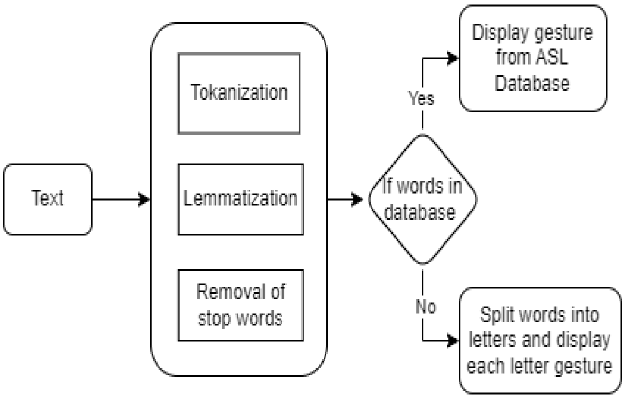

Text
Tokanization
Lemmatization
Removal of stop words
If words in database
Yes
No
Display gesture from ASL Database
Split words into letters and display each letter gesture

Figure 2 Animation Generation

2.1.5. Hand Gesture Recognition using CNN

In addition to text/audio to ASL conversion, our project incorporates hand gesture recognition using a CNN. This component enables real-time detection and interpretation of hand gestures, allowing seamless interaction with users. The process involves the following steps:

2.1.6. Input Processing

The CNN takes in 12,000 colored images of six different hand gestures, dividing them into small regions to analyze patterns. [8-10]

2.1.7. Feature Extraction

Four layers of convolution is used to determine characteristics that include borders and shapes, whereas pooling layers which follows each layer, lessen spatial dimensions while focusing on significant data.

2.1.8. Pattern Learning

Through multiple layers, the CNN learns increasingly complex patterns, discerning unique characteristics of each hand gesture class.

2.1.9. Classification

Five fully connected layers, each followed by a dropout layer, interpret the learned features, assigning probabilities to each gesture class using the SoftMax function.

2.1.10. Training and Validation

The model trains on a dataset of a total 12000 images in 6 classes namely, “Hello”,” I love You”,” Okay”,” Thankyou”,” Yes”,” No”. The folders are divided into 80% and 20% for training and validation respectively and parameters are adjusted to minimize errors while validating its performance on unseen data. [11], Figure 3 Explains System Architecture.

2.1.11. Model Saving

After training, the CNN saves its architecture and trained weights for later use in recognizing hand gestures in real-time applications.

2.1.12. Real-Time Detection

Once trained, the CNN model is capable of real-time detection of hand gestures from video input. It processes each frame of the video, identifies hand gestures, and provides output labels indicating the recognized gestures and reads it aloud using TTS library ‘pyttsx3’. [12]

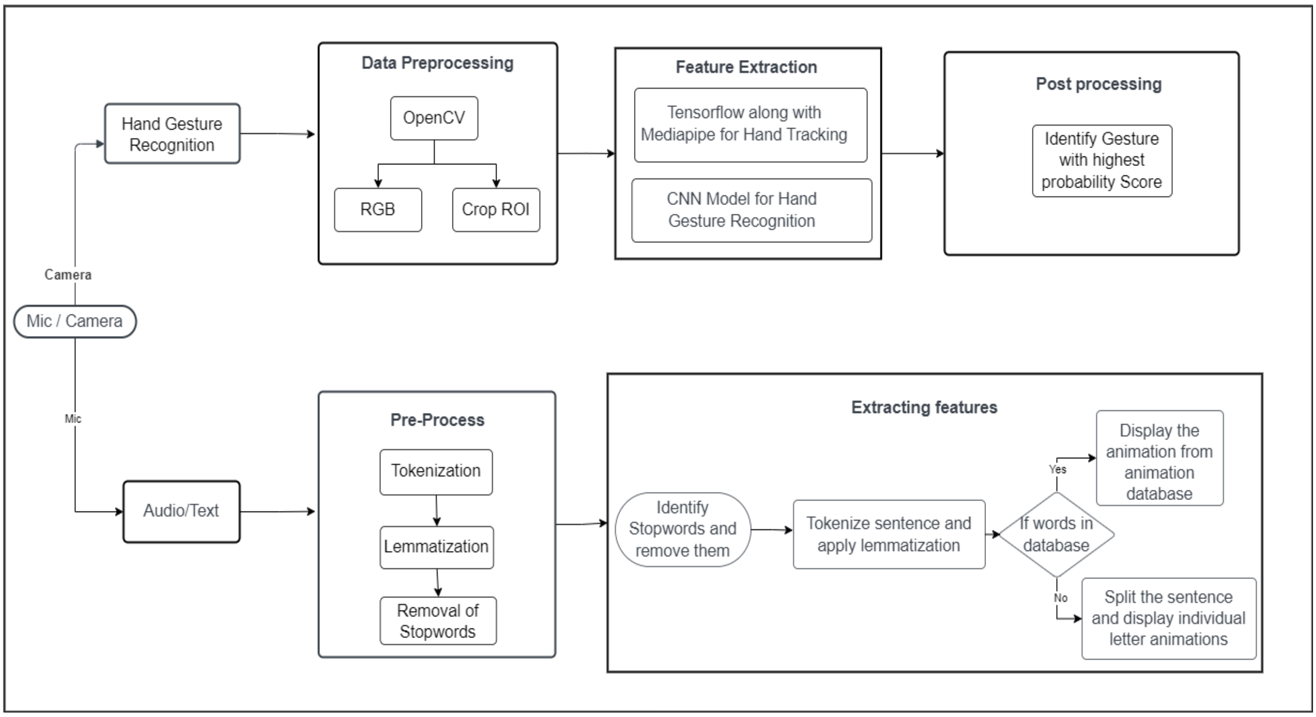
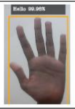
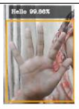
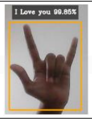
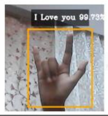
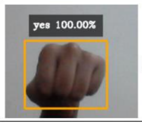
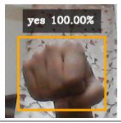
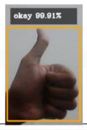
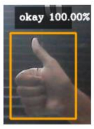
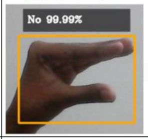
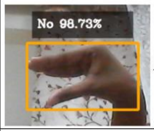
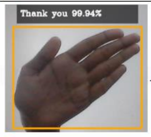
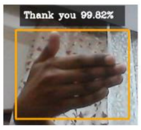

Hand Gesture Recognition
Data Preprocessing
OpenCV
RGB
Crop ROI
Feature Extraction
Tensorflow along with Mediapipe for Hand Tracking
CNN Model for Hand Gesture Recognition
Post processing
Identify Gesture with highest probability Score
Camera
Mic / Camera
Mic
Audio/Text
Pre-Process
Tokenization
Lemmatization
Removal of Stopwords
Extracting features
Identify Stopwords and remove them
Tokenize sentence and apply lemmatization
If words in database
Yes
No
Display the animation from animation database
Split the sentence and display individual letter animations

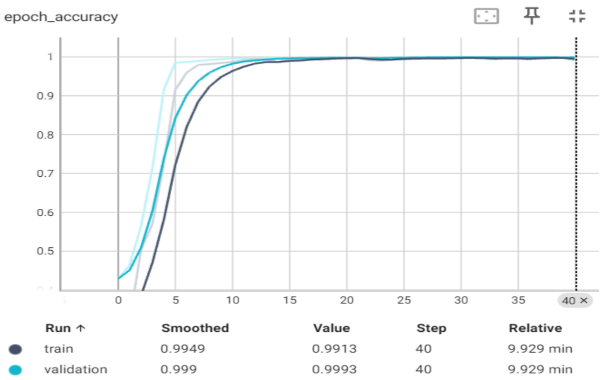
Figure 3 System Architecture

3. Results and Discussion

The text/audio to ASL conversion component effectively processes multilingual inputs, employing tokenization, lemmatization, and removal of stop words to enhance language processing efficiency. ASL animations generated from the converted text/audio accurately represent the intended messages through mapped gestures and movements. Furthermore, the hand recognition component, which is fueled by a CNN design, achieves effective instantaneous recognition as well as interpretation of hand motions using a camera feed as shown in the Table 1. The model that has been trained offers a high accuracy of range 98% to 100% as shown in the Figure 4, enabling effortless communication with clients. Evaluation and testing confirm the system's capacity for better inclusion and accessibility in channels of communication among individuals with hearing impairments. Through iterative development and user feedback, the system continues to evolve, striving to meet the diverse needs of its users and improve their communication experiences. The resultant screenshots are below under output screenshots from Figure 5 to 16. [13]

Table 1 Interpretation of Various Hand Motions in Plain and Varied Backgrounds

Gesture Names	Gesture Accuracy (Plain background)	Gesture Accuracy (Varied background)
HELLO	 Accuracy : 99.96%	 Accuracy : 99.86%
I LOVE YOU	 Accuracy : 99.85%	 Accuracy : 99.73%
YES	 Accuracy : 100%	 Accuracy : 100%
OKAY	 Accuracy : 99.91%	 Accuracy : 100%
NO	 Accuracy : 99.99%	 Accuracy : 98.73%
THANK YOU	 Accuracy : 99.94%	 Accuracy : 99.82%



Run ↑	Smoothed	Value	Step	Relative
● train	0.9949	0.9913	40	9.929 min
● validation	0.999	0.9993	40	9.929 min

Figure 4 The Graph Shows Epochs and Accuracy Taken on X-Axis and Y-Axis Respectively

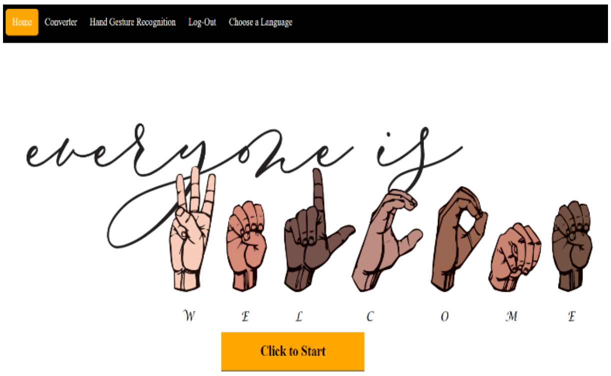
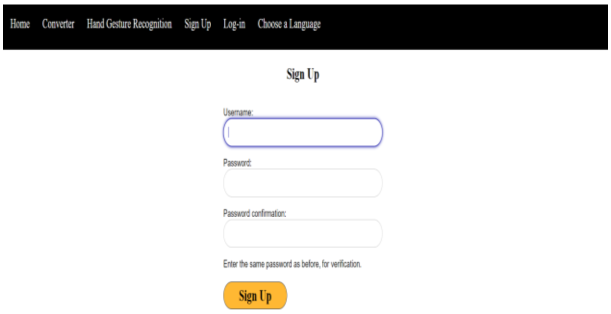


Figure 5 Home Page



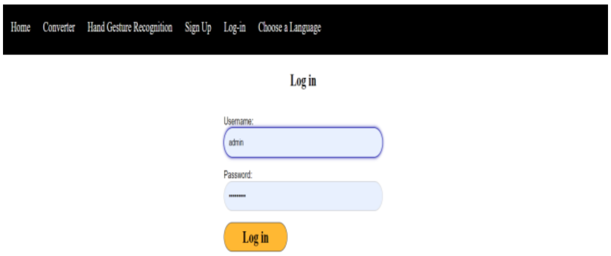
Username:

Password:

Password confirmation:

Enter the same password as before, for verification.

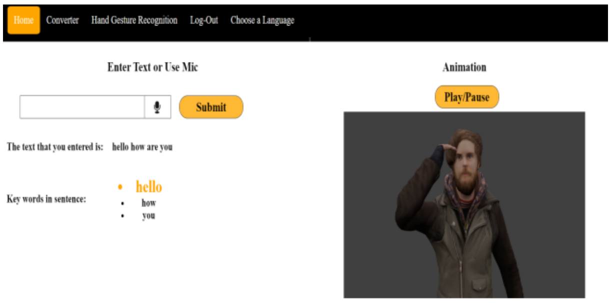
Figure 6 Sign Up Page



Username: admin

Password: ••••••••

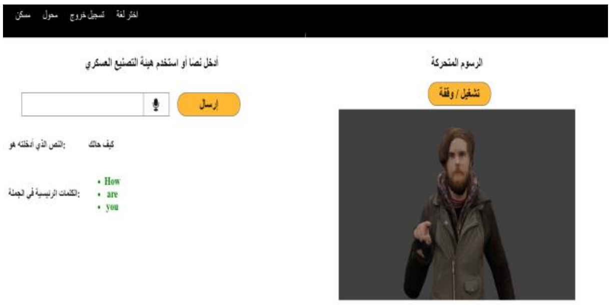
Figure 7 Login Page



The text that you entered is: hello how are you

Key words in sentence:
- hello
- how
- you

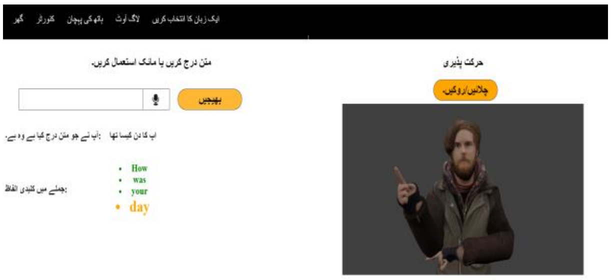
Figure 8 Text Conversion of “Hello, how are you” in English



النص الذي أدخلته هو: كيف حالك

الكلمات الرئيسية في الجملة:
- How
- are
- you

Figure 9 Text Conversion of “How are you” in Arabic



آپ نے جو متن درج کیا ہے وہ ہے: ب کا دن کیسا تھا

جملے میں کلیدی الفاظ:
- How
- was
- your
- day

Figure 10 Text Conversion of “How was your day” in Urdu

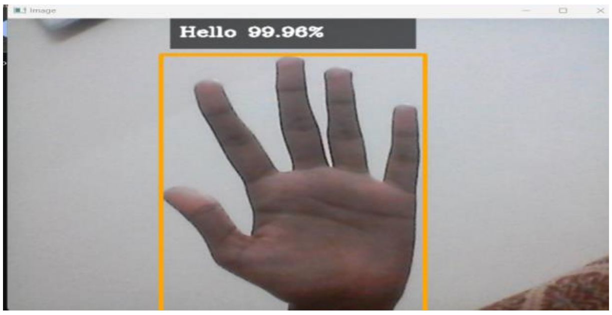


Figure 11 Interpretation of the Hand Motion for “Hello”

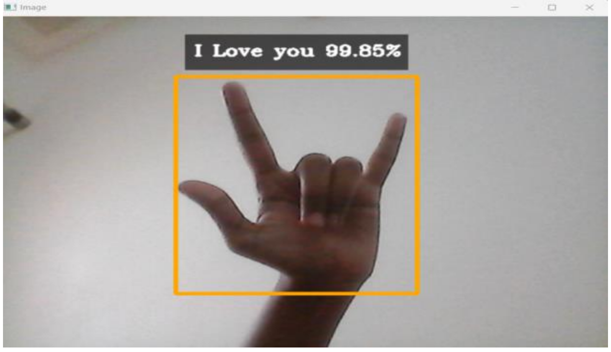


Figure 12 Interpretation of the hand motion for "I love you"

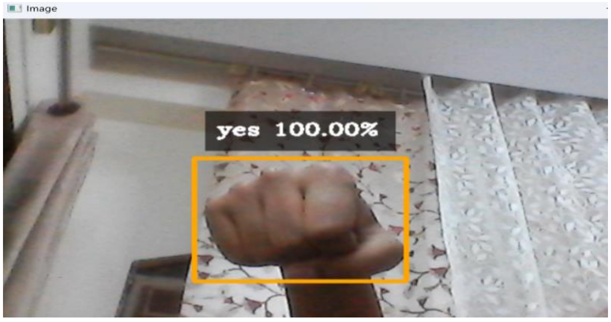


Figure 13 Interpretation of the hand motion for "Yes"

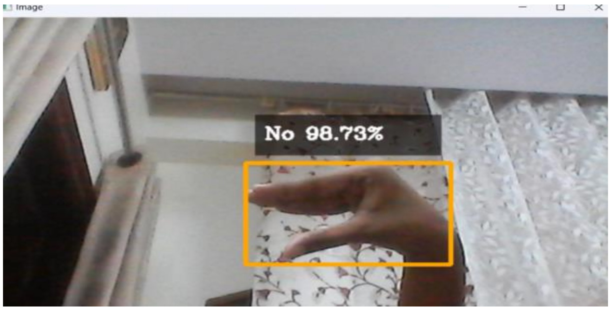


Figure 14 Interpretation of the hand motion for "No"

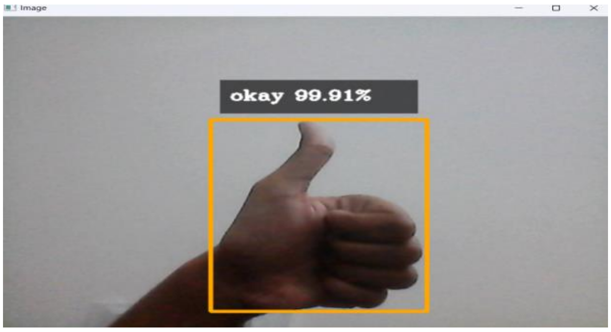


Figure 15 Interpretation of the hand motion for "Okay"

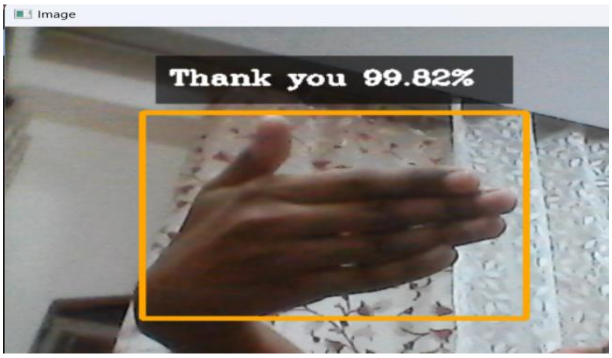


Figure 16 Interpretation of the hand motion for "Thank You"

Conclusion

This research provides an all-encompassing approach that combines gesture-based identification and text/audio-to-sign Language (ASL) interpretation to boost communication accessibility for individuals with hearing disorders. The technique successfully interprets text/audio information to ASL motion and interprets gestures made with hands in actual time by utilizing intricate methods involving Convolutional Neural Networks, or CNNs, and machine learning algorithms providing us with an accuracy that ranges from 98% to 100% respectively. The methodology encompasses data collection, initial processing, feature extraction, model learning, and real-time detection, culminating in a methodical approach to system development. The effectiveness of ASL interpretation and motion tracking might be enhanced, the information set could be extended to incorporate a broader spectrum of motions and language versions, and new aspects like the detection of facial emotions could be investigated to enrich the interaction competence. [14]

References

[1]. R. K, P. A, P. K S, S. Sasikala and S. Arunkumar, (2023) "Hardware Implementation of Two Way Sign Language Conversion System," (IHCSP), BHOPAL, India, 2023, pp. 322-326, doi: 10.1109/IHCSP56702.2023.10127188.

[2]. J. Hu, Y. Liu, K. -M. Lam and P. Lou, (2023) "STFE-Net: A Spatial-Temporal Feature Extraction Network for Continuous Sign Language Translation," IEEE Access, doi: 10.1109/ACCESS.2023.3234743.

[3]. Reddy, B. R., Rup, D. C., Rohith, M., & Belwal, M. (2023, March). Indian sign language generation from live audio or text for tamil. In 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 1507-1513). IEEE.A. Dixit et al, (2022)"Audio to Indian and American Sign Language Converter using Machine Translation and NLP Technique,"(ICICICT), pp. 874-879, doi: 10.1109/ICICICT54557.2022. 9917614.

[4]. T. A. Siby, S. Pal, J. Arlina and S. Nagaraju, (2022) "Gesture based Real-Time Sign Language Recognition System," (CSI), Trivandrum, India, pp. 1-6, doi: 10.1109/CSI54720.2022.9924024

[5]. Kothadiya, Deep, et al. (2022) "Deepsign: Sign language detection and recognition using deep learning." Electronics 11.11: 1780.

[6]. K. R. Prabha, B. Nataraj, B. Thiruthanikachalam, S. Surya NarayananS and M. Vishnu Prasath, (2023) "Audio to Sign Language Translation using NLP," 2023 (STCR), Sathyamangalam, India, pp. 1-4, doi: 10.1109/STCR59085.2023.10397050.

[7]. K. J.C and D. Nagarajan, (2023) "Real Time Automated Sign Language Recognition and Transcription with Audio Feedback," IEEE (ICCST), Pune, India, pp. 1-6, doi: 10.1109/ICCST59048.2023.10474276.

[8]. Lopez, J. A., Murtagh, I., & Castilho S., (2023)" Spoken Language to Irish Sign Language Machine Translation: A Linguistically Informed Approach".

[9]. Chandarana, N., Manjucha, S., Chogale, P., Chhajed, N., Tolani, M. G., & Edinburgh, M. R. M. (2023, October). Indian Sign Language Recognition with Conversion to Bilingual Text and Audio. In 2023 International Conference on Advanced Computing Technologies and Applications (ICACTA) (pp. 1-7). IEEE.

[10]. Shin, Jungpil, et al., (2023) "Korean sign language recognition using transformer-based deep neural network." Applied Sciences 13.5: 3029.

[11]. Tarrés, L., Gállego, G. I., Duarte, A., Torres, J., & Giró-i-Nieto, X., (2023) "Sign language translation from instructional videos". In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5624-5634).

[12]. De Coster, Mathieu, et al., (2023) "Towards the extraction of robust sign embeddings for low resource sign language recognition." arXiv preprint arXiv:2306.17558.

[13]. Tmar, Z., Othman, A., & Jemni, M., (2013)". A rule-based approach for building an artificial English-ASL" (March) corpus. IEEE (pp. 1-4).

[14]. T. Vichyaloetsiri and P. Wuttidittachotti, (2017) "Web Service framework to translate text into sign language," (CITS), pp. 180-184.