Real-Time Sign Language and Audio Conversion Using AI

Boobal A
Department of Computer Science and Engineering
Kalasalingam Academy of Research and Education
Virudhnagar, TamilNadu,India
boobal4002@gmail.com

Loyola jasmine J
Department of Computer Science and Engineering
Kalasalingam Academy of Research and Education
Virudhnagar,TamilNadu,India
loyolajasminej@gmail.com

Chennama Charan Kesava Reddy
Department of Computer Science and Engineering
Kalasalingam Academy of Research and Education
Virudhnagar, TamilNadu, India
charankesavareddy@gmail.com

Challa Akshay Reddy
Department of Computer Science and Engineering
Kalasalingam Academy of Research and Education
Virudhnagar,TamilNadu, India
akshaychalla0923@gmail.com

Chaluvadi Bala Venkata Sai Rohith
Department of Computer Science and Engineering
Kalasalingam Academy of Research and Education
Virudhnagar,TamilNadu,India
rohithchaluvadi2003@gmail.com

2024 International Conference on Communication, Control, and Intelligent Systems (CCIS) | 979-8-3315-2820-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/CCIS63231.2024.10932061

***Abstract*—The appearance of audio conversion and the appearance of gesture language have greatly advanced communication technologies, especially for hearing impairment and blind. These double -purpose technologies promote communication between these groups, provide opportunities for expression and speech behind. The technology uses advanced artificial intelligence algorithms to provide real-time speech and sign language translation to improve accessibility and inclusivity. It is essential to protect data privacy, maintain the accuracy of AI models to avoid misinterpretations, and address ergonomic issues associated with long-term use. Technology converts audio entry into a gestic signal that can be seen by avatars using computer vision and natural language processing. There is a depth learning model that improves voice recognition and the acquisition of signs, and is fully adapted to many languages and sign dialects. Addressing such health and safety issues is essential to advancing the communication and optimization of this advanced technology.**

Keywords*— Sign language translation, Real-time gesture recognition, Speech-to-sign language, Sign language to audio, Deep learning, Convolutional Neural Networks (CNNs), Natural language processing (NLP), Audio-to-sign language conversion, Machine learning, Gesture recognition, Computer vision in sign language, Django, MySQL, Indian sign language (ISL)

I. INTRODUCTION

Using activities for the transfer of different sign language meanings is the main means of communication for deaf and dumb people or people who are difficult to hear. Nevertheless, many sign language users have suffered extreme social exclusion due to the difficulty of communicating between them and the hearing crowds.

Despite the natural ability, 466 million people (5% of the world's population) have speech disorders, leading to limited social isolation and economic opportunities.

The communication barrier is estimated to cost $ 750 billion each year. From the perspective of employers, this also affects productivity. To solve this problem, we plan to provide a conducive environment and avoid communication breakdowns. Spontaneous, immediate communication. By promoting all connections and capabilities, this pragmatic paradigm is designed to increase availability and inclusion.

II. SIGN LANGUAGE TRANSLATOR

A Sign Language Translator using real-time AI technology acts as an innovative tool to overcome communication gaps between sign language users and non-sign language users. By combining image processing and machine learning, it records sign language motions through cameras or sensors and turns them into spoken or written words, facilitating fluid communication. This real-time system has issues such as providing high accuracy, accounting for the complexity of sign languages, and securing user privacy. It must also adapt to multiple sign languages and cultural contexts to cater to diverse users, improving accessibility in educational, social, and public services.

III. AUDIO TO SIGN LANGUAGE

An audio-to-sign language converter translates spoken words into sign language, aiding deaf and hard-of-hearing individuals. Using Automated Speech Recognition (ASR), it transcribes sounds in real-time, then employs digital avatars or animations to portray the signs. Natural Language Processing (NLP) helps capture context, while computer vision aids in comprehending gestures, providing smooth communication across many domains like education, healthcare, and daily encounters. This technology supports inclusivity by facilitating broader involvement and interaction for those with hearing impairments.

IV. RELATED WORK

To improve accuracy and efficiency, recent advances in audio-to-sign language translation systems have focused on a combination of machine learning and natural language processing (NLP) techniques. Efficient systems handle complex sign language grammars and provide instant translation capabilities using large databases and structured language processing. In addition, web application trends using user -friendly interfaces and immediate visual feedback to ensure smooth communication significantly improves the availability of people who do not hear and immediate visual feedback [1]. The literature review focuses mainly on one-way solutions that limit communication for people with speech and hearing impairments, highlighting the shortcomings of two-way sign language translation systems. While interest in languages such as Indian Sign

Language is growing, there are not many quick fixes available, indicating the urgency of improving this area.[2]. Current speech recognition technologies frequently disregard real-time translation into Indian Sign Language (ISL), highlighting a need for more accessible alternatives. Moreover, the limited availability of sign language instruction in India amplifies communication issues, underlining the significance of novel applications that employ developments in machine learning and image processing [3].

Recent improvements in audio-to-sign language converters enable real-time communication for the deaf by merging speech recognition and machine learning to translate spoken language into animated gestures [4]. This literature review focuses on innovative applications of deep learning, specifically CNN and NLP, that convert sign language to text and speech with 97.6% accuracy, significantly improving hearing by combining gesture recognition with neural synthesis of text-to-speech communication between hearing and deaf people [5]. Advancements in Sign Language Recognition focus on voice-to-text translation and animations, enhancing comprehension for students with hearing impairments while tackling issues including feature extraction and speaker variability [6]. This literature analysis identifies communication challenges for the hearing-impaired in sub-Saharan Africa, notably Nigeria, due to resource limits. It underlines the need for a dataset for Nigerian Sign Language and machine learning for real-time sign-to-speech synthesis to increase social inclusion and opportunity for those with hearing disabilities [7].

Current text-to-sign language systems are constrained by a lack of comprehensive sign language corpora, particularly in Indian Sign Language (ISL), which inhibits effective communication for the hearing handicapped. The proposed project intends to enhance this process by combining grammar rules and real-time video outputs, making it more user-friendly and accessible for both deaf and hearing users while addressing shortages in ISL resources [8]. This literature analysis provides a novel method for transforming sign language into text with intonation markup for voice synthesis, focused on the Kazakh language. It displays great accuracy in gesture recognition and highlights major societal benefits, facilitating communication for those with impairments and allowing adaption across different sign languages [9]. Advancements in multimodal generative modelling focus on text-to-image and text-to-video, with the Make-An-Audio system tackling text-to-audio issues with a prompt-enhanced diffusion model for high-fidelity audio creation [10].

V. OBJECTIVES

Instant Sign Language Translator was created with the aim of bridging the communication gap between sign language users and non-sign language users by providing fast, accurate and user-friendly translation services. The goals are as follows:

- Enable seamless communication: Provide a system that makes it easier for people who use sign language to communicate with people who do not use sign language, thereby improving social inclusion and accessibility in a variety of settings, including public spaces, education and daily life. - Ensure accurate translation: Use advanced image processing and machine learning techniques, especially convolutional neural networks (CNN), to ensure accurate translation. This will reduce errors and misunderstandings by accurately recognizing sign language gestures and translating them into spoken or written language.

Implement instant processing: To provide immediate feedback and eliminate noticeable pauses in discussions, implement a system that can instantly understand and process gestures.

To serve a world user base and address unique regional has a difference in sign language, translators must be able to support a variety of sign languages and dialects.

Manage security and privacy: Establish strict data protection procedures to protect user privacy and ensure secure handling of visual materials, taking care to address any potential concerns about the collection and use of personal data.

Prioritize user comfort: Ensure that technology is ergonomically designed to provide a comfortable user experience, reduce physical stress, and ensure long-term use of the system.

VI. PROPOSED WORK

A. *Sign to Audio Recognition Algorithm*

The recommended technique of audio identification is used by conventional neural networks (CNN) to classify hand movements in certain categories, such as “well ", "peace", 'thumb' and others. The process begins with the collection of photos of different hand movements and classifications in training and test catalogues. The goal of data augmentation is to make the training set more diverse. In addition to the top layer, forward learning is effectively implemented using pre-trained Mobile Net models. A sequential architecture with a global averaging layer and a dense layer with SoftMax activation for classification is used to fine-tune the model.

The model is then trained using data generators and evaluated using metrics such as accuracy, precision, and F1 score. The trained model will be detained in the future implementation of actual applications and thereby realize a signature conversion for real time to get communication assistance.

B. *Audio to Sign Recognition Algorithm*

The suggested method for audio to sign recognition takes spoken words through the speech recognition library, transforming them into text via Google’s Speech API. The algorithm then matches the identified text to predetermined sign language movements (represented as images or GIFs) saved in a database. The audio input will be first recorded, then after its converted into text, and then compared to a predefined list of sentences. If a match is found, then the associated gesture picture or GIF will be displayed. Then the technology links each spoken letter to its corresponding sign language gesture.

While the system does not include standard deep learning training, it relies on the accuracy of the voice recognition and the predetermined gesture database. The system continuously checks voice input and reproduces the correct gesture instantly. It runs until the user decides to leave, which is an important tool for communication between people with hearing impairments and others.

C. System Architecture

The diagram demonstrates a system that detects sign language motions using a Convolutional Neural Network (CNN). It starts by recording real-time footage of a person signing through a camera. A single frame is retrieved from this movie and processed with segmentation techniques to isolate the hand gesture from the background. This isolated hand image is then put through a CNN trained on an Indian Sign Language (ISL) dataset, where the network's layers recognize and classify characteristics unique to each move (Figure 1). The system outputs a label that matches to the recognized ISL gesture, giving an instant translation of the sign language into text or speech. This provides real-time communication for persons with speech or hearing impairments, boosting accessibility and inclusivity.

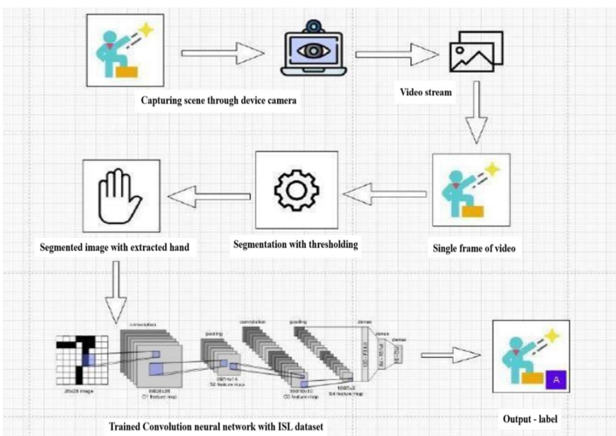

Fig 1:- System architecture

D. Software workflow

The suggested system provides seamless communication between voice, text, and sign language, commencing with user verification to assure secure access. After logging in, users can utilize modules for various translations: converting audio to sign language via speech recognition and text-to-sign conversion, translating sign language gestures to audio using a Convolutional Neural Network (CNN), and providing a detailed pathway from audio to text to sign. First, type the audio input, convert it to the text, and then compare it to the list of predetermined sentences. If you find the appropriate item, the associated gesture image or GIF will appear. The technology connects each spoken letter with its corresponding sign language gesture.

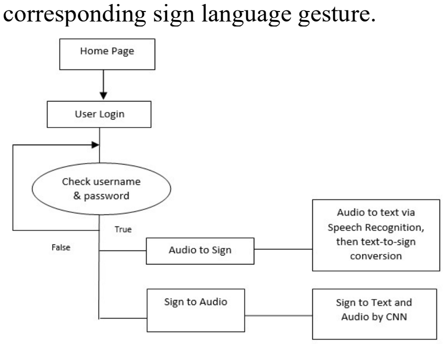

Fig 4:- Flow Diagram

VII. MODEL TRAINING & RESULTS

A. Datasets used

A dataset including roughly 5,076 samples (Figure 2) was utilized for the training and testing of the sign language to audio conversion system. The dataset comprises diverse hand movements, allowing the Convolutional Neural Network (CNN) to precisely identify and categorize ISL signals, so enabling real-time communication via text or audio output.

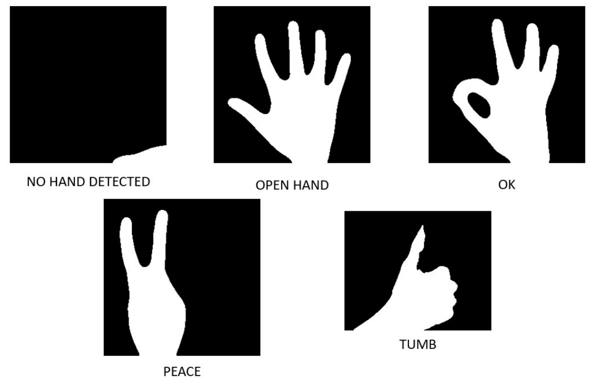

Fig 2:- Datasets for sign language

B. Accuracy and loss analysis

This accuracy graph provides Training and Validation Accuracy across 20 epochs. Both accuracies approach high levels, with training accuracy swiftly improving and validation accuracy remaining stable. A modest dip around the 17th epoch has low influence, showing effective learning and robust generalization with little overfitting. (Figure 3).

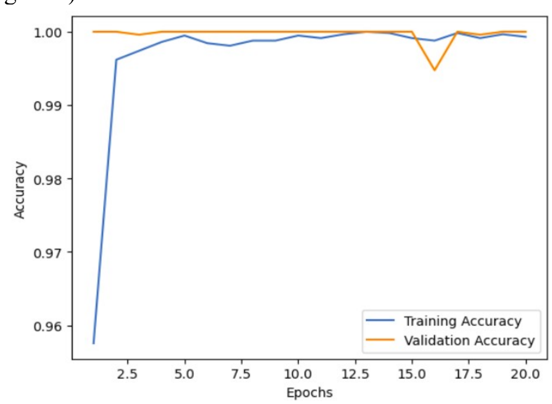

Fig 3:-Graph (accuracy vs epochs)

This graph illustrates Training and Validation Loss over 20 epochs. The quick drop in training loss followed by stabilization suggests good learning. The low, steady validation loss reflects high generalization, with minimal overfitting. A modest increase at epoch 17 doesn't influence the model's overall performance (Figure 4).

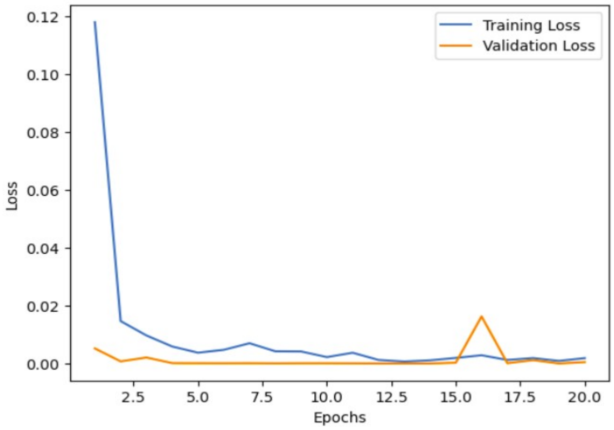

Fig 4:- Graph (loss vs epochs)

The confusion matrix displays flawless classification outcomes in every category, devoid of any misclassifications. The projected labels for "No Hand Detected," "Okay," "Open Hand," "Peace," and "Thumb" match all real labels (Figure 5).

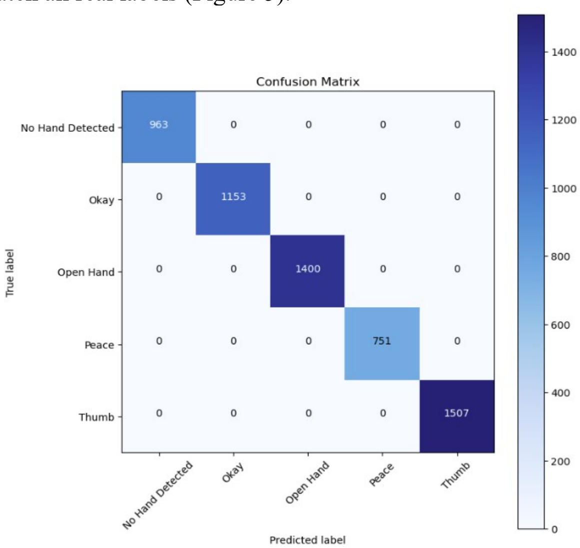

Fig 5:- Confusion matrix

The output metrics reveal an exceptional model performance with 98% accuracy, precision, recall, and F1-Measure, indicating highly accurate predictions across all tested data. (Table 1).

Metric	**Accuracy**
Accuracy	0.979
Precision	0.968
Recall	0.998
F-Measure	0.988

Table 1:- Evaluation table

C. *Home Page*

This web application (Fig6) focuses on overcoming communication barriers for the deaf and hard of hearing by using advanced artificial intelligence technology to convert sign language speech and sign language to speech in real time. The platform has a simple user interface with reception, connection and contact settings.

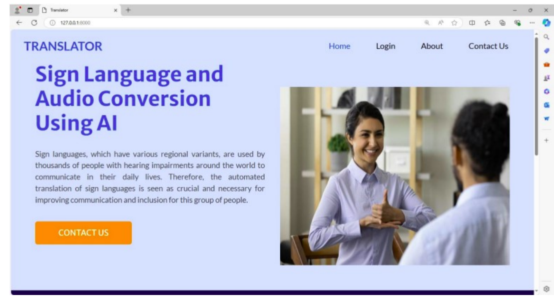

Fig 6:- Home page

D. *User Registration*

User registration process for real-time speech sign language and speech sign language application (Figure 7). The user's recording procedure is an important factor in the program designed to secure the user's safe access and personalized experience. The registration page reflects important user information for creating an account and simplifying user maintenance. If you fill out the registration form, the user can send information and create an account. The application will check the provided email address so that users can access important emails and update.

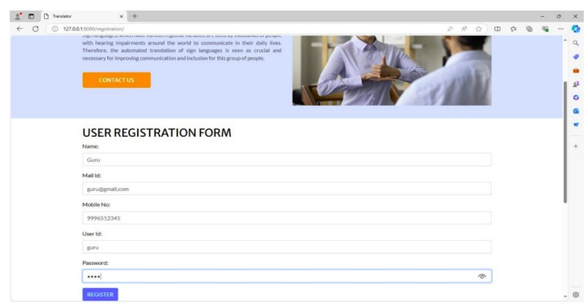

Fig 7:- User Registration page

E. *User login*

The login page serves as the gateway for users to access their accounts within the program (Figure 8). This important component ensures secure and efficient user authentication. This allows users to use the language conversion system function for audio and gestures. After entering his username and password, the user submits his entry to the system and accesses his account. After submission, the system verifies the credentials against the saved data. If the credentials provided are valid, the user will be redirected to the main dashboard where they can utilize the app's features, if the credentials are invalid the user will be prompted to enter the information again.

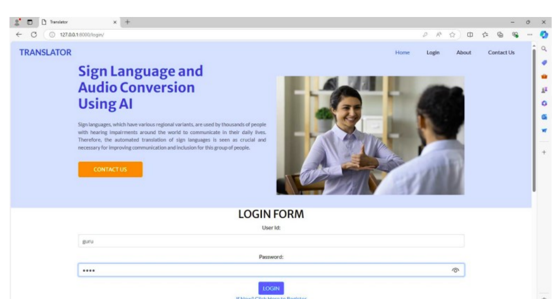

Fig 8:- User login page

F. User home

Two primary options are highlighted by the application: audio signatures and voice signatures. To help people communicate better by reducing slurs in various social and educational contexts, use the audio signature logo to translate sign language into spoken words (Figure 9). Clients can choose to use real-time motion detection to generate audio output or submit their own branded videos. On the other hand, users can use audio-to-sign language software to translate spoken words into animated sign language gestures. This feature makes it easier and more natural for both hearing and deaf people to communicate. Both features are easy to use and provide a seamless experience for any user looking to improve their communication.

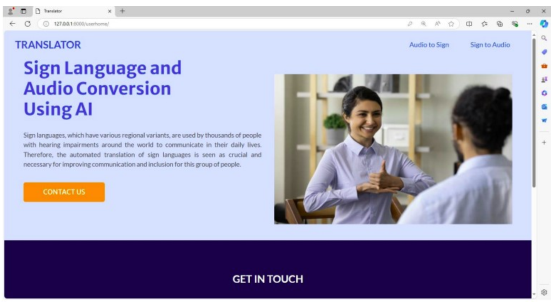

Fig 9:- User home page

G. Audio to sign language

In the sound of a language conversion tool, improved natural language treatment algorithms are used to speak the language of transforming the language of Indian sign language (ISL) (Figure 10). By accurate analysis of the speaking phrases, this technology transforms them into comparable ISL signals so that those who are difficult to listen to can talk effortlessly. This translation process plays an important role in bridging communication gaps and facilitating easier communication between deaf people and the wider community.

Fig 10:- User Interface for audio to sign

The implementation of this feature underscores the necessity of accessibility in communication technology. By providing real-time translation, the device not only boosts understanding but also promotes social inclusion for those with hearing problems (Figure 11). Future developments will focus on integrating regional dialects and contextual differences in ISL, significantly expanding the system's effectiveness and reach, ensuring it satisfies the unique demands of users across different regions..

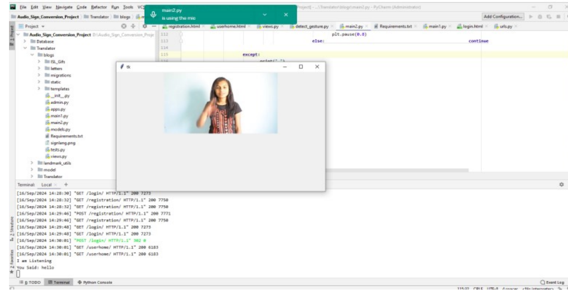

Fig 11:-Translating word hello to sign language

H. Sign language to audio

The speech converter sign language translates into speaking sentences of Indian sign language (ISL) movement using sophisticated machine learning methods. People who are deaf or hearing people can communicate more effectively, thanks to the ability of the system to recognize several Isl signals that use complex gesture recognition algorithm and effectively convert these movements into a similar audio output. This ability is important to remove obstacles to communication and allow deaf people to interact with the hearing community more freely.

This feature, which gives the means of hearing impairment to verbal manifestations, emphasizes the need for inclusion in communication technology. This technology promotes social interaction and inclusion, and improves understanding by instantly converting ISL into speech (Figure 12). language and ensuring that it can handle more features and expressions to meet the diverse communication needs of users.

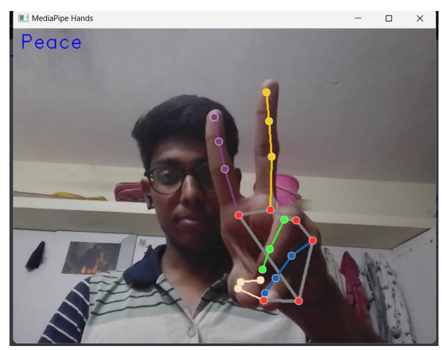

Fig 12:- Sign language to Audio for Peace

VIII. CONCLUSION

This essay claims that the integration of voice, text, and sign language in communication systems based on artificial intelligence is an innovative step to increase the inclusiveness of blindness, hearing impaired, and stupid groups. Thanks to these technologies that promote the actual communication of fluids, people with disabilities have become completely involved in various life, such as social interactions, government services, work, and education. Demanding barriers to communication promotes equality, empathy, and social integration in addition to increasing independence.

In addition, these communication systems can be improved by active advanced technologies such as Sparkling Neural Networks (CNN). CNN is effective for spatial and visual treatment, and as you know, greatly improves speed and efficiency. Victims can be a big deal in the modelling of models that can convert gesture languages into speeches. With more datasets and better training methods, these models will become more accurate and efficient, allowing users to reliably create subtle, easy-to-understand translations. As these systems develop, they create a completely and -connected space for everyone, improve the life of the disabled, and create a more fair society. We will provide it to you.

IX. REFERENCES

[1] Yadav, Anju, Rahul Saxena, Bhavna Saini, Vivek K. Verma, and Vibhav Srivastava. "Audio to Sign language translator web application." In 2021 International Conference on Computational Performance Evaluation (ComPE), pp. 321-326. IEEE, 2021.

[2] Grover, Yuvraj, Riya Aggarwal, Deepak Sharma, and Prashant K. Gupta. "Sign language translation systems for hearing/speech impaired people: a review." In 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM), pp. 10-14. IEEE, 2021.

[3] Tiwari, Rishin, Saloni Birthare, and Mayank Lovanshi. "Audio to Sign Language Converter." IJRASET47271 (2022): 11-02.

[4] Bharghavi, Ms Pokala, and K. Venkata Rao. "ADVANCED AUDIO-TO-SIGN LANGUAGE CONVERSION AND INTERPRETATION SYSTEM." International Journal of Advanced Research in Computer Science 11, no. 6: 214-219.

[5] Duraisamy, P., A. Abinayasrijanani, M. A. Candida, and P. D. Babu. "Transforming Sign Language into Text and Speech through Deep Learning Technologies." Indian Journal of Science and Technology 16, no. 45 (2023): 4177-4185.

[6] Mehta, Abhishek, Kamini Solanki, and Trupti Rathod. "Automatic translate real-time voice to sign language conversion for deaf and dumb people." Int. J. Eng. Res. Technol.(IJERT) 9 (2021): 174-177.

[7] Kapoor, Parul, Rudrabha Mukhopadhyay, Sindhu B. Hegde, Vinay Namboodiri, and C. V. Jawahar. "Towards automatic speech to sign language generation." arXiv preprint arXiv:2106.12790 (2021).

[8] Katariya, Ashmi, Vaibhav Rumale, Aishwarya Gholap, Anuprita Dhamale, and Ankita Gupta. "Voice to Indian Sign Language Conversion for Hearing Impaired People." SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology 12, no. SUP 2 (2020): 31-35.

[9] Amangeldy, Nurzada, Aru Ukenova, Gulmira Bekmanova, Bibigul Razakhova, Marek Milosz, and Saule Kudubayeva. "Continuous Sign Language Recognition and Its Translation into Intonation-Colored Speech." Sensors 23, no. 14 (2023): 6383.

[10] Huang, Rongjie, Jiawei Huang, Dongchao Yang, Yi Ren, Luping Liu, Mingze Li, Zhenhui Ye, Jinglin Liu, Xiang Yin, and Zhou Zhao. "Make-an-audio: Text-to-audio generation with prompt-enhanced diffusion models." In International Conference on Machine Learning, pp. 13916-13932. PMLR, 2023.

[11] Buttar, Ahmed Mateen, Usama Ahmad, Abdu H. Gumaei, Adel Assiri, Muhammad Azeem Akbar, and Bader Fahad Alkhamees. "Deep learning in sign language recognition: a hybrid approach for the recognition of static and dynamic signs." Mathematics 11, no. 17 (2023): 3729.

[12] Abdullahi, Sunusi Bala, and Kosin Chamnongthai. "IDF-Sign: addressing inconsistent depth features for dynamic sign word recognition." IEEE Access (2023).

[13] Wei, Fangyun, and Yutong Chen. "Improving continuous sign language recognition with cross-lingual signs." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 23612-23621. 2023.

[14] Patel, Bhavinkumar Devendrabhai, Harshit Balvantrai Patel, Manthan Ashok Khanvilkar, Nidhi Rajendrakumar Patel, and Thangarajah Akilan. "ES2ISL: an advancement in speech to sign language translation using 3D avatar animator." In 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 1-5. IEEE, 2020.