JSRT JOURNAL

PAGES: 27-35
01/07/25
JOURNAL OF SCIENTIFIC RESEARCH AND TECHNOLOGY VOLUME-3 ISSUE-7 JULY

A Novel Approach for Real-Time Audio-to-Sign Language Translation using Naive Bayes classifier and Natural Language Processing Technique

Poornima[1], Dr. Shilpa B Kodli [2]

[1] *PG Student, Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University, CPGS Kalaburagi, Karnataka, India. poornimagouda444@gmail.com*

[2] *Assistant Professor, Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University, CPGS Kalaburagi, Karnataka, India. shilpakodli@gmail.com*

ABSTRACT

Real-time translation from spoken language to sign language, crucial for the deaf and hard of hearing, is difficult. employing machine learning advances, this research suggests employing Naive Bayes to instantaneously translate audio input into sign language motions. The initiative provides precise translations between spoken and visual language to improve accessibility and inclusion. The system generates sign language animations from audio inputs using natural language processing and deep learning. Language extraction, tense and context classification, and proficient sign language synthesis are key aspects. The research innovates by using a Naive Bayes model trained on varied spoken language and sign language gesture datasets to ensure robust performance across contexts and user situations. To enhance realism and comprehension, the system integrates a 3D animated avatar to render the translated signs visually. A custom-built gesture mapping mechanism aligns each keyword with its corresponding sign language motion. The tool operates in real-time, supporting both sentence-level and word-level translations. This approach contributes to bridging communication gaps and supports inclusive digital interaction for the hearing-impaired community.

Keywords: Sign Language, Hearing, Machine Learning, Instantaneously, Accessibility, Precise Translations.

I. INTRODUCTION

Deaf and hard of hearing people use sign language to communicate via gestures and visual signals. Despite its crucial relevance, difficulties continue in real-time translation from spoken languages to sign language, preventing efficient communication for many within this society. This work strives to overcome these issues by providing a unique technique that leverages machine learning, notably the Naive Bayes algorithm, to construct a system capable of transforming audio input into sign language motions in real time. This thorough overview analysis the landscape of sign language identification systems, tracking their progress from conventional approaches to the newest breakthroughs in machine learning and computer vision. Globally, sign language is one of the main languages for those who cannot communicate verbally. Despite its global presence, not many people understand it or use it. In 2020, 48 million people in the United States alone experience some form of hearing loss, with less than 500,000 about 1% – of them that drive sign language regularly. While mild hearing loss can be remedied with hearing aids and rehabilitation, these solutions may often be too expensive. Individuals can alternatively learn sign language. Hand gestures are a form of non-verbal communication used by individuals in conjunction with speech to communicate. With the increasing use of technology, hand-gesture recognition is considered an essential aspect of Human-Machine Interaction (HMI), allowing the machine to capture and interpret the user's intent and respond accordingly. As technology becomes easier to use and accessible, many people can likely perform simple commands with computer devices, such as typing text and video streaming [1].
In recent years, there has been a growing emphasis on developing technologies that bridge the communication gap between the hearing-impaired and the general population. Neog et al. proposed a system that translates spoken or written input into Indian Sign Language using Natural Language Processing (NLP) techniques. Their work demonstrates the potential of using syntactic and semantic analysis to generate accurate gesture-based output. The system effectively processes both speech and text inputs, converting them into animated sign language. This approach contributes significantly to inclusive communication by reducing the reliance on human interpreters. Their results highlight the importance of NLP and gesture generation models in real-time assistive communication tools. [2]. To address communication challenges faced by the deaf and hard-of-hearing communities, Tauqeer

Safdar et al. proposed Sign4PSL, a real-time translation system designed. Their architecture utilizes natural language processing techniques to analyze input text and map it to corresponding sign language gestures. The system incorporates a 3D avatar to display accurate PSL signs, ensuring clarity and cultural relevance. This work highlights the importance of regional sign language systems and sets a foundation for developing scalable, real-time translation tools across different linguistic contexts [3]. Recent advancements in computer vision and machine learning have significantly improved the automation of sign language recognition systems. A robust approach for detecting American Sign Language (ASL) gestures and generating corresponding textual output in real time. Their work utilized deep learning models to process hand gestures and translate them into meaningful text, facilitating better communication for individuals with hearing and speech impairments. The system focused on enhancing recognition accuracy and reducing processing time. This research highlights the potential of integrating gesture recognition with natural language generation to support inclusive communication technologies. Such efforts pave the way for more accessible and efficient assistive solutions in human-computer interaction [4]. The research analyses significant difficulties such as diversity in sign gestures and contextual variables effecting identification accuracy. It highlights current research topics, including dataset generation, model building, and assessment measures, seeking to better system performance and usability. This study addresses the use of deep learning techniques in sign language recognition, evaluating state-of-the-art models, datasets, and assessment methodologies. It explores the efficiency of deep neural networks in capturing complicated gesture patterns and overcoming problems such as differences in hand forms and motions.
Remainder of the paper is structured as: Section 2 summarizes about some of the existing work carried out by various researchers. Section 3 provides proposed system; section 4 discusses results achieved by proposed method. Lastly section 5 discusses conclusion and future enhancement.

II. LITERATURE SURVEY

In this article, we have examined and evaluated the work of several researchers in order to accurately identify Real-Time Audio-to-Sign Language Translation utilizing the Naive Bayes classifier and Natural Language Processing Technique.

Pankaj Sonawane et al. [5] proposed a a system aimed at converting spoken English into Indian Sign Language (ISL) to facilitate communication between individuals with and without hearing impairments. The approach incorporates Natural Language Processing (NLP) to recognize spoken input, which is then transformed into ISL gestures using a predefined gesture dataset. The system features integrated speech-to-text processing and gesture mapping to enable real-time ISL animations. To improve translation precision, methods such as tokenization and keyword extraction were applied. Designed specifically for everyday conversational use within the Indian context, the system's performance was validated through testing with a range of spoken inputs.

Shubham Thakar et al. [6] implemented a real-time sign language to text conversion system by applying transfer learning. They leveraged pre-trained convolutional neural networks (CNNs) to boost recognition accuracy and minimize training duration, making their system efficient for real-time execution. The primary focus was on enhancing the classification of gestures while keeping computational demands low. The proposed model demonstrated high responsiveness and accuracy during live testing, underlining the value of transfer learning in optimizing sign language recognition solutions.

Orbay et al. [7] addressed the poor performance in end-to-end models under limited data conditions by integrating adversarial learning, multitask learning, and transfer learning techniques. Their work explored both 3D-CNN and 2D-CNN architectures designed to interpret hand shapes and human poses, aiming to discover more effective tokenization strategies. They employed diverse sign language datasets across multiple tasks and introduced a novel frame-level tokenization approach that operates independently of the target domain, significantly outperforming previous state-of-the-art methods.

K. R. Prabha et al. [8] developed a system that translates spoken audio into sign language using Natural Language Processing (NLP) techniques. Their approach involves converting speech to text, applying semantic analysis, and subsequently generating the corresponding sign language gestures. Contextual relevance and accuracy in gesture mapping were improved using machine learning techniques. The primary objective of their system is to facilitate communication between the hearing-impaired and those without hearing disabilities. The solution proved effective in delivering real-time audio interpretation and gesture output.

Kohsheen Tiku et al. [9] utilized a pre-existing dataset (ASL Kaggle) and uses Canny Edges for image processing and SVM for gesture classification, achieving high accuracy in real-time sign language to text and speech conversion. And also presented a Real-time Conversion of Sign Language to Text and Speech they have been used dataset for this paper is the ASL Kaggle dataset, which contains 3000 images for every alphabet of the English vocabulary and Canny Edges, SVM (Support Vector Machine). They achieved 98.65% accuracy from the model.

Aruna Bhat et al. [10] implemented a deep learning-based model for converting sign language gestures into text. The system utilizes Convolutional Neural Networks (CNN) to extract features from hand gesture images and classify them effectively. The model was trained on a custom ASL dataset and demonstrated high accuracy in gesture recognition. The approach emphasizes real-time performance and minimal latency, making it suitable for practical applications. The study contributes to enhancing communication tools for individuals with hearing or speech impairments.

Tiya Ann Siby et al. [11] explored a gesture-based real-time sign language recognition system aimed at enhancing communication for the speech and hearing impaired. The system uses a camera to capture hand gestures and applies deep learning techniques for accurate recognition. It focuses on real-time processing to ensure interactive and immediate translation of sign language into text. The approach improves recognition speed and accuracy while maintaining low computational complexity. The system was validated using various gesture datasets, demonstrating its potential for practical deployment in assistive technologies.

III. PROPOSED SYSTEM

The lack of efficient real-time translation from spoken language to sign language for people who are deaf or hard of hearing is the issue this study attempts to address. In many situations, effective communication is made difficult by existing systems' lack to provide quick and accurate interpretations. The grammatical and contextual complexities of sign language exacerbate this issue, making it hard for current technology to gather and evaluate data in real-time. Furthermore, relying on human interpretation or delayed translation services makes it more difficult for users to be included and accessible in social, professional, and educational contexts.

3.1 System Architecture

In system design, the real-time audio-to-sign-language translation tool starts by analysing input received in either text or audio format. Audio inputs undergo first translation into text utilizing powerful voice recognition technologies built in Natural Language Processing (NLP). Each word inside the translated text is then cross-referenced against a database comprising pre-recorded sign language videos. When a match is detected, the related video is obtained and smoothly merged into the output. For terms without pre-existing video representations, the system dynamically develops sign language animations by collecting appropriate gestures.
In the proposed system architecture, the real-time audio-to-sign-language translation tool is designed to process both text and audio inputs seamlessly. When an audio input is provided by the user, it undergoes a speech-to-text conversion using advanced voice recognition algorithms integrated with Natural Language Processing (NLP) techniques. This conversion ensures that spoken language is accurately transcribed into its textual equivalent, preserving the semantic structure of the original speech. Once the input is in textual form, whether from direct user entry or converted audio, the system performs linguistic analysis to break down the sentence into individual words or meaningful components. Each extracted word is then checked against a comprehensive sign language database that contains pre-recorded video clips corresponding to commonly used words or phrases in sign language. If a match is found, the appropriate video segment is retrieved and queued for output.

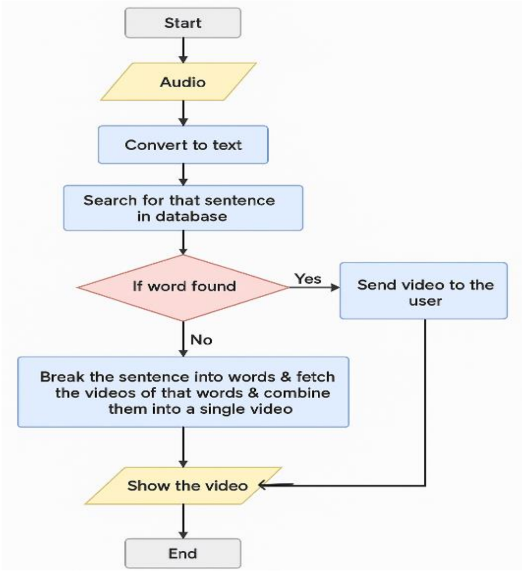

Figure 1: System Architecture of Audio to Sign Language Detection

To ensure continuity and comprehension, the system dynamically concatenates these individual sign language clips to form a coherent visual translation of the entire sentence. In cases where the input word or phrase does not exist in the database, the system initiates a gesture synthesis module. This module utilizes pre-trained gesture models or avatar animation engines to generate a custom sign language animation, based on the morphological and syntactic structure of the unknown term. The final output is a visually smooth and contextually accurate sign language representation, displayed through a 3D animated avatar or video playback module. This real-time processing flow allows the system to facilitate effective communication between hearing and hearing-impaired individuals, bridging the gap through intuitive and adaptive sign language translation.

1. Start and Audio Input: The system process initiates when the user begins interaction by providing an audio input. This can be achieved either through real-time voice input via a microphone or by uploading a pre-recorded audio file. The system captures this audio and prepares it for processing. This step acts as the entry point for converting spoken language into sign language.

2. Convert to Text: After the audio input is received, the next step involves converting the spoken words into textual form. This is done using Automatic Speech Recognition (ASR) technology. ASR tools analyse the speech signal, identify phonemes, and map them to corresponding words. NLP (Natural Language Processing) techniques are optionally applied to clean, correct, and standardize the transcribed text, making it ready for further processing.

3. Search for the Sentence in Database: The converted text is then checked against a predefined sign language video database to see if a video corresponding to the exact sentence already exists. This database contains pre-recorded sign language videos for commonly used phrases and sentences. The purpose of this step is to reduce processing time and improve fluency by utilizing complete sentence videos whenever possible.

4. If Word Found (Conditional Check): This decision-making step determines the next action based on the sentence match. If the exact sentence is found in the database, the system takes the "Yes" path, quickly retrieving and sending the associated video to the user. This results in an efficient, natural, and grammatically correct sign language output. If the sentence is not found (taking the "No" path), the system moves to the next step where it decomposes the sentence into individual words.

5. Break the Sentence into Words and Fetch Videos: When no full sentence match exists, the system performs a word-level decomposition of the sentence. Each word is searched individually in a word-level sign language video database. The corresponding videos for each recognized word are then retrieved. After gathering all the word-level clips, the system uses video stitching techniques to concatenate the videos in the correct sequence. This ensures that the resulting sign language video conveys the original meaning of the sentence as accurately as possible.

6. Show the Video: Once the sign language video is ready whether retrieved directly as a sentence or generated by combining individual word videos it is presented to the user. This step serves as the visual output of the system, helping users understand the original spoken content through sign language.

7. End: After displaying the sign language video, the system completes the current session or awaits further input. This marks the end of one complete cycle of audio-to-sign-language conversion.

3.2. Methodology Used

1) Problem Definition and Requirement Analysis: The primary goal is to develop a real-time system that translates spoken audio into sign language to assist individuals with hearing impairments. The system must accurately process live audio/text input and generate corresponding sign gestures instantly. Key user needs include fast response time, high gesture accuracy, and usability in noisy environments. The solution should also be adaptable to various communication scenarios such as classrooms, hospitals, and public places.

2)Data Collection and Preprocessing: The system gathers diverse datasets of spoken language and corresponding sign language motions, including real-time audio and gesture recordings. Audio data is processed to extract linguistic features, while gesture videos are labeled and standardized. Deep learning models are trained on this multimodal data to improve recognition accuracy. A 3D animated character created using Blender visually represents the gestures. Powered by advanced machine learning algorithms, the system accurately interprets both text and audio inputs, even in challenging conditions like background noise.

3)Feature Extraction and NLP Processing: Natural Language Processing (NLP) techniques are applied to tokenize spoken or converted text into meaningful units like words or phrases. These tokens are further analyzed to identify parts of speech, tense, and contextual meaning. This step helps in understanding the grammatical structure and intent of the input. Extracted features are essential for accurate translation, classification, or response generation in downstream applications.

4) Naive Bayes Model Development: The Naive Bayes algorithm is implemented to classify and predict sign language gestures from extracted language features. The model is trained using pre-processed datasets, learning the relationship between text or speech inputs and corresponding gestures. It leverages probability-based learning to make accurate predictions. This lightweight model is suitable for real-time applications due to its simplicity and speed.

5) Integration with Django Framework: A Django-based web application is developed to host the real-time sign language translation system. It includes user-friendly interfaces for audio input and gesture output display. Django ensures smooth backend management and scalability. The platform allows seamless user interaction with dynamic content updates.

6) System Testing and Validation: The system undergoes thorough testing to evaluate accuracy, responsiveness, and stability in various linguistic contexts. Tests include different accents, sentence structures, and noisy environments. User feedback and expert reviews are gathered to enhance usability and functionality. Results help in refining the model and interface.

7) Documentation and Reporting: Comprehensive documentation is created to capture each development phase, including methods, algorithms, and technical details. This includes model training steps, system architecture, testing outcomes, and integration workflows. Clear reporting ensures future maintenance, upgrades, and research scalability.

IV. RESULTS AND DISCUSSIONS

Figure 2: Homepage

The above figure 2 shows the home page of audio to sign language tool which provides navigation in the site and overview of the website which contains all tags such as sign up, login, about us and contact.

Figure 3: Login page

The above figure 3 shows the login page where we have to insert the username and password to get login into the system.

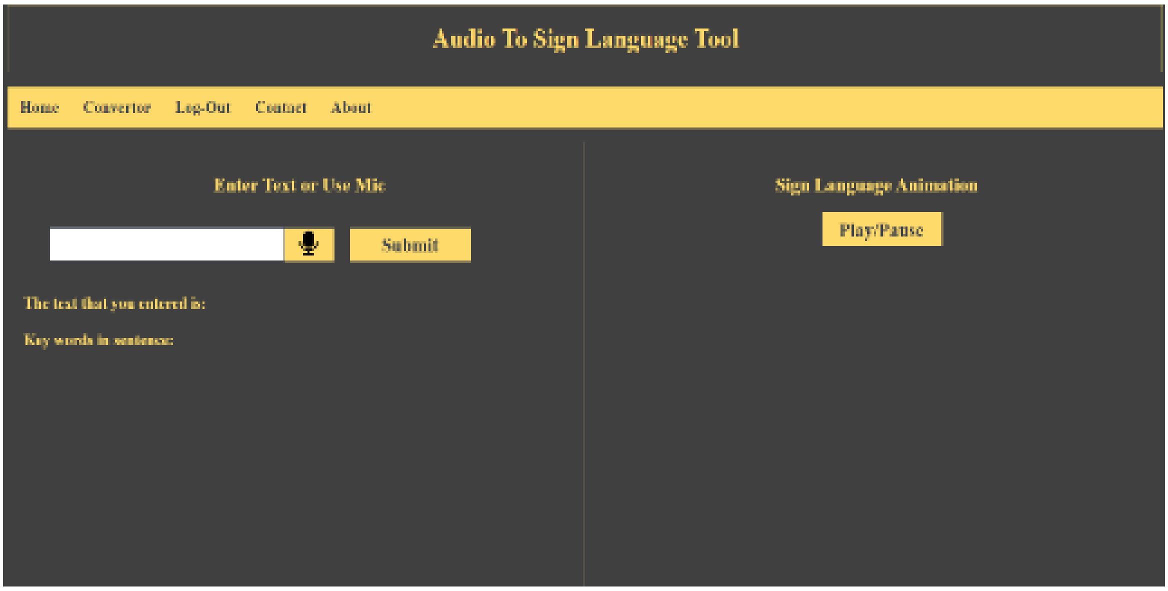

Figure 4: Enter input in text format

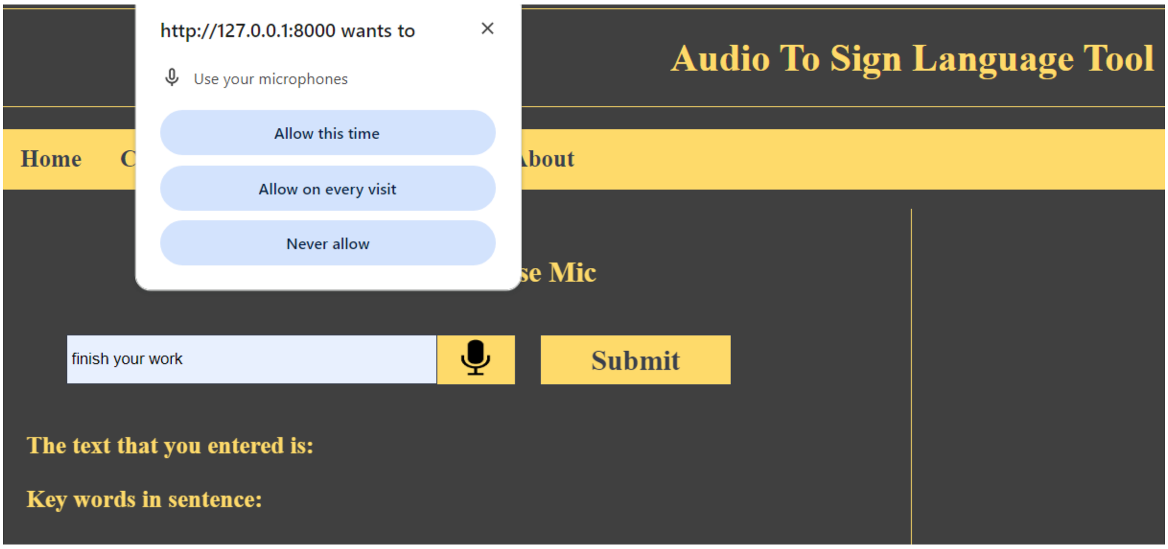

Figure 5: Enter the input in audio format

The above figure 4 and figure 5 shows two options for uploading: one in text form and another in audio form. By submitting your input through either option, you will receive the corresponding result in sign language. the image illustrates the interface of an Audio to Sign Language conversion tool requesting microphone access. The spoken phrase "finish your work" is captured and transcribed into text using speech recognition. After submission, the system identifies key terms from the input sentence. These key terms are then used to produce the relevant sign language gestures for visual translation.

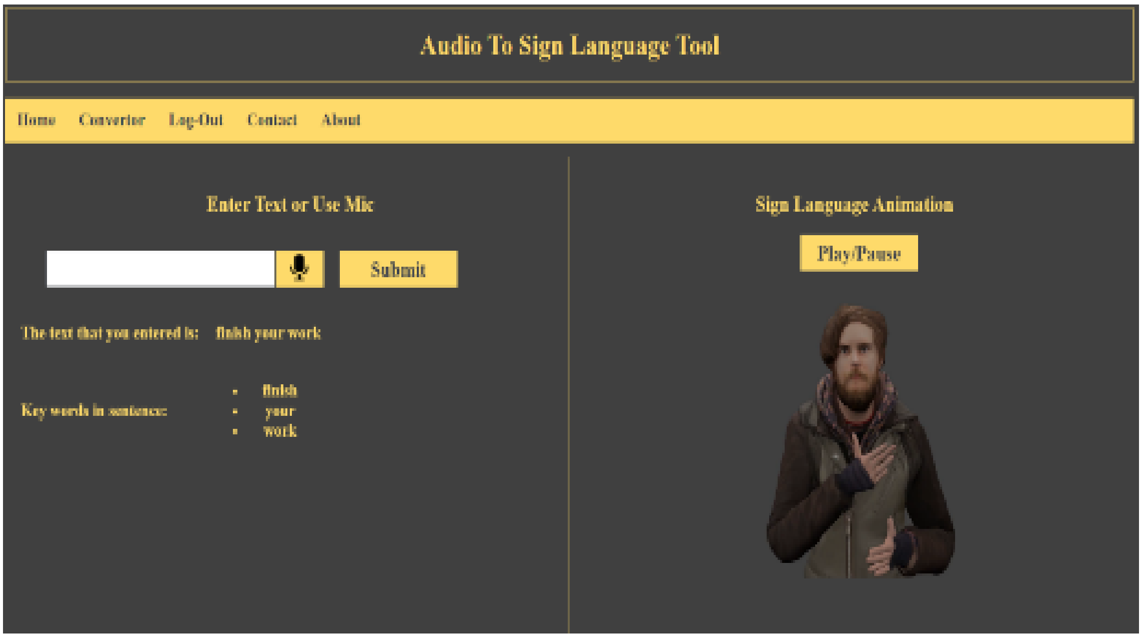

Figure 6: Results

The above figure 6 shows the results of the text or audio to sign language conversion. The sentence **"finish your work"** is shown, with individual keywords extracted and listed below. On the right, a 3D animated avatar performs sign language gestures based on the extracted keywords. The system allows users to control playback of the sign animation using the Play/Pause button.

V. CONCLUSION

The research has successfully constructed a robust sign language recognition system employing sophisticated machine learning algorithms. Through deep learning models trained on a wide dataset of sign language motions, the system demonstrated excellent accuracy in detecting both text and audio inputs, efficiently discriminating right movements even in tough settings like background noise. Key technologies created include a powerful identification engine and user-friendly interface, aimed at boosting accessibility for the hearing-impaired population and overcoming communication barriers. Moving further, extending the dataset for greater model generalization, incorporating real-time recognition capabilities, and researching adaptive learning approaches are key. This study not only enhances communication accessibility but also establishes a platform for larger applications in assistive technology and human-computer interaction, promising substantial improvements in the area. The project's findings illustrate its potential to change accessibility technology, giving a scalable solution that benefits people and enterprises alike. By integrating machine learning developments, the system not only increases accuracy but also promotes user engagement via intuitive interfaces.

REFERENCES

1. Jacky Li, Jaren Gerdes, James Gojit, Austin Tao, Samyak Katke, Kate Nguyen, and Benyamin Ahmadnia. “Sign Language Recognition and Translation: A Multi-Modal Approach using Computer Vision and Natural Language Processing” *2023 In Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pages 658–665, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.
2. A. P. Neog, A. Kalita, and N. Pandiyarajan, “Speech/Text to Indian Sign Language using Natural Language Processing,” *Indian Journal of Computer Science and Engineering (IJCSE)*,pp. 550–559,2023,doi: 10.21817/indjcse/2023/v14i3/231403030.
3. T. Safdar, M. A. Javed, M. Sajjad, S. W. Baik and A. M. Siddiqui, "A Real-Time Automatic Translation of Text to Sign Language," *The Journal of Supercomputing*, pp. 4582–4605, 2021. doi: 10.1007/s11227-020-03315-6
4. Y. Farhan, A. Ait Madi, A. Ryahi and F. Derwich, "American Sign Language: Detection and Automatic

Text Generation," *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, Morocco, 2022, pp. 1-6, doi: 10.1109/IRASET52964.2022.9738061.

5. Pankaj Sonawane, Karan Shah, Parth Patel, Shikhar Shah, Jay Shah**,** "Speech to Indian Sign Language (ISL) Translation System," *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, Greater Noida, India, 2021, pp. 92-96, doi: 10.1109/ICCCIS51004.2021.9397097.
6. S. Thakar, S. Shah, B. Shah and A. V. Nimkar, "Sign Language to Text Conversion in Real Time using Transfer Learning," *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)*, Bangalore, India, 2022, pp. 1-5, doi: 10.1109/GCAT55367.2022.9971953
7. A. Orbay and L. Akarun, "Neural Sign Language Translation by Learning Tokenization," *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, Buenos Aires, Argentina, 2020, pp. 222-228,
doi: 10.1109/FG47880.2020.00002.
8. K. R. Prabha, B. Nataraj, B. Thiruthanikachalam, S. Surya Narayanan, S and M. Vishnu Prasath, "Audio to Sign Language Translation using NLP," *2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR)*, Sathya Mangalam, India, 2023, pp. 1-4, doi: 10.1109/STCR59085.2023.10397050.
9. Kohsheen Tiku, Jayshree Maloo, Aishwarya Ramesh, Indra R, "Real-time Conversion of Sign Language to Text and Speech," *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA),* Coimbatore, India, 2020, pp. 346-351, doi: 10.1109/ICIRCA48905.2020.9182877.
10. A. Bhat, V. Yadav, V. Dargan and Yash, "Sign Language to Text Conversion using Deep Learning," *2022 3rd International Conference for Emerging Technology (INCET)*, Belgaum, India, 2022, pp. 1-7,
doi: 10.1109/INCET54531.2022.9824885.
11. T. A. Siby, S. Pal, J. Arlina and S. Nagaraju, "Gesture based Real-Time Sign Language Recognition System," *2022 International Conference on Connected Systems & Intelligence (CSI)*, Trivandrum, India, 2022, pp. 1-6, doi: 10.1109/CSI54720.2022.9924024.