Sign2speech: Enabling Communication Beyond Barriers

Aadhya Dwivedi[1], Ankit Kumar[2], Aryan Srivastav[3], Sinha Neha[4]
Noida Institute of Engineering and Technology

***Abstract*—The approach, implementation, challenges, and any kind of improvements that can be applied to the system are covered in this paper. Despite the fact that the need of communication is one of the fundamental human's rights, that millions of people in our society with speech and hearing impairments struggle to interact with other daily. Our study presents Sign2Speech, it is a software that helps people to communicate easily by recognising the sign languages or hand gestures and translate them into voice and text.**

***Index Terms*—AI-powered translation, accessibility, gesture recognition, computer vision, deep learning, sign language recognition, text-to-speech (TTS), real-time processing, inclusivity, and communication assistive technology.**

I. INTRODUCTION

Language is the key component of the human interaction because it allows people to exchange their ideas, feelings, and information through words. Sign language is used for people with speech and hearing impairment; however, it is not commonly understood by the general public, and it is an essential part of the individuals with speech and hearing impairments. This communication gap obstructs the everyday interactions, which includes professional and educational settings, and social gathering.

Sign language is an arranged and organised visual language that communicates through land gestures, body posture, and facial expressions to convey their message to the others. Still, due to many people not knowing the sign language despite it being widespread, people with speech and hearing impairment are isolated and often not included in the group. Their exists multiple methods to bridge the gap between signers and non-signers, such as human interpreters and text-based communication, but they also have drawbacks in terms of accessibility, cost, and real-time engagement.

Artificial Intelligence (AI) and Machine Learning have opened up many new approaches to solve this issue. Sign2Speech uses these technologies to create a real-time system, that can recognise sign language and by analysing them, it converts the gestures into voice and text message for the non-signers to understand. By using Natural Language Processing (NLP), Computer Vision Techniques, and Deep Learning Models, our software makes sure to convert the hand gestures accurately and effectively, by this it enhances the communication for the sign language users.

This paper investigates the development and implementation of the Sign2Speech, highlighting the methods for speech combination, gesture analysing, and real-time processing. The system's aim is to provide a scalable, reasonably priced, and easy to navigate that can be used into everyday communication devices. We have also discussed issues and future advancements to improve accuracy and adaptability.

Our system aims to advance in society where every person with speech and hearing impairments can easily engage in social, educational, and professional contexts by eliminating language barriers and promoting simple communication medium. Our goal is to contribute in this growing system to promote the widespread use of AI-powered sign language recognition systems.

II. OBJECTIVES

Our system Sign2Speech aims to develop an intelligent system that can recognise sign language and hand gestures to convert them into text and speech in order to improve communication between individual who are not familiar with sign language and need help to communicate, and who have hearing or speech impairments. By using deep learning and computer visions, the system provides real-time, accurate, and

efficient gesture recognition, ensuring seamless and natural communication. Additionally, this system aims to enhance accessibility, inclusivity, and social integration by bridging the communication gap with a scalable and approachable solution.

III. MOTIVATION

Communication is a necessary for humans to interact with others, but many people around the world with speech and hearing impairments struggle to communicate with those who are unfamiliar with sign language. This language barrier causes social exclusion and limits the interaction of these people, be it in healthcare, education, and employment. Despite the existence of solutions, like interpreters, they are not always practical or available in real-time scenarios. By using artificial intelligence and computer vision to create a user-friendly, real-time sign language translation system, Sign2Speech aims to help individuals with speech and hearing impairments. By bridging the communication gap between sign people and non- sign people, it allows them to be independent, and be included in society.

IV. LITERATURE SURVEY

S.No.	Paper/Study	Methodology	Strengths	Limitations
1.	S. Starner et al. (1998)	Hidden Markov Models (HMM) with gloves-based sensors	High accuracy for structured gestures	Requires wearable sensors, not real-time
2.	C. Vogler & D. Metaxas (2001)	3D motion tracking with computer vision	3D motion tracking with computer vision	Computationally expensive
3.	T. Starner (2003)	Vision-based American Sign Language (ASL) recognition using neural networks	Sensor-free, real-time processing	Limited vocabulary size
4.	K. Kadir et al. (2015)	CNN-based image classification for sign recognition	High accuracy with deep learning	Requires large labeled datasets
5.	Y. Huang et al. (2018)	LSTM with CNN for sequential gesture recognition	Captures temporal dependencies in gestures	High computational cost
6.	X. Li et al. (2020)	Transformer-based vision models for sign language recognition	More robust feature extraction	Requires high-end hardware for inference
7.	R. Patel et al. (2022)	Real-time sign language recognition with mobile deployment	Efficient and user-friendly application	Limited accuracy under poor lighting

V. PROBLEM FORMULATION & PROPOSED WORK

- Problem Statement:

Communication barriers between the hearing-impaired community and non-sign language people have a significant impact on social interactions, education, and career opportunities. Existing solutions, such as human interpreters and text communication, are either expensive or ineffective in the moment. Despite previous researches on sign language recognition using various sensors and computer vision-based methods, challenge to achieve high-accuracy, real-time processing, and adaptability still remain unsolved.

The primary challenges include:

1. Sign Gestures' Complexity: Hand posture, movement, and expression variations make recognition difficult.
2. Real-time Processing: Accurate, low-latency recognition is crucial for real-world usability.
3. Environmental Factors: Background noise, lighting, and obstructing all effect recognition performance.
4. Scalability: To accommodate a large number of sign languages and dialects, a accurate and adaptable model is required.

- Proposed Solution:

To address these challenges, we introduce Sign2Speech, real-time sign language recognition system that translates sign gestures into text and speech using deep learning techniques. The system's primary components are as follows:

1. Gesture Recognition Module: Using CNNs and Neural Networks, this module extracts the temporal and spatial features of hand movements. It uses various methods for hand tracking and feature extraction. It increases resistance to variations in hand size, and orientation.
2. Translation & Speech Synthesis Module: This module converts recognised gestures into pre-written words and sentences using Natural Language Processing (NLP). It integrates a Text-to-Speech engine, to translate recognised text into speech that sounds natural.
3. System Architecture & Deployment: The system is designed to be a real-time application that is accessible via mobile devices and the internet. It uses TensorFlow for model implementation. To optimise for low-latency processing, model compression techniques like pruning were applied.

VI. METHODOLOGY

From data collection to the final voice synthesis, the system Sign2Speech is made up of several interconnected modules. To guarantee precise and effective sign language translation, the system uses a structured approach that consists of data collection, pre-processing model training, and text-to-speech conversion.

- Information Collection:

Data collection is a crucial step in building a trustworthy model for sign language recognition. This comprises: Dataset Selection: The selected dataset of sign language motions consists of thousands of labelled images and videos showing various signals. There are various databases are among the public datasets that are currently in use. Building a Custom Dataset: to increase accuracy, hand and body gestures are captures in real-time using a camera/webcam. Each sign is photographed or recorded from a range of backgrounds, lighting conditions, and angles to confirm resilience. Data augmentation: Techniques like noise addition, flipping, rotation, and variation are used to diversify the datasets in order to enhance the model.

- Getting the system ready:

The collected data is pre-processed to remove noise, standardise input, and extract useful features. This includes: Background Removal: There are various libraries used to extract hand movements from backgrounds. Hand Tracking and Segmentation: A deep learning model is used to pinpoint crucial finger and palm locations in order to guarantee precise gesture identification. Feature Extraction: Key features such as motion detection, finger locations, and hand orientation are extracted using CNN-based feature.

- Model Training:

The system is built around a deep learning-based gesture recognition model. A combination of CNNs and neural networks are used for classification: CNN-based Feature Extraction: A pre-trained CNN model is used to extract characteristics from sign imageless for Gesture Sequence detection: Since sign language requires successive hand motions, LSTM's ability to capture temporal dependencies ensures accurate detection of dynamic gestures. Training & Optimisation: The model is trained on labelled sign data using optimiser in order to attain good classification performance.

- Text and Speech Conversion:

Once a sign has been identified, the relevant text and audio output are generated: Sign-to-Text Mapping: Each recognised sign is mapped to a dictionary of

words or sentences. Text-to-spoken Synthesis: Google Text-to-Speech are used to convert the recognised text into a natural-sounding spoken output. Context Awareness: NLP techniques are combined to ensure grammatically correct phrase generation when multiple indications are recognised sequentially.

VII. IMPLEMENTATION

The Sign2Speech system is a real-time system that combines deep learning computer vision, and speech synthesis technologies, with three components: Speech synthesis, text conversion, and gesture recognition.

- Components of Hardware:

Sign2Speech uses common, easy to access hardware to guarantee usability accessibility: Camera Module: Real-time hand gestures are captured by a standard webcam/mobile camera. Processing Unit: A mobile device or a laptop with enough processing power for deep learning inference. For faster processing, an environment with GPU acceleration is preferred. Speaker & Microphone: For optional voice-based user interaction and speech output.

- Architecture for Software:
- Acquiring and Preparing Images The camera captures frames in real-time at a pre-set frame rate (FPS). The image is resized and converted to a RGB or greyscale based on the model's requirements. OpenCV is used for background removal, hand segmentation, and edge detection. A bounding box and a hand tracking algorithm guarantee stable input for recognition.
- Deep Learning for Recognising Gestures A CNN is used to extract features from hand gestures. Transformer models are used to capture motion-based signs, where gestures involve movement rather than still positions. The model is trained using a special dataset of commonly used sign language gestures, such as those in ASL (American Sign Language) and ISL (Indian Sign Language). A classification layer maps recognised gestures to their corresponding textual meanings. The model is optimised for fast inference on mobile devices.
- Text conversion and language processing Once identified, a gesture is mapped to the corresponding word or phrase from a sign-to-text dictionary. A Natural Language Processing (NLP) module enhances sentence structure for continuous gesture recognition. Context-awareness is used to dynamically construct coherent sentences and manage multi-word phrases.
- Speech Synthesis for Text-to-Speech Conversion The generated text is fed into a Text-to-Speech (TTS) engine. The system provides voice output with programmable pitch, speed, and volume to enhance clarity. Users can choose from a range of languages and voice tones for more customisation.

- System deployment

We configured the system in a number of ways to enable Sign2Speech: Web application: Flask/Django can be used for the backend part of the system, and ReactJS for the interactive frontend of the system. Mobile Application: React Native/Flutter can be used in the development of the mobile application for the system, which supports both iOS and Android. Edge Computing Devices: Optimised models run on low-power devices like Raspberry Pi for portable applications.

- User interface and Interaction

The user-friendly interface of the system has the following features: Live gesture recognition provides real-time feedback on detected gestures. Speech Output Controls: Users can adjust the volume, switch on and off audio playback, and change voice settings. Translation History: Stores, for later use, previously recognised gestures and their textual translations. Support for Multiple Languages: Allows users to choose from a variety of output languages and sign language dialects.

VIII. CHALLENGES

Gesture Complexity: Improved feature extraction techniques make it easier to identify some minute variations in gestures. Lighting Variations: Adaptive image processing ensures accuracy in a range of lighting scenarios. Real-time Processing: Optimised

model compression and GPU acceleration guarantee lower latency.

IX. APPLICATIONS

The Sign2Speech system has many real-world applications that enhance inclusivity and accessibility in a range of domains:

Medical Care: It enables efficient communication between medical staff and the patients who have difficultly due to speech and hearing impairments. It is useful in emergency scenarios where it is essential to get to the patient's needs immediately.

Learning: It facilitates communication between educators and learners, fostering inclusive classrooms, allowing every child to be able to participate in the class. It makes it easier for those who have hearing loss to integrate into the classroom environments. Accessibility & Public Services: It makes government buildings, banks, and customer service departments more accessible. In public places like train stations and airports, it automatically translates the sign language to the textual language or voice message.

Integration at Work: It helps business create inclusive work environments for employees with hearing or speech impairments. It makes it easier to interpret meetings in sign language and communicate professionally.

Social Engagement & Daily Conversation: By overcoming the communication barrier in social settings, it makes interactions between sign language users and non-users smoother. It can be integrated into smartphones and wearable technology to offer real-time translations.

Entertainment & Media: By providing live events, podcasts, and videos in sign language, it improves accessibility in digital content. It can interact with users of sign language through virtual assistants.

Law enforcement and emergency response: It aids emergency responders, including police and firefighters, in understanding and assisting those who struggle with communication. It helps ensure fair access to legal services for the deaf and mute individuals.

X. FUTURE SCOPE

Multi-Language Support: Expand recognition to ASL, BSL, ISL with flexible training for regional variations. Improved Accuracy: Use Vision Transformers and multi-model learning, i.e., video + sensor input, for better recognition. Real-time Optimisation: To process data faster and offline, use edge computing and model quantisation. Context-aware Translation: Use Natural Language Processing (NLP) and adaptive learning in sentence building to improve accuracy.AR and Haptic Feedback: Use AR for real-time visualisation and getting live feedbacks to enhance communication & Smart Assistants: Use signs to enable communication with IoT devices and smart home assistants. User accessibility: Improve user interfaces and develop personalised gesture models to support a range of disabilities.

XI. CONCLUSION

To improve the communication medium for the speech and hearing-impaired individual, our method efficiently converts sign language gestures into text and speech by using computer vision and deep learning, our approach effectively translates sign language gestures into text and speech, facilitating seamless communication with non-sign language users. Real-time processing, adaptability to varying lighting conditions, and user-friendly deployment all contribute to practicality in real-world scenarios. Despite positive results, there is still opportunity for development in areas such as gesture complexity, environmental variations, and vocabulary expansion. Future developments like support for multiple sign language. Sign2Speech may be able to bridge the communication gaps by empowering individuals with disabilities and promoting an accessible and inclusive society.

REFERENCES

[1] "Real-time sign language recognition: A survey," by S. P. Maity, S. Paul, and P. Ghosh, IEEE Access, vol. 9, pp. 115398-115414, 2021. doi: 10.1109/ACCESS.2021.3105467.

[2] In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371-1375, Dec. 1998, T. Starner and A. Pentland, "Real-time American Sign Language recognition

using desk and wearable computer-based video." doi: 10.1109/34.735811.

[3] "Deep learning-based sign language recognition: A comprehensive review," by P. Kaur, M. Kumar, and R. Sharma, IEEE Access, vol. 10, pp. 3645-3663, 2022. Doi: 10.1109/ACCESS.2022.3141598.

[4] "Sign language recognition using sub-units," Journal of Machine Learning Research, vol. 13, no. 1, pp. 2205-2231, 2012; H. Cooper, E. Ong, N. Pugeault, and R. Bowden.

[5] "Sign language recognition: A deep survey" by A. Rastgoo, K. Kiani, and S. Escalera, Expert Systems with Applications, vol. 164, p. 113794, March 2021. doi: 10.1016/j.eswa.2020.113794.