Enhancing Sign Language Detection for Improved Accessibility

UDAY SHEKAR BIRADAR[1], MAHESH D M[2], TEJKUMAR S[3], HEMALATHA D[4]
[1, 2, 3, 4] *Presidency University*

Abstract- ***A technical device called "Sign Language Detection with Voice" attempts to help people with hearing problems communicate more effectively. This project aims to recognize and translate sign language motions into spoken or written language by applying cutting edge AI and machine learning. This project has the potential to improve inclusivity and accessibility, close the gap between the hearing and deaf communities, and demonstrate the ability of technology to promote harmony and understanding. Through our app, we will be able to gather and analyze patient data to identify prevalent diseases, provide quick and detailed diagnostic reports, and ensure easy and seamless interaction with the app. By concentrating on these results, we may significantly enhance the availability and quality of healthcare in rural areas, hence enhancing the general well-being of the rural populace. This effort is a big step toward making the world more accessible and inclusive for people who have hearing loss. It has the potential to enhance the quality of life for individuals who depend on sign language as their major means of communication by translating and understanding sign language into spoken or written forms. We hope to contribute to a society where communication is unrestricted and where everyone, regardless of hearing ability, can connect and interact with the world more easily by utilizing the most recent developments in AI and voice recognition.***

I. INTRODUCTION

In a changing world where technology is constantly changing how we perceive inclusion and accessibility, the "Sign Language Detection with Voice" project stands out as a trailblazing endeavor. Through the creative integration of cutting-edge AI and machine learning technology, this ground-breaking project aims to eliminate communication barriers faced by people with hearing impairments. The project intends to translate sign language movements into spoken or written language by utilizing these state-of-the-art techniques to recognize and understand them with ease. By doing this, it seeks to close the long-standing divide that exists between the hearing and deaf communities, illuminating the revolutionary potential of technology in promoting a more inclusive and connected world.

Essentially, this concept imagines a time where artificial intelligence is able to both understand and interpret the complex language of sign into more comprehensible forms. This goal is important not only for technology but also because it represents a significant advance in the direction of building a society in which people with hearing loss can interact with others and integrate into society at large.

This technical advancement has a significant wider influence and expands healthcare accessibility to rural places. The project aims to gather and process patient data by creating a user-friendly mobile application, which would provide quick and thorough diagnostic results. Through tackling communication barriers and expanding its advantages to the field of healthcare, the project aims to significantly enhance overall well-being, especially for marginalized communities.

The initiative is a technological milestone that has the potential to change social standards as we embrace the cutting edge of AI and voice recognition technology. It imagines a society in which everyone can easily connect with the outside world, regardless of their hearing ability, and communication breaks down barriers. In this regard, the "Sign Language Detection with Voice" project is a monument to the revolutionary potential of invention, providing a window into a future when technical growth, understanding, and unity will foster inclusivity.

II. PROBLEM STATEMENT

The goal of the Sign Language with Voice project is to provide natural-sounding voice creation in addition to an accurate sign language detection system. This program, which is geared toward deaf and hard-of-hearing people as well as interpreters, attempts to provide an approachable solution that will improve the accessibility of communication for those with hearing impairments.

III. OBJECTIVE

Giving the deaf and hard of hearing people a sophisticated means of communication is the main goal of the Sign Language Recognition System project. The goal of the research is to create a highly accurate system that can recognize sign language motions using machine learning. A two-pronged research approach will be used to make sure the system is both technically solid and specifically designed to satisfy the demands of each individual user. The intended audience's communication challenges will be well understood through the application of qualitative research techniques.

The accuracy of sign language recognition technology will then be evaluated and improved through the application of quantitative research approaches. Implementing a comprehensive system that makes the most of these technologies' strengths will require the integration of well-known frameworks like TensorFlow, Keras, OpenCV, and MediaPipe.

The main goal is to close the gap between spoken and sign language for the hearing challenged, improving their ability to communicate. The prioritization of real-time recognition capabilities will facilitate smooth and fluid talks, hence promoting dynamic interaction. The incorporation of sophisticated artificial intelligence methods is essential for attaining recognition efficiency and accuracy, guaranteeing that the system functions as a state-of-the-art and intuitive solution. In the end, our research hopes to make a substantial contribution to communication accessibility and inclusion, enabling people with hearing loss to participate effectively in a world where communication obstacles are common.

IV. METHODOLOGY

The system uses a vision-based method. Since all of the signs are portrayed with bare hands, there is no need for any artificial gadgets for interaction.

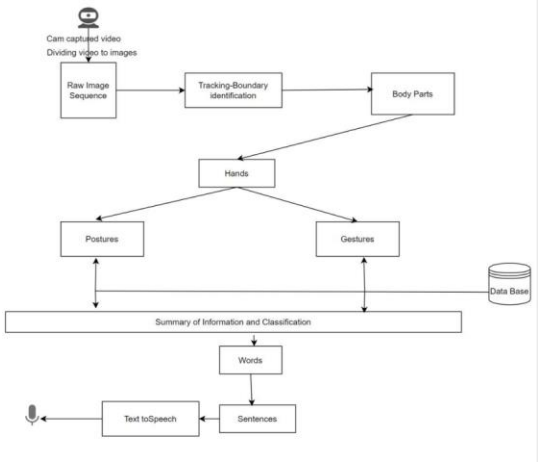
We looked for pre-made datasets for the project but couldn't find any in the form of raw images that met the software specifications. As a result, we decided to develop our own data set. To create the dataset, Open Computer Vision (OpenCV) library of python is used. To accurately recognize the sign gestures and translate them into text & then convert the text to speech, our proposed method comprises three stages: data preprocessing and feature extraction, data cleaning and labelling and gesture recognition & speech translation. Data preprocessing and feature extraction are carried over by the MediaPipe framework. Here, features from the face, hands, and body are extracted as keypoints and landmarks using built-in data augmentation techniques from sequence of input frames taken from a web camera. In stage 2, the extracted keypoints from stage 1 are saved in a file to identify and remove the null entries from the data, after which data labelling follows. In stage 3, the cleaned and labelled gestures are trained and classified by our LSTM model for sign language recognition in the form of text on the screen. The displayed text is then converted to speech. The three stages of the proposed methodology are elaborately discussed below. System Architecture

Long Short-Term Memory (LSTM) is an advanced type of recurrent neural network (RNN) designed to address the limitations of traditional RNNs in capturing long-term dependencies within sequential data. Introduced by Hochreiter and Schmidhuber in 1997, LSTMs are particularly effective in tasks such as natural language processing, time series analysis, and speech recognition.

At the core of an LSTM unit is a memory cell that stores and updates information over time. This cell state is regulated by three essential components: input, forget, and output gates. The input gate controls the inflow of new information into the cell state, the forget gate determines what information to discard, and the output gate influences the hidden state, which is the short-term memory or output of the LSTM.

The input, forget, and output gates employ sigmoid activation functions to regulate the flow of information, while the cell state is updated using a combination of the input gate and a candidate value calculated using the hyperbolic tangent function. This architecture enables LSTMs to avoid the vanishing gradient problem, allowing them to capture and learn long-range dependencies in data.

During training, the network learns optimal parameters for these gates through backpropagation and gradient descent, adjusting the weights and biases to improve its ability to remember or forget information as needed. The result is a powerful model capable of maintaining context and handling sequential data with extended dependencies, making LSTMs especially valuable in various applications requiring the analysis and generation of sequential pattern



- Cam captured video Dividing video to images: Refers to video footage captured by a camera. Involves extracting individual frames from the video, converting it into a sequence of images
- Ram Image Sequence: Represents the sequence of images obtained directly from the video without additional processing
- Tracking-Boundary Identification: Involves tracking objects or regions across frames and identifying the boundaries of these objects within images.
- Postures: Likely refers to recognizing and analyzing different body postures or positions within the video frames.
- Hands: Suggests a focus on detecting and understanding hand movements or positions within the video.
- Summary of Information and Classification: Implies the aggregation and classification of information extracted from the video frames, possibly involving categorization or labeling.
- Words/Text to Speech/Sentences: Indicates processing textual information within the video, possibly converting text to speech or extracting sentences.
- Gestures: Involves recognizing and interpreting gestures made by individuals in the video
- Body Parts: Suggests analyzing and identifying various body parts within the video frames.
- Database: Likely refers to storing and organizing the processed information, possibly for future retrieval or analysis.

COCLUSION

In summary, the development of a machine learning and cutting edge technology-based sign language recognition system is critical to removing obstacles to communication for those who are hard of hearing. This project is evidence of the dedication to equality and accessibility in communication, going beyond technological innovation. The deliberate integration of qualitative and quantitative research approaches highlights the comprehensive approach adopted to guarantee that the system not only satisfies technical requirements but is also acutely aware of the real demands and difficulties encountered by the community of the deaf and hard of hearing.

The central objective of this project is to empower individuals with hearing impairments, providing them with a robust and user-friendly means of communication. Through accurate sign language recognition and natural-sounding speech generation, the system aspires to bridge the gap between the deaf and the hearing worlds. The potential impact of this development extends far beyond the realm of technology; it holds the promise of facilitating the full participation of deaf and hard-of-hearing individuals in various aspects of society. As communication is a cornerstone of human connection, the success of this project can significantly enhance the overall quality of

life for those who rely on sign language as their primary mode of expression.

It's great to see initiatives that leverage technology for social good and contribute to larger societal objectives like diversity and inclusivity. Using innovation to address the needs of marginalized communities is a powerful way to create a more just and compassionate society. The emphasis on prioritizing the needs of these communities is crucial for fostering positive change and ensuring that technology is harnessed for the benefit of all.

This initiative sets an excellent example of how technological developments can be tapped into to make a meaningful impact. By focusing on inclusivity, particularly in addressing the needs of individuals with hearing disabilities, it contributes to creating a more inclusive world. The ability of technology to enhance human interaction and participation, regardless of physical abilities, is indeed a beacon of progress.

As we look to the future, initiatives like these demonstrate the potential of technology to bridge gaps, break barriers, and contribute to a more equitable and interconnected global community. It highlights the positive role that innovation can play in shaping a better and more inclusive world for everyone.

REFERENCES

[1] M. Al-Qurishi, T. Khalid and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," in IEEE Access, vol. 9, pp. 126917-126951,2021, doi: 10.1109/ACCESS.2021.3110912.

[2] H. Luqman, "An Efficient Two-Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion," in IEEE Access, vol. 10, pp. 93785-93798, 2022, doi:10.1109/ACCESS.2022.3204110.

[3] Z. Wang et al., "Hear Sign Language: A Real-Time End-toEnd Sign Language Recognition System," in IEEE Transactions on Mobile Computing, vol. 21, no. 7, pp. 2398-2410, 1 July 2022, doi: 10.1109/TMC.2020.3038303.

[4] E. Rajalakshmi et al., "Multi-Semantic Discriminative Feature Learning for Sign Gesture Recognition Using Hybrid Deep Neural Architecture," in IEEE Access, vol. 11, pp. 2226-2238, 2023, doi: 10.1109/ACCESS.2022.3233671.

[5] J. DelPreto, J. Hughes, M. D'Aria, M. de Fazio and D. Rus, "A Wearable Smart Glove and Its Application of Pose and Gesture Detection to Sign Language Classification," in IEEE 21Robotics and Automation Letters, vol. 7, no. 4, pp. 10589-10596, Oct. 2022, doi: 10.1109/LRA.2022.3191232.

[6] K. Kudrinko, E. Flavin, X. Zhu and Q. Li, "Wearable Sensor-Based Sign Language Recognition: A Comprehensive Review," in IEEE Reviews in Biomedical Engineering, vol. 14,pp. 82-97, 2021, doi: 10.1109/RBME.2020.3019769.

[7] S. Khetavath et al., "An Intelligent Heuristic Manta-Ray Foraging Optimization and Adaptive Extreme Learning Machine for Hand Gesture Image Recognition," in Big Data Mining and Analytics, vol. 6, no. 3, pp. 321-335, September 2023, doi:10.26599/BDMA.2022.9020036.

[8] S. Turaev et al., "Review and Analysis of Patients ' Body Language From an Artificial Intelligence Perspective," in IEEE Access, vol. 11, pp. 62140-62173, 2023, doi:10.1109/ACCESS.2023.3287788.

[9] Milan Patel,"American Sign Language Detection", in IEEE Access, 2023

[10] G. Park, V. K. Chandrasegar and J. Koh, "Accuracy Enhancement of Hand Gesture Recognition Using CNN," in IEEE Access, vol. 11, pp. 26496-26501, 2023, doi:10.1109/ACCESS.2023.3254537.