

SILENT WHISPERS: A Real-Time Sign Language to Text Translator

Pratiksha Santosh Shenkar¹, Ms. Pooja Amin²

^{1,2} *Department of Information Technology, Chikitsak Samuha's S.S. & L.S. Patkar College of Arts & Science, and V. P. Varde College of Commerce & Economics*

Abstract—Individuals who are mute or hard of hearing often struggle to communicate effectively with the broader public, primarily because not everyone understands sign language. This creates a barrier in conveying their thoughts and needs in everyday interactions. The communication gap between sign language users and those unfamiliar poses significant challenges in various aspects of daily life. To address this issue, we proposed to develop a robust and efficient sign language detection system using Python. The objective of this project is to design and implement a machine learning-based system capable of recognizing and translating sign language gestures into written text in real-time.

It aims to develop a real-time sign language translator that converts sign language gestures into text. Leveraging Python for gesture recognition, Scikit-learn for machine learning, and Jupyter for data analysis and model training, the system provides accurate and efficient translation. The front, designed with React.js and supported by Node.js for development and deployment, offers an intuitive user interface. This integrated approach enhances communication accessibility for the deaf community by providing a seamless and interactive translation experience.

Keywords: Real-time translation, Gesture recognition, Sign-Language Detection.

I. INTRODUCTION

Silent Whispers is a real-time sign language-to-text translation system developed using computer vision and machine learning. It aims to bridge communication gaps between the deaf or hard-of-hearing community and those unfamiliar with sign language. The system captures hand gestures via webcam and processes them using a trained CNN model to display accurate text output. Core technologies include gesture recognition, computer vision, and human-computer interaction. An intuitive user interface ensures accessibility for users with varying technical skills.

The project emphasizes ethical design by addressing cultural nuances and avoiding biases. Key objectives include enhancing inclusivity, improving accessibility, and supporting seamless communication. It is designed to work in educational, professional, and social settings. The system is currently limited to one sign language and static gestures. Environmental factors and hardware limitations may impact recognition accuracy. Accessibility is enhanced through user-friendly design and real-time feedback. User data is protected through strong privacy and security measures. Applications include classrooms, healthcare centers, public services, and daily communication. The system showcases technological innovation in assistive communication tools. Ultimately, Silent Whispers aims to promote inclusivity and empower deaf individuals through technology.

II. LITERATURE REVIEW

Previous studies have explored various methods for recognizing sign language, ranging from sensor-based gloves to vision-based systems. While sensor-based approaches offer precision, they are often expensive and intrusive. Vision-based systems, particularly those utilizing machine learning, have gained popularity due to their non-invasive nature and scalability. CNNs are widely recognized for their strength in classifying images and have been effectively utilized in recognizing patterns in ASL-based gesture input. Our work builds upon these findings by integrating CNN models with a web application to enable real-time gesture recognition.

III. METHODOLOGY

The system architecture of Silent Whispers consists of three primary components: gesture acquisition, gesture recognition, and text display. A webcam captures hand

gestures, which are preprocessed and fed into a trained CNN model. The model outputs the predicted alphabet corresponding to the hand sign. The backend, built-in Flask (Python), handles model inference and communicates with the React-based frontend to display the translated text in real time. The dataset used is a modified version of the ASL alphabet dataset with augmentation techniques applied to improve generalization.

IV. SYSTEM ARCHITECTURE

The system comprises a client-server model. The frontend captures video input and sends frames to the backend via REST API. The backend processes each frame through the CNN model and returns the predicted character. The front end updates the UI dynamically to show recognized signs as text. The CNN model includes multiple convolutional layers followed by pooling, ReLU activation, and fully connected layers. Technologies used include OpenCV for image capture, TensorFlow/Keras for model training, Flask for backend API development, and React for the user interface.

V. RESULTS & DISCUSSION

The CNN model was trained on a dataset of American Sign Language alphabets. The model achieved an accuracy of 96.8% on the validation set, demonstrating robust performance in recognizing static hand gestures. During testing, the system performed well under different lighting conditions and hand orientations. The web interface responded with minimal latency, making it suitable for real-time communication. A usability study with ten participants revealed that users found the system intuitive and useful for daily conversations with hearing-impaired individuals.

VI. CHALLENGES & LIMITATIONS

One major challenge faced was the accurate classification of similar-looking hand gestures such as 'M' and 'N'. Environmental factors like poor lighting and complex backgrounds also affected model accuracy. The current system is limited to static signs from the ASL alphabet and does not support dynamic gestures or full-word recognition. Additionally, performance depends on the quality of the webcam and

the processing power of the device running the application.

VII. FUTURE SCOPE

Future enhancements include expanding the model to support dynamic gestures and full sign language phrases. Also, in the future just like Google Meet, we will be creating an app that will be mostly used by people who use sign language frequently and those people who communicate with people who use sign language. This app will be similar to Google Meet but will have features like clicking on a button e.g. named "Translate" which will give the output when the mute/deaf will do gestures, it will be captured by the camera and using the appropriate algorithm and data uploaded in the database the output in the form of text will be shown to the opposite person's screen who doesn't know sign language. This will benefit both parties where the mute/deaf and the person who is unaware of sign language in a way they don't need to hire a translator for communication and there will be no miscommunication from the end.

VIII. CONCLUSION

The Silent Whispers project successfully achieves its goal of translating sign language gestures into text, bridging the communication gap between the deaf and hearing communities. Through the integration of machine learning, computer vision, and a user-friendly web interface, this web application enables real-time sign language recognition and accurate text conversion. The project ensures accessibility and inclusivity, making digital communication more equitable. Additionally, its cloud-based architecture enhances flexibility, allowing users to interact seamlessly across various devices.

The testing phase, which included functional, non-functional, black-box, and white-box testing, validated the robustness of the system. Performance and scalability tests confirmed that the system can handle multiple user inputs efficiently, ensuring a smooth experience. Moreover, cost estimation using the COCOMO model provided insights into the project's development efforts, making it a feasible and practical solution. Through continuous testing and modifications, the accuracy and efficiency of the

application were optimized, resulting in a reliable and responsive system.

In the future, Silent Whispers can be expanded to support multiple sign languages, integrate text-to-speech functionality, and implement AI-driven predictive text for better user experience. Enhancements in gesture recognition, deep learning models, and mobile compatibility will further refine its performance. By leveraging technology for social good, Silent Whispers has the potential to make digital interactions more inclusive, fostering effective communication for individuals with hearing impairments.

REFERENCE

- [1] Chen, S., Mulgrew, B., & Grant, P. M. (1993). *A method for clustering in digital communication channel equalization using radial basis function neural networks*. IEEE Transactions on Neural Networks, 4, 570–578.
- [2] Duncombe, J. U. (1959). *Infrared navigation—An early feasibility assessment (Part I)*. IEEE Transactions on Electron Devices, ED-11, 34–39.
- [3] Lin, C. Y., Wu, M., Bloom, J. A., Cox, I. J., & Miller, M. (2001). *Robust public image watermarking resistant to rotation, scaling, and translation*. IEEE Transactions on Image Processing, 10(5), 767–782.
- [4] Grassnoted. (n.d.). *ASL Alphabet Dataset*. Kaggle. Retrieved from: <https://www.kaggle.com/datasets/grassknotted/asl-alphabet>