GestureSpeak: A Real-Time Sign Language Interface

Harsh Patel [1], Hanil Shah[2], Harsh Sharma[3], Shekhar Hande[4], Sham Hange[5], Makarand Upkare[6]
[1,2,3,4,5,6] *Vishwakarma Institute of Technology, Pune*

***Abstract*— The system, promises to recognize gestures in real time and translate them into sign language. The goal for the system is to act as a liaison for the people who can communicate using sign language and those who cannot by translating their gestures into spoken English. From the point of view understanding using it, the system aids inclusivity while reducing communication gaps within the society. The system does not only incorporate innovative technology but also addresses the social problem of lack and barriers to communication that the people suffering from the disabilities face on a daily basis. This guarantees that rational solution has been provided to improve access and ease of communication which highlights the urgency of supporting the community by providing socially diverse services.**

***Index Terms*— Empowerment, Inclusive Communication, Real-Time Gesture Interpretation, Sign Language Translation, Societal Impact.**

I. INTRODUCTION

Imagine hands that speak volumes through silent storytelling. Millions of people use sign language, and for them, communicating can frequently feel like a silent battle that separates them from the spoken world. By implementing a real-time sign language translation system that gives voice to gestures and creates understanding bridges, this project aims to dismantle that invisible wall.

This system serves as a connection medium in addition to being a technical tool. It converts expressive hand gestures into comprehensible spoken language in real time. Imagine using the seamless beauty that this technology brings to daily life to have meaningful conversations with coworkers, order coffee with style and ease rather than using a scrawled note, or share a quiet joke with a friend.

Fundamentally, this is a movement to break down barriers to communication and create a more inclusive social fabric rather than merely a tale of innovation. In this idealized world, relationships flourish between various communities and everyone, regardless of background or ability, has the chance to express themselves with confidence.

Simple conversations like ordering a favorite drink, sharing sincere words, or whispering secrets that were previously suppressed can now be openly expressed. Beyond the technical details, our research explores the human side of this transition. In the future, communication will be enhanced by the expressive language of the hands rather than being limited to spoken words. It does this by amplifying silent voices and fostering a symphony of empathy.

The distinction between hearing and non-hearing starts to erode in that future, giving way to a dynamic web of connections where each gesture conveys a message and each conversation broadens comprehension.

II. LITRETURE REVIEW

The article identifies issues with the deaf community in communication and promotes solutions that are inclusive. It addresses limitations during emergencies and daily lives as a result of communication barriers. Digital solutions such as AI-based sign language translators are being developed to assist deaf people in digital as well as physical spaces. A particular project is explained, involving a sign language translator using a machine learning pipeline that can recognize more than 500 sign language gestures. The objective of the project is to enable the deaf and reduce gaps in sign language recognition technology for social development. The conclusion discusses the implications of this study and offers possible future direction in this research area.[1]

The system employs special sensor-equipped gloves to record hand gestures and various machine learning methods to transform the gestures into written or spoken words. The system's data consists of a dataset of data relating to hand movements captured using gloves. The dataset contains 96 words or classes for various signs, and machine learning algorithms such as k-Nearest Neighbor, Decision Tree Classifier, and Neural Network are employed in classifying and interpreting the hand movements into words. The objective is to develop a device that enables deaf and mute individuals to communicate easily with people who don't know sign language, without the use of a human interpreter.[2]

Emphasis on computer vision methods, which encompass constructing interfaces based on the way people process visual data. The aim is to construct a system able to effectively identify sign language gestures through such parameters as finger orientations, hand shape, and complexion. The writing stresses difficulties in designing a vision-based interface that is applicable worldwide because of differences in sign language, but it is positing that focusing on a specific group or country is more reasonable. The essay summarizes past approaches and methods of research and is intended to offer insights for future research on sign language recognition systems [3].

The emphasis is placed on Indian Sign Language (ISL), that is employed by deaf people in India. The paper introduces an automatic method for recognizing static gestures in the Indian sign language alphabet, specifically for 17 English alphabet letters. The method includes the employment of skin color statistics to separate the hand, image binarization, and feature extraction methods. These characteristics are then employed for classification and identification using Artificial Neural Network and SVM (Support Vector Machine). The aim is to develop a system that is capable of easily and effectively identifying Indian sign language gestures in real-time so that deaf people can easily communicate.[4]

This article highlights the importance of communication for both sides, but at the same time reveals the challenges that occur when hearing-impaired individuals try to speak with deaf or mute people. The main issue is the poor understanding of sign language, which often deaf individuals use. A sign language translation application is in the process of development in an attempt to bypass this barrier. This application will convert the 26 alphabets of Bahasa Isyarat Malaysia (Malaysian Sign Language) with OpenCV image processing technology. Enhancing and making communication between people with disabilities and non-disable people is the ultimate aim.[5]

This paper explains a glove specially designed to assist in communication by deaf or silent people. Following the American Sign Language Standard, the glove is able to read sign language gestures and convert them to words. It has an accelerometer to monitor hand positions and sensors to measure bending and finger movement. They made the glove learn about different sign language movements and translate them into letters in real time with a method known as Principal Component Analysis (PCA). The letters and words are spoken aloud and displayed on the glove when it is paired with an Android phone via Bluetooth. The glove is accurate 92% of the time, which is quite good.[6]

The emphasis in sign language recognition (SLR) studies in the last 20 years has shifted from focusing on single signs to the grammatical and syntactic structures of sign language. With a focus on linguistic intricacies, the authors pose the Sign English Translation (SLT) challenge, which involves translation of spoken English to sign language. The authors introduce the RWTH-PHOENIX-Weather 2014T dataset, a publicly available Continuous SLT corpus, through Neural Machine Translation (NMT) in end-to-end and pretrained cases. This website facilitates Neural SLT testing through the provision of weather announcements in German Sign Language as well as glossaries and translations. Quantitative and qualitative results are reported for various SLT configurations, proposing interesting directions for future work in this new field. Although end-to-end tokenization networks translated with 9.58 and 18.13 BLEU-4, the upper bound translation quality reached 19.26 BLEU-4 on levels frames and gloss, respectively.[7] Translation accuracy is significantly improved when signs are individually recognized, as per previous research on sign language translation.

Gloss-level tokenization is used in the most recent translation techniques. The paper presents a novel transformer-based method that learns Continuous Sign Language Recognition and Translation end-to-end at the same time. This is accomplished without the need for timing information by combining translation and recognition into a single architecture through the use of a Connectionist Temporal Classification (CTC) loss. The evaluation performed on challenging PHOENIX14T dataset illustrates state-of-the-art performance surpassing existing models. Translation networks outperform both gloss to speech and sign video to speech models; in some conditions, they even double performance (9.58 vs. 21.80 BLEU-4). The paper further offers new transformer network baseline performances for a range of sign language translating tasks.[8] Sign Speaker is an innovative American Sign Language (ASL) recognition system using inexpensive, portable hardware such as smartphones and smartwatches.

It utilizes the smartphone loudspeaker to translate sign signals that are captured on the smartwatch. The results of the test are astounding: 99.2% ratio of detection, 99.5% reliability in sign recognition, and a very small 1.04%-word error rate in running sentence recognition are reported when translating an eleven-word sentence within only 1.1 seconds. Its mission is to grant the deaf population reliable, real-time communication.[9] The Sign4PSL application to assist in better hearing-deaf communication is presented in this research.

It addresses the challenge individuals unfamiliar with Sign Language (SL) face in interpreting daily sentences. Using a virtual signing character, Sign4PSL translates English sentences into Pakistan Sign Language (PSL), making it possible to communicate. It is portable, cross-platform, and offers offline text translation. Sign4PSL effectively conveys stories when practiced on deaf students, showing its potential in facilitating inclusive communication and bridging the difference between individuals who are deaf and those who are not.[10] AI-based technologies can significantly enable deaf or hearing-impaired individuals to interact with other communities and facilitate social inclusion.

Various applications have emerged as a result of recent developments in sensing technology and AI algorithms to suit the demands of these communities. This survey summarizes the most up-to-date techniques in sign language capturing, recognition, translation, and depiction, pointing out their merits and drawbacks. It also investigates applications and discusses difficulties in the area. The survey proposes directions for future research to lead future researchers in the development of sign language technologies, prioritizing the potential of AI to overcome communication barriers and improve the social integration of the deaf and hearing-impaired population.[11] The main research papers analyzed various methods of sign language recognition and translation.

One research paper relies on gloves with sensors, which record hand movements, and another on computer vision to parse command and positioning. According to some research, linguistic aspects are emphasized to provide improved grammatical accuracy in translation, while others investigate various classification techniques for improved recognition. Another study proposes handheld devices such as smartphones and smartwatches for real-time communication. Each of the papers confronts specific challenges like data collection, speed of processing, and differences in sign language that all work towards the overall objective of enhancing accessibility and communication among the deaf.

III. METHODOLOGY

The suggested system offers an effective, real-time translation platform by combining traditional sign language interpretation with a larger library of user-defined gestures. From data collection to deployment and continuous system improvement, the methodology is divided into several structured phases. Below is a description of each step.

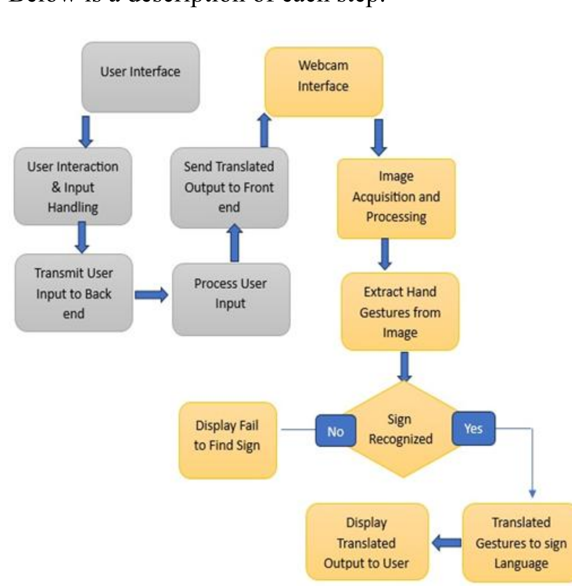

Fig 1 "Sign Language Recognition Process Flow"

Collect Data: A large collection of sign language videos was assembled in order to create an extensive and representative dataset. Standard signs and their corresponding written or spoken expressions were among them. Diverse gestures and expressions were included thanks to cooperation with sign language communities and domain experts. Enhanced contextual translation was made possible by the introduction of custom shorthand gestures in addition to commonly known signs, such as the thumbs-up sign, which is used to convey entire sentences.

Data Preprocessing: To eliminate background noise and standardize resolution and lighting variations, the raw video data was pre-processed and standardized. Important characteristics were manually annotated, such as hand gestures, facial expressions, and upper body movements. In order to streamline repetitive usage and enhance system familiarity, a gesture history module was added, enabling users to refer to previously executed signs.

Feature Identification: Critical components like hand trajectories, finger articulation, facial cues, and general posture were identified and isolated using visual processing techniques. The recognition of unique gesture patterns was made easier by frame-by-frame motion tracking and segmentation. The system extends the translation capabilities beyond traditional sign language by supporting both standardized and customized gesture input.

Gesture Classification: For gesture recognition, a classification method based on proximity was applied. Thirty reference images were saved for every gesture, and distance metrics were used to compare them to fresh input. In real-time situations, this method offered dependable performance with low latency. The chosen method was especially well-suited for deployment in resource-constrained environments because of its interpretability, simplicity, and low computational requirements.

System Training: The structured dataset was used to train the system to translate gesture inputs into spoken or textual outputs. Custom gestures were given extra thought, and in order to guarantee accurate recognition, specialized sample sets were needed. By comparing input samples to stored references, the classification mechanism made it possible to quickly identify gestures.

Evaluation Phase: The accuracy and responsiveness of the system were assessed using a different test dataset. Standard metrics like accuracy, precision, recall, and F1-score were used to gauge performance. Both standard and custom gestures were used in the evaluation, showcasing the system's versatility and efficiency. The classification method's lightweight design allowed for extensive real-time testing without performance degradation.

Refinement Process To improve accuracy and generalization, post-evaluation adjustments were made. As needed, dataset augmentation, system optimization, and parameter tuning were implemented. To improve user interaction, a web interface was also created, offering features like the ability to review gesture history, tooltips for explanations, and access to additional video resources for user support and training.

Deployment Stage: The finished system was made available on a variety of platforms, including mobile and web-based interfaces. Users can customize the gesture input to suit their own preferences thanks to the design's emphasis on accessibility, customization options, and easy navigation. To guarantee reliable performance, the solution was tested in a range of environmental settings.

Continuous Enhancement: A cycle of continuous improvement was created to keep the system relevant and increase its capabilities. This covers frequent system testing, incorporating user feedback, and updating data on a regular basis. New gestures and functional improvements can be added quickly and effectively without requiring a total redesign thanks to the modular architecture.

IV. FUTURE SCOPE

Future advancements will have the system render the translated signs as more natural and fluid, imitating how sign language looks like in actual conversations. Future improvements will also have the system able to interpret different sign languages, including ASL, BSL, CSL, LSF, and others, making it adaptable for use by users from different regions. The addition of facial expressions and body language will help capture emotional context and enhance the quality of communication even further. For daily, real-time use, the system can be integrated into wearable devices such as earbuds or smart glasses. Haptic feedback can be implemented to allow for the delivery of verbal responses to signers through touch, thus enabling more efficient two-way communication. In school settings, the system can be utilized to help deaf kids learn sign language better. Programs can also be developed to help sign language interpreters so they can perform improved in real-time applications. Collaboration with schools and businesses will help integrate the system for workplace application and learning support. The overall objective is to make communication more accessible to all. Working with various groups in various regions of the globe will make the system open and culturally adaptable, and translating software from spoken to sign language can facilitate bridging gaps in communication in various cultures.

V. CONCLUSION

Our sign language technology from text is more than a technical fix, a voice raised above the void of silence. It creates uninterrupted communications between sign language users and listeners by whispering meaning in the vocabulary of handshapes. Cumbrous pencil fumbling and slow explanation no longer interrupt conversations. Orders of coffee, nuanced discussions, and jokes all ride smoothly on this unseen bridge of spoken word. Rather than focusing on just technological innovation, our project aims to empower individuals, promote inclusivity, and break down barriers. This revolution insinuates a future wherein hands sign and all motion is meaningful. In the future, people with varying skill sets will work together in harmony to create a web of understanding and bonding. The voiceless are given voice in this chorus, and dialogue dances around a world that day by day is becoming more intertwined.

VI. ACKNOWLEDGEMENT

We would like to genuinely express our gratitude to all who helped us develop our AI-based sign language translator. We have worked together on this and are incredibly thankful for the guidance and support we have attained. We are extremely thankful to Prof. Makarand Upkare Sir, our mentor, whose input and encouragement have significantly impacted the way of our project. Their sage advice and steadfast encouragement have significantly pushed forward our knowledge of AI technology and language translation. The volunteers who made the time to test and provide feedback are especially appreciated. They have been instrumental in the system's development and to making sure it is applicable in real-world applications. Given that every participant in the projects contributed greatly towards the effective deployment of the AI-powered sign language interpreter, we would like to thank them for the dedication and cooperation. Their cooperation and commitment have contributed immensely to the success of the projects. We also thank the scientific and academic community for its contribution in ensuring that an environment conducive to creativity and learning is fostered. The progress of the project has been significantly facilitated by the input and expertise that have been provided by our fellow colleagues. We would like to thank our families, specifically, for their support and understanding throughout our experience. We deeply appreciate their unshakable belief in our capabilities, and their encouragement has been motivating.

REFERENCE

[1] W. Qin, X. Mei, Y. Chen, Q. Zhang, Y. Yao, and S. Hu, "Artificial Intelligence for Sign Language Translation - A Design Science Research Study," *Communications of the Association for Information Systems*, vol. 53, no. 1, pp. 545-573, 2023.

[2] P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra, "A multimodal framework for sensor-based sign language recognition," *Neurocomputing*, vol. 259, pp. 21-38, 2017.

[3] S. S. Rautaray and A. Agrawal, "Vision-based hand gesture recognition for human-computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1-54, 2015.

[4] J. Singha and K. Das, "Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 2, pp. 188-195, 2013.

[5] R. Zainuddin, N. S. Azman, and R. Sahak, "Development of Bahasa Isyarat Malaysia (BIM) Translation Application Using OpenCV," *International Journal of Advanced Computer*

Science and Applications, vol. 10, no. 9, pp. 323-328, 2019.

[6] K. Lee, J. Lee, and M. Kim, "Development of a Sign Language Translation System Using Wearable Device," *IEEE Access*, vol. 6, pp. 40120-40128, 2018.

[7] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neural Sign Language Translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7784-7793.

[8] H. Zhou, W. Zhou, Y. Zhou, and H. Li, "Spatial-Temporal Multi-Cue Network for Continuous Sign Language Recognition," in *AAAI Conference on Artificial Intelligence*, 2020, pp. 13009-13016.

[9] C. Wang, X. Zhang, X. Zhu, Z. Chen, Y. Xiang, and W. Zou, "SignSpeaker: Real-time, Low-cost Sign Language Translation on Mobile Devices," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1-14.

[10] S. Kausar, M. Y. Javed, and S. Haroon, "Sign4PSL: A Communication Bridge between Hearing-Impaired and Normal People Using Pakistan Sign Language," *IEEE Access*, vol. 7, pp. 180435-180447, 2019.

[11] Y. Zhang, S. Yang, J. Xiao, S. Shan, and X. Chen, "Artificial Intelligence Technologies for Sign Language," *Journal of Artificial Intelligence Research*, vol. 69, pp. 1234-1256, 2021.

[12] Wadhawan, S., & Kumar, P. (2021). "A Deep Learning Approach for Indian Sign Language Recognition Using CNN and LSTM." *International Journal of Computer Applications*, 183(19), 12-18.

[13] Koller, O., Camgoz, N. C., Ney, H., & Bowden, R. (2020). "Weakly Supervised Learning with Multi-Stream CNN-LSTM-HMMs to Discover Sequential Parallelism in Sign Language Videos." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(9), 2209-2222.

[14] Ding, L., Martinez, J. A., & Pugeault, N. (2022). "Real-Time Vision-Based Sign Language Recognition Using Deep Neural Networks." *Neurocomputing*, 481, 1-16.

[15] Momeni, L., & Evans, R. (2019). "Challenges in Sign Language Recognition: A Comprehensive Review of Data Collection and Recognition Methodologies." *Journal of Machine Learning Research*, 20(1), 1-27.

[16] Zhu, Z., Huang, C., Xu, X., & Bai, L. (2019). "Hand Gesture Recognition with Depth and Skeletal Data for Sign Language Translation." *IEEE Transactions on Human-Machine Systems*, 49(2), 233-245.