Website for seamless translation of sign language using KNN and TensorFlow

V Karthick, Associate Professor
Department of CSE
Rajalakshmi Engineering College
Chennai, India
vkarthick86@gmail.com

Tamil Priya V, UG Student
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701282@rajalakshmi.edu.in

Subhikshaa S, UG Student
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701264@rajalakshmi.edu.in

***Abstract*—** **The Sign Language Translator is a user-friendly website designed to train the sign language that acts as a bridge for the communication between users. The core functionality lies in capturing and training individual signs with their corresponding meanings. It proceeds with a start and stop gesture to maintain a flow of end and beginning of the translation. It uses a minimum of thirty samples to get trained with more accuracy in recognizing the sign language shown. The start and stop gesture is collected and trained. The next step includes the signs that need to be trained with their meaning with atleast thirty samples each. This translator depicts how a machine learning model works by training the samples and testing on the data given by the user. It detects the sign by using the trained samples and gives the result as text with computer generated speech in the background. One can sign multiple words with one gesture and copy the translated text with a click of a button. Its purpose is to allow users to communicate more effectively with their computers and other people. The use of KNN (K-Nearest Neighbours) algorithm contributes to gesture recognition of sign language when trained with samples. Users can initiate and conclude the translation process through start and stop gestures trained by them thus giving a controlled flow. Previous research work in this area has models that can translate trained symbols in 2.3 to 3.1 seconds with a confidence level of 84%. We targeted to focus on this time interval, confidence level and also showcase the project with a straight-forward user interface. It also serves more as an educational tool in the view of communication between deaf people and the normal people providing a way to vanish the indifferences. This translator additionally provides the feature of copying the text translated from recognizing the sign language which offers functionalities such as sharing it with others or integrating it to a document. The auditory reinforcement of the translated text is another notable feature of this translator offering a clear vocal representation.**

Keywords*—** ***Sign Language translation, computer generated speech, KNN, auditory reinforcement, gesture recognition

I. INTRODUCTION

Sign language is a medium of communication used by people who cannot hear or have any difficulty in hearing and processing oral communication [1]. Sign languages themselves contain a set of lexicons and their own conventions for representing a word or letter. They essentially utilize the visual learning practice [2] to convey the ideas to people. Often general people find it difficult translating their normal ideas, speech or talks to sign language to be understood by the people who have difficulty in hearing. There are people who can translate the normal oral communications and speeches to sign languages simultaneously and deliver them to people with hearing disability seamlessly [3]. They work as Sign language Interpreters and get paid. But we can find that these people who work as interpreters are very few in number. In an idea to automate this process and help people use this technique to communicate with people with disabilities, we have come up with this solution of a commonly accessible website that translates sign language in a swift way. This website can be used by a wide range of people at ease as the visual user interface is designed with utmost care to provide a rich user experience with less complexity in screen navigation. Understanding and translating gestures [4] to words is the ultimate work that drives our research. Also for people who cannot talk as well as hear use sign language as their principal mode of communication [5]. For them, the presence of voice assistants everywhere we counter, may be difficult to access or in most of the cases they will not help. They need a mechanism which can convert their thoughts or ideas to text as instantly as possible to interact with common people and to bridge the gap between themselves and normal users. These people can also make use of our website and get their thoughts in textual format instantly which they can also copy and use to search, browse or communicate through social media platforms. They can also train their own words and sentences priorly and can customize the process according to them.

The website requires the users to train samples for the start and stop symbols priorly while opening it. Then they can be used to train words or sentences and later translate them. Normal people who cannot understand sign languages can use this platform to get to know what the person with hearing or vocal challenges is trying to convey in seconds or milliseconds. They can use this as a mode of communication to collaborate and empathize with them easily. This website can also be used by deaf and hard of hearing people within their communities effectively for enhancing their communication thereby building strong relationships within the community [6]. The additional computer based voice system provides a voice based assistance which reads the translated texts aloud. This can be used as a second layer of verification for confirming the texts or ideas as they have been translated properly or not. This can also enhance the overall user experience and helps all the people connect more with the platform and use it with ease. The start and stop gestures ensure proper flow of translation.

The work by Prashant and Khushboo [7] demonstrates the recognition of various ISL hand gestures and alphabets. They have performed detailed analysis of the human-computer interaction approaches and other image processing techniques and have identified some key difficulties including costs and budgets. Their system uses OpenCV and python in training the model. They tried to provide an answer for almost all the underlying questions they have identified. Utilizing color techniques, they have tried to provide a stable solution.

Srivastava, Amisha Gangwar, Richa Mishra, Sudhakar Singh [8] in their research on sign language recognisers, have concentrated more in the recent advancements and have confined their research to the latest possible date of publishing. They have considered various inclusion and exclusion parameters for filtering the studies. Their inclusion criterias were based on the language of the paper, the algorithm used with the paper and the keywords and their exclusion criterias were based on the papers that did not implement deep learning models and also the ones that were not fully accessible.

Siming He [9] in his research work has studied the sign language applications and have developed a hand locating, sign language recognition system based on popular and familiar signs using neural networks. Using LSTM coding and decoding, he has established a sign language recognition framework which handles and processes the data. They aim to increase the recognition accuracy. This paper uses a hand locating network to overcome the problem of RGB color recognition.

Yang sang, Mengru Liu and Feilu Wang [10] have employed the use of pressure sensors and have integrated 7 different categories of gestures, each category or group having a set of gesture symbols together in a wrist band. Using the response time of the sensors, they have developed their solution.

Achenbach, Laux and Muller [2] has demonstrated the use of hand gloves in recognising hand symbols. They have employed multiple classifiers that can filter out samples and non contributing features to train an efficient model.

Pathan, Biswas and Yasmin [11] have used a multi-headed convolutional network in recognising known words through hand gestures. They have used a variety of datasets in training the model for basic symbols and sentences.

Gonzalez [12] has concentrated his research on the area of processing images digitally for training the model in predicting results of hand movements. They claim that proper training using digital image processing techniques can render fruitful results.

Cozza, Fiasca and Martini [1] have gone through the mechanism behind hearing loss and how people with hearing difficulty process our oral communication. Their study about the working underlying the deaf people have provided a different perspective in this area of research.

Ewe, Lee and others [13] have attempted to limit the overfitting problems on sign language recognition of US datasets using Random Forest algorithms. They have also analyzed the research papers published before them and have mentioned their complexities.

Kakizaki, Miah, Hirooka and Shin [14] have developed a solution for the community of deaf people in japan using MediaPipe. They have used RGB cameras for carefully collecting the dataset. They have mentioned extending this work as a global project in their future work.

Ngugen and Wang [15] have highlighted the use of latitude and longitudinal axes for location detection. They have predicted the speech using the angles in facial movement and have proven a significant accuracy level. They claim that smaller changes in facial angles can also be detected and can prove helpful in the overall recognition process.

However, all these solutions on the whole lacked a commonly accessible and desirable user interface with a focus on speedy translation like websites with simple yet powerful and desirable User Interface which motivated us to create one.

II. MATERIALS AND METHODS

The website requires 30 samples for the start and stop symbols. Subsequently the users can also train modals for different signs or symbols each with 30 or more samples. The TensorFlow object detection API detects and uses all the samples to translate signs seamlessly in the next step. The portal also provides a vocal representation which is also an additional advantage. After providing the samples for start and stop symbols. The website provides the option for adding gestures with a label. Each of them requires a minimum of 30 samples for training with each gesture. All the images should be taken in a well lighted environment to increase the efficiency and accuracy of the model. It also works with low lighting environments however confidence level of prediction may vary. An active copy/paste functionality is available for the users to copy which requires the browser permissions for pasting.

Hardware requirements of the project include a working operating System which can be either Windows or Linux with a minimum RAM of 4 GB and a Secondary Storage of 256 GB. The Processor should be either Intel Core i5 (8th generation or higher) or AMD Ryzen 5 (3rd generation or higher). Additionally, for executing the project successfully a Web Camera along with audio speakers are required.

Software requirements include the Internet browser (Chrome/Edge/Mozilla Firefox) with a stable internet connection. An operational WebCam and audio drivers for speakers. For building the project a code editor, preferably Visual Studio Code is essential. For development Node JS should be installed in Visual Studio Code.

III. EXISTING SYSTEM

Existing system includes sign language translation [16] using OpenCV, Deep learning algorithmic techniques and libraries to train the model with a specific set of dataset or symbols and uses it to test separate samples. Several researchers also employ perception neuron modal to translate sign languages of various communication languages [17]. Vision based systems have also been proposed for image distinction [18]. Image processing has to be very accurate in translating them to words. However there are also significant challenges [19] posed over this.

Sign language detection systems with motion initiation [20] through angular movements where the position and orientation of objects have a direct influence on the output and automatic classification of symbols using reptile search algorithms [21] have also been used. There are also systems that have employed electromyographic signals in the recognition process and have been developed or trained to detect known signs and symbols [22]. These systems used a large number of pre-existing datasets that were confined to a significant number of unique gestures. Scientists and researchers have also dwelled into the working of the brain [23] that acknowledges sign languages, hand gestures swiftly in people with hearing difficulties. These observations have also significantly impacted the model building and training steps. Software bots for automatically generating signs and gestures [24] has also been an important area of research which is implemented in very rare models.

IV. PROPOSED SYSTEM

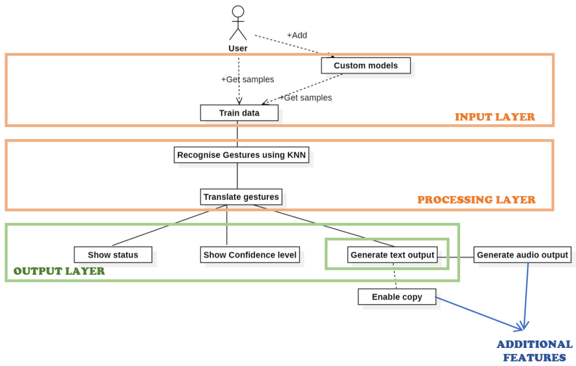

Figure 1. Architecture of the proposed website for sign language translation

Our proposed solution aims to provide a flexible method or platform where users of any category can train the model with customized signs and can assign values to their signs. The use of adobe illustrator files ensures high quality designs are being rendered. They process the input symbols received from the user and transform them into suitable formats for training. The system uses KNN algorithm and TensorFlow libraries to detect and translate the signs. Training the model for 75 rounds and combining it with stratified k-fold cross validation, confidence level of prediction can be significantly increased. As our solution is employed as a website which can be made available anywhere in the world and can be used anytime. It proves to be a reliable solution with high user desirability and at the same time ensures feasibility. Users can directly start to train the model with a start and stop symbol. For each symbol the model requires a minimum of 30 samples and more than 30 is also accepted. While the process of translation begins the user should start with the start symbol ensuring the start of the translation process and can show different gestures. If it is already trained for the shown gestures, it will be generated below the camera space and also simultaneously a audio will be played deciphering or reading the generated text aloud for the users to hear and also thereby confirm.

V. METHODOLOGY

Our system uses KNN (K- Nearest Neighbor Algorithm) for training the model with the given image samples and recognising them. The following steps are included to describe the flow of training the samples and recognizing the signs respectively.

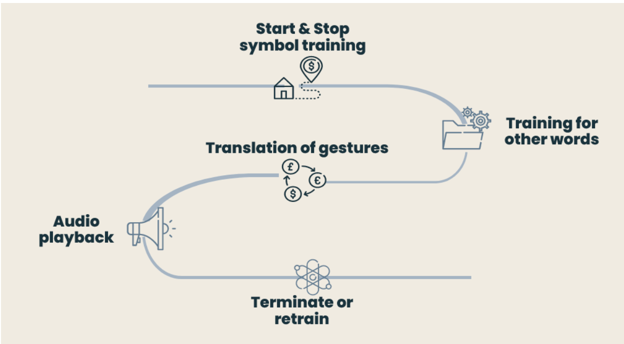

Figure 2. Overview of the methodology

A. Data collection:

The initial step of the website is to collect sample data that is getting images of signs from the user to train the start and stop gesture. It requires a minimum of 30 samples to train each of the start and stop gestures. These pre required signs are collected to proceed to the next step to train other signs. The start and stop signs serve as markers to indicate beginning and end of a sentence to determine the prediction of sign language as a conversation.

B. Adding custom words:

The model can be trained with samples for their own words or sentences and can view their added words with their trained gestures below the training area. All these new words or sentences require a minimum of 30 samples as mentioned earlier.

All these samples can be continuously captured or any faulty image taken in between can be clear according to a last in first out fashion.

C. Translation initiation:

Once done with adding samples for their corresponding gesture, users can click on the start translation button to initiate the translation process. They have to start with the start gesture trained in the previous step to get started with the process.

D. Translation process:

During the translation process, each predicted word with the confidence level appears straight before the video location. For models trained in a well light environment and moderately lit environment, confidence level will be much closer to 100%. In most of the cases, it achieves a confidence level of 99%.

E. Audio generation:

While the translation is being seamlessly done, an additional voice that converts the text to speech also gets triggered. It reads aloud the content being translated after the start symbol in a seconds time.

F. Live status availability:

Whether translation is being done or not, there will be a status message pop up that shows what is happening with the model. The users can use this as a quick hold for getting through what is happening at the second with the model. It shows if it is translating or not at the left corner of the screen.

G. Translation completion:

Showing the stop symbol to the translator marks the end of the translation process. Anytime during this or the previous stage, users can go back and add new words and train gestures by providing live samples. Until the user returns back to the homepage, all the trained data is retained in and to other navigable pages.

VI. RESULTS

Thus our website was able to provide a swifter response than the other systems that we referred to. Also as the website has an intuitive User interface that can be accessed by people of any age, the target audience for our project has been expanded to a greater extent.

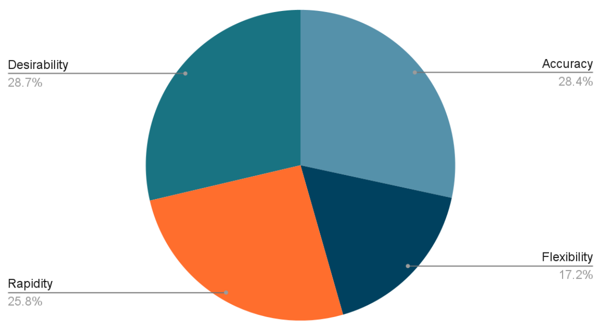

Figure 3. Features that were observed in our solution

The metrics that we focussed on for transition which was the time taken to recognise and translate the gesture was reduced and we achieved a greater accuracy with a quicker and yet stable response time.

VII. DISCUSSION

The previous works that we considered mostly were constrained to specific categories of people being the audience or target users. Also, their model worked for a common vocabulary dataset and does not allow for any customisation options.

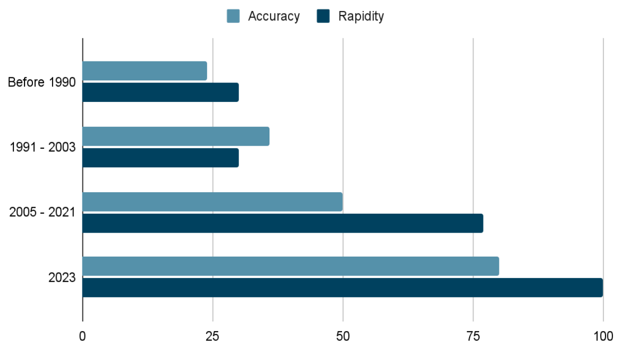

Figure 4. Comparison of previous research and their results (Accuracy and Rapidity in prediction) over the years

Our platform is a straightforward solution developed by considering people with hearing difficulties. We have included copy and paste sharing features. The audio system that was an auxiliary add on enhanced the overall efficiency. These additional features ensure user satisfaction and pain-free access.

Our solution can also be equipped with storing and saving functionalities as a future work so that it can be even more helpful for people while used on a daily basis.

VIII. CONCLUSION

Thus our solution aims at providing a reliable and easy-to-use solution that is accessible for all people indifferent to having any hearing discomforts. The facilities employed for easier transfer of recognised text makes it trouble-free in sharing information. Our goal is to provide an efficient medium or translator that can be trained and used for rapid translation for any person according to their own needs and we were able to see remarkable results such as a rapid response in the range of 5 to 6 microseconds which is much faster than the previous models we studied.

IX. REFERENCES

[1] A. Cozza, V. M. Di Pasquale Fiasca, and A. Martini, “Congenital Deafness and Deaf-Mutism: A Historical Perspective,” *Children*, vol. 11, no. 1, Dec. 2023, doi: 10.3390/children11010051.

[2] P. Achenbach, S. Laux, D. Purdack, P. N. Müller, and S. Göbel, “Give Me a Sign: Using Data Gloves for Static Hand-Shape Recognition,” *Sensors* , vol. 23, no. 24, Dec. 2023, doi: 10.3390/s23249847.

[3] J. Pataky, E. C. Demalis, J. Shelly, K. Miller, Z. M. Moore, and M. E. Vidt, “Use of a factor analysis to assess biomechanical factors of American Sign Language in native and non-native signers,” *J. Biomech.*, vol. 165, p. 112011, Mar. 2024.

[4] S. Escalera, I. Guyon, and V. Athitsos, *Gesture Recognition*. Springer, 2017.

[5] A. Camurri and G. Volpe, *Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers*. Springer, 2011.

[6] E. Bergeron, R. Valdez, C. J. Moreland, R. Wang, T. Knight, and P. Kushalnagar, “Community Health Navigators for Cancer Screening Among Deaf, Deafblind, and Hard of Hearing Adults Who Use American Sign Language,” *J. Cancer Educ.*, Feb. 2024, doi: 10.1007/s13187-024-02416-x.

[7] Verma, Prashant, and Khushboo Badli. "Real-Time Sign Language Detection using TensorFlow, OpenCV, and Python." *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* 10.V (2022): May2022.

[8] “Sign Language Recognition System using TensorFlow Object Detection API” Accessed: May 06, 2024. [Online]. Available: https://arxiv.org/pdf/2201.01486

[9] S. He, “Research of a sign language translation system based on deep learning,” in *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, IEEE, Oct. 2019. doi: 10.1109/aiam48774.2019.00083.

[10] Y. Song, M. Liu, F. Wang, J. Zhu, A. Hu, and N. Sun, “Gesture Recognition Based on a Convolutional Neural Network-Bidirectional Long Short-Term Memory Network for a Wearable Wrist Sensor with Multi-Walled Carbon Nanotube/Cotton Fabric Material,” *Micromachines (Basel)*, vol. 15, no. 2, Jan. 2024, doi: 10.3390/mi15020185.

[11] R. K. Pathan, M. Biswas, S. Yasmin, M. U. Khandaker, M. Salman, and A. A. F. Youssef, "Sign language recognition using the fusion of image and hand landmarks through multi-headed convolutional neural network," *Sci. Rep.*, vol. 13, no. 1, p. 16975, Oct. 2023.

[12] Gonzalez, *Digital Image Processing,2/e*. Pearson Education India, 2002.

[13] E. L. R. Ewe, C. P. Lee, K. M. Lim, L. C. Kwek, and A. Alqahtani, "LAVRF: Sign language recognition via Lightweight Attentive VGG16 with Random Forest," *PLoS One*, vol. 19, no. 4, p. e0298699, Apr. 2024.

[14] M. Kakizaki, A. S. M. Miah, K. Hirooka, and J. Shin, "Dynamic Japanese Sign Language Recognition Throw Hand Pose Estimation Using Effective Feature Extraction and Classification Approach," *Sensors* , vol. 24, no. 3, Jan. 2024, doi: 10.3390/s24030826.

[15] J. Nguyen and Y. C. Wang, "A Classification Model Utilizing Facial Landmark Tracking to Determine Sentence Types for American Sign Language Recognition," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2023, pp. 1–4, Jul. 2023.

[16] M. Perea-Trigo, C. Botella-López, M. Á. Martínez-Del-Amor, J. A. Álvarez-García, L. M. Soria-Morillo, and J. J. Vegas-Olmos, "Synthetic Corpus Generation for Deep Learning-Based Translation of Spanish Sign Language," *Sensors* , vol. 24, no. 5, Feb. 2024, doi: 10.3390/s24051472.

[17] Y. Gu, H. Oku, and M. Todoh, "American Sign Language Recognition and Translation Using Perception Neuron Wearable Inertial Motion Capture System," *Sensors* , vol. 24, no. 2, Jan. 2024, doi: 10.3390/s24020453.

[18] F. Amber, *Vision Based Sign Language Identification System Using Facet Analysis*. GRIN Verlag, 2014.

[19] L. Kane, B. K. Dewangan, and T. Choudhury, *Challenges and Applications for Hand Gesture Recognition*. IGI Global, 2022.

[20] M. Gil-Martín, M. Villa-Monedero, A. Pomirski, D. Sáez-Trigueros, and R. San-Segundo, "Sign Language Motion Generation from Sign Characteristics," *Sensors* , vol. 23, no. 23, Nov. 2023, doi: 10.3390/s23239365.

[21] H. Alsolai, L. Alsolai, F. N. Al-Wesabi, M. Othman, M. Rizwanullah, and A. A. Abdelmageed, "Automated sign language detection and classification using reptile search algorithm with hybrid deep learning," *Heliyon*, vol. 10, no. 1, p. e23252, Jan. 2024.

[22] A. Ben Haj Amor, O. El Ghoul, and M. Jemni, "Sign Language Recognition Using the Electromyographic Signal: A Systematic Literature Review," *Sensors* , vol. 23, no. 19, Oct. 2023, doi: 10.3390/s23198343.

[23] Å. Elwér and J. Andin, "Geometry in the brain optimized for sign language - A unique role of the anterior superior parietal lobule in deaf signers," *Brain Lang.*, vol. 253, p. 105416, May 2024.

[24] Y. Verma and R. S. Anand, "Gesture generation by the robotic hand for aiding speech and hard of hearing persons based on indian sign language," *Heliyon*, vol. 10, no. 9, p. e29678, May 2024.