English Text to Sign Language Translation Using Python and CNN

Manisha N. More, Jainil D. Bhanushali, Aditya Bhandari, Aadi Bhangdiya,
Yash Bhandari and Yogini Bhamare
Department of Computer Engineering, Vishwakarma Institute of Technology (VIT), Upper Indira Nagar, Bibwewadi, Pune, Maharashtra, India

Keywords: CNN, Flask, NLP, Flask API, Tensor Flow, UI, Web Development, Tokenization.

Abstract: One of the most essential components of human existence is communication. It helps us communicate our ideas and feelings, which strengthens our bonds with other people. We propose our project, which focusses primarily on the deaf and mute community, with this important factor in mind. Any sign language, meanwhile, is a complicated language that confronts many difficulties nowadays, such as a lack of trained interpreters and low general public understanding. We use state-of-the-art technology, namely Convolutional Neural Networks (CNNs) and Natural Language Processing (NLPs), to decode Sign Language in order to solve this problem. This cutting-edge device helps hearing and non-hearing people communicate with each other. Our system takes user-provided text and converts it into matching sign language films that are kept in a database curated by the authors, enhancing accessibility and inclusivity for the deaf community.

1 INTRODUCTION

Sign language is utilised to facilitate communication between such a vast number of followership members and regular people. The community of people who are deaf or hard of hearing use sign language, which is a visual language. (M.M Mohan Reddy and G. Soumya, 2021) It is one of the most crucial forms of communication since it uses body language, face expressions, and hand gestures to convey message. In education, social interactions, and accessibility, sign language is essential because it successfully closes the communication gap between the hearing and the deaf. (Sang-Ki Ko, et.al., 2019) However, sign language specialists are not widely available throughout the world. Engineers, programmers, and others are working on a number of initiatives to eradicate this.

One of the most effective way to tackle this issue is the integration of advanced technologies so that people can smoothly use it by sitting at their home. (H.Y.H Lin and N. Murli, 2022) This project is an effort of the same. Advanced technologies such as Convolutional Neural Networks (CNN), Flask, and Natural Language Processing (NLP) are used in this project. These new technologies can recognise hand movements and textbook, converting them into accurate sign Language for flawless communication. Our project helps in enabling users to communicate without any issue with ease.

The primary aim of this paper is to analyse, understand, and explore the development and application of sign language in bridging the gap between the hearing and deaf communities. (H.Y.H Lin and N. Murli,2022) Our project's main vision is to give users a smooth and intuitive communication experience, enhancing availability and inclusivity. This will help them to communicate better with others without putting in much effort. (Kulkarni,et.al., 2021)

2 LITERATURE REVIEW

2.1 Gesture-to-Text Translation Using SURF for ISL

Using the Speeded-Up Robust Features (SURF) model for feature extraction, Tripathi et al. suggested a gesture-to-text translation system (K. M. Tripathi, et.al., 2023). To enhance gesture identification, their method makes use of image preprocessing techniques such feature extraction, edge detection, and skin masking. For classification, a Bag of Visual Words (BoVW) model was used, and it achieved an accuracy range of 79% to 92%. (P. Verma and K. Badli, 2022)

More, M. N., Bhanushali, J. D., Bhandari, A., Bhangdiya, A., Bhandari, Y. and Bhamare, Y.
English Text to Sign Language Translation Using Python and CNN.
DOI: 10.5220/0013879900004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 2*, pages 189-194
ISBN: 978-989-758-777-1

However, rather than fully translating text to signs, the main focus of this study is gesture-to-text conversion.

2.2 Speech to ISL Translator Using Rule-Based Methods

Using a rule-based methodology, Jadhav et al. created a speech-to-ISL translation system that translates spoken words into ISL gestures (K. Jadhav, et.al., 2021). Using voice-to-text conversion, the system first translates speech into text before mapping the text to matching ISL signs that are kept in a database. Although this approach guarantees grammatical accuracy, it is not flexible enough to translate in real time and has trouble with complicated sentence patterns. (R. Harini,et.al., 2020)

2.3 Speech to ISL Translator Using NLP

An NLP-based method for speech-to-ISL translation was presented by Bhagat et al. (D. Bhagat, et.al., 2022), who used parsing, tokenization, and lemmatization techniques to divide sentences into signable parts. Their technology ensures smoother communication by retrieving relevant ISL indicators from a specified database. (K. Yin and J. Read, 2020) However, its scalability and robustness are challenged by out-of-vocabulary words and a lack of ISL datasets.

2.4 Translating Speech to ISL Using Natural Language Processing

In order to translate spoken words into ISL animations, Sharma et al. investigated natural language processing (NLP) methods such tokenization, parsing, and part-of-speech tagging (P. Sharma,et.al., 2022). Their technology takes audio input, translates it into text, and then compares it to ISL sign videos that have already been captured. (N. C. Camgoz, et.al., 2023) Although this method enhances translation at the sentence level, real-time implementation is challenging due to its dependence on a fixed video database.

2.5 Speech to ISL Translation Using Kinect-Based Motion Capture

Sonawane and colleagues created a Kinect-based ISL translation system that uses Unity3D to convert human motion into 3D animated ISL signs (P. Sonawane, et.al., 2021). The technology is quite engaging because it shows ISL indications in real-time visually. However, its limited accessibility and portability due to its reliance on Kinect technology prevent it from being widely used. (S. Clifford M. Murillo,et.al., 2021)

2.6 Deep Learning-Based ISL Translator Using CNN

Table 1: Comparison of Methods.

Author	Methodology	Database used	Accuracy	Limitations
Tripathi et al. (2023)	SURF-based recognition of gestures	42000 images used (alphabets and numbers)	79%-92%	Restricted to conversion from gestures to text (K. M. Tripathi, et.al., 2023)
Jadhav et al. (2021)	Rule-based speech-to-ISL translation	Pre-determined ISL dictionary	Not specified	Struggles with complex sentences, and real time use (K. Jadhav, et.al., 2021)
Bhagat et al. (2022)	NLP-technique that convert speech to ISL	Finite ISL Dataset	Not specified	Faced issue with limited ISL dataset, struggles with out-of-vocabulary words (D. Bhagat, et.al., 2022)
Sharma et al. (2022)	NLP-technique that convert speech- to-video ISL system	Pre-recorded ISL videos database	Not specified	Dependent on pre-recorded ISL videos, not working in real-time (P. Sharma,et.al., 2022)
Sonawane et al. (2021)	Kinect-based ISL translation	Captured motion data from kinetic	Not specified	Requires specialized Kinect hardware, limits accessibility (P. Sonawane, et.al., 2021)
Bagath et al. (2023)	CNN-based gesture recognition	Large dataset (not specified)	High Accuracy (not numerically defined)	Requires large datasets, high computational power (B. S, D. Varshini, and J. G., 2023)

A CNN-based ISL recognition system that uses deep learning models to convert hand gestures into text was proposed by Bagath et al. (B. S, D. Varshini, and J. G., 2023). The system uses a Convolutional Neural Network (CNN) to classify hand motions into ISL signs after preprocessing them using computer vision techniques. table 1 shows the Comparison of methods Despite the great accuracy of this model, real-world implementation is difficult due to its dependency on big labelled datasets and significant computational resources. (Z. Liang, et.al., 2023).

3 METHODOLOGY

Through this initiative we want to help deaf people. The primary goal is to create an English text to sign language translator. It will be useful to both deaf as well as normal people. There are a lot of programmers or engineers working on new ways on interaction between normal human and a hard to hear human. This project is an effort to remove the communication barrier between normal people and hard of hearing people. In this the user types the sentence or word or number and the system will give its correct sign gesture in the form of a video. The system is a combination of Convolutional Neural Networks (CNN), web frameworks (Flask API), Python, Natural Language Processing (NLP), and front-end concepts to create this translation tool. (S. Clifford M. Murillo,et.al., 2021) A use of dataset containing all the videos is done in this project. These videos are resized, renamed and somewhat edited before actual use. We haven't included audio input because of speech recognition challenges, background noises, accents etc. Also, the language we speak and the sign language have different grammar, so it may not produce meaningful results. For the time being this system only accepts text inputs and we can integrate audio input also in the future.

The main component which we used in this project Convolutional Neural Networks (CNN). It is very useful in video processing tasks especially when converting text to sign gestures. Based on the input text CNN just predicts the correct sign gesture in in (R. Harini,et.al., 2020). Next, we designed the system in a webpage so that it will be user friendly. The UI is made through basic web development components HTML (for structure), CSS (for styling), Javascript (functionality and connecting Flask to the webpage). Flask is used as a backend framework which look after the main model and user interface. An API is created in Flask to accept input from user, this input is passed through CNN model and the corresponding sign gesture is displayed on user interface in the form of video. The Javascript takes inputs in front end from user and send request to Flask API in real time. The CNN model is connected with Flask backend. This is where the user input is processed and sign gesture is displayed on the UI. Natural Language Processing (NLP) is used in tokenization, grammar conversion (rearranging words in proper order in sign gesture as it is different from the grammar we speak).

First user inputs English text in the textbox, clicks on translate. Now, as part of text pre-processing step in NLP text is converted to lowercase for better matching to correct sign gesture, consistency and reducing unnecessary duplicates. Also, Lemmatization of text is done using NLP itself. It means grouping together different forms of words so that there is consistency in text processing like in (Kulkarni,et.al., 2021). This is sent to Flask via HTTP and then to CNN for text processing. The model translates the text to accurate sign gesture in the form of videos from our database. Videos are present of letters, words, numbers. If user types a word or sentence which is not present in our dataset, then the concatenated version of the letters or numbers is shown (S. Clifford M. Murillo,et.al., 2021). Figure 1 shows the Working of system flowchart. This concatenation can be in the form of letters or numbers as well (Dr. S V.A Rao,et.al., 2022).

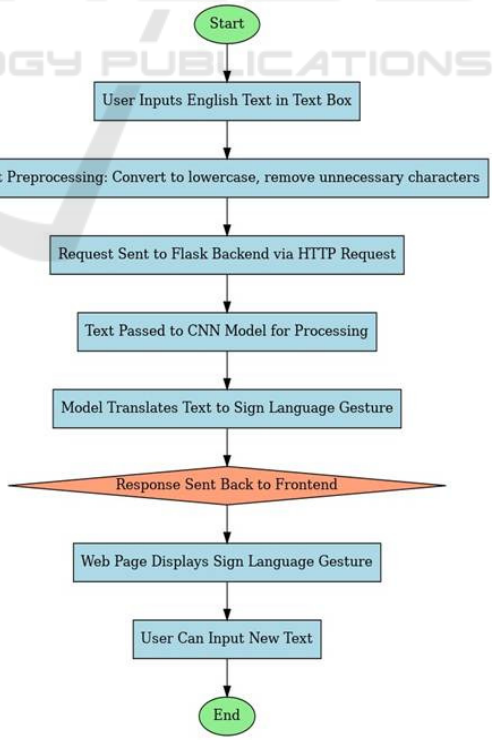

Figure 1: Working of System Flowchart.

This video thus processed is shown on the webpage where user can interact with it. The dataset used in this is majorly animated one with American sign language. This dataset has letters, numbers and some important words we use in our day-to-day life primarily in ASL... All videos are animated in this dataset. We took this dataset online from Kaggle. It contains labelled sign gestures videos. We somewhat modified it for pre-processing steps in our project.

The final created system is tested properly before actual use. Its accuracy and effectiveness are also tested. During this project several challenges are imposed. First is variation in sign gesture across different regions. We have to look that the UI we build is easy to understand for those who don't know much about technology.

So, this methodology provides an easy approach to build an English text to sign language translator. In the future we can improve the accuracy, we can expand the dataset as well as more languages can be added to it.

4 MAJOR MATHEMATICAL EQUATIONS

The text processing is done which is known as tokenization, meaning breaking a sentence into words or words into letters. This is done with the help of NLP techniques:

$$T = \{W1, W2, W3, \ldots . . Wn \quad (1)$$

Where,

T = Tokenized sentence or word.

Wi = individual words or letters.

A word for example "Hello" is typed, it will be treated as a sequence of letters: {'h', 'e', 'l', 'l', 'o'}. Then the lengths of the respective letters are checked: If h is 2 sec, e is 3 sec and l is 1 sec, the total time to display the concatenated video is given by their sum

$$\sum_{i=1}^{n} Duration\ of\ letters\ (or\ words)i \quad (2)$$

Therefore, for 5 letter clips

Total time = Duration 1 + Duration 2 + …. + Duration 5

2) Flask processing equation

Total duration to process a request to Flask is

$$T\ total\ =\ T\ preprocess\ +\ T\ inference\ +\ T\ response \quad (3)$$

Where:

T preprocess**:** Time taken for processing of text (NLP)

T inference: Time taken for CNN model

T response: Time taken to send the final sign gesture to the user

All the calculations of time are in milliseconds (ms).

3) Word vector embedding:

After converting letters and words to lowercase, it is still text. They should be converted to numbers for processing by machine learning models. Therefore, we use vectors for this.

Every word or letter is converted to vectors of numbers. This help models to understand the connection between words.

Equation is

$$E(w) = W \cdot v + b \quad (4)$$

E(w)**:** Final numerical representation of word

v: input word vector (each word has this vector which is a fixed size set of numbers).

W: Weight matrix (transform input word vector to some other form which has more meaning).

b: Bias term (small value added so that models can understand better).

How this all works together:

Step 1. User enter English text

Step 2. Conversion of text to Lowercase

Step 3. Tokenization of words

Step 4. Conversion of words to vectors

Step 5. Pass these vectors for sign language conversion to the machine learning models.

5 RESULTS AND DISCUSSION

In this work of English text to sign language translation, use of various modern techniques is done such as CNN, Flask API, NLP etc. A very basic UI design made up of HTML, CSS, Javascript is also created which is completely user friendly. The figure 2,3,4,5,6 shows the results.It provides an easy approach for people who are needy for this but don't know about technology much. Thus, this system helps to bridge the gap between a normal person and a hard to hear person.

The processing time and display of final result take a few seconds. For small words or letters or numbers it takes time below 5 seconds and if the words are big or user enters a sentence, the corresponding sign language is shown in 15 seconds or less than that. The accuracy of this project is 100%

Here are some of the results of the system:

Figure 2: Screenshot of Result.

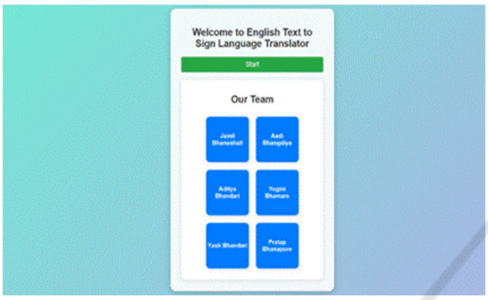

Figure 3: Screenshot of Result.

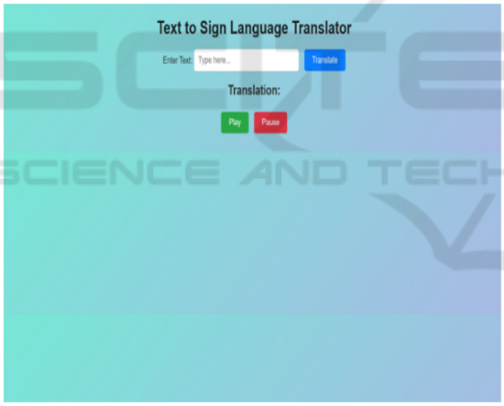

Figure 4: Screenshot of Result.

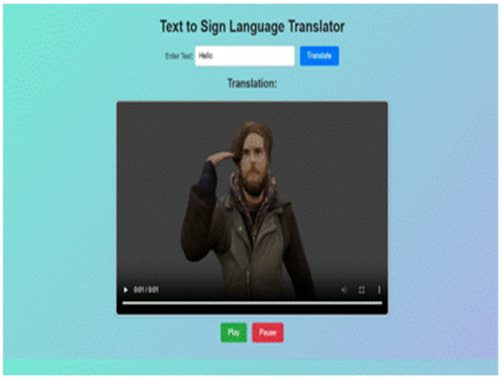

Figure 5: Screenshot of Result.

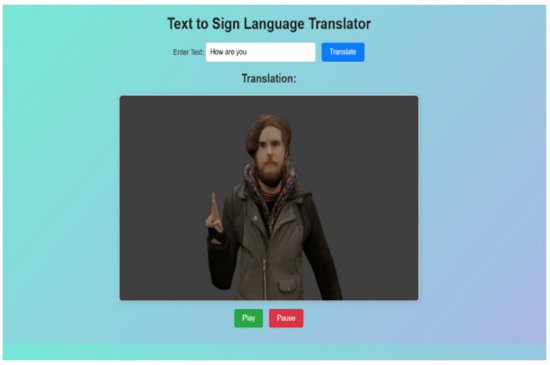

Figure 6: Screenshot of Result.

6 FUTURE SCOPE

This project has a wide range of development ideas and future advancements. One is to expand the dataset to include more words, phrases and even sentences to improve translation. Next improvement we can do is to provide multiple languages so that more people who are actually needy can use this without any language barrier. In the future we can take audio inputs as well making it a real time speech recognition system. Gesture based input can also be integrated using camera and lenses. Computer vision and technologies like media pipe can recognize live movements giving real time translation. Further we can also develop a mobile app for this project with all enhancements. This app will be available for Android as well as iOS.

7 CONCLUSIONS

The current ISL translation systems use a variety of methods, including rule-based strategies, natural language processing, machine learning, and gesture recognition. Although organized text-to-sign translation is provided via rule-based and natural language processing techniques, they frequently lack real-time adaptability. High accuracy is offered by deep learning models, especially CNN-based techniques, however they come with a high computational cost and training data requirement. Although they require specific technology, gesture-based techniques like Kinect motion capture improve real-time engagement. For a more complete ISL translation system, future studies should concentrate on hybrid strategies that combine natural language processing, machine learning, and 3D animation.

ACKNOWLEDGEMENTS

We are very thankful to Vishwakarma Institute of Technology, Pune for helping us by giving this fantastic platform and needed resources to develop this project. We are also thankful to our faculty mentors and teachers for their continuous help, valuable inputs and guidance throughout the whole process of research and developments. Their support is very crucial role in making our project.

Additionally, we are also express our gratitude to our college, class friends and collaborators for their constructive feedback and discussions which help us to modify and improve our project and refine our approach.

REFERENCES

"Speech To Indian Sign Language (ISL) Translation System," 2021 International Conference on Computing,

A. Kulkarni, A.V Kariyal, Dhanush V, P.N Singh, "Speech to Indian Sign Language Translator," Proceedings of the 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021), Atlantis Press, volume 4, 2021

B. S, D. Varshini, and J. G. M, "Sign Language Translator Using Python," International Journal of Scientific Research in Engineering and Management (IJSREM), vol. 7, no. 8, pp. 1-3, 2023.

Based on Python," Material Science and Technology, vol. 22, no. 3, pp. 227–237, Mar. 2023.

Communication, and Intelligent Systems (ICCCIS), IEEE, pp. 92-93, 2021.

D. Bhagat, H. Bharambe, and J. Joshi, "Speech to Indian Sign Language Translator," International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), vol. 2, no. 3, pp. 538-541, 2022.

Dr. S V.A Rao, K.S Karthik, B. Shiva Sai, B. Sunil Kumar, N Madhu, "Audio to Sign Language Translator Using Python," Neuroquantology, vol. 20, issue 19, pp. 5461-5468, Nov. 2022.

H.Y.H Lin, N. Murli, "BIM Sign Language Translator Using Machine Learning (TensorFlow)," Journal of soft computing and data mining, vol. 3, no. 1, pp. 68–77, Jun. 2022.

K. Yin and J. Read, "Better Sign Language Translation with STMC-Transformer," arXiv preprint, arXiv:2004.00588, 2020.

K. Jadhav, S. Gangdhar, and V. Ghanekar, "Speech to ISL (Indian Sign Language) Translator," International Research Journal of Engineering and Technology (IRJET), vol. 8, no. 4, pp. 3696-3697, 2021.

K. M. Tripathi, P. Kamat, S. Patil, R. Jayaswal, S. Ahirrao, and K. Kotecha, "Gesture-to-Text Translation Using SURF for Indian Sign Language," Applied System Innovation, vol. 6, no. 35, 2023. DOI: 10.3390/asi6020035.

M.M Mohan Reddy, G. Soumya, "Sound to Sign Language Translator

N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, "Sign Language Transformers: Joint End-to-End Sign Language

P. Sharma, D. Tulsian, C. Verma, P. Sharma, and N. Nancy, "Translating Speech to Indian Sign Language Using Natural Language Processing," Future Internet, vol. 14, no. 253, pp. 1-17, 2022. DOI: 10.3390/fi14090253.

P. Verma, K. Badli, "Real-Time Sign Language Detection Using TensorFlow, OpenCV, and Python," International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 10, issue 5, May 2022.

P. Sonawane, P. Patel, K. Shah, S. Shah, and J. Shah,

R. Harini, R. Janani, S. Keerthana, S. Madhubala and S. Venkatasubramanian, "Sign Language Translation," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 883-886. DOI: 10.1109/ICACCS48705.2020.9074370

Recognition and Translation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 10023–10033.

S. Clifford M. Murillo, M. Cherry Ann E. Villanueva, K. Issabella M. Tamayo, M. Joy V. Apolinario, M. Jan D. Lopez, "Speak the Sign: A Real-Time Sign Language to Text Converter Application for Basic Filipino Words and Phrases," Central Asian Journal of Mathematical Theory and Computer Sciences, Volume: 02 Issue: 08, Aug 2021

Sang-Ki Ko, Chang Jo Kim, H Jung, C Cho, "Neural Sign Language Translation Based on Human Key point Estimation," International Journal of Computer Vision, vol. 9, no. 13, pp. 456–467, Jul. 2019.

V. Swathi, C.G Reddy, N. Pravallika, P. Varsha Reddy, Ch. Prakash Reddy, "Audio to Sign Language Converter Using Python," Journal of Survey in Fisheries Sciences, 10(1), 2023, 2621-2630.

Z. Liang, H. Li, and J. Chai, "Sign Language Translation: A Survey of Approaches and Techniques," Electronics, vol. 12, no. 12, p. 2678, June 2023.