International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.096

Sign Language Translator for Speech Impaired

Dr Ananthajothi.K *Computer Science and Engineering Rajalakshmi Engineering College*
Chennai, India
ananthajothi.k@rajalakshmi.edu.in

Srisai R
Computer Science and Engineering
Rajalakshmi Engineering College Chennai, India srisai.r.2019.cse@rajalakshmi.edu.in

Vikram R
Computer Science and Engineering
Rajalakshmi Engineering College Chennai, India
vikram.r.2019.cse@rajalakshmi.edu.in

Srivathsan S
Computer Science and Engineering
Rajalakshmi Engineering College Chennai,
India srivathsan.s.2019.cse@rajalakshmi.edu.in

***Abstract*—There are several sign recognition algorithms that produce word output to facilitate communication between mute persons and non-mute people. Our suggested approach focuses on translating sign language into text and voice. With the aid of Natural Language Processing (NLP) techniques, which include automatic word completion and sentence creation, appropriate English sentences are framed with expected words before being transformed to speech. Live video is used as input, then video is framed and segmented. For monitoring hand and facial gestures, Media Pipe Holistic algorithm is utilized.**

Keywords—Natural Language Processing, Holistic Media Pipe

I. Introduction

People with hearing loss frequently utilize as a communication method, sign language is used. A sign language is nothing more than a collection of varied hand signals created by varying hand shapes, movements, and orientations, together with face expressions. 34 million of the 466 million people with hearing loss in the world's population are children. Deaf individuals have minimal to no hearing abilities. They communicate using sign language. Around the world, many sign languages are used by people. They are quite few in number when compared to spoken languages. Indian Sign Languages is the name of the sign language used in India (ISL). There aren't many deaf-specific schools in poor nations. Gestures are naturally performed by humans.

Gestures are made intentionally, as signals or indicators, or unconsciously to communicate attitudes or intentions. Even though they entail the entire body's movements, studies frequently concentrate on the arms and hands since they are necessary for movement and dialogue. Facial expressions play a significant part in communication and are also regarded as gestures. Most everyday human actions or activities involve gestures, which take part in human interaction through either completing speaking or standing in for someone else for uttered words in situations requiring silent communication (under water, in noisy environments, in secret, etc.) or for those who have hearing impairments.

Between the speaking community and the deaf and mute population, there has always been a communication barrier. This is particularly clear when there is an emergency. A human interpreter is typically used in order to translate. However, everyone cannot afford a professional translator, and in times of crisis, the presence of a human interpreter cannot be guaranteed. This undertaking tries to remove this obstacle. The process of image processing and machine learning are used to complete the assignment. For picture categorization and recognition, very potent tools like the process of processing images and machine learning are frequently used.

Image processing is concerned with the image, its elements, and the actions taken to extract information from it. The research into algorithms and statistical information that are used to carry out tasks utilizing different data patterns and conclusions is known as machine learning. A camera will be used to capture the images of sign language for this project. The features are then retrieved from the photos using image processing. These photos are then compared using the datasets that are readily available, and the indications are deciphered using deep learning. The information is shown on a display so that individual seeing a deaf or mute person can comprehend their sign language

This project seeks to bridge the communication gap between a person with a physical impairment and a person who is specially challenged by identifying symbolic expressions through photographs and translating them into voice or text. To provide output in the form of words, various sign recognition algorithms are applied. Our suggested approach focuses on translating sign language into speech. Natural language processing techniques, such as automatic word completion and sentence creation, are used to construct appropriate English sentences with expected terms. Live video is used as the input, then video is framed and segmented. For tracking hand and facial gestures, Media Pipe Holistic algorithm is employed.

II. Literature Survey

There are four basic steps in the implementation: Enhancement and segmentation of images 2. Detection of orientation Extraction of Feature 4. Categorization.

In [1], The system uses video sequences from which convolutional neural networks were built and adds photographs of the train data. Predictions for individual frames were made using the Trained Convolutional Neural Network model in order to provide a sequence of predictions. The system was created for the purpose of translating Indian sign language and producing legible output, the proposed system creates a trustworthy communication interpretation programmed Static sign-language motions were identified using a deep learning system. This algorithm was developed for use with a Raspberry Pi. A camera was used to capture the photos, after which the features were retrieved using HSV filtering.

DOI: rp-9788770040723.096

The values for the skin tones were chosen to distinguish the hands from the background. The image was then changed to a grayscale format in order to recover the hand's characteristics. Following this, the groupings of these images were separated into training and test groups, and the model was fed with them. The proposed model had two dense layers, the first of which served as the activation layer and the second of which served as the tanh layer [2].

The suggested solution makes use of a hand glove to recognize unique patterns and gestures. They employed Flex Sensors. Processing was carried out using Arduino Mega. To mimic the user's hand movements and patterns, a glove was used. Flex sensors are positioned here along the thumb and fingers. The sensors determined how much the fingers were bent and produced a voltage fluctuation as an output [3].

The technology designed produced a variety of words, voice, and gestures. This suggested method let normal people and people with disabilities communicate with one another. By using a high-definition camera to record various hand motions, a training dataset is built. A minimum of ten photos are taken for each gesture. The median filter is used to denoise the collected images. By removing the backdrop information, the hand gesture alone was segregated for better training [5].

For this system's implementation, the suggested system offered audio as input, and pattern matching techniques were utilized to match the audio. The results displayed the indicators for audio matchups. Second, for the system to record the Sign movements, a web camera was required. The result of the identified sign language gesture in audio form was shown [5].

[6] The device provided real-time audio, making it simpler for those with hearing loss to communicate with others who can hear them. If the phrase is saved in the database, the system will be able to recognize it in the sentence and present sign language videos that are relevant to it. Almost a thousand films, a combination of the author's own recordings and freely accessible lectures by ISL professors, were stored in the database.

The proposed device is a personal communication system designed to assist people who are deaf, dumb, and/or blind in communicating with others. The device is adaptable and incorporates various technologies such as text-to-voice conversion, OCR, and speech-to-text conversion to enable users to interact with each other through text or spoken language. For dumb individuals, the device offers a text-to-voice conversion feature that reads out the text for them. Visually impaired individuals can use the device's OCR feature to read text, and the e-speak technology can read the text aloud for them. Deaf individuals can use the device's speech- to-text conversion feature to read what others are saying or communicate using written text that can be converted to speech for others. Blind individuals can also use the device's OCR feature to read text or paragraphs. [7].

[8] Using a three-dimensional hand point as input, the hand pose estimator precisely and successfully regresses the hand joint locations. A folding-based encoder is used in the proposed model to fold a converted into the appropriate joint coordinates from a 2D hand skeleton. Folding was guided by multi-scale characteristics that included global and joint-wise local features for increased estimation accuracy.

The "intelligent robotic arm voice-controlled assistance device for physically challenged person" was part of the planned system. The joystick was used to control a robotic arm that was installed on a rail system. The movements were recognized as input by both the gloves and the camera. The gestures were converted into speech with the use of the pre-defined dataset, which had a gesture corresponding to each word [9].

The proposed device enabled interaction between the blind, deaf, and dumb by using a glove-based communication interpreter system. A three-axes movement tracking inertial sensor was part of the glove system. (x,y,z). Data from the accelerometer sensor was analyzed by the microcontroller before being sent over Bluetooth to the mobile device, where it was converted into text or speech [10].

[11] The suggested approach made it possible for a deaf person to interact with hearing people without the use of an interpreter. The authors made use of the Tensorflow library and the Keras framework. The camera captured the image, compared it to the previously recorded photographs, and then assigned it to the appropriate pre-defined character. The result was created once the words were placed together into a phrase that made sense as a sentence.The suggested technique uses self-comparisons to accurately estimate the hand positions on three publicly available hand pose datasets. For the ICVL and NYU datasets, the suggested method outperformed methods that were based on 2D CNN, and it also did well on the MSRA dataset. The suggested technique determined the mesh vertices of each individual human part, including the body, hands, and face regions, in a 3D environment [12].

The proposed model makes use of CNN and Attention-Based Hierarchical RNN to quickly identify URLs. They were able to develop an accurate model to recognize phishing URLs by integrating these URL models using a three-layer CNN [13].

The first stage in developing sign language detecting systems was accurate hand segmentation. The hand in the original image was recognised by the authors in the proposed study, and they gave this part of the image to the multi-class classifier, which can identify static motions. The proposed approach first built the data collection, converting each image into a feature vector (X) and giving each one a label that matches to the sign language alphabet given by (Y). The projected classifiers correctly identified 65% of the aforementioned images [14].

[15] The purpose of this approach was to close communication gap between those with hearing and speech impairments and the general public. The available alternatives either don't provide real-time data or provide data with a low level of precision. This system produced positive results for both parameters. 33 hand motions were

DOI: rp-9788770040723.096

identified, along with a few ISL gestures. A smartphone camera captured frames of Sign Language, which were then transferred for processing to a distant server. It is user-friendly because no additional hardware, such as gloves is required. A Grid-based Feature Extraction approach is used to represent the pose of the hand in the image.

III. METHODOLOGY

A. Data Generation

Data generation refers to the process of creating or generating data for use in various applications, such as machine learning, data analytics, and artificial intelligence. This data can be created using a variety of methods, including simulations, surveys, experiments, and data scraping.

Data generation is an important aspect of many applications, particularly in the fields of artificial intelligence and machine learning, where large amounts of data are required to train models and algorithms. However, it is important to ensure that the data generated is of high quality and is representative of the real-world situations being modeled or analyzed.

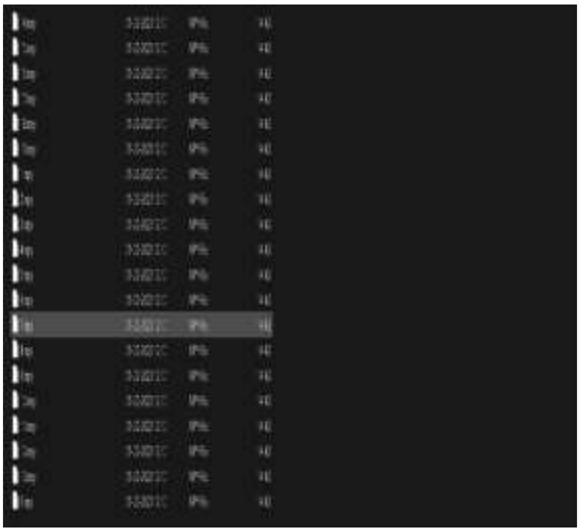

Fig. 1. Data Generation

Sign Translator application is divided into 3 modules:

1) Handgesture Recognition

The input image is pre-processed and the Region of Interest is extracted. The media pipe holistic framework extracts body pose coordinates, hand pose coordinates and face pose coordinates. Then the coordinates are transformed into NumPy arrays. For each gestures up to 30 NumPy arrays are already generated and collected in respective class and the model has been trained with the collected data. So that the gestures will be recognized and corresponding words will be produced.

There are several approaches to hand gesture recognition, including using depth cameras, 2D cameras, and machine learning algorithms. Depth cameras, such as Microsoft's Kinect, use infrared technology to capture depth information, making it possible to distinguish between the hands and other objects in the environment. 2D cameras, on the other hand, rely on image processing techniques to detect and track the hands.

Machine learning algorithms, such as convolutional neural networks (CNNs), are commonly used for hand gesture recognition. These algorithms are trained on large datasets of hand gesture images, allowing them to recognize a wide variety of gestures. They can also adapt to different lighting conditions, camera angles, and hand positions.

Hand gesture recognition has numerous applications, including controlling devices without physical contact, assisting people with disabilities, and improving human-robot interaction.

Fig. 2. Handgesture Recognition

2) Gesture to text and speech

The process of gesture to text and speech typically involves using a camera to capture hand gestures, which are then interpreted by computer vision algorithms. These algorithms can recognize different types of hand gestures, such as pointing, waving, or making a specific hand shape, and translate them into text or speech output. One common application of gesture to text and speech technology is sign language recognition.

From the hand gesture recognition module, the keywords are produced. Using the Long Short Term Memory Natural Language Processing (LSTM NLP) model, meaningful statement from the produced words will be framed. Also using the face pose marks and body pose marks the proper meaningful expressive sentences can be generated with proper tenses.

DOI: rp-9788770040723.096

Fig. 3. Gesture To Text And Speech

3) Speech Or Text To Gestures

From the Mic the voice will be recorded and sent to Natural Language Processing (NLP) model. The Speech will be processed by Natural Language Processing (NLP) model. Then the keywords are extracted. For each word animated gesture video are made and stored. The extracted words will be mapped to the respective video file. Finally, all the mapped videos will be merged and produced as a single video which can be understood by Speech impaired people.

Fig. 4. Speech or text to gestures

IV. RESULTS AND ANALYSIS

The classification model's performance is measured in terms of how well it can properly identify instances that belong to a particular class using precision, recall, and F1 score. These metrics are necessary for evaluating the efficacy of a machine learning model and deciding whether to use it in practical applications. The Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture used in deep learning. LSTM is particularly useful for processing sequential data, such as speech, natural language, and time- series data. LSTM can produce an accuracy of 100% on train data and 98% on test data.

Fig. 5. Classification Report

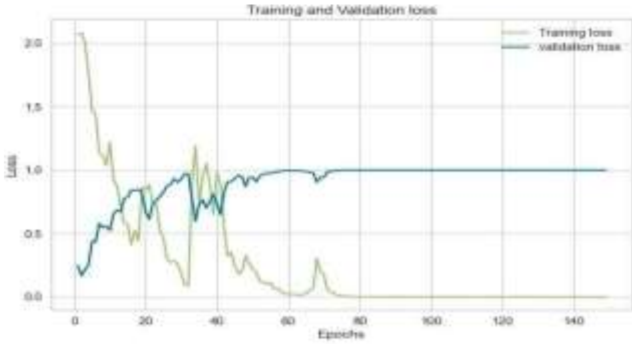

Fig.6. Training and Validation Loss

V. CONCLUSION

In this study, the existing systems are examined, and a proposed application is created to address the drawbacks of the earlier systems. After studying a number of algorithms and works that give output results with various degrees of accuracy, it is clear that the current systems are either unable to recognize hand gestures accurately or struggle to translate them into speech. The majority of the way the current system operates doesn't happen in real time.

REFERENCES

[1] Abougarair, Ahmed and Arebi, Walaa, "Smart glove for sign language translation," International Journal of Robotics and Automation, vol. 8, pp. 109-117, 2022, 10.15406/iratj.2022.08.00253.

[2] M. Shree, 2022, "Sign Language Conversion to Speech with the Application of KNN Algorithm".

[3] Tan, Gloria and Khairu, Nur and Tan, Chi Wee and Ung, Ling and Leng, Ngo and Eri, Zeti, "An Assistive Bahasa Malaysia Sign Language Translator Using Convolutional Neural Network," 2022.

[4] Joshi, Prof and Desale, Shraddha and Gunje, Shamali and Londhe, Aditi, "A Survey on Sign Language Translation Systems," International Journal for Research in Applied Science and Engineering Technology, vol. 10, pp. 519-525, 2022, 10.22214/ijraset.2022.41295.

[5] D. Karthika and Kumar, L. Ashok, "Indian Sign Language Recognition Using Deep Learning Techniques," International Journal of Computer Communication and Informatics, vol.4, pp. 36-42, 2022, 10.34256/ijcci2214.

[6] Dhanabalan, S. S., Sitharthan, R., Madurakavi, K., Thirumurugan, A., Rajesh, M., Avaninathan, S. R., & Carrasco, M. F. (2022). Flexible compact system for wearable health monitoring applications. Computers and Electrical Engineering, 102, 108130.

[7] Ke, Soong and Mahamad, Abd and Saon, Sharifah and Fadlilah, Umi and Handaga, Bana, "Malaysian Sign Language Translator for Mobile Application," 2022, 10.1007/978-981-16-8129-5_139.

[8] K Mohammed, "American Sign Language Translator," International Journal for Research in Applied Science and Engineering Technology," vol. 10, pp. 3430-3433, 2022, 10.22214/ijraset. 2022.41924.

DOI: rp-9788770040723.096
[9] Baktash, Abdullah and Mohammed, Saleem and Yahya, Ammar, "Sign language translator: Web application based deep learning," AIP Conference Proceedings, vol. 2398, p. 050020, 2022, 10.1063/ 5.0093367

[10] Mishra, Harshita and Sharma, Mansi and Ali, Muskan and Chaudhary, and Shivani, "Audio to Indian Sign Language Translator," International Journal for Research in Applied Science and Engineering Technology, vol. 10, pp. 3904-3907, 2022, 10.22214/ ijraset.2022.43269.

[11] Gomathy, V., Janarthanan, K., Al-Turjman, F., Sitharthan, R., Rajesh, M., Vengatesan, K., & Reshma, T. P. (2021). Investigating the spread of coronavirus disease via edge-AI and air pollution correlation. ACM Transactions on Internet Technology, 21(4), 1-10.

[12] Badarch, Luubaatar and Ganbat, Munkh-Erdene and Altankhuyag, Otgonbayar and Amartuvshin, Togooch, "Mongolian Sign Language Recognition Model," ICT Focus, vol. 1, pp. 1-9, 2022, 10.58873/sict.v1i1.27.

[13] Ferreira, Gustavo and Bettencourt, and Nuno. "Healthcare-Oriented Portuguese Sign Language Translator," 2023, 10.1007/978-3-031-27499-2_79.

[14] Kadam, and Sanjay, IJRESM V3 I4 103-sign language, 2022.

[15] M. Shree, "Sign Language Conversion to Speech with the Application of KNN Algorithm, 2022.